SCHOOL OF ECONOMIC & BUSINESS SCIENCES

Research Report

Data Scientist: Using a Competency Based Approach to Explore an Emerging Role

Submitted by:

Naseema Nosarka

1558925

Susan Benvenuti

Date: 20 September 2018

PLAGIARISM DECLARATION

I declare that this is my own unaided work and is, to the best of my knowledge and belief, original, except as acknowledged in the text.

I have read and understood the Senate Policy on Plagiarism

I understand that plagiarism is the "failure to acknowledge the ideas or writing of another" or "presentation of the ideas or writing of another as one's own" - whether such failure to acknowledge the ideas or writings of others is intentional or unintentional.

I understand what is expected of me in terms of referencing style and how to appropriately acknowledge the ideas or writing of others.

I am aware of the consequences of plagiarism.

Name and Surname: Naseema Nosarka

Signed: NNosarka

Date: 20/09/2018

TABLE OF CONTENTS

PLAGIARISM DECLARATION 2

LIST OF TABLES 6

LIST OF FIGURES 6

LIST OF ABBREVIATIONS 7

ACKNOWLEDGEMENTS 8

ABSTRACT 9

1 INTRODUCTION 11

1.1 BACKGROUND 11
1.2 PROBLEM STATEMENT 12
1.3 PURPOSE OF THE STUDY 14
1.4 INTENDED CONTRIBUTIONS OF THIS STUDY 14
1.5 LIMITATIONS OF THIS STUDY 14

2 LITERATURE REVIEW 16

2.1 INTRODUCTION 16
2.2 DATA SCIENTIST ROLE 16
2.3 DATA SCIENTIST SKILLS 18
2.4 DATA SCIENTIST SKILLS GAP 21
2.5 DATA SCIENCE IN THE BANKING AND INSURANCE INDUSTRY 22
2.6 COMPETENCY 23
2.7 COMPETENCY MODELS 24
2.8 COMPETENCY LIFE CYCLE 30
2.9 SUMMARY 30

3 CONCEPTUAL FRAMEWORK 32

3.1 INTRODUCTION 32
3.2 RESEARCH FRAMEWORK 32
3.3 SUMMARY 35

4 RESEARCH METHODOLOGY 37

4.1 INTRODUCTION 37

4.2 RESEARCH APPROACH AND PARADIGM 37
4.3 RESEARCH METHODOLOGY AND DESIGN 38
4.4 DATA COLLECTION METHODS 38
4.5 DATA ANALYSIS METHODS 42
4.6 RESEARCH EVALUATION METHODS 44
4.7 ETHICAL CONSIDERATIONS 45
4.8 SUMMARY 45

5 RESULTS AND DISCUSSION 47

5.1 INTRODUCTION 47
5.2 PROFILES OF THE CONTRIBUTING DATA SCIENTISTS 47
5.3 THE ROLE OF THE DATA SCIENTIST 49
5.4 SUMMARY OF THE ROLE OF A DATA SCIENTIST 55
5.5 COMPETENCIES OF A DATA SCIENTIST 58
5.6 SUMMARY OF THE COMPETENCIES OF A DATA SCIENTIST 80
5.7 GAPS IN COMPETENCIES OF A DATA SCIENTIST 82
5.8 RECOMMENDATIONS FROM DATA SCIENTISTS 84
5.9 SUMMARY 86

6 CONCLUSION 88

6.1 RESEARCH QUESTIONS AND FINDINGS 88
6.2 MIGRATION OF ROLES 91
6.3 SUMMARY 92

7 REFERENCES 95

8 APPENDIX 102

8.1 APPENDIX A: LITERATURE REVIEW 103
SEARCH STRINGS AND QUALITY CRITERIA 106
8.2 APPENDIX B: SUMMARY OF THEMES FOR DATA SCIENTIST ROLE FROM THE LITERATURE 108
8.3 APPENDIX C: LITERATURE REVIEW MAPPED TO RESEARCH MODEL 109
8.4 APPENDIX D: ETHICAL CLEARANCE CERTIFICATE 113
8.5 APPENDIX E: PARTICIPATION LETTER (SPL) 114
8.6 APPENDIX F: PARTICIPATION CONSENT FORM (SPCF) 115
8.7 APPENDIX G: CONSENT FORM FOR INTERVIEW RECORDING (CIR) 116
8.8 APPENDIX H: INTERVIEW QUESTIONS 117
8.9 APPENDIX I: EXAMPLE OF INTERVIEW DATA 120
8.10 APPENDIX J: A VIEW OF THE TRANSCRIBED DATA WITH CODES DONE IN EXCEL 122

8.11 APPENDIX K: SUMMARY OF THEMES FOR DATA SCIENTIST ROLE FROM THE RESULTS 123
8.12 APPENDIX L: COMPETENCY PROFILES FROM INTERVIEWS 124
8.13 APPENDIX M: TECHNICAL COMPETENCY CATEGORY 130
8.14 APPENDIX N: DATA SCIENCE ORCHESTRATION COMPETENCY CATEGORY 133
8.15 APPENDIX O: ORGANISATIONAL KNOWLEDGE CATEGORY 135
8.16 APPENDIX P: BUSINESS INTERPERSONAL COMPETENCY CATEGORY 136
8.17 APPENDIX Q: DATA SCIENTIST FUNDAMENTAL COMPETENCY CATEGORY 138
8.18 APPENDIX R: DATA SCIENTIST COMPETENCY MODEL TABLE 141
8.19 APPENDIX R: TURNITIN PAGE 144

LIST OF TABLES

Table 1: Literature Review Data Scientist Themes 17
Table 2: Data Scientist Skills Gap 22
Table 3: Data Scientist Profiles 49
Table 4: Summarised Themes describing the Role of a Data Scientist 56
Table 5: Technical Competency from Results and Literature 62
Table 6: Data Science Orchestration Competency Category 66
Table 7: Organisational Knowledge Competency Category 68
Table 8: Business Interpersonal Competency Category 73
Table 9: Data Scientist Fundamental Competency Category 79

LIST OF FIGURES

Figure 1: Pyramid view of Competencies 25
Figure 2: Competency Model of a Business Process Analyst 26
Figure 3: Skill and information system knowledge classification 28
Figure 4: Competency Model of a Business Intelligence Analyst 29
Figure 5: Big Data Professional Competency Model 29
Figure 6: Competency Categories 32
Figure 7: Research Framework 35
Figure 8: Research Flow 39
Figure 9: Data Scientist Competency Model 59
Figure 10: Data Scientist Competency Model 81

LIST OF ABBREVIATIONS

IT	Information Technology
BI	Business Intelligence
DS	Data Scientist
SQL	Standard Query Language
AI	Artificial Intelligence
IS	Information Systems
BI&A	Business Intelligence and Analytics

ACKNOWLEDGEMENTS

I would like to thank my wonderful husband Riaz for taking care of our little ones Mohamed Rayyan and Zara Safia while I was trying to complete this dissertation and for providing me with support and love during this period. I would like to thank my children who were my motivation throughout this time. I would like to thank my father and my sister for motivating me to consider doing my Masters which is something I have always wanted to do. I would like to thank my mother for always being there for me and for providing unconditional love and support. I would also like to thank my parents in law for all their support during this period.

I would like to thank my supervisor Susan for all help, advice and wisdom which led to the completion of this dissertation. Your guidance has helped shape this piece of work immensely. I am truly grateful to have had you as a supervisor and really appreciate all that you have done over the past year.

I would like to thank my lecturer Judy for the course on Trends in IT which opened my world to the endless possibilities of technologies and the future. This course was a stepping stone which led to my dissertation topic. I would like to thank lecturers Jean Marie, Jason and Mitch for the course on Research Methods and Advancements in IS research which were very enlightening and also for the support and advice given over the past 2 years.

I would also like to thank my friends, colleagues and my manager for all the support and encouragement given during this time.

ABSTRACT

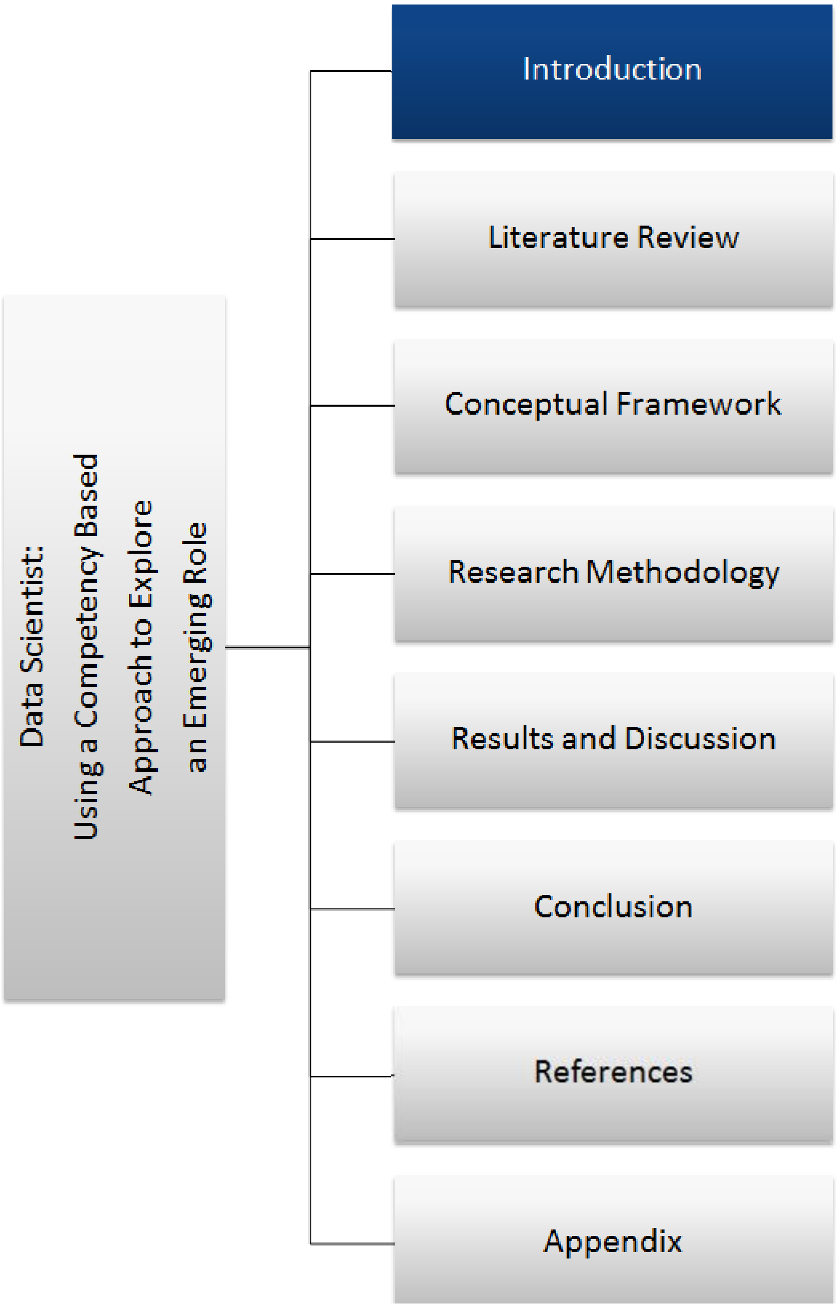
Purpose: The aim in this research study was to explore the role and competencies of Data Scientists in South Africa as the role starts to emerge. Due to the newness of the role, jobs in this sphere are currently being filled by skilled professionals moving from other related areas. Knowledge and skills for Data Scientists were explored in order to examine the role of a Data Scientist and the competencies they should have.

Design/methodology/approach: The studies that have been published on the role of a Data Scientist are limited as the field of Data Science is still new. Therefore the design of the research was exploratory and used qualitative methods. Data gathered for this research was analysed using thematic analysis. The study used respondents drawn from the banking and insurance industries as they are amongst the first to employ Data Scientists in the real sense of the term in South Africa. Six Data Scientists were interviewed.

Originality/value: Research that focuses on the role of Data Scientists especially in South Africa is limited as most of the research has taken place in developed countries. There is also limited research on the role of a Data Scientist within the banking and insurance industry. This study contributes to practitioner and research knowledge by exploring the emerging role of a Data Scientist in the South African context.

Practical implications: This research improves our understanding of the knowledge and skills Data Scientists should have within the banking and insurance industry. This research adds insight by highlighting the role that Data Scientists are currently undertaking by providing information on the specific skills that they report as required. This research can help in the shaping of education and developing the required skills for individuals who intend to pursue the career path of a Data Scientist as well as help managers hire the right people for the position of a Data Scientist.

Keywords: *Data Scientist, Data Science Competencies, Competency Models, Data Scientist Skills*


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

1 INTRODUCTION

1.1 Background

Large amounts of data are being collected, stored and analysed by organisations which is now referred to as Big Data (Chatfield, Shlemoon, Redublado & Rahman, 2014; Zikopoulos, Parasuraman, Deutsch, Giles & Corrigan, 2012). Big Data can be defined as sets of data that are large and complex and cannot be processed and analysed by traditional technologies. Big Data is characterised by the 3 V's which are defined as the rise in **V**olume of the data, the **V**elocity with which the data arrives, and the **V**ariety of different forms the data takes (Power, 2014). Big Data can provide insights which can transform the way we look at the world by changing how organisations process, store and use data (Abbasi, Sarker & Chiang, 2016; Cai & Zhu, 2015). Organisations understand that in order to be successful and gain a competitive edge against competitors, the importance of studying customer behaviour to meet their goals is vital (Halaweh & Massry, 2015). This has reformed the relationships that organisations have with customers and has enabled a greater move towards customer centricity (De Mauro, Greco & Grimaldi, 2015; Wamba, Akter, Edwards, Chopin & Gnanzou, 2015). Big Data can be used to achieve this demand and increase profits especially in competitive environments (Trifu & Ivan, 2014; Davis, 2014).

Organisations are putting people and processes in place to take advantage of Big Data opportunities (Chen, Chiang & Storey, 2012). Value from Big Data is derived through the use of Analytics (Watson, 2014). By providing Analytics, organisations can extract meaning, hidden patterns, surprising correlations and unexpected connections that can add value to the business (Ozkose, Ari, & Gencer, 2015; Vahn, 2014). The use of Big Data is responsible for adding to the increasing demand for people who can use and **analyse** Big Data which is producing a new set of Analytics and skills (Chen et al., 2012; De Mauro et al., 2015; Power, 2014). The person who is tasked with the role of analysing Big Data is known as the **Data Scientist** (Chen et al., 2012).

A **Data Scientist** is able to manage the lifecycle of data and is an expert who extracts meaningful value from Big Data. A Data Scientist is able to bridge the communication gap between business and Information Technology (IT), put forward proposals to measure value, model data, visualise the solution and automate the process (Chatfield et al., 2014; Kim, Zimmermann, DeLine & Begel, 2016).

Big Data is creating a large number of jobs and changing existing ones. Gartner predicted in 2012 that around 4.4 million IT jobs would be created worldwide by 2015 in order to meet demands to use Big Data (Patil & Davenport, 2012). Due to the growth of Big Data environments, the demand for Data Scientists has increased significantly (Gupta, Goul & Dinter, 2015). McKinsey Global Institute did a study in 2011 and made a prediction that in the United States there would be a shortage of up to 190,000 people with skills required for Big Data analysis by 2018 (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, 2011).

Currently in South Africa, there is a low rate of adopting Big Data due to the shortage of the skills required to work with this data (van Zyl, 2014). This was confirmed by the Head of Analytics at a leading financial institution who stated that one reason for the low rate of adopting Big Data in the banking and insurance industry is due to the shortage of Data Scientists, a role which requires people with the correct Statistical and Mathematical capabilities, to turn data into insights and actions through the use of Statistical models (van Zyl, 2014). Although there is currently a low rate of adoption, the banking and insurance industries with the Data Scientists currently employed, are leading in advancements in technology and increasing rates of customer satisfaction through the use of Big Data analysis in South Africa when compared to other industries (van Zyl, 2014).

1.2 Problem Statement

The adoption of Big Data creates a requirement for specialist analytic skills different to those used in Business Intelligence (BI) (Debortoli, Müller & vom Brocke, 2014). This has resulted in the emergence of the role of a Data Scientist (Chen et al., 2012; Chatfield et al., 2014; Wixom, Ariyachandra, Douglas, Goul, Gupta, Iyer, Kulkarni, Mooney, Phillips-Wren & Turetken, 2014). Data Science is a discipline that is not yet as well defined as professions such as Business Analysis. Organisations are having difficulty finding and developing people with the appropriate skill and experience (Chen et al., 2012; Chatfield et al., 2014; Demchenko, Belloum, Los, Wiktorski, Manieri, Brocks, Becker, Heutelbeck, Hemmje & Brewer, 2016). A further complication facing organisations is the uncertainty around the required skillset to look for when hiring Data Scientists.

Due to the emerging adoption of Big Data in the banking and insurance industries in South Africa, an associated need for the knowledge and skills necessary to analyse this valuable information has increased (Moyo, 2015). As often typical with an emerging economy, South Africa lags somewhat behind the developed world both in adopting technology and developing the associated knowledge and skills (Mamabolo, 2017; Reddy, Bhorat, Powell, Visser & Arends, 2016). According to the Africa Data Forum, there is a shortage of skills in Data Science caused by the increasing demand for Analytics for decision making by organisations (Mamabolo, 2017). According to Wixom et al. (2014), in order for organisations to attract the right talented people, the **role** and required **competencies** of an effective Data Scientist need to be clarified. In order to reduce the gap between skills required and existing Data Science skills, competencies for a Data Scientist needed to be defined.

The need for research to determine the specialised set of skills and competencies thus arose which could potentially help South African organisations take advantage of Big Data opportunities (Isik, Jones & Sidorova, 2013; Wixom et al., 2014). This study therefore examined the knowledge and skills of Data Scientists within the banking and insurance industry in South Africa to identify competencies for a Data Scientist. In comparison to other industries both in South Africa and globally, the banking and insurance industry has been at the forefront of innovation and technology adoption (van Zyl, 2014). The products and offerings using Big Data that have been released into the South African market are from the leading organisations within the banking and insurance industry namely ABSA with its tool Barclays Pronto, Nedbank with its tool Market Edge, and Discovery with its tool to reduce fraudulent claims which has saved the company over 25 million dollars (ABSA, 2015; Malinga, 2016). This research was therefore focussed on the roles and competencies of Data Scientists within organisations in the banking and insurance industry.

Research Questions:

The following questions were explored in this study:

- *What is the emerging role of the Data Scientist in South Africa (in the banking and insurance industries)?*
- *What skills and competencies do Data Scientists currently working in South Africa possess?*
- *What is the gap (if any) between required and current Data Scientist skills and competencies?*

Research Objectives:

Based on the research questions above, the objectives of this research were defined as follows:

1. To describe the emerging role of a Data Scientist as demonstrated by respondents working in those roles (within the banking and insurance industry in South Africa).
2. To explore the skills and competencies of Data Scientists (within the banking and insurance industry in South Africa).
3. To determine the nature of any existing gap between required and current skills and competencies as reported.
4. To highlight recommendations by Data Scientists relating to education, training and development.

1.3 Purpose of the Study

Due to the growing adoption of Big Data in South Africa (particularly in banking and insurance industries), there is an associated need for the skills necessary to analyse this valuable information. Thus far, anecdotal evidence suggests Data Scientists working in South Africa are generally not formally trained as Data Scientists, but have migrated from other related fields in Information Systems/Information Technology or from Mathematical and Statistical backgrounds.

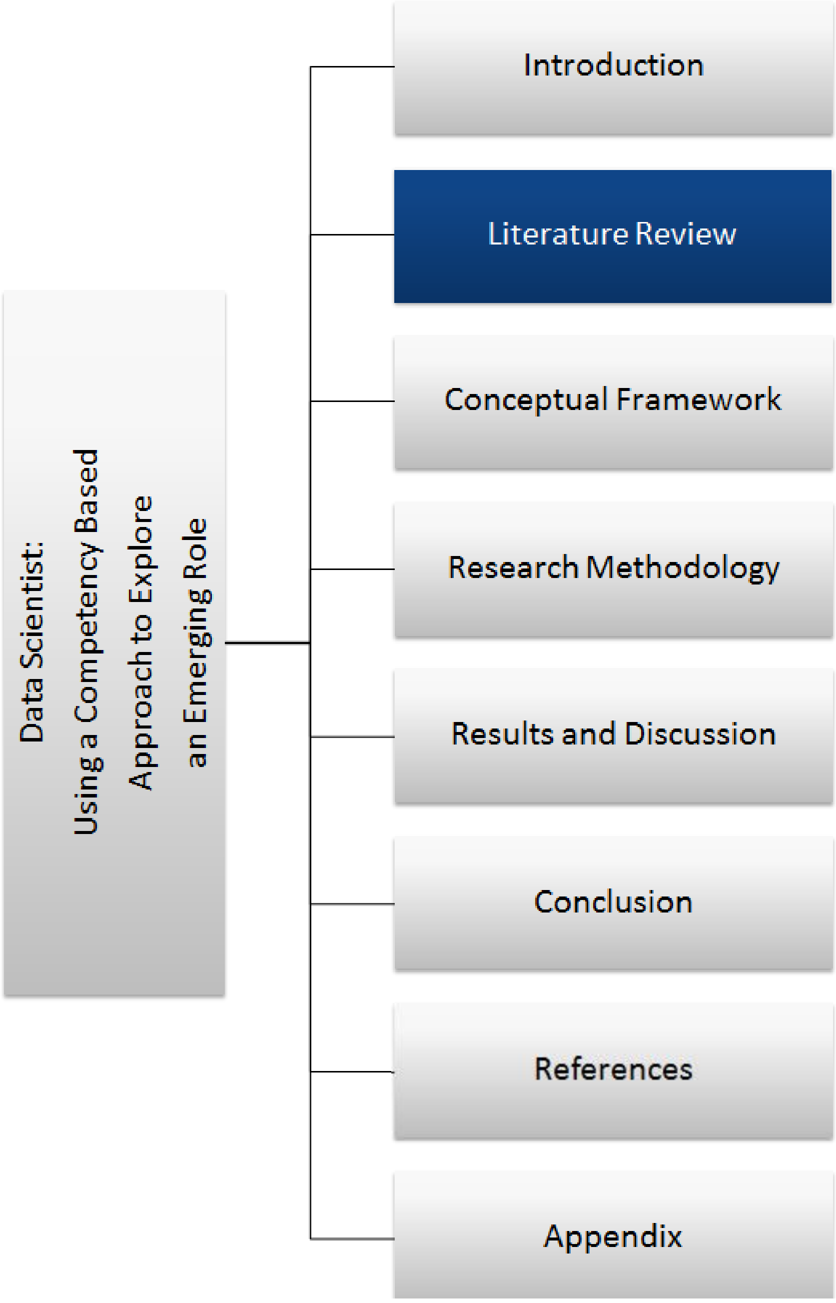
The purpose of this study was to explore the role and competencies of Data Scientists in South Africa, given both the newness of the role and the limited number of established formal educational programmes in this country. While international studies exist (Demchenko et al., 2016; Manieri, Brewer, Riestra, Demchenko, Hemmje, Wiktorski, Ferrari & Frey, 2015) which detail the competencies required for a Data Scientist and the skills they should have, this study looked to extend the existing research by using analysis of local data to develop an understanding of the skills and competencies for Data Scientists that is relevant in the South African job market.

1.4 Intended Contributions of this Study

- The research adds to the emerging Data Science body of knowledge in South Africa by examining the roles and competencies of Data Scientists working in South Africa.
- The findings of this research will help benefit organisations in their quest to hire Data Scientists, as many organisations are having difficulty obtaining the person deemed a Data Scientist. This could lead to an increase in adoption of Big Data technologies.
- The Data Scientist competencies identified in this research can provide guidance in the development of skills and in education for those wanting to pursue a career in Data Science.
- Educational institutions can specifically cater to these needs by offering degrees based on the role and competencies examined in this research which could potentially decrease the skill shortages of Data Scientists that South Africa faces.

1.5 Limitations of this Study

The initial sampling size to be interviewed was 10 Data Scientists taking into account the small number of formal Data Scientist positions currently in South Africa. The researcher managed to gain access to 6 Data Scientists for this study. Another limitation is the possible skills transferability of Data Scientists within different industries. Banking and insurance industries were chosen as they have been at forefront of innovation and technology in comparison to other industries in South Africa (van Zyl, 2014) and the role and competencies examined in this study might not fully overlap with those of a Data Scientist within a different industry.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

2 LITERATURE REVIEW

2.1 Introduction

This chapter looks at the role of the Data Scientist as well as the knowledge and skills that a Data Scientists should have by examining the literature. Gaps in skills that a Data Scientist should have are discussed. Competencies are then discussed as a way in which the knowledge and skills can be defined for a role. Competency models and the competency life cycles are discussed as way to define the Data Scientist role.

2.2 Data Scientist Role

Even though Data Scientist is seen as the newest job title in the technology field according to Patil and Davenport (2012), there is limited clarity on what the role and competency of a Data Scientist is in the literature (Chatfield et al., 2014; Watson, 2014). Many companies have created definitions on what the role of a Data Scientist is, based on the organisation's needs (Kim & Lee, 2016; Kim et al., 2016).

The term 'Data Scientist' was mentioned in 2008 by DJ Patil who was the first chief Data Scientist hired in the White House and J Hammerbacher (Patil & Davenport, 2012). The role of the Data Scientist is used to refer to people who work with Big Data problems. According to Wixom et al. (2014), Data Scientists should have a unique blend of skills. They should solve problems by interpreting and using data to create information and then use their Communication skills to persuade business to apply that information when making decisions.

The role performed by Data Scientists according to Song and Zhu (2016) is detailed below:

1. A focus on extracting information from the data that can be actioned to solve problems (Chen et al., 2012; Chiang, Goes & Stohr, 2012; Dhar, 2013; Carter & Sholler, 2016; Dinter, Douglas, Chiang, Mari, Ram & Schoder, 2015).
2. The ability to ask the right questions keeping in mind the goals of business and the ability to evaluate these questions against the findings (Abbasi et al., 2016; Chatfield et al., 2014; McAfee, Brynjolfsson, Davenport, Patil & Barton, 2012; Provost & Fawcett, 2013).
3. Identifying the correct requirements (Davenport, 2012).
4. Locating and working with pertinent data (Chen et al., 2012; Chiang et al., 2012; Dhar, 2013; Carter & Sholler, 2016; Dinter et al., 2015).
5. Using advanced technologies (Patil & Davenport, 2012; Wixom et al., 2014)
6. Ability to work with subject matter experts (Patil & Davenport, 2012).

7. Ability to analyse, visualise and describe results and problems (Chen et al., 2012; Chiang et al., 2012; Dhar, 2013; Carter & Sholler, 2016; Dinter et al., 2015).
8. Be able to automate data to drive the making of informed decisions (Patil & Davenport, 2012; Wixom et al., 2014).

These skills enable Data Scientists to answer questions relating to large datasets to make informed decisions (Abbasi et al., 2016; Chatfield et al., 2014; Kim et al., 2016; McAfee et al., 2012). By designing experiments and sampling methodologies in large data sets, Data Scientists can turn data and information from large data sets into meaningful and actionable insights and communicate results (Chatfield et al., 2014; Shah, 2017). They are skilled in using and exploring Information Technology, and have the necessary training to use advanced technologies and unusual data sources with predictive Analytics to gain insights and make discoveries (Abbasi et al., 2016; Chatfield et al., 2014; McAfee et al., 2012; Provost & Fawcett, 2013).

Examples of actionable insights derived through analytics are (Abbasi et al., 2016; Chatfield et al., 2014; Kim et al., 2016; McAfee et al., 2012):

1. Analyse all units of stock to derive optimal prices which lead to increased profits
2. Analyse social media data to determine changes in demand and new market trends
3. Identify top customers using different statistical methods
4. Determine behaviour that is fraudulent by using clickstream analysis
5. Mitigating risk by being able to automatically recalculate risk portfolios
6. Analyse customer information to determine new strategies to acquire and retain customers.

Themes were identified in the literature review (Appendix A) and shown in Appendix B. Table 1 shows the themes that were identified.

Table 1: Literature Review Data Scientist Themes

Theme From Literature	Source	No. in Appendix A
Answers questions in large datasets to make informed decisions	Abbasi et al. (2016) Chatfield et al. (2014) McAfee et al. (2012) Provost & Fawcett (2013)	1, 4, 12, 15
Designing experiments and sampling methodologies in large data sets	Brynjolfsson et al. (2011) Davis (2014)	2, 7,
Turn data and information into meaningful and actionable insights	Chen et al. (2012) Chiang et al. (2012) Dhar (2013)	5, 6, 9
Turn data and information into meaningful and actionable insights and communicate results	Carter & Sholler (2016) Dinter et al. (2015)	3, 10

Skilled in using and exploring Information Technology	Davenport (2012)	8
Turn data and information from large data sets into meaningful and actionable insights	Kim & Lee (2016) Miller (2014) Waller & Fawcett (2013) Watson (2014)	11, 13, 17, 19
Necessary training and curiosity to make discoveries in Big Data	Patil & Davenport (2012)	14
Using advanced technologies and unusual data sources with predictive Analytics to gain insights	Wixom et al. (2014)	20

Of the 20 definitions of roles gathered from the literature review which were summarised into 8 themes, the prominent theme was “Turn data and information into meaningful and actionable insights” (Chen et al., 2012; Chiang et al., 2012; Dhar, 2013). Two studies added to this theme with the skill of Communication: “Turn data and information into meaningful and actionable insights and communicate results” (Carter & Sholler, 2016; Dinter et al., 2015) Four studies added to the first theme to include large datasets “Turn data and information from large data sets into meaningful and actionable insights” (Kim & Lee, 2016; Miller, 2014; Waller & Fawcett, 2013; Watson, 2014). By combining this theme with the theme “Using advanced technologies and unusual data sources with predictive Analytics to gain insights” found in a study by Wixom et al. (2014) whose research was based on a survey done to gather information from faculty and industry partners, and a theme that had a high number of studies “ Answers questions in large datasets to make informed decisions” (Abbasi et al., 2016; Chatfield et al., 2014; McAfee et al., 2012; Provost & Fawcett, 2013) a definition of a role was summarised:

A Data Scientist with the use of advanced technologies and large structured and unstructured data sources can turn data into meaningful information to answer the right questions to make more informed decisions leading to actionable insights.

2.3 Data Scientist Skills

The skills that a Data Scientist needs have been reviewed in the literature and a full list can be seen in Appendix A. The skills are as follows:

Providing Actionable Insights, Decision Making

A Data Scientist is able to derive actionable insights from the data for the organisation. Actionable insights have been mentioned by Chen et al. (2012), Chiang et al. (2012), Dhar (2013), Dinter et al. (2015), Kim & Lee (2016), Miller (2014), Schoenherr & Speier-Pero (2015), Waller & Fawcett (2013), Wamba et al. (2015), Watson (2014) and Wixom et al. (2014). By providing the insights needed by organisation, decisions can be made more effectively (Abbasi et al., 2016; Chatfield et al., 2014; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015).

Answer Questions Asked and Answer Questions not asked

A Data Scientist has the ability to work with data in such a way that questions that business asks are answered. During this exploration of data, Data Scientists can also discover certain patterns in the data which lead to the possibility of answering questions that have not been asked by the business (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; McAfee et al., 2012; Provost & Fawcett, 2013; Watson, 2014; Chiang et al., 2012; Dinter et al., 2015).

Tools and Programming Languages used to Work with Data

Python and R, Hadoop, SQL, Machine Learning and Artificial intelligence (AI) have been mentioned in the literature as some of the ways in which data is analysed (Abbasi et al., 2016; Davenport, 2012; Schoenherr & Speier-Pero, 2015; Brynjolfsson, Hammerbacher & Stevens, 2011; Dhar, 2013; Kim & Lee, 2016; Miller, 2014; Sagiroglu & Sinanc, 2013). According to an international framework that was created in the United Kingdom, the skills for a Data Scientist that are important are Big Data analysis tools such as Python, Scala, Java, C++, Machine Learning algorithms and Hadoop (Demchenko et al., 2016; Kim et al., 2016; Manieri et al., 2015). A majority of the literature reviewed mentioned the importance and need for programming skills in Data Scientists (Abbasi et al., 2016; Brynjolfsson et al., 2011; Dhar, 2013; Dinter et al., 2015; Kim & Lee, 2016; Lewis, 2017; Patil & Davenport, 2012; Provost & Fawcett, 2013; Watson, 2014; Wixom et al., 2014).

Analytics, Predictive Analytics, Predictive Model, Problem Solving

Data Scientists should have the ability to apply Analytic methods on data such as being able to build predictive and prescriptive models to solve problems the organisation has. Through the use of Analytics, discoveries can be made in the data. Analytics refers to the tools, methods and results to get information from the data and to get knowledge from the information (Abbasi et al., 2016; Chatfield et al., 2014; Chiang et al., 2012; Davenport, 2012; Davis, 2014; Dinter et al., 2015; Kim & Lee, 2016; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014).

System, Business Knowledge and Information Technology

By having system knowledge, business knowledge and knowledge of Information Technology, the Data Scientist is able to provide and develop solutions to the business that are relevant (Brynjolfsson et al., 2011; Carter & Sholler, 2016; Chiang et al., 2012; Davenport, 2012; Abbasi et al., 2016; Chen et al., 2012; Davenport, 2012; Kim & Lee, 2016; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014).

Communication

Data Scientists need to possess the ability to effectively communicate thoughts, ideas and solutions verbally so that business understands and can act on the information provided (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; Chen et al., 2012; Davenport, 2012; Davenport & Dyché, 2013; McAfee et al., 2012; Patil & Davenport, 2012; Wamba et al., 2015; Watson, 2014; Wixom et al., 2014).

Computer Science, Mathematics and Statistics

Data Scientists should have the ability to use Statistical and Mathematical techniques on data to discover new patterns and insights. With the use of Computer Science, solutions can be developed and implemented (Chatfield et al., 2014; Miller, 2014; Carter & Sholler, 2016; Dhar, 2013; Kim & Lee, 2016; Brynjolfsson et al., 2011; McAfee et al., 2012).

Data Management Process

The ability to work with data is an important skill for Data Scientists. The following steps are part of the process in order gain insights which need to take place:

- Data Analysis
  Analysing large sets of data to produce the insights required by business and have an understanding of how the data relates to the organisations outcomes (Davis, 2014; Dhar, 2013; Dinter et al., 2015; Kim et al., 2016; McAfee et al., 2012; Patil & Davenport, 2012; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Watson, 2014)
- Data Management Process
  A Data Scientist has the ability to develop strategies for the management of data which can then be implemented which will be responsible for the collection of data, storage of data, preservation of data and the availability of data so that it can be used in future processes (Cai & Zhu, 2015; Kim et al., 2016; Wixom et al., 2014).
- Data Mining
  The Data Scientist has the ability to extract meaning from raw data through the use Computational and Statistical methods (Cai & Zhu, 2015; Dinter et al., 2015; Kim & Lee, 2016; McAfee et al, 2012; Miller, 2014; Patil and Davenport, 2012; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Watson, 2014).

- Data Sourcing

  Wixom et al. (2014) mentions the use of new and unusual sources of data. The Data Scientist is able to work with data from different sources and is able to bring it together using different data techniques and models to create a solution (Watson, 2014).
- Data Transformation

  A Data Scientist has the ability to transform data by converting data from one format into another to be used in models to derive insights (Cai & Zhu, 2015; Carter & Sholler, 2016; Chatfield et al., 2014; Chen et al., 2012; Chiang et al., 2012; Kim & Lee, 2016)
- Data Visualisation

  The Data Scientist needs to have the ability to visually present the data in such a way that business would be able to easily understand the solution needed (Chatfield et al., 2014; Dinter et al., 2015; McAfee et al, 2012; Schoenherr & Speier-Pero, 2015; Wamba et al., 2015)

2.4 Data Scientist Skills Gap

The literature highlights skills that a Data Scientist should have. In the papers reviewed, many skills that are deemed important and that are rarely found in Data Scientists have been listed in Table 2. The themes in the table 2 which were identified from the skills that are rarely found in Data Scientists were Communications skills and Business Knowledge.

Communication skills were mentioned by Chiang et al. (2012) who states that Data Scientists need to be able to communicate their finding to management. Davenport (2012) adds to this statement by mentioning the need to be able to speak to customers as well. Kim and Lee (2016) further stipulate that Data Scientists need to have good interpersonal skills.

Business Knowledge is referred to as business acumen, domain expertise, business management, business domain expertise, or business domain knowledge or business value in the literature (Chiang et al. 2012; Dinter et al., 2015; Davenport, 2012; Dhar, 2013; McAfee et al., 2012; Miller, 2014; Watson, 2014; Wixom et al., 2014). Business Knowledge was seen as an important skill for the Data Scientist. Business Knowledge is crucial in understanding the problems that business has and whether or not the solutions created by the Data Scientist will be of value (Chiang et al., 2012). The lack of Business Knowledge might result in solutions that are not optimal for business (Wixom et al., 2014; Yin & Kaynak, 2015). Watson (2014) mentions the lack of Business Knowledge in Data Scientists and mentions that Business Knowledge is seen as a rare skill to find in a Data Scientist. Dhar (2013) mentions the need to integrate Computer Science, Engineering and Business Management.

The outcome of this skill would be software and processes that have been created to directly meet the needs of business. It is an important skill but is split in the literature as Business Knowledge and Computer Science and Programming skills (Chiang et al. 2012; Dinter et al., 2015; Davenport, 2012).

Table 2: Data Scientist Skills Gaps

No in Literature Review	Source	Skills not easily found	Theme
5	Chen et al. (2012)	Data Scientists need to interact with business and domain experts	Communication Skills
6	Chiang et al. (2012)	Need for strong communication skills Understand business outcomes of their analysis and be able to communicate findings to management and be able to tell the story of the data.	Communication Skills
8	Davenport (2012)	Business acumen and the ability to communicate with managers and customers	Business Knowledge
9	Dhar (2013)	Integration of computer science, engineering and business management	Business Knowledge
10	Dinter et al. (2015)	Data Scientist story telling technique is difficult to find and teach	Communication Skills
11	Kim & Lee (2016)	Data Scientists should be well rounded with expert skills and interpersonal skills	Communication Skills
12	McAfee et al. (2012)	Domain expertise with data science	Business Knowledge
13	Miller (2014)	Deep and broad skills across both technical and business domains	Business Knowledge
16	Schoenherr & Speier-Pero (2015)	Communication skills	Communication Skills
19	Watson (2014)	Many Data Scientists don't have degrees in business and to be effective they need to understand the industry in which they work - lack of business domain knowledge	Business Knowledge
20	Wixom et al. (2014)	Lack business value of data	Business Knowledge

2.5 Data Science in the Banking and Insurance Industry

The Africa Data Forum has been training Data Scientists in Africa for the past 3 years. Their CEO mentions that the shortage of Data Science skills is known in Africa and that Africa is full of resources that need to be developed in order to meet the demand for Data Scientists (Mamabolo, 2017).

Russom (2011) mentions that one of the main barriers when adopting Big Data and Analytics in an organisation is insufficient staffing and skills for the Data Scientist role.

According to Patil and Davenport (2012), one of the main challenges facing organisations is to find and attract talented people and a way to ensure that they are productive. In South Africa, one of the reasons for the small increase in adopting Big Data is due to the shortage of the skills required to work with this data (van Zyl, 2014).

The Head of Analytics at a leading financial institution recently stated that one of the reasons for the lack of adoption of Big Data within the banking and insurance industry in South Africa is due to shortages of Data Scientists in South Africa which require people to have the correct Statistical and Mathematical capabilities to turn data into insights and actions through the use of Statistical models (van Zyl, 2014). By creating a competency model, the skills needed to be successful in Data Scientist role can be highlighted (McClelland, 1973).

2.6 Competency

Characteristics of people which are related to successful performance are called competencies according to Boyatzis (1998). A competency is the capability to use and apply knowledge and skills, personal characteristics and behaviours to perform work tasks or perform successfully in a certain role (Campion, Fink, Ruggeberg, Carr, Phillips & Odman, 2011). Competencies also highlight abilities and the behaviours that are expected (Campion et al., 2011). Competencies described as "knowledge, skills and attributes" are needed to effectively meet professional demands (Campion et al., 2011). **Knowledge** is information that is based on facts that is required for a specific job. A **skill** is an ability that has been developed and learned through education and training. Skills enable a person to accomplish tasks in the most effective way (Bassellier & Benbasat, 2004). **Attributes** are motives, traits, personal characteristics, values and the way in which people think that affect a person's behaviour (Bassellier & Benbasat, 2004).

Organisations started using competency based approaches in the 1970's. A psychologist from Harvard, David McClelland introduced the concept of competency in 1973 to help the United States information agency selection procedure by not only focusing on intelligence tests and school grades but to also focus on competencies, as intelligence tests and school grades did not predict successful job performance (Campion et al., 2011). In his paper he argues that the tests used to hire people were not effective in measuring whether a person would be successful in a role (McClelland, 1973). The use of the competency based approach has been increasing since its inception (Campion et al., 2011). The competency names used in this research were adopted from different competency libraries such as the Harvard competency library, the INFORMS competencies for a certified professional in Analytics and from the research conducted by Sonteya and Seymour (Harvard University, 2015; Nestler, Levis & Klimack, 2012; Sonteya & Seymour, 2012).

2.7 Competency Models

A competency model is used to identify sets of competencies (the knowledge, skills and attributes) that are related, for a specific occupation or role which are necessary for successful performance in an organisation (Campion et al., 2011).

There have been different approaches to creating competency models such as:

- Developing a model for a particular job. This can be done through interviews and observation (Hawk, Kaiser, Goles, Bullen, Simon, Beath, Gallagher & Frampton, 2012). Hawk et al. (2012) used interviews to compare skills of IT service providers and customers. The categories determined in the studies were technical skills, business domain skills, project management skills and sourcing and buying skills.
- Making changes to an existing competency model for a similar role (Sonteya & Seymour, 2012)
- A model created from competency libraries where competencies can be chosen for a specific role as done by Bassellier and Benbasat (2004) where a questionnaire was used which had pre-defined skills aimed at determining the business competency of IT professionals.

Researchers within Information Systems (IS) have created competency models for IT managers, IT professionals and other roles within IT (Matook & Maruping, 2014). Models were created using online job advertisements, competency skill libraries and interviews. Lee and Lee (2006) analysed over 500 job advertisements to create a skill set library for managers in the IT field. The resulting competency categories were business, system and technical skills. Viittala (2005) used a pyramid form (Figure 1) derived from the iceberg model to visualise the competencies of managers in his study.

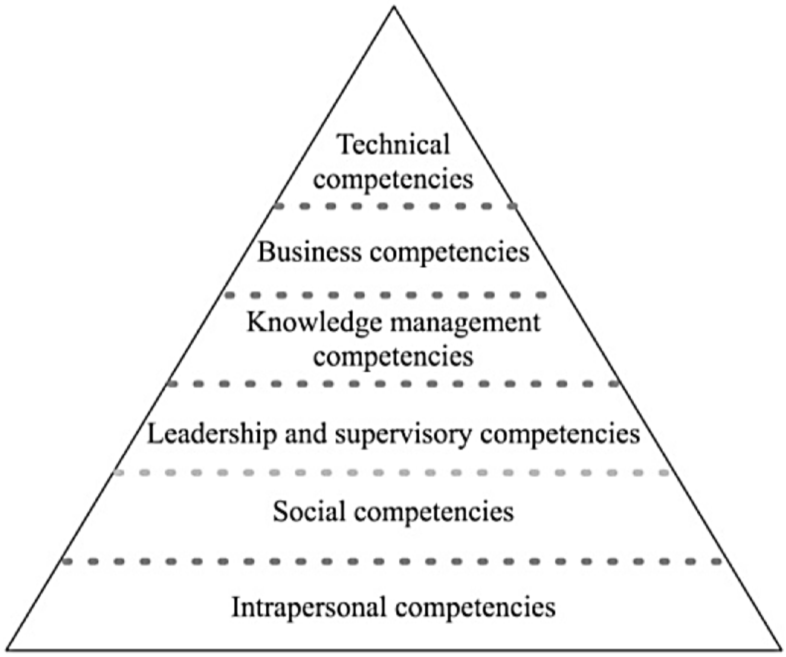


Figure 1: Pyramid view of Competencies (Viitala, 2005)

The model shows the main competency categories for a manager with competencies closer to the top being more education and experience driven which can be more easily obtained rather than the lower level competencies which are behaviour based.

This study used 794 responses from an internet survey conducted to determine development areas of managers in Finland and then to compare these areas with existing manager competency models.

Sonteya and Seymour (2012) used Viitala's (2005) model as a base when they did a study to determine the competencies of a Business Process Analyst. Their study resulted in a competency model of a Business Process Analyst after interviewing 10 people within business process management. Due to the number of themes that had emerged, the Attride-Stirling method was used to group the themes into categories (Sonteya & Seymour (2012). The competency categories are shown in Figure 2. The main categories created were similar to the layers in Viitala's (2005) study.

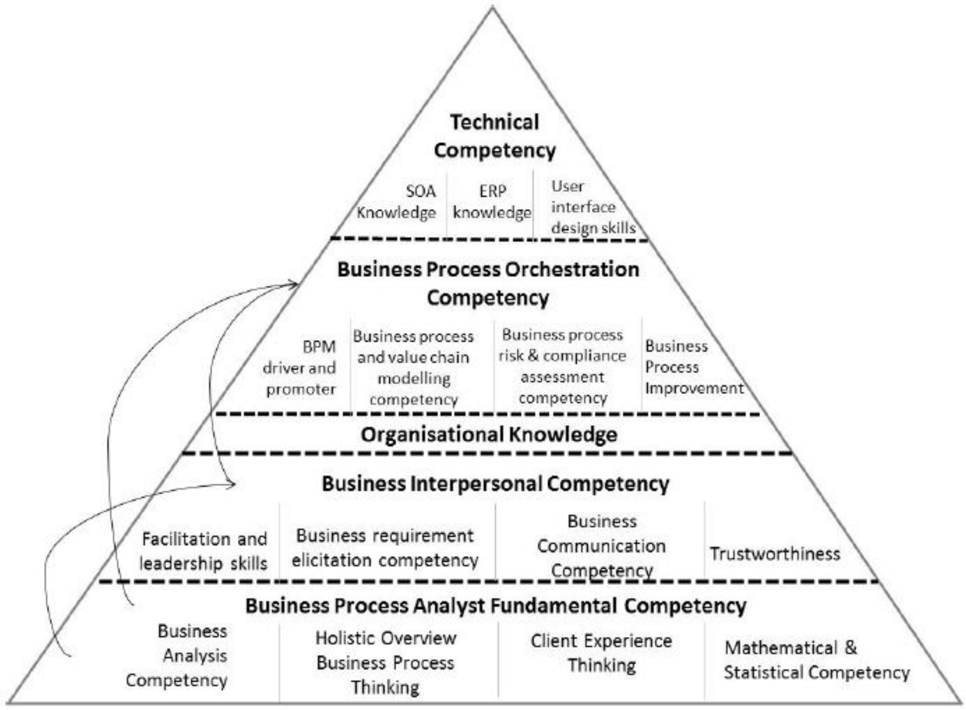


Figure 2: Competency Model of a Business Process Analyst (Sonteya & Seymour, 2012)

Five categories for a Business Process Analyst were created (Sonteya & Seymour (2012). These categories are described below along with the competencies that fall within these levels:

Business Process Analyst Fundamental Competency category (Sonteya & Seymour, 2012)

This category forms the foundation of the Business Process Analyst and contains the following competencies:

- Competency in business analysis (Business Analysis Competency)
- Holistic overview for business process thinking
- Thinking in terms of client needs (Client Experience Thinking)
- Mathematical and Statistical competency

Business Interpersonal Competency category (Sonteya & Seymour, 2012)

This category formed the second layer of the model and is similar to Viitala's (2005) Social competency (Sonteya & Seymour, 2012). It contained the following competencies:

- Trustworthiness
- Business communication competency
- Business requirement elicitation competency
- Facilitation and leadership skills

Organisational Knowledge category (Sonteya & Seymour, 2012)

This category was placed on the same level by Sonteya & Seymour (2012) as knowledge management competencies in Viitala's (2005) model.

Business Process Orchestration Competency category (Sonteya & Seymour, 2012)

This category fell on the same level by Sonteya & Seymour (2012) as business competency in Viitala's (2005) model and includes the following competencies:

- BPM driver and promoter
- Business process and value chain modelling competency
- Business process improvement competency
- Business process risk and compliance assessment competency

Technical Competency category (Sonteya & Seymour, 2012)

This category fell in the same location as Viitala's (2005) model and includes the technical abilities required by a Business Process Analyst.

- SOA (Service Orientated Architecture) knowledge
- ERP (Enterprise Resource Planning) knowledge
- User interface design skills

These categories along with the competencies that fell within each one made up the competency model seen in Figure 2.

Todd, Mckeen and Gallupe (2005) examined the changes in knowledge and skills of Information Systems professionals over 20 years by examining job advertisements. They looked at Information System managers, Programmers and System Analysts. Through the use of content analysis, three categories were created for skills: technical skills, business skills and system knowledge skills.

These categories were further broken down shown in Figure 3:

Class	Category	Description
Technical Knowledge	1. Hardware	Mainframe, mini, and personal computers. Other devices such as storage devices, controllers, printers, and other peripherals plus networks.
	2. Software	Application systems, operating systems, packaged products (such as database, graphics, word processing), networking software and languages.
Business Knowledge	3. Business	Functional expertise (such as finance, marketing) and industry expertise (such as retail, mining).
	4. Management	General management skills including leadership, project management, planning, controlling, training, and organization.
	5. Social	Interpersonal skills, communication skills, personal motivation and ability to work independently.
Systems Knowledge	6. Problem Solving	Creative solutions, quantitative skills, analytical modelling, logical capabilities, deductive/inductive reasoning, innovation.
	7. Development Methodology	Knowledge of systems development methodologies, systems approach, implementation issues, operations and maintenance issues, general development phases, documentation, and analysis/ design tools/techniques.

Figure 3: Skill and Information System knowledge classification (Todd, Mckeen & Gallupe, 2005)

Chen et al. (2012) did a study based on the top journals that had published Business Intelligence and Analytics articles. Using the search term "Business Intelligence, Business Analytics and Big Data", a total number of 3602 articles were found. Business Intelligence and Analytics (BI&A) was structured into three sections namely BI&A 1.0 (information based on a database that has structured data), BI&A 2.0 (data that is web based and unstructured), and BI&A 3.0 (mobile and sensor data). Chen et al. (2012) indicate that BI&A is actually called Data Science in organisations. With this in mind, the skills that surfaced for the BI&A professional in this study were referred to as skills necessary for a Data Scientist. The skills identified in this study are: analytical information technology, communication, domain and business knowledge skills.

Debortoli et al. (2014) compared Business Intelligence skills and Big Data skills. A competency classification for Business Intelligence professionals and Big Data professionals emerged from this study. The analysis of job advertisements from online portals was used to create the models in this study. A method called latent semantic analysis which is a text mining technique that can perform automatic content analysis was used.

The Business Intelligence competencies are displayed in Figure 4. The two higher level competency groups were business and IT (information technology).

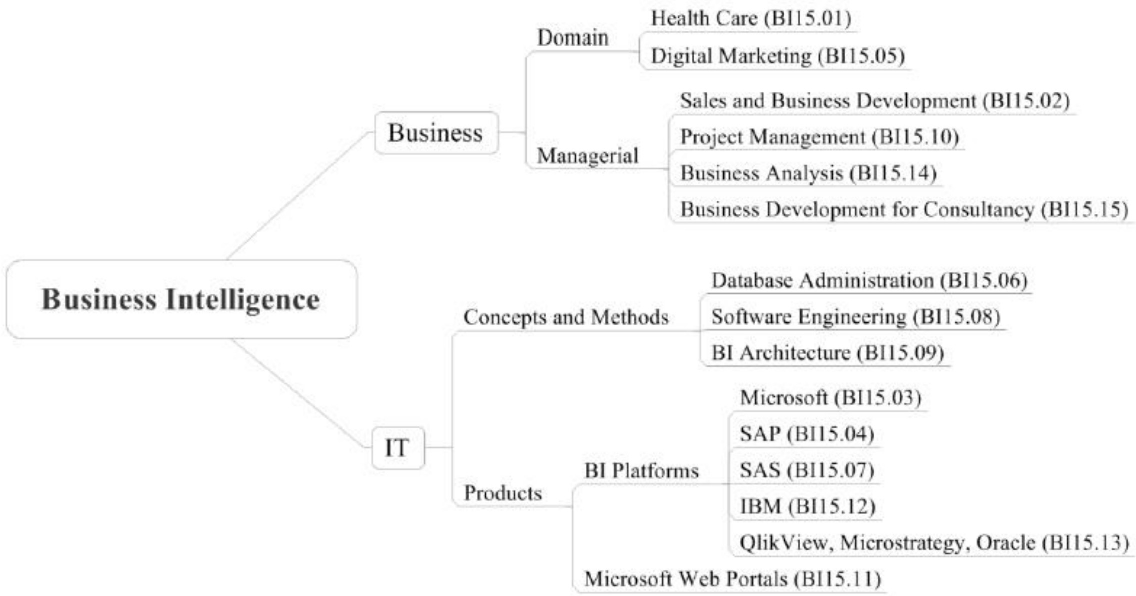


Figure 4: Competency Model of a Business Intelligence Analyst (Debortoli et al., 2014)

The Big Data competencies identified through their analysis are represented in Figure 5.

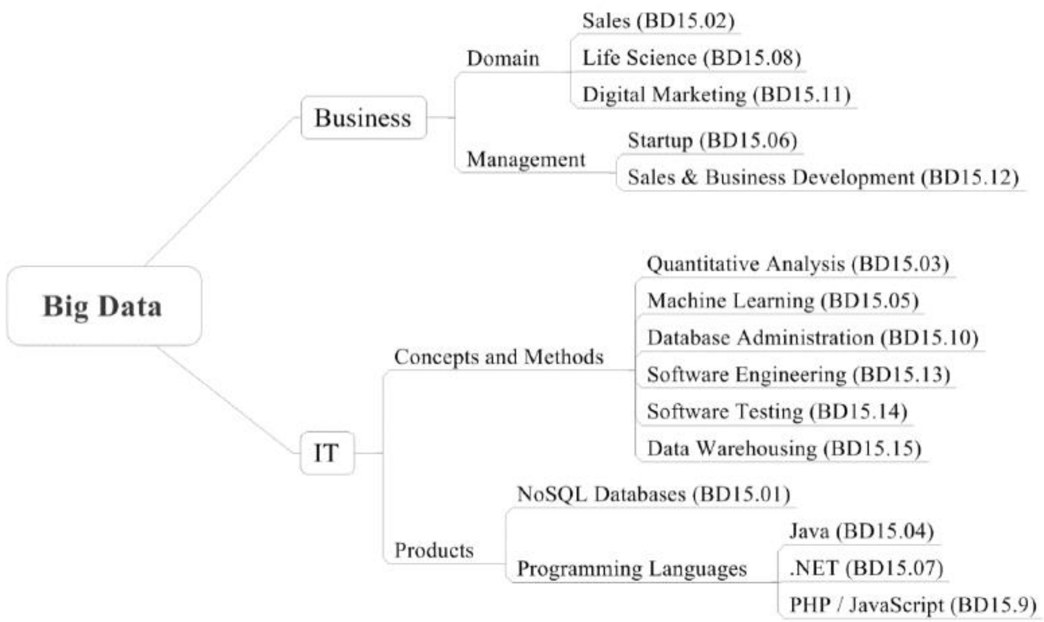


Figure 5: Big Data Professional Competency Model (Debortoli et al., 2014)

2.8 Competency Life Cycle

The competency life cycle was used by Draganidis and Mentzas (2006) as a way to determine competencies of the Data Scientist. Draganidis and Mentzas (2006) defined the competency life cycle as a series of four steps taken in order to identify, develop and continuously enhance an individual's competencies for a particular role. The four steps were defined as:

Step 1: Competency Mapping gives an organisation an outline of all competencies that are required to meet objectives set out by the organisation as well as requirements of job roles (Draganidis & Mentzas, 2006; Campion et al., 2011). This step relates to the first research question on the role of a Data Scientist.

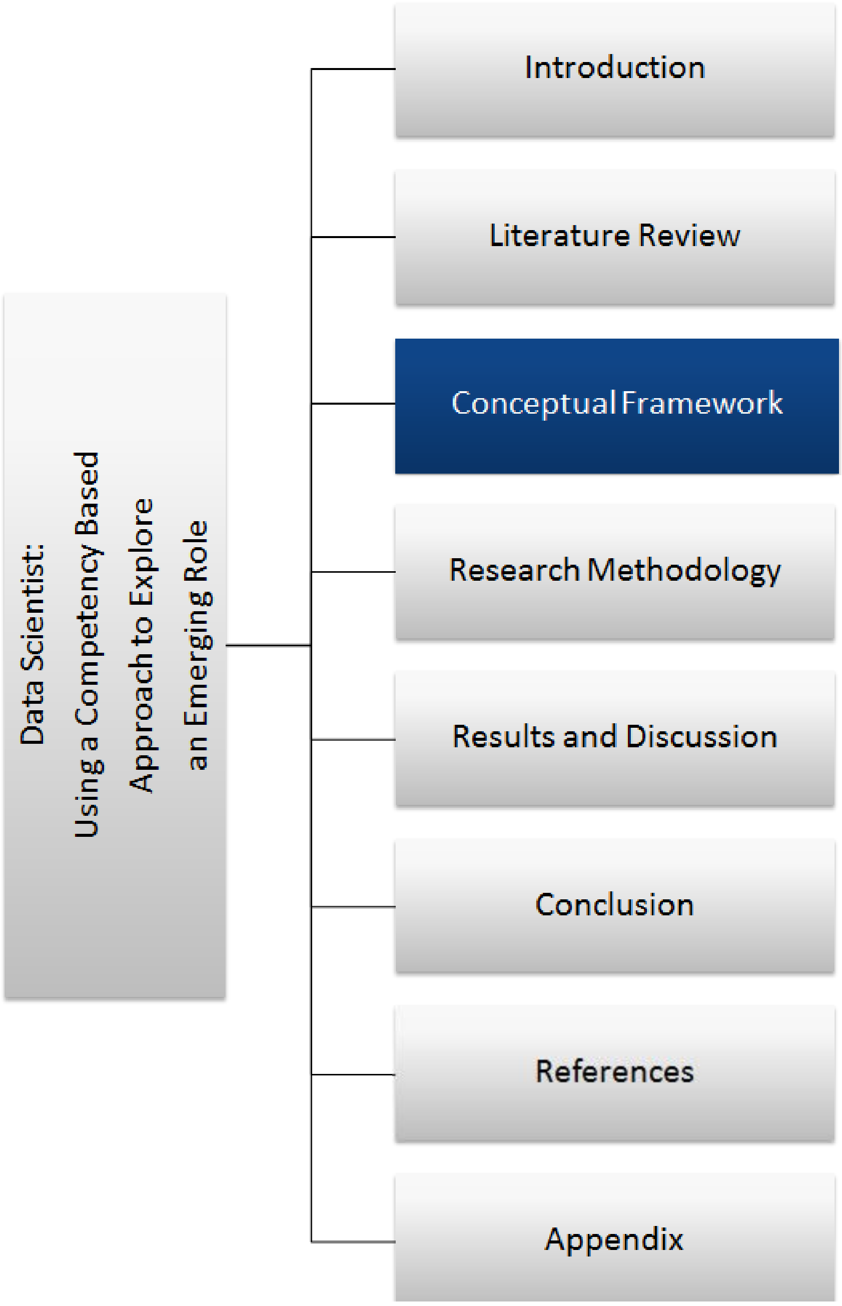
Step 2: Competency Diagnosis provides a view of the current situation of competencies that each individual has. This is the phase where a skills gap analysis can be performed to determine the gap between competencies that are **required** for the role and the **current** competencies that an individual has (Draganidis & Mentzas, 2006). This step relates to the second and third research question on the competencies of a Data Scientist and the gaps that exist between the current and required role.

Step 3: Competency Development is the area that deals with aligning an individual's competencies with the desired level which would have been identified in the first two phases and the skills gap analysis (Draganidis & Mentzas, 2006).

Step 4: Competency Monitoring is the last phase and deals with the monitoring of results from the competency lifecycle (Draganidis & Mentzas, 2006).

2.9 Summary

This chapter discussed what a Data Scientist role is as defined in the literature as well as the associated knowledge and skills of a Data Scientist. The gap in knowledge and skills was also discussed as identified in the literature. Competencies were then defined and discussed according to the different competency models that have been created in the literature. Lastly the competency life cycle was introduced which is one of the ways in which competencies can be identified. As identifying competencies is a way to define knowledge and skills of a particular role, this study focussed on exploring the knowledge and skills of Data Scientists which resulted in the development of competencies for a Data Scientist.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

3 CONCEPTUAL FRAMEWORK

3.1 Introduction

The adoption of Big Data in South Africa (particularly in banking and insurance industries) highlights the need for specialist skills. The aim of this study was to examine the roles and competencies for Data Scientists as an emerging role relevant in South Africa.

Competency models as discussed in the literature review were chosen for this study as they have been used to identify the competencies (skills, knowledge and attributes) necessary to perform specific roles and to enable organisations to identify successful behaviours that can increase the performance of the organisation (Campion et al., 2011; Draganidis & Mentzas, 2006; Debortoli et al., 2014).

For this study, due to the newness of the role, the knowledge and skills of the Data Scientists were examined through interviews to identify the competencies of a Data Scientist. The knowledge and skills of Data Scientists were grouped to derive competencies which were then further grouped into categories created by Sonteya and Seymour in their Business Process Analyst study (2012) to create the competency model.

3.2 Research Framework

The research framework used in this study was based on the model created by Sonteya and Seymour (2012) for categories as shown in Figure 6 below.

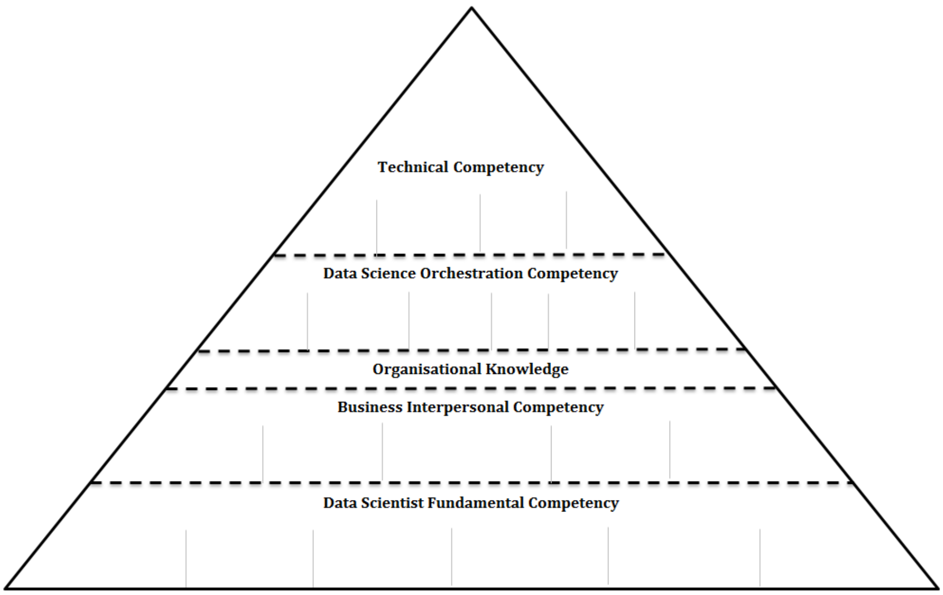


Figure 6: Competency categories

The model created by Sonteya and Seymour was used to determine the competencies of a Business Process Analyst. The categories shown in the model are now discussed as well as the reason for the shape of the model.

Technical Competency category

- This category includes the different technical knowledge and skills required for a Data Scientist. According to the literature review, a Data Scientist needs to be able to use the latest technology and tools to perform his role (Davenport, 2012). Technology knowledge of Hadoop and Machine Learning is important for big sets of data (Abbasi et al., 2016).
- This category also depicts the knowledge and skills that a Data Scientist has specialised in.
- According to Viitala (2005), the higher a category in the triangle, the more closely it is related to education and experience which is the reason for the Technical Competency category to be located at the top.

Data Science Orchestration Competency category

- This Data Science Orchestration Competency category was adapted from the category Business Process Orchestration Competency category in Sonteya and Seymour's research (2012). The Data Science Orchestration Competency category encompasses all the competencies that Data Scientists need to have in order to provide the actions needed to deliver correct and optimal results to business (Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015).
- This category was located below the Technical Competency category as it is also a mix of the experience and knowledge and skills that a Data Scientist has obtained. More experience which cannot be easily obtained provides the Data Scientist with a better understanding of how to extract meaning from the data.

Organisational Knowledge category

- Data Scientists need to possess Organisational Knowledge which refers to the Data Scientist level of knowledge of the all the different processes that occur within the organisation. By understanding the processes, the Data Scientist will have a holistic view which would enable the Data Scientist to provide solutions to cater for the organisation requirements (Kim & Lee, 2016).

- According to Viitala (2005) this layer is a bridge between the cognitive categories which fall in the top two layers of the triangle and the more social and fundamental knowledge categories which are seen in the bottom two layers of the triangle.

Business Interpersonal Competency category

- This category highlights Data Scientists' need to bridge the gap in Communication between business, IT and other departments, and also how they communicate the results of their work to drive solutions within the organisation (Schoenherr & Speier-Pero, 2015).
- Data Scientists need to be able to identify and ask the right questions to gather the correct requirements from business keeping in mind the goals of business (Abbasi et al., 2016; Chatfield et al., 2014; McAfee et al., 2012).
- Presentation and Communication skills are needed for this level (Wixom et al., 2014; McAfee et al., 2012)
- This category falls into the bottom of the triangle as the ability to form and maintain relationships with business or any of the stakeholders is more of a social attribute of a person and is not a skill that can be easily obtained through learning.

Data Scientist Fundamental Competency Category

- This Data Scientist Fundamental Competency category was adapted from the category Business Process Analyst Fundamental Competency category in Sonteya and Seymour's research (2012). This level depicts that a Data Scientist needs to have the capability to extract meaning and insights from the data and be able to work with the data lifecycle understanding the process of how to access data, prepare it and process it (Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015).
- This level identifies the critical basic skills needed by a Data Scientist and should include:
  - Analytic skills - Data Scientists perform Analytics, evaluate, and visualize (Chen et al., 2012; Chiang et al., 2012; Dhar, 2013; Carter & Sholler, 2016; Dinter et al., 2015). Data Scientists also need to be able to work with different analytic models and Predictive Analytics (Miller, 2014; Kim & Lee, 2016; Davenport, 2012).
  - Mathematics and Statistics skills (Dhar, 2013; Miller, 2014; Kim & Lee, 2016).
  - Systems knowledge and skills (Chen et al., 2012; Chiang et al., 2012; Dhar, 2013; Carter & Sholler, 2016; Dinter et al., 2015).

The research framework is depicted in Figure 7 below:

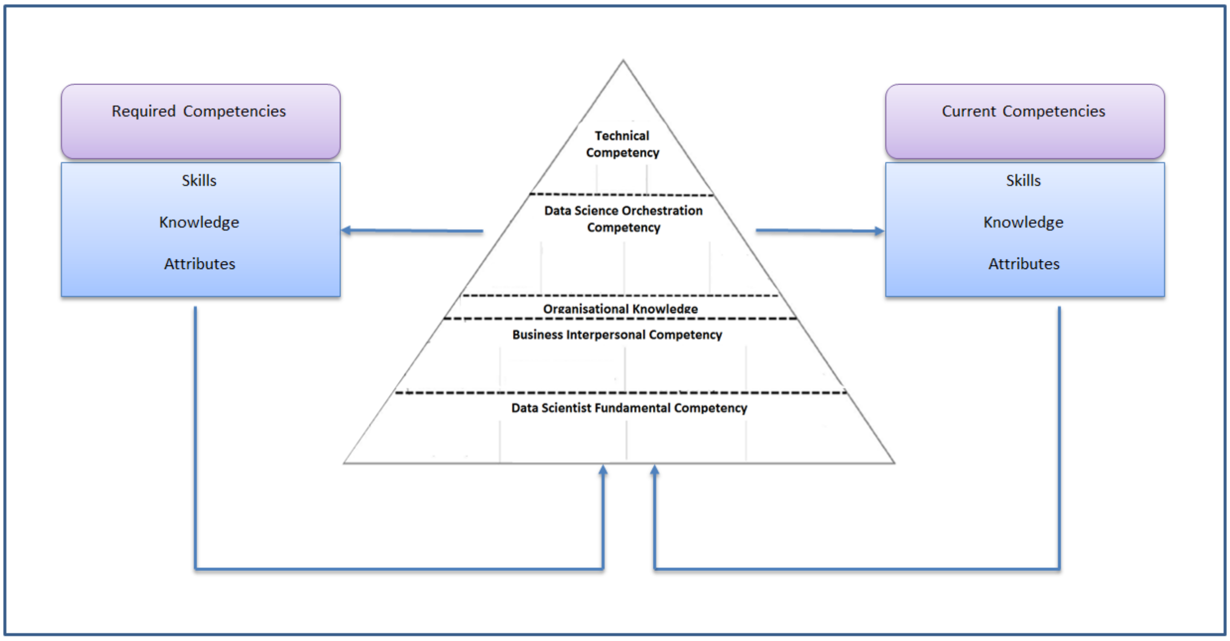
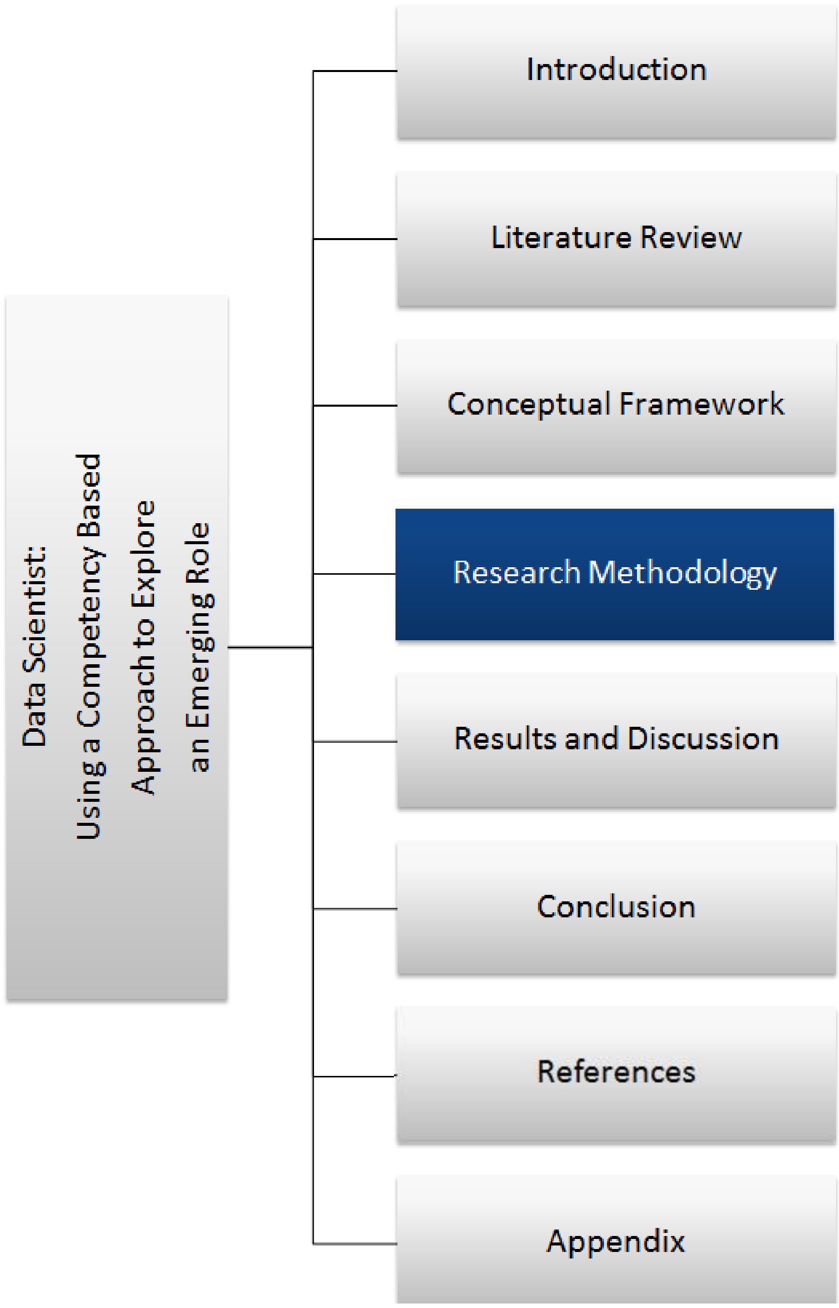


Figure 7: Research Framework

- The literature review resulted in the structure for the competency model as well as the categories that were defined. This was pivotal in determining the questions that were posed to the Data Scientists in the interviews and also important in how the questions were grouped.
- The results and the analysis of the data gathered from the interviews then fed back into the competency model as can be seen by the arrows.

3.3 Summary

This chapter defined the research framework that was used for this study. Based on Sonteya and Seymour's research for a Business Process Analyst, two of the categories were adapted for the Data Scientist role which was the Data Science Orchestration Competency category and the Data Scientist Fundamental Competency category (2012). This chapter also defined each layer of the research framework which is summarised and completed through data gathered from Data Scientists in chapter 5.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

4 RESEARCH METHODOLOGY

4.1 Introduction

This chapter will discuss the research approach and methodology that were chosen to conduct this research. The data collection methods are discussed which includes the sample size and instrument used to collect data. The chapter also discusses how the data was analysed, the ways in which this research was evaluated and ethical considerations.

4.2 Research Approach and Paradigm

Scientific research consists of both theory and observation which are seen as the pillars of research (Bhattacherjee, 2012). The theory level deals with concepts about social phenomena or natural phenomena and the relationships between the phenomena. The observation level, also known as the empirical level, observes reality and tries to see whether the theoretical concepts and relationships reflect reality. The aim of the observation level is to refine and redevelop theories which better reflect the observed reality (Bhattacherjee, 2012).

The two main paradigms or philosophies in research are positivist and interpretivist (Hussey & Hussey, 1997). A paradigm is a collection of beliefs. It is a worldview, the relationship between the individual and the world and the parts that make up the world.

The interpretivist paradigm was selected for this study. The purpose of interpretivism is to provider deeper meaning in the research and to look at areas previously not studied (Walsham, 1995). An interpretive study can provide valuable insights and can help determine research questions which could be used for future studies (Collis & Hussey, 2003; Gioia & Pitre, 1990). Interpretivist methods use an inductive approach which starts by looking at and analysing data and then refining or redeveloping a theory about the research area (Bhattacherjee, 2012). This study was documented at each step so that in future the study can be replicated.

This research was exploratory as the aim was to explore the roles and competencies of Data Scientists and due to this an interpretivist approach provided the deeper understanding needed for this study (Bhattacherjee, 2012). Exploratory research also provides high quality information that could potentially help in determining the prominent issues that could be further investigated (Creswell, 2013).

4.3 Research Methodology and Design

Choosing the right way to design the research is necessary to determine how to collect data and the instrument to be used (Creswell, 2013).

The research design for this study was an exploratory study and according to Bhattacherjee (2012), exploratory research is frequently used in areas that are new and not yet established. Exploratory research design is necessary to gain a better understanding of an area in which little is known. Exploratory research can be used to discover and report relationships among different aspects of phenomena in new areas such as the field of Data Science.

The Data Scientist role is new having emerged alongside Big Data and thus far few studies have aimed at investigating this role. In this study, the aim was an exploration of the role of the Data Scientist to examine the competencies required. As the role and competencies have not yet been fully established in the literature, an exploratory study was chosen to best suit this research area as more information is needed to understand the role of a Data Scientist as well as the competencies they need to possess (Creswell, 2013; Sekaran & Bougie, 2013).

The exploratory research design is characterised by qualitative methods (Saunders, Lewis & Thornhill, 2009). Qualitative methods gather information through the use of case studies, secondary information and interviews for analysis for research that is exploratory (Sekaran & Bougie, 2013). As discussed in the literature review, previous research on the development of competencies also used qualitative methods to gather data which supports the choice of qualitative methods (Sonteya & Seymour, 2012; Hawk et al., 2012).

4.4 Data Collection Methods

In order to determine competencies in previous studies, interpretive qualitative research methods with semi structured interviews were used (Hawk et al., 2012; Sonteya & Seymour, 2012). Interviews were used in this study to explore the role of a Data Scientist as well as the competencies they should have. Interviews enable probing of the research area to get a deeper understanding (Kothari, 2004). This research entailed understanding the role and looking at the required and current skill set to identify competencies of Data Scientists.

The steps for this research are shown in Figure 8:

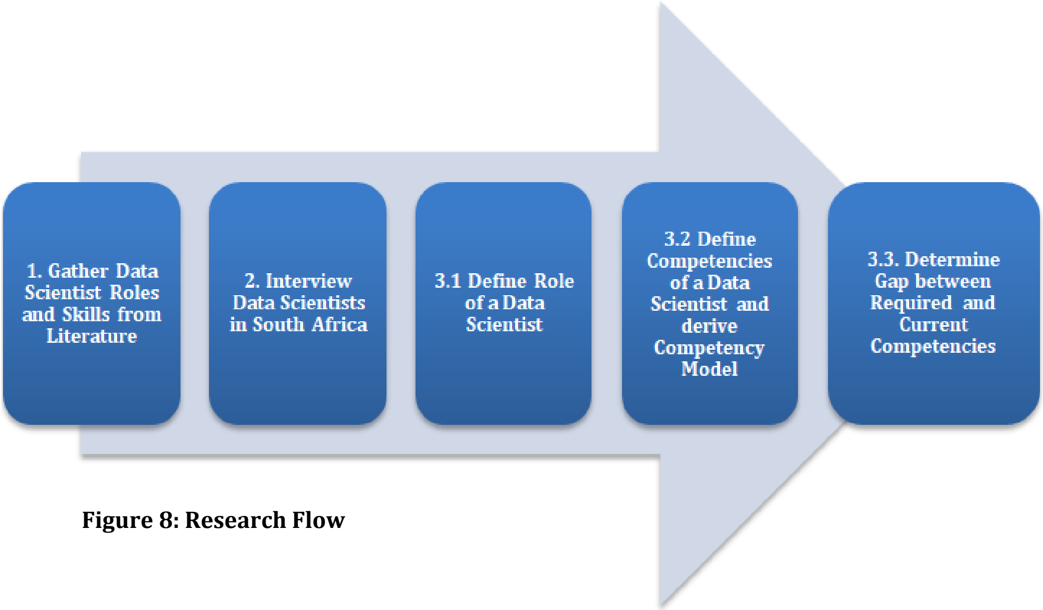


Figure 8: Research Flow

1. Gather Data Scientist Roles and Skills from Literature

The literature was used to determine the roles of Data Scientists in organizations (Kothari, 2004) as seen in Appendix A. Several types of roles were discussed. Themes were identified for the roles shown in Appendix B and then a summarised role was defined. This was the first task that was done in this study to identify the role a Data Scientist fulfils. Knowledge and skills were also discussed and identified in the literature review. The literature review was also mapped to the research framework and can be seen in Appendix C.

2. Interview Data Scientists in South Africa

Categories that were identified in the literature were used to group the interview questions in Appendix H. Data was gathered using semi structured face to face interviews which allowed for unanticipated information. A sample of Data Scientists and people in the role of a Data Scientist within the banking and insurance industry were interviewed. The interviews were between 30 minutes to an hour long. Interviews were recorded with prior consent from the interviewees. Recording the interviews assisted the researcher to ensure accuracy of the data collected during the interview. The purpose of the interviews was to examine the role a Data Scientist plays within the banking and insurance industry and the competencies the Data Scientist would need to have to carry out tasks.

3.1. Define Role of a Data Scientist

Themes were analysed and categorised from the interview data to understand the role of a Data Scientist seen in Appendix K. Themes were then discussed according to the literature and conclusions were drawn.

3.2. Define Competencies of a Data Scientist and derive Competency Model

Competencies were identified from the knowledge and skills gathered in the interviews shown in Appendix L. Competency models were used by Viitala (2005) and Sonteya and Seymour (2012) to map skills, knowledge and behaviours. Viitala (2005) and Sonteya and Seymour (2012) derived the models based on the literature they researched. They then mapped the interview responses to the literature (Viitala, 2005; Sonteya & Seymour, 2012). This method was used to derive the competency model for this study which was based on the model created by Sonteya and Seymour (2012). The competencies which were identified were mapped to the research framework derived from Sonteya and Seymour's research (2012) and discussed in chapter 5.

3.3. Determine Gap between Required and Current Competencies

In order to determine if any gaps in roles, knowledge and skills of interviewees existed, questions in the interview were split between what the interviewees thought was required of a Data Scientist and secondly questions were asked about their current role. This highlighted certain gaps which were then discussed according to the literature. Questions were also asked to the Data Scientist on skills they wish they had which further identified more areas for development.

4.4.1 Sampling and Respondents

A population is a group of organisations, people or events having the same characteristics that the researcher would like to study (Bhattacherjee, 2012; Sekaran & Bougie, 2013). The research population was Data Scientists (or people playing that role) employed in the banking and insurance industry in South Africa. The study used respondents drawn from the banking and insurance industry as they are amongst the first to employ Data Scientists in the real sense of the term in South Africa.

Purposive sampling was used to select a sample of Data Scientists. The knowledge of the respondents and the purpose of the study are applicable to purposive sampling methods (Bhattacherjee, 2012). Purposive sampling was chosen as a sampling method as respondents needed to be Data Scientist or playing the role of a Data Scientist to add value to this research.

Their knowledge, skills and experience added to the depth in this study. Due to the sparse population of Data Scientists within South Africa, contacts within the banking and insurance industry were used to locate Data Scientists. Snowballing was also used by asking people working within the banking and insurance industry if they knew of Data Scientists within their organisation (Bhattacherjee, 2012). The study initially aimed to obtain a sample size of 10 Data Scientists within the banking and insurance industry. Due to the emerging nature of this role, a limited population of Data Scientists exists in South Africa and of this population 6 Data Scientists were interviewed. Fortunately, during the data analysis stage, it emerged that theoretical saturation had been reached during interview number 6. According to Sekaran and Bougie (2013), theoretical saturation is reached when processes (interviews) are repeated and no new data arises.

A profile of all the interviewed Data Scientists is presented in chapter 5 to provide a background on the interviewees in this study.

4.4.2 Interview Design and Administration

In order to design the items for the interview which was used to gather information for this study, a rigorous process was followed to ensure that the questions being asked in the interview most accurately aligned with the research aims (Saunders et al., 2009) as detailed in Appendix H. The Data Scientist Fundamental Competency identifies the critical basic skills needed by a Data Scientist. The following questions were asked:

- What are your thoughts on developing or training people to work in the Big Data Arena?
- Who should we be training to move into those roles? (E.g. Developers, BI Analysts, Statisticians, etc.)
- What type of education ideally should a typical Data Scientist have?
- What type of experience would you value in a Data Scientist?
- In your opinion, which of the following areas are important for a Data Scientist to be educated/ trained in?

Interviews were administered by firstly notifying the Data Scientist of the interview. A meeting time was set up with the Data Scientist. A participation letter was given to the Data Scientist at the start of the interview shown in Appendix E. A brief introduction as to why the current study was being researched was given to the Data Scientist. Participation consent forms shown in Appendix F and permission to record the interview shown in Appendix G were then signed by the Data Scientists. The researcher described what the terms knowledge, skills and attributes as used in the study were to the Data Scientist to ensure that the responses were aligned to the questions. The interview consisted of 12 questions broken down into two sections. The first section was demographic questions in which the respondent was asked their age and gender. The second section was based firstly on the required role and competencies of a Data Scientist and secondly the current role and competencies of the Data Scientist. Within these sections, the questions were grouped according to the 5 competency categories defined in the research framework:

- Technical Competency
- Data Science Orchestration Competency
- Business Interpersonal Competency
- Organisational Knowledge
- Data Scientist Fundamental Competency

4.5 Data Analysis Methods

Thematic analysis was used to analyse data collected from interviews (Creswell, 2013; Silverman, 2015). Creating themes for the data is considered to be as difficult as coding as it requires the reflection on the meaning of the data gathered from interviewees (Saldana, 2016). A theme is a short sentence or an extended phrase which identifies what a certain piece of data is about and what it means (Saldana, 2016). According to Saldana (2016), theming data is applicable to interviews. Boyatzis (1998) mentions that themes can be identified directly from observed information. Several qualitative researchers have recommended the analysing and labelling of certain parts of the data with a themed statement rather than a shorter code and as such theming the data provides a brief profile of the data with a description that follows (Saldana, 2016). It is a way to categorise data into a statement that is reflective of a group of ideas that are repeated. The steps to analyse the data were as follows:

- Interview recordings were transcribed in Microsoft Word shown in Appendix I and then moved to Excel during the initial analysis of data. An example of a recording is "A Data Scientist is someone who extracts value out of data, whether available or derived. They would proactively extract information from various data sources and analyse it for a better understanding of how their business or operations performs. After such an analysis, they would typically aim to automate certain processes within the company, thereby making work easier and more streamlined."
- Comments were then put in to code the data shown in Appendix J. The comments that were taken out from the previous example were: Extracts meaning from Data, Data Sources, Data Analysis and Automate processes.
- Themes and emergent themes were derived from the comments, which were used to combine knowledge, skills and attributes. In the example, the comment Data Analysis

emerged into the Data Analytics and Visualisation theme which formed part of the Data Scientist Fundamental competency.

- The researcher became immersed in the data to identify themes (competencies) that the data contained.
- Thirty four themes were initially identified from the data obtained on the role of Data Scientists and can be seen in Appendix K. These themes were then grouped into eleven themes. An example of the grouping process is the grouping of three initial themes Data Analysis, Data Visualisation and Analytics into the single theme: Data Analytics and Visualisation.
- The competency names were drawn from different competency libraries and previous research (Harvard University, 2015; Nestler, Levis & Klimack, 2012; Sonteya & Seymour, 2012; Strengell, 2017). Due to the specific nature of this study on the role of Data Scientists and the limited studies that have been conducted, some competency names were adapted from original competency names.
- Themes (competencies) that were identified were examined again to map these into the categories that were created in the research framework (Saldana, 2016). The Data Analytics and Visualisation theme was mapped to the Data Scientist Fundamental competency category.
- The final output is shown in Appendix L.
- The role and competencies that were identified in the literature review were also mapped to the research frameworks categories and are shown in Appendix C. This information was used to discuss the findings of this research in relation to literature in chapter 5.

4.6 Research Evaluation Methods

In qualitative studies according to Lincoln and Guba (1985) the following requirements were used to ensure quality of the research to ensure the research procedure is done correctly:

Credibility is how trustworthy and believable the research is (Bhattacherjee, 2012; Creswell, 2013). This study provides thick description on the interviewees in order to increase credibility. Rich descriptions through the creation of profiles for each Data Scientist were created. Triangulation methods are used in this study. Sonteya and Seymour (2012) used interviews to gather data to derive competencies of Business Process Analysts. This study examined competencies from the use of literature on the Data Scientist role as well as from interviews to ensure that triangulation occurs from the data sources which will lead to credibility. The duration of the interviews were noted as well as the number of interviews that took place. Interviews were recorded which allowed for accurate transcription and analysis instead of relying only on notes and interpretations. If the notes are not written up early, there is a chance to misinterpret explanations and to miss key points in the interview and also there is a possibility of mixing up interviews. The notes were written up on the same day as the interview as well as during the interview process which further enhanced the reliability and interpretation of the information to reduce information bias.

Transferability refers to how this research can be transferred to other contexts. The research for Data Scientist within the banking and insurance industry was used as a base to determine competencies of a Data Scientist. Rich descriptions on the context were provided so that future researchers will have enough detail to conduct this research in different contexts. Assumptions and processes were detailed.

Dependability refers to the same phenomenon being investigated by different researchers who come to the same conclusion for example. All steps taken in this research were documented in detail to allow a researcher to repeat the study.

Confirmability refers to how well the study represents the ideas and experiences which can be confirmed by others. This was done by asking interviewees to confirm whether or not they agree with how their results are interpreted. Full details were discussed in the research as well as how all categories were created and conceptualized.

4.7 Ethical Considerations

The ethical considerations that the research followed are stated below. An ethical clearance certificate was attained from the Ethics committee at the university shown in Appendix D.

1. Participation of the interviewees were voluntary and free from harm

Interviewees were made aware that their involvement in this study is voluntary. A Participation letter (Appendix E) was given to each interviewee. The letter informed them that they were free to leave the study at any point without consequences, and that there would be no negative consequences to their decision to participate in the process or not. This is important to prevent the interviewees from feeling threatened to fill in a survey (Golden-Biddle & Locke, 1993; Harrington, 1996). In this study consent forms (Appendix F and Appendix G) were given to each interviewee prior to each interview. This form was filled in by each respondent thus:

- Gaining consent ensured that the researcher did not force the interviewees to participate (Golden-Biddle & Locke, 1993; Harrington, 1996).
- The interviewees did not feel any obligation to complete the interview as there were no consequences for not completing the interview which also ensured that the interviewees had free will (Bhattacherjee, 2012, Saunders et al., 2009).

2. The identity of the interviewees in a study should be kept confidential and anonymous

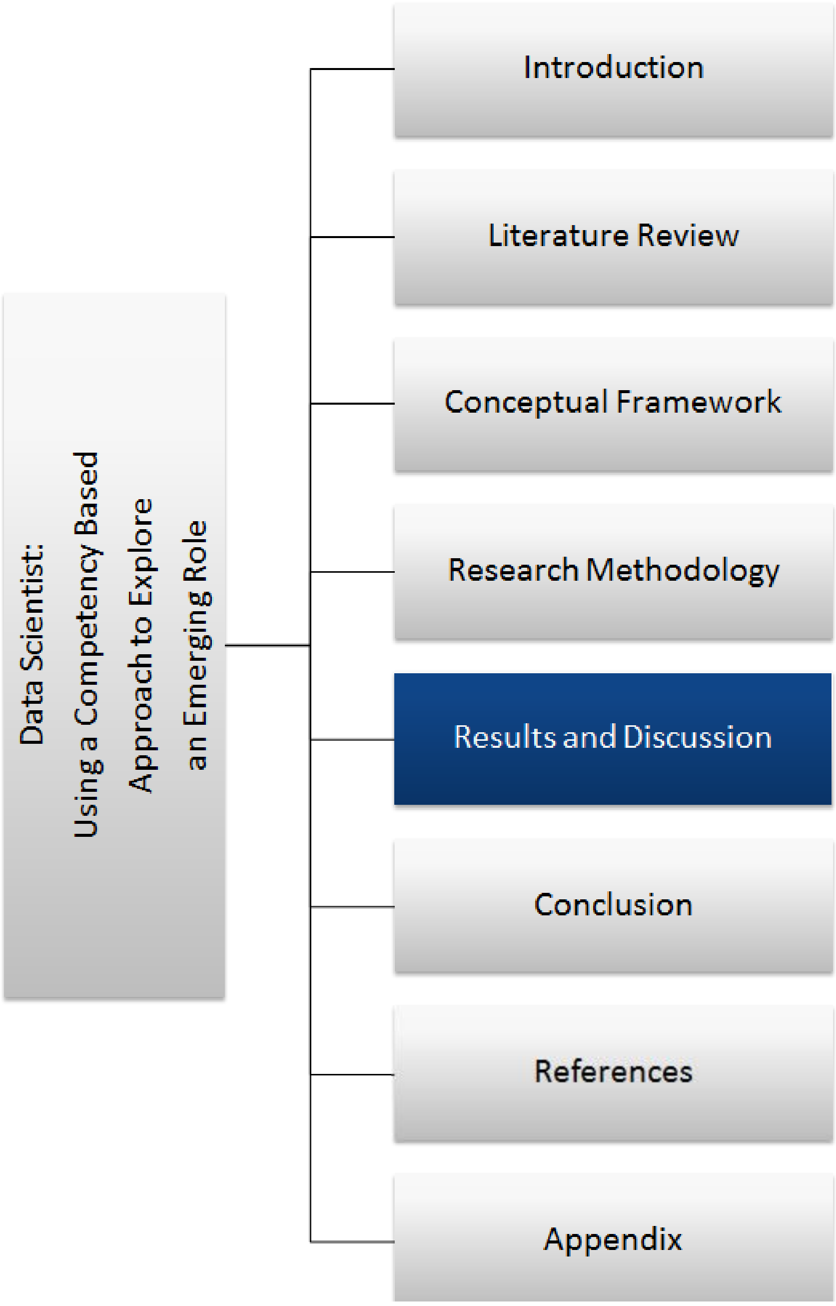
Pseudonyms for the interviewees (e.g. DS 1 is Data Scientist 1) were created to ensure that the identities were kept private and to protect the interviewees in a study. This principle is important to ensure that interviewees feel that they are able to be truthful knowing that the information they provide will remain confidential (Golden-Biddle & Locke, 1993).

3. The researcher should disclose information on the study to the interviewees

Interviewees were made aware of the aim of the research being conducted, potential gains and how it can provide benefits. This information can be beneficial in the participation of the individual in the study (Golden-Biddle & Locke, 1993).

4.8 Summary

This chapter discussed the research methods that this study would follow. To summarise, the research was exploratory as the role of the Data Scientist is new and research on their competencies is limited and not yet fully established. Semi structured interviews were used to gather data which was then analysed using thematic analysis. This chapter also discussed methods to evaluate the research and ethical considerations. The next chapter will discuss the results that were gathered from the interviews and discussed against literature.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

5 RESULTS AND DISCUSSION

5.1 Introduction

The purpose of this research was to explore the role of the Data Scientist and by doing so to identify the competencies they have. The interviewees were asked a range of questions to determine firstly what they perceived was required of a Data Scientist in terms of the both the role and required competencies, and secondly to determine the actual role played and competencies that they currently possess. This enabled the researcher to examine the degree to which the current competencies aligned to the required competencies. Possible recommendations were given by the Data Scientists to aspiring Data Scientists and those employing them on routes into this career and educational requirements.

The first four chapters of this dissertation focussed on the introduction to the problem surrounding the shortage of skilled Data Scientists, the literature reviewed around the Data Scientist role, competencies and competency modelling as well as the research framework and research methods that were used in obtaining data for this research.

This chapter presents the findings that emerged from the data that was collected and analysed using the research framework and methodology presented in chapters 3 and 4. The profile is given of the Data Scientists interviewed in order to give readers a context against which to interpret the findings. Thereafter, the 3 research questions are discussed in relation to the findings obtained from the interviews and the literature on the role of a Data Scientist (research question 1), competencies of a Data Scientist (research question 2) and the gaps identified (research question 3).

5.2 Profiles of the Contributing Data Scientists

This qualitative study was conducted using data which was collected through semi structured interviews. Pseudonyms for the interviewees (e.g. DS 1 is Data Scientist 1) were created to ensure that the identities were kept private.

Three of the Data Scientists were in the 35 - 45 age group. The other three Data Scientists were under 30. Data Scientists interviewed have a minimum of a Bachelor's degree in a field of study with additional courses taken by four of the Data Scientists in the Data Science field. All the Data Scientists have an academic background that contains some level of Mathematics and Statistics. One of the Data Scientists has a degree in Computer Science BSc (Computer Science) and is currently studying towards his honours degree in Computer Science. Two of the Data Scientists have a degree in Mathematical Statistics, one a MSc (Mathematical Statistics) and the other BSc (Mathematical Statistics), one Data Scientist a degree in Actuarial Science - BSc (Actuarial Science), another an honours degree in Computational and Applied Mathematics - BScHons (Computational and Applied Mathematics) and the last an honours in Accounting degree - BComHon (Accounting).

All interviewees were fairly new in the role of the Data Scientist. DS 2 had the most experience having been a Data Scientist for the last 3 years and has a background in Predictive Analytics having completed his Master's degree in Mathematical Statistics. In a Data Scientist hiring guide published as part of an e-book by SAS, 4 Data Scientists were interviewed. The background of the first Data Scientist was a Bachelor of Science in Mathematics and then a Master's degree in Statistics. This Data Scientist employed by SAS started their career as a Research Analyst and moved on to a role as a Statistician and has now been in the role of Data Scientist for 2 years. This background is similar to DS2 having been in a Statistician position first and a Predictive Analyst before becoming a Data Scientist.

DS 1 was initially a Data Analyst having moved through several areas within the organisation in their 15 years in the role and was then hired as a Data Scientist. He has been in the role of a Data Scientist for 1 year and has been working on automating predictive and prescriptive models for the organisation. He is also involved in projects aiming to increase sales for the organisation by investigating data from different sources and providing insight to business.

DS 3 was previously a senior Actuarial Analyst before becoming a Data Scientist which is a position that he has held for 2 years. DS 4 has 20 years of experience in the banking and insurance industry and was instrumental in the creation of several processes aimed to improve business in his previous company. He has been in the role of a Data Scientist for just over a year. DS 5 is new to working environment having recently graduated from university. He is currently employed as a Data Scientist. DS 6 was previously in a senior Data Analytics role and was in the role of a Data Scientist for 1 year before becoming a senior Data Scientist.

It is interesting to note that all the Data Scientists interviewed were male. This may be coincidental, but gender could be looked at in a future study. Table 3 summarises the detailed profiles of the Data Scientists interviewed in this study.

Table 3: Data Scientist Profiles

Profile of Interviewee DS 1	
Age	37
Gender	Male
Job Title	Data Scientist
Educational Background	Bachelor of Computer Science. Studying towards Honours degree in Computer Science as well as Machine Learning and Natural language process courses - BSc Hon (Computer Science)
Experience	15 years' experience as a Data Analyst and 1 year in the role of a Data Scientist
Role and Responsibilities	Source, transform data to answer business questions. Recycle data to answer future directed questions.
Profile of Interviewee DS 2	
Age	27
Gender	Male
Job Title	Predictive Analyst and now a Data Scientist
Educational Background	Masters in Mathematical Statistics, Honours in Mathematical Statistics, BSc in Mathematical Statistics
Experience	3 years' experience as a Data Scientist
Role and Responsibilities	To create and maintain ongoing predictive models for customer behaviour, as well as to perform ad-hoc data-analysis either as a first go on new datasets to go into production or to quickly find answers to questions being asked in business
Profile of Interviewee DS 3	
Age	29
Gender	Male
Job Title	Data Scientist
Educational Background	BSc (Actuarial Science) plus Actuarial exams, Data Science courses
Experience	5 years in the banking and insurance industry as an Actuarial analyst and 2 years as a Data Scientist consulting within the banking industry
Role and Responsibilities	Find marketing levers to improve profitability.
Profile of Interviewee DS 4	
Age	37
Gender	Male
Job Title	Data Scientist
Educational Background	Honours in Computational and Applied Mathematics, BSc in Computational and Applied Mathematics
Experience	20 years' experience in banking and insurance and 1 year as a Data Scientist
Role and Responsibilities	Provide answers to business regarding bonuses. Predictive modelling
Profile of Interviewee DS 5	
Age	25
Gender	Male
Job Title	Data Scientist
Educational Background	Degree in Mathematical Statistics.
Experience	1 year experience
Role and Responsibilities	Development, Business Intelligence, Business and Systems Analysis, Project and Programme Management and Solutions design and Architecture
Profile of Interviewee DS 6	
Age	43
Gender	Male
Job Title	Senior Data Scientist
Educational Background	Honours in Computer Science, Statistics, Bachelor Degree in Accounting and Economics
Experience	20 years' experience in banking and insurance of which 10 years in a Senior position
Role and Responsibilities	Responsible for building & supporting new analytical initiatives as well as operational reporting. Projects focus on machine learning, predictive and classification outcomes in an agile environment using Big data.

5.3 The Role of the Data Scientist

The role of the Data Scientist is explored in this section. Interviewees spoke firstly about their perception or understanding of the role that a Data Scientist is generally required to perform and then spoke of how that description aligned to their current role. According to a description of the role of a Data Scientist posted by MIT (Massachusetts Institute of Technology), Data Scientists are responsible for building models of complex problems, discovering insights and opportunities by using various techniques and tools, and communicating the results (Massachusetts Institute of Technology, 2015). Thirty four themes were identified from the data obtained on the role of Data Scientists and can be seen in Appendix K. These themes were then grouped into eleven themes which are discussed in Section 4.5.

According to the results, the role a Data Scientist plays is becoming critical to the ongoing success of an organisation and is broad in scope. The aspects of the role of the Data Scientist which emerged as fundamental were **Data Analysis**, **Data Sourcing and Tools** and thirdly **Understanding Business Needs**. These three themes are seen as the core role of a Data Scientist and are discussed first followed by the other themes. The various aspects of the role are now discussed through each theme.

5.3.1 Data Analysis

Data Analysis, identified as one of the core roles of a Data Scientist, is the process of getting information from data by using different Analytic methods (Abbasi et al., 2016; Chatfield et al., 2014; Chiang et al., 2012; Davenport, 2012; Davis, 2014; Dinter et al., 2015; Kim & Lee, 2016; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014). Data Analysis was mentioned throughout the interviews. DS 1 mentions that his role is to transform data through Data Analysis to answer business questions. DS 5 mentioned that it is through Data Analysis that the insights are gained. DS 2 mentions that one of his roles is to perform Data Analysis on new sets of data. Without the ability to analyse data which is a fundamental role of the Data Scientist according to Kim & Lee (2016), the Data Scientist would not be able to able to answer business questions, provide insights or perform any of the other roles such as improving sales and customising products. Data Analysis would therefore be expected to feature as a core competency.

5.3.2 Data Sourcing and Tools

Data Sourcing and Tools, identified as one of the core roles of a Data Scientist, is the process to source and work with data from different sources by using various tools to bring data together using data techniques and models to create a solution (Abbasi et al., 2016; Brynjolfsson et al., 2011; Dhar, 2013; Dinter et al., 2015; Kim & Lee, 2016; Patil & Davenport, 2012; Provost & Fawcett, 2013; Watson, 2014; Wixom et al., 2014). DS 1 states that data sourcing as well as the tools to use to source data is part of his role as a Data Scientist. DS 5 states "I know how to work with different data structures and manipulating them to serve my purpose, as well as how to get data from different sources." DS 2 states that "The range of problems can be broad and often relate to predicting individual customer's behaviour, or finding levers to use to improve sales, but the method of finding the solution is consistently to **look at data and use the necessary tools** (be it statistical or otherwise), to find the solution. This statement highlights the core nature of data sourcing and being able to use the tools through which combining data from different sources can be achieved. DS 4 mentions that complexities arise in the data as the data is from different sources. He also states that a Data Scientist would "proactively extract information from **various data sources** and analyse it for a better understanding of how their business or operations perform." This statement highlights the importance of being able to combine data from different sources in order to understand how all the different operations within a business works to provide meaningful insight.

5.3.3 Understanding Business Needs

Understanding Business Needs has been identified as one of the core roles of a Data Scientist. A Data Scientist by having system knowledge, business knowledge and knowledge of Information Technology, must understand business needs and provide and develop solutions to the business that is relevant (Brynjolfsson et al., 2011; Carter & Sholler, 2016; Chiang et al., 2012; Davenport, 2012; Abbasi et al., 2016; Chen et al., 2012; Davenport, 2012; Kim & Lee, 2016; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014). Data Scientists need to ask business the right questions in order to get the right requirements to deliver a solution that business wants. DS 4 states that "A Data Scientist needs to understand the business' needs and develop analytical tools that meet those objectives. Real value comes from delivering the results that match the actual needs of the bank." A Data Scientist should ask the right questions in order to meet business's needs. DS 5 states that the Data Scientist should engage with different departments regarding their requirements. By doing this, a Data Scientist can gather the actual requirements in order to develop solutions that would meet the objectives of the organisation. These statements from the Data Scientists further reinforce the centrality of Understanding Business Needs to their role.

5.3.4 Find Patterns in Data and Provide Actionable Insights

A Data Scientist is able to derive actionable insights from the data for the organisation. Actionable insights have been mentioned by Chen et al. (2012), Chiang et al. (2012), Dhar (2013), Dinter et al. (2015), Kim & Lee (2016), Miller (2014), Schoenherr & Speier-Pero (2015), Waller & Fawcett (2013), Wamba et al. (2015), Watson (2014) and Wixom et al. (2014). DS 3 states that his role as a Data Scientist is to "Find patterns in the data that can be used to make a positive change in the business." From this statement it can be seen that by finding patterns in the data it can lead to insights for the organisation. He further mentions that the insights that come from analysis can be used to identify issues before they become problems. Data Scientists, by being able to create models using data, are able to discover insights and also identify opportunities and problems for business (Kim & Lee, 2016; Wamba et al., 2015). DS 5 states the role of a Data Scientist is to "Be able to analyse data and bring actionable insights". This statement further establishes the importance of this theme as the insights gathered from the data are used to provide insights that can be actioned by the organisation.

5.3.5 Answer Business Questions

A Data Scientist must work with data in such a way that questions that business asks are answered (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; McAfee et al., 2012; Provost & Fawcett, 2013; Watson, 2014; Chiang et al., 2012; Dinter et al., 2015). Answering Business Questions came through strongly in each interview. The interviewees felt that Data Scientists should be able to answer questions posed to them by the organisation. DS 1 mentions that Data Scientists can help organisations answer questions they have as well as future related questions. According to DS 2, Data Scientists attempt to solve a broad range of business problems by looking at the data around the problems. DS 4 states that real value comes from delivering the results that meet the needs of the business. This theme therefore relates to the role of Data Scientists in solving complex business questions through the answers provided by their output.

5.3.6 Answer Questions not asked

During the exploration of data phase, Data Scientists can discover certain patterns in the data which lead to the possibility of answering questions that have not yet been asked by the business (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; McAfee et al., 2012; Provost & Fawcett, 2013; Watson, 2014; Chiang et al., 2012; Dinter et al., 2015). According to DS 1 "A Data Scientist can help organisations answer questions they have not asked." Data Scientists can bring about new ways of doing tasks in an organisation with the use of their skills and techniques and in the process can make discoveries leading to solving questions that were not even asked previously (Watson, 2014). DS 1 was the only one who mentioned that Data Science could answer questions not asked. This statement although not mentioned by the other interviewees, is in line with the literature which mentions the possibility of discovering patterns in the data that no one had identified.

5.3.7 Automate Business Processes

The Data Scientist needs to identify and subsequently advise on the automation of certain business processes (Davenport & Dyché, 2013; Kim et al., 2016; Watson, 2014). DS 5 stated, "Data Scientist is someone who extracts value out of data, whether available or derived. They would proactively extract information from various data sources and analyse it for a better understanding of how their business or operations perform. After such an analysis, they would typically aim to **automate certain processes** within the company, thereby making work easier and more streamlined."

Data Scientists, with their skills of programming and technical skills (see competency discussions later) can create new processes to automate manual existing ones (Kim et al., 2016). DS 5 further suggests that with the process of automating processes, manual labour within the organisation will be reduced. Processes could run more efficiently saving time and labour. DS 4 supports this by mentioning that the Data Scientist through the **automation of business processes** can save time and labour for the business.

5.3.8 Create and Maintain Predictive Models

Data Scientists by applying their Analytic skills and methods on data are able to build predictive and prescriptive models to solve problems the organisation has (Abbasi et al., 2016; Chatfield et al., 2014; Chiang et al., 2012; Davenport, 2012; Davis, 2014; Dinter et al., 2015; Kim & Lee, 2016; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014). The role of Data Scientists should include creating and maintaining predictive models according to statements made by DS 1, DS 2, DS 4, DS 5 and DS 6. DS 2 mentions that the predictive models that are built for their organisation are used to predict customer behaviour. Being able to predict future outcomes emerges as an important role of the Data Scientist. DS 4 mentions that in order to answer business questions, Data Scientists are building Mathematical and Statistical predictive models. DS 5 spoke of how important his role is in building predictive models. DS 6 mentioned that predictive models can be used to predict future trends. Prescriptive models are also built which can let business know the prescribed steps needed to take for the future predicted outcome. This theme relates to how a Data Scientists can solve complex business problems by building and maintaining predictive models.

5.3.9 Customise Products

The customising of products can emerge as a result of work done by Data Scientists. By examining the information that Data Scientists provide to business, products can be customised to meet customers' needs (Davenport & Dyché, 2013; De Mauro et al., 2015; Wamba et al., 2015). DS 5 states that a Data Scientist should analyse data and bring insights that would help customise banking and insurance products. This statement by DS 5 is in line with the literature. DS 2 mentions that one of his roles is to inform management of how customers are reacting either to new products that are being introduced or products that are being customised, to enable the organisation to predict the outcome of the product. The customising of products as a theme for the Data Scientist role is due to the Data Scientists background of being in the banking and insurance industry and might also be common only for a Data Scientist working in a retail industry or in any industry where a product is supplied to a customer as it is also a theme that arises in the literature (Ziora, 2015).

5.3.10 Enhance and Predict Customer Behaviour and Engagement

Organisations understand that in order to be successful and gain a competitive edge against competitors, the importance of studying customer behaviour to meet the organisations goals is vital (Halaweh & Massry, 2015). This has changed the relationships that organisations have with customers and has enabled a greater move towards customer centricity which can be achieved through the outputs provided by the Data Scientist (Davis, 2014; Davenport & Dyché, 2013; De Mauro et al., 2015; Wamba et al., 2015; Trifu & Ivan, 2014; Ziora, 2015.). According to DS 2 who states "My role is to create and maintain ongoing predictive models for customer behaviour", part of his role is to ensure that management is aware of how customers are behaving or reacting. Through this awareness, the engagement between business and customers can be enhanced to increase sales. DS 2 further establishes the importance of this theme by stating that, "Data Scientists can improve customer interactions with the organisation by using predictive models to identify customers who may have unmet (financial) needs or who would be well suited to certain products by using data from customer traffic through portals to improve automated call-centre experience or website experience which would then result in an enhanced customer experience." This illustrates the critical role of a Data Scientist in being able to enhance and predict customer behaviour and engagement.

5.3.11 Improve Sales and Profitability

Improving Sales and Profitability can emerge as a result of the work done by the Data Scientist. Sales and profitability can be increased through the steps 5.3.4 (Find Patterns in Data and Provide Actionable Insights), 5.3.9 (Customise Products) and, 5.3.10 (Enhance and Predict Customer Behaviour and Engagement) which was discussed under the respective headings in this section. DS 3 states that his role is to help management better understand the source of profitability and how this can be changed. Through analysis of the data and creation of predictive models, sales and profitability can be increased (De Mauro et al., 2015; Trifu & Ivan, 2014).

5.4 Summary of the Role of a Data Scientist

The role of the Data Scientist was discussed in the literature review. Themes that were identified for the role were combined to summarise the role of the Data Scientist according to the literature reviewed. The role of the Data Scientist is then discussed according to the results gathered from the interviews. Themes that were identified from the interview data were combined to summarise the role of a Data Scientist according to the findings of the study. Differences in both summarised roles are discussed.

5.4.1 Summary of the Literature

By examining the role of the Data Scientist as discussed in the literature, what was evident were the different roles a Data Scientist could perform that came through from the various definitions for the role of a Data Scientist. A definition was summarised to encompass all roles described in the literature. This definition was derived by combining the themes that were identified in the literature review shown in Appendix B:

"A Data Scientist with the use of advanced technologies and large structured and unstructured data sources can turn data into meaningful information to answer the right questions to make more informed decisions leading to actionable insights."

5.4.2 Summary of the Interview Data

34 initial themes were identified from the interviews for the role of a Data Scientist and can be seen in Appendix K. These were then grouped into 11 themes as shown in Table 4 which were discussed and described in section 5.3 to provide a description of the role of a Data Scientist.

Table 4 provides a summarised description of the 11 themes discussed which indicates how diverse the role of a Data Scientist is. The core roles have been made bold as those roles are fundamental to a Data Scientist from which the other roles can be achieved.

Table 4: Summarised Themes describing the Role of a Data Scientist

Summarised Themes	**Description**
Data Analysis	Ability to analyse data to answer business questions.
Data Sourcing and Tools	Ability to use data from various sources and transform using different tools to manipulate the data
Understanding Business Needs	Being able to ask business the right questions in order to get the right requirements to deliver what business requires
Find Patterns in Data and Provide Actionable Insights	Being able to find patterns in the data and providing insights to business that can be actioned
Answers Business Questions	Able to answer questions by using various tools and analysis in large datasets to make informed decisions
Answer Questions not Asked	Should be able to show business insights that were not aware of based on the data
Automate Business Processes	Should be able to automate business processes by building models that can be used
Create and Maintain Predictive Models	Ability to create predictive models and maintain to answer business questions
Customise Products	Ability to customise products offered by business based on customer needs
Enhance and Predict Customer Behaviour and Engagement	Building predictive models to enhance customer behaviour and engagement
Improve Sales and Profitability	With the use of predictive model, being able to improve sales and profitability

From the results obtained in this study, a role of a Data Scientist in South Africa is shown below by combining the 11 themes discussed. The core themes that were discussed are placed first in the definition. This definition of a role is applicable to the banking and insurance industry and can be defined as follows:

*A Data Scientist is someone who **understands business needs** and is able to find patterns in the data through the use of **data sourcing and tools** to **analyse data**, automate business processes and answer business questions in large data sets to make informed decisions and provide actionable insight. A Data Scientist also creates and maintains predictive models and is able to answer questions that have not been asked which leads to the customisation of products, enhancement of customer behaviour and engagement and an improvement in sales and profitability.*

5.4.3 Discussion of the Role

The role of a Data Scientist as identified through the themes in the literature emphasised a general overview of a Data Scientist. "Being able to use advanced technologies" came out in the literature review as a role of a Data Scientist. This was similar to Data Sourcing and Tools found in the interview results obtained. Data Scientists according to the literature made use of both structured and unstructured data while the interview results obtained focused more on structured data. DS 1 and DS 5 were the only ones who worked with both structured and unstructured data. Both roles created in the literature and from the interview results obtained in this study aimed to help business make more informed decisions as well as being able to answer questions posed by business. Both roles' output aimed at providing actionable insight for the business.

The additions to the role came through the specification of this study in choosing the banking and insurance industry. The use of predictive models, customisation of products and enhancement of customer behaviour was emphasised in the interview results obtained as these are the aims within the banking and insurance industry when dealing with customers as well as increasing sales and profitability. According to Trifu and Ivan (2014) and Davis (2014) one of the benefits of data which is analysed by Data Scientists is to increase profits especially in competitive environments. The interview results obtained in this research align with the results gathered from the literature for the role of a Data Scientist and further adds to the literature to get to a role that is specific to Data Scientists within the banking and insurance industry by the addition of this statement in the role created using interview data summarised in this study:

Data Scientist also creates and maintains predictive models and is able to answer questions that have not been asked which leads to the customisation of products, enhancement of customer behaviour and engagement and an improvement in sales and profitability.

The role summarised in this study is more specific to the banking and insurance industry which might not be applicable to other industries.

The next section will discuss the competencies which are needed to achieve the role summarised for a Data Scientist.

5.5 Competencies of a Data Scientist

A competency as defined in the literature review is a combination of "knowledge, skills and attributes" which are required to successfully meet work demands (Campion et al., 2011). Following on from the examination of the role of a Data Scientist, the competencies of Data Scientists were explored in this research. Step 1 and 2 of the competency life cycle discussed in the literature review were used to establish what Data Scientist practitioners believe the competencies required for a Data Scientist are and which competencies they currently possess. This was then reviewed against the literature on Data Scientist competencies.

Knowledge, skills and attributes that were gathered in this study were grouped to form **competencies**. Competencies were identified by determining what each "knowledge, skill or attribute" was applicable to. The competency names were adopted from different competency libraries such as the Harvard competency library, the INFORMS competencies for a certified professional in Analytics and from the research conducted by Sonteya and Seymour (Harvard University, 2015; Nestler et al., 2012; Sonteya & Seymour, 2012; Strengell, 2017). These competencies were then grouped according to the five categories that were identified in the research framework used in the study conducted by Sonteya and Seymour (2012). These five categories are:

- Technical Competency category
- Data Science Orchestration Competency category (based on **Business Process Orchestration Competency category)**
- Organisational Knowledge category
- Business Interpersonal Competency category
- Data Scientist Fundamental Competency category (based on **Business Process Analyst Fundamental Competency category)**

The categories and competencies which were identified in this study were then put together to create a competency model with each layer placed according to Sonteya and Seymour's (2012) research.

The Data Scientist competency model presented in Figure 9 shows each category together with the competencies that fall within the category identified within the results and literature. Asterisks have been put in to show the following:

- One asterisk shows competencies that appeared in the literature but were not mentioned by the respondents (*). If the asterisk is preceded with a hash (#), this indicates that a specific knowledge, skill or attribute appeared in the literature but was not mentioned by the respondents, not necessarily the competency as a whole.
- Two asterisks show competencies that were mentioned by respondents but not mentioned in the literature (**). If the asterisks are preceded with a hash (#), this indicates that a specific knowledge, skill or attribute was mentioned by the respondents but was not mentioned in the literature.
- Three asterisks shows the gaps between required and current competencies of the Data Scientists identified by respondents (***)

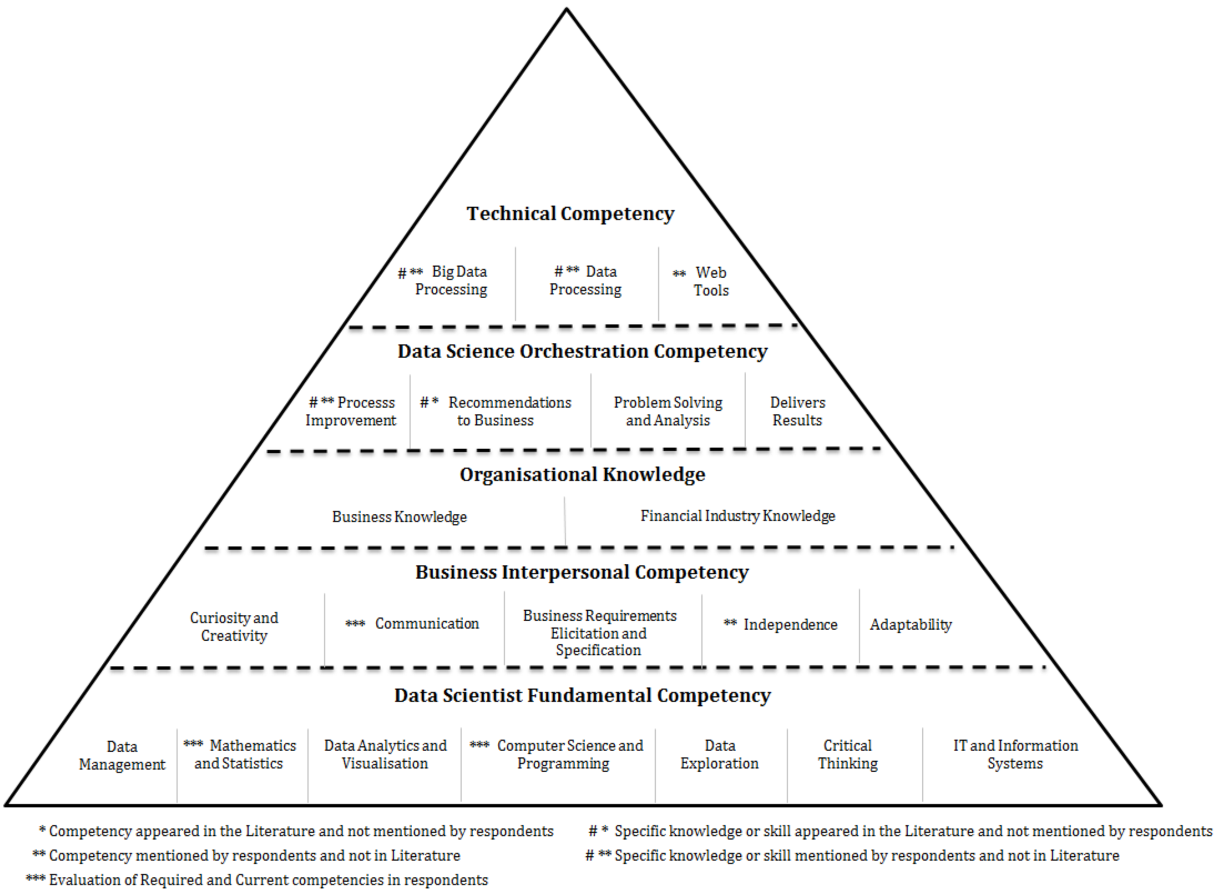


Figure 9: Data Scientist Competency Model

A full view of all competencies identified from the literature and grouped according to the research framework is shown in Appendix C. Appendix L contains the results of the interviews mapped to the research framework. Each category will now be discussed.

5.5.1 Technical Competency category

The Technical Competency category includes the various technical knowledge and skills required by a Data Scientist. According to the literature review, a Data Scientist needs to use the latest technology and tools to perform his role (Davenport, 2012; Kim & Lee, 2016; Song & Zhu, 2016). The competencies grouped under the Technical Competency category are:

- Big Data Processing
- Data Processing
- Web Tools

Each of these competencies is now discussed further in relation to the literature. All the competencies that fell under the Technical Competency category as well as the associated knowledge and skills the Data Scientist interviewees have can be seen in Appendix M.

a. Big Data Processing

The analysis of Big Data can extract meaning, hidden patterns, correlations and unexpected connections that can add value to the organisation (Ozkose et al., 2015; Vahn, 2014). The competency Big Data Processing includes the ability to use the tools and languages that are used to analyse Big Data such as Python, R and Hadoop. According to the literature Hadoop, Machine Learning, Python and R are the tools mentioned for use by a Data Scientist and are mainly for the analysis of Big Data (Abbasi et al., 2016; Davenport, 2012; Schoenherr & Speier-Pero, 2015; Brynjolfsson et al., 2011; Dhar, 2013; Hu, Wen, Chua & Li, 2014; Kim & Lee, 2016; Miller, 2014).

In the results, R and Hadoop were the common tools used among the Data Scientists as they were mentioned by 5 out of the 6 Data Scientists and was also linked back to the literature (Abbasi et al., 2016; Davenport, 2012; Schoenherr & Speier-Pero, 2015). Following on from R and Hadoop is Machine Learning and Python which was found in 4 of the 6 Data Scientists and is also mentioned in the literature for the analysis of Big Data (Abbasi et al., 2016; Davenport, 2012; Schoenherr & Speier-Pero, 2015; Brynjolfsson et al., 2011; Dhar, 2013; Kim & Lee, 2016; Miller, 2014; Wen et al., 2014).

The Big Data Processing competency also refers to an understanding and ability in artificial intelligence, Facebook Marketing, Google Analytics, Natural Language processing and neural networks which were mentioned in the interviews. Kim and Lee (2016) highlight these tools in their study with the exception of Google Analytics which is not found in the literature review. In the literature, the importance of having technical programming skills was mentioned (Carter & Sholler, 2016; Chen et al., 2012; Patil & Davenport, 2012; Schoenherr & Speier-Pero, 2015; Wamba et al., 2015; Wixom et al., 2014).

Overall, this research aligns to the literature for the processing of Big Data with the only exception being the use of Google Analytics which was mentioned by 1 Data Scientist during their interview and did not emerge from the literature reviewed. As a result two asterisks with a preceding hash have been placed next to the Big Data Processing competency indicating that specific knowledge or skill was not found in the literature and not the competency as a whole which did align to the literature.

b. <u>Data Processing</u>

The competency Data Processing is made up of tools and procedures that are used by the Data Scientist for the management of data and consists of knowledge and expertise relating to languages and programmes such as SQL, Excel, Java, Qlikview, Power BI, Tableau, Mathematica and Matlab. SQL emerged as the most used language by this sample of Data Scientists with them all using SQL for data analysis and transformation. DS 4 mentions that Data Scientists should possess the ability to clean and integrate, analyse and then transform data, and suggests that these are the core skills required. Kim and Lee (2016) identified SQL, Java and Matlab as the most important skills for data analysis and claim that SQL, Java and Matlab are part of the top 20 skills that a Data Scientist should have, based on their study which explored the definition of a Data Scientist within different industries (2012). Mathematica was mentioned by one of the interviewees in this study who had a background in Applied Mathematics, but was not mentioned in the literature.

The most common tools amongst the Data Scientist interviewees in this study for data visualisation were Qlikview, Power BI and Tableau. DS 5 mentioned the importance of skilling oneself in data visualisation software, like Power BI or Tableau. This statement highlights the importance of being able to use data visualisation tools in the work of a Data Scientist. The literature mentions Tableau, Power BI, SQL and Excel as the popular tools to work with data visualisation (Dinter et al., 2015; Kim & Lee, 2016; McAfee et al., 2012; Schoenherr & Speier-Pero, 2015).

Overall this competency aligned with the literature. The only exception was the tool Mathematica which was not mentioned in the literature. Two asterisks with a preceding hash have been placed next to the Data Processing competency indicating that specific knowledge or skill Mathematica was not found in the literature and not the competency as whole which did align to the literature.

c. <u>Web Tools</u>

This competency was not identified in literature and was only mentioned by 1 responding Data Scientist. Web Tools competency is the ability to create an online solution for business with the aim of solving a specific business problem. DS 5 was the only one who used the tools and languages that fell within this competency. During his time at a leading scientific research centre, he built a predictive analytic website that looked at real life problems where he was exposed to tools including Bootstrap, CSS, Django, HTML and JavaScript. This competency is indicated with two asterisks in the competency model indicating that this competency was mentioned by respondents and not found in the literature. This competency might be of lesser importance when compared to other competencies, and organisations might not need to focus as much on this competency when hiring a Data Scientist.

Table 5 shows a summary of the Technical Competency category and the competencies that were identified with a description. The knowledge and skills that were identified as part of each competency in the interviews and literature are all shown.

Table 5: Technical Competency from Results and Literature

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Technical Competency	**# ** Big Data Processing**	The ability to use skills and tools to work on data to answer business questions	Hadoop	Abbasi et al. (2016)
			R	Davenport (2012) Schoenherr & Speier-Pero (2015)
			Machine Learning	Abbasi et al. (2016) Brynjolfsson et al. (2011) Dhar (2013) Kim & Lee (2016) Miller (2014)
			Python	Davenport (2012) Schoenherr & Speier-Pero (2015)
			Artificial Intelligence	Dhar, 2013
			Facebook Marketing	Kim & Lee (2016)
			# ** Google Analytics	Not mentioned in the literature
			Natural Language Processing	Kim & Lee (2016)
			Neural Networks	Kim & Lee (2016)
			Technical programming skills	Carter & Sholler (2016) Chen et al. (2012) Patil & Davenport (2012) Schoenherr & Speier-Pero (2015) Wamba et al. (2015) Wixom et al. (2014)
	# ** Data Processing	The ability to use data tools for analysis,transfromation and visualisation of data	SQL	Kim & Lee (2016)
			Excel	Schoenherr & Speier-Pero (2015)
			Java	Kim & Lee (2016)
			# ** Mathematica	Not mentioned in the literature
			Programming on Matlab	Kim & Lee (2016)
			Qlikview	Dinter et al. (2015)
			Power BI	McAfee et al. (2012)
			Tableau	Schoenherr & Speier-Pero (2015)
	**** Web Tools**	The ability to build web applications using certain tools and applications	Bootstrap	Not mentioned in the literature
			CSS	Not mentioned in the literature
			Django	Not mentioned in the literature
			HTML	Not mentioned in the literature
			Javascript	Not mentioned in the literature

* Competency appeared in the Literature and not mentioned by respondents

\# * Specific knowledge or skill appeared in the Literature and not mentioned by respondents

** Competency mentioned by respondents and not in Literature

\# ** Specific knowledge or skill mentioned by respondents and not in Literature

*** Evaluation of Required and Current competencies in respondents

5.5.2 Data Science Orchestration Competency category

The Data Science Orchestration Competency category encompasses all the competencies that Data Scientists have in order to provide the actions needed to deliver correct and optimal results to business (Kamioka & Tapanainen, 2014; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015). The competencies grouped under the Data Science Orchestration Competency category are:

- Process Improvement
- Recommendations to Business
- Problem Solving and Analysis
- Delivers Results

Each of these competencies is now discussed with reference to knowledge, skills and attributes that have been classified as making up each competency. All competencies that fall under the Data Science Orchestration competency category as well as the associated knowledge and skills the Data Scientists possess can be seen in Appendix N. From the results, the common competencies of a Data Scientist within this category were Process Improvement, Recommendations to Business, and Problem Solving and Analysis. These competencies are discussed first followed by the other competency within this category.

a. <u>Process Improvement</u>

The Process Improvement competency is the ability of the Data Scientist to create new processes to replace existing ones resulting in reducing manual labour and making processes more efficient for the organisation in response to requirements raised by the business (Schoenherr & Speier-Pero, 2015; Saltz, 2015). The competency Process Improvement was common amongst the Data Scientists. According to Chiang et al. (2012), the challenge for Data Scientists is to create the capability for business to be able to understand what is happening in the data and to be able to interpret that information to take advantage of the opportunities. Through Process Improvement, opportunities can be identified in the organisation (Dinter et al., 2015; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015). Provost and Fawcett (2013) state that Data Science involves "principles, processes, and techniques or understanding" through the automation of business processes which further enhances the importance of being able to create or improve processes. Data Scientists are tasked with creating new processes, automating manual processes and enhancing existing ones. DS 1, DS2, DS 4, DS 5 and DS 6 all mentioned the creation or improvement of a process in the interview.

DS 2 states the benefits of Process Improvement: "Data Scientists can help in optimising processes and improving on customer interactions using predictive models to identify who would be well suited to certain products." The only skill that was mentioned by the Data Scientists which did not form part of the literature reviewed was the skill global optimisation. This could be due to the specific nature of the industry in which the sample was obtained.

The research aligns to literature for Process Improvement as this study had a similar outcome with the only exception being the skill global optimisation. As a result two asterisks with a preceding hash have been placed next to the Process Improvement competency indicating that specific knowledge or skill was not found in the literature and not the competency as a whole which did align to the literature.

b. Recommendations to Business

Recommendations to Business is the ability of the Data Scientist to answer business questions through exploring the data, solving problems and providing the insights needed by organisations to make effective decisions (Kim & Lee, 2016). Most of the literature fell within this competency (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; McAfee et al., 2012; Provost & Fawcett, 2013; Chen et al., 2012; Chiang et al., 2012; Dhar, 2013; Dinter et al., 2015; Kim & Lee, 2016; Miller, 2014; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Watson, 2014). Data Scientists help business answer questions either related to their customers or products and have the ability to make recommendations based on the answers they find in the data (Kamioka & Tapanainen, 2014; Poleto, de Carvalho & Costa, 2015; Provost & Fawcett, 2013). DS 5 mentions that a Data Scientist should be able to analyse data and bring actionable insights. DS 3 states that he helps management to better understand the source of profitability with his analysis. Both examples highlight the Data Scientist making Recommendations to Business in order to increase either sales or profitability for the organisation.

The Data Scientists interviewed spoke of making Recommendations to Business and did not mention Decision Making. In the literature review, Decision Making was identified as one of the skills of a Data Scientist. By providing the insights needed by organisation, **decisions** can be made more effectively (Abbasi et al., 2016; Chatfield et al., 2014; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015). The next step from making **Recommendations to Business** is **Decision Making** which are decisions based on the recommendations provided by the Data Scientists (Chen et al., 2012). Data Scientists in the literature were involved in the Decision Making phase as Abbasi et al. (2016) mention that Data Scientists have to work closely with Analysts in the Decision Making phase. This could imply that the Data Scientists in this study are either not involved in Decision

Making and their work ends once recommendations are provided to business or that Decision Making is implied when they refer to recommendations made to business.

The research aligns to the literature for Recommendations to Business as this study had the same views with the only exception being Decision Making which was not mentioned by the interviewees.

Due to the absence of the knowledge or skill Decision Making by the Data Scientists during the interviews, the competency Recommendations to Business has one asterisk with a preceding hash placed next to the Recommendations to Business competency indicating that specific knowledge or skill Decision Making was mentioned in the literature and was not mentioned by the interviewees and not the competency as a whole which did align to literature (Abbasi et al., 2016; Miller, 2014; Schoenherr & Speier-Pero, 2015).

c. <u>Problem Solving and Analysis</u>

The competency Problem Solving and Analysis is the ability to analyse problems and solve problems within the organisation which is an important skill for a Data Scientist (Dhar, 2013; Kim & Lee, 2016; Provost & Fawcett, 2013). According to Kim and Lee (2016), a Data Scientist is an expert who solves problems by their ability to discover insights in datasets. The Data Scientist is able to break down problems into smaller parts that can be analysed (Dhar, 2013). They are able to come up with many solutions to the problem and can identify the information which is needed to effectively solve the problem (Dhar, 2013). The Data Scientist further provides the problem analysis and the recommended solution rather than only specifying the problem (Provost & Fawcett, 2013).

DS 2 encapsulates the problem solving skill when he mentioned the following: "Data Scientists attempt to solve a broad range of problems in business and the method of finding the solution is consistently to look at the data and use the necessary tools (be it Statistical or otherwise), to find the solution." This statement confirms the Data Scientist's ability to analyse and solve problems. DS 2, DS 3, DS 4 and DS 5 all mentioned problem solving as part of their tasks as well as finding problems when looking in the data. The competency Problem Solving and Analysis aligned to the literature.

d. <u>Deliver Results</u>

The Deliver Results competency is the ability to provide results that lead to solutions that make an improvement in the way an organisation operates (Abbasi et al., 2016). This competency contained the following skills which were mentioned by Data Scientists which included Enhance Customer Behaviour, Enhance Customer Experience, Improve Profitability, Predict Customer Behaviour and Enhance User Experience which highlight the value a Data Scientist adds. DS 2, DS 3 and DS 6 mentioned that the results of their work enhanced the behaviour of customers with the predictive models that they created. Profitability was increased in the organisation for DS 3.

Table 6 shows a summary of the Data Science Orchestration Competency category and the competencies that were identified, with a description. The knowledge and skills that were identified as part of each competency in the interviews and literature are shown. In the interview results, the competencies that were common for a Data Scientist in this category were Process Improvement, Recommendations to Business and Problem Solving and Analysis.

Table 6: Data Science Orchestration Competency Category

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Data Science Orchestration Competency	**# ** Process Improvement**	The ability to create new processes and optimisise current processes to reduce manual labour	Deliver new model	Dinter et al. (2015)
			Automate processes	Provost & Fawcett (2013)
			Create new processes	Schoenherr & Speier-Pero (2015)
			Dynamically Adjust Solutions	Schoenherr & Speier-Pero (2015)
			Global Optimisation	Not mentioned in the literature
			Process Optimisation	Schoenherr & Speier-Pero (2015)
	# * Recommendations to Business	The ability to answer business questions and make recommendations and provide insights	Answer Business Questions	Abbasi et al. (2016) Carter & Sholler (2016) Chatfield et al. (2014) McAfee et al. (2012) Provost & Fawcett (2013)
			Actionable Insights	Chen et al. (2012) Chiang et al. (2012) Dhar (2013) Dinter et al. (2015) Kim & Lee (2016) Miller (2014) Schoenherr & Speier-Pero (2015) Waller & Fawcett (2013) Wamba et al. (2015) Watson (2014)
			Answer Questions not asked Drive Results and Change Help management understand profitabilty Validate Business Assumptions	Watson (2014)
			Decision Making Not mentioned by Respondent	Abbasi et al. (2016) Miller (2014) Schoenherr & Speier-Pero (2015)
	Problem Solving and Analysis	The ability to idetify and resolve problems and propose solutions	Problem Solving Problem Identification Solution Design	Dhar (2013) Kim & Lee (2016) Provost & Fawcett (2013)
	Deliver Results	The ability to provide insights that lead to solutions that make an improvement in the way an organisation interacts with its customers	Enhance Customer Behaviour	Abbasi et al. (2016)
			Enhance Customer Experience	
			Improve Profitability	
			Predict Customer Behaviour	
			User experience	

* Competency appeared in the Literature and not mentioned by respondents
# * Specific knowledge or skill appeared in the Literature and not mentioned by respondents
** Competency mentioned by respondents and not in Literature
# ** Specific knowledge or skill mentioned by respondents and not in Literature
*** Evaluation of Required and Current competencies in respondents

5.5.3 Organisational Knowledge category

The Organisational Knowledge category includes the application of the Data Scientists' extensive and specialised knowledge and skills to accomplish the organisations' objectives effectively (Wamba et al., 2015). Data Scientists need to be able to identify and ask the right questions to gather the correct requirements from business, while keeping in mind the goals of business (Abbasi et al., 2016; Chatfield et al., 2014; McAfee et al., 2012). Data Science is highly context based and most problems cannot be solved without strong Organisational Knowledge (Waller & Fawcett, 2013). This is the reason why most companies prefer professionals that have a substantial amount of experience in a similar domain, even if it was not directly correlated to Data Science tasks (Patil & Davenport, 2012; Davenport, 2012; Dinter et al, 2015, Kim & Lee, 2016). The competencies grouped under the Organisational Knowledge category are:

- Business Knowledge
- Financial Industry Knowledge

Each of these competencies is now discussed with reference to knowledge, skills and attributes that have been classified as making up each competency. All competencies that fall under the Organisational Knowledge category as well as the associated knowledge and skills the Data Scientists possess can be seen in Appendix O.

a. <u>Business Knowledge</u>

Data Scientists need to possess Business Knowledge which is the ability of a Data Scientist to understand the different processes that occur within the specific organisations context in which the Data Scientists works (Wixom et al., 2014). According to Provost and Fawcett (2013),"Data-science projects require close interaction between the scientists and the business people responsible for the decision making." This way of working allows the Data Scientist to familiarise themselves with the business processes and also highlights the importance for Data Scientists to have Business Knowledge in the company that they work in to understand how the business operates. By understanding the processes, the Data Scientist will have a holistic view which would enable the Data Scientist to provide solutions to cater for the organisation requirements (Kim & Lee, 2016). DS 3 mentioned solving business problems is only possible through the understanding of what is actually required. Through in-depth discussions with business, the Data Scientist is able to clarify what business wants. The following papers in the literature spoke of the importance of having Business Knowledge (Chen et al., 2012; Davenport, 2012; Kim & Lee, 2016; Waller & Fawcett, 2013;

Wamba et al., 2015; Wixom et al., 2014). This competency aligned to the literature as both the literature and the results gathered highlight the importance of Business Knowledge.

b. Financial Industry Knowledge

Data Scientists need to have Financial Industry Knowledge (Davenport, 2012; Kim & Lee, 2016; Waller & Fawcett, 2013). Data Scientists need to understand the financial processes and procedures as well as having knowledge on all products and services in the industry that they reside in (Waller & Fawcett, 2013). DS 4 states that his financial knowledge of the industry has made him a stronger resource in his team. Financial Industry Knowledge is the knowledge that the Data Scientists have gathered. Within this study, the knowledge of the Data Scientists which were interviewed, was specific to the banking and insurance industry. Each Data Scientist mentioned their Financial Industry Knowledge as 5 of the 6 Data Scientists had been within the industry prior to becoming Data Scientists. Although they were fairly new in the role of a Data Scientist, their previous positions were in the same industry enabling the advantage of gaining the experience and knowledge with this industry. This competency aligned with the literature and is applicable to organisations within the financial industry and would not be applicable to organisations within the health industry, as an example.

Table 7 shows a summary of the Organisational Knowledge category and the competencies that were identified with a description. The knowledge and skills that were identified as part of each competency in the interviews and literature are shown.

Table 7: Organisational Knowledge Category

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Organisational Knowledge	**Business Knowledge**	Specialised business knowledge	Business Knowledge	Chen et al. (2012)
	Financial Industry Knowledge	Applies Financial knowledge and skills and specialised knowledge of the Industry	Financial Knowledge	Davenport (2012) Kim & Lee (2016) Waller & Fawcett (2013) Wamba et al. (2015) Wixom et al. (2014)

5.5.4 Business Interpersonal Competency category

The Business Interpersonal competency category refers to the Data Scientists' ability to communicate their work to drive solutions within the organisation (Patil & Davenport, 2012). Presentation and Communication skills are needed for this category (Wixom et al., 2014; McAfee et al., 2012). This category falls into the bottom of the triangle as the ability to form and maintain relationships with business or any of the stakeholders is more of a social attribute of a person and is not necessary a skill that can be easily obtained through learning (Viitala, 2005). The competencies grouped under the Business Interpersonal competency category are:

- Curiosity and Creativity
- Communication
- Business Requirements Elicitation and Specification
- Independence
- Adaptability

All the competencies that fall under the Business Interpersonal competency category as well as the associated knowledge and skills the Data Scientists possess can be seen in Appendix P. Each of these competencies that have been classified as making up the Business Interpersonal competency category is now discussed. From the results, the common competencies of a Data Scientist within this category were Curiosity and Creativity, Communication and the Business Requirements Elicitation and Specification competency. These competencies are discussed first followed by the other competencies within this category.

a. <u>Curiosity and Creativity</u>

This competency was not found within the competency libraries (Harvard University, 2015; Nestler et al., 2012; Sonteya & Seymour, 2012; Strengell, 2017). This competency was discovered in a study published online by the World Economic Forum (World Economic Forum, 2015). The Curiosity and Creativity competency is the desire and ability to ask questions and be inquisitive as well as being able to generate new and innovative ideas to solve problems (World Economic Forum, 2015). This competency is seen as a crucial competency for the 21st century for professionals (World Economic Forum, 2015). According to the literature, Data Scientists come up with new ideas and are always exploring different possibilities with the data to go beyond the surface of the data and to find meaning through exploration and an inquisitive personality (Van der Aalst, 2014; Schoenherr & Speier-Pero, 2015). According to Patil and Davenport (2012), a Data Scientist most dominant trait should be "an intense curiosity, a desire to go beneath the surface of a problem, find the questions at its heart, and distil them into a very clear set of hypotheses that can be tested." From this statement it is evident that the nature of a Data Scientist is one that strives to understand, to discover patterns and relationships and to find insights within the data. Curiosity and Creativity was the most common competency among the Data Scientists interviewed. DS 1 mentioned that he had a passion when working with data. He is very curious when it comes to problem solving and his abilities help him break problems into smaller parts to explore.

Data Scientists have an inquisitive nature and are always seeking new possibilities in the data (Schoenherr & Speier-Pero, 2015). DS 5 mentions that he is inquisitive, likes challenges and also is willing to learn new things and further mentions that Data Scientists are passionate about working with data, concurred in the statement by DS 4 who indicates that a Data Scientist is someone who has an "appetite for data and are inquisitive to analyse data further." DS 4 comments further that "Data Scientists tend to be **curious**, and would then **explore** data and data sources" meaning that Data Scientists explore data due to their curiosity. DS 4 also states "I'm very inquisitive to understand and research any information that I find interesting." The Curiosity and Creativity competency aligned to the literature.

b. Communication

Communication competency is the Data Scientist's ability to effectively communicate thoughts, ideas and solutions verbally so that business understands and can act on the information provided (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; Chen et al., 2012; Davenport, 2012; McAfee et al., 2012; Patil & Davenport, 2012; Wamba et al., 2015; Watson, 2014; Wixom et al., 2014). Data Scientists need to be able to communicate their results in a way that business understands in order to move towards actionable insights or problem resolution (Patil & Davenport, 2012). They also need good presentation skills in the way they portray the results to business, as well as being able to convey a visualisation of the solution to business (Wamba et al., 2015).

Most of the literature reviewed in this study, when mapped to the categories in the research framework, fell within the Communication competency (Abbasi et al., 2016; Carter & Sholler, 2016; Chatfield et al., 2014; Chen et al., 2012; Davenport, 2012; McAfee et al., 2012; Patil & Davenport, 2012; Wamba et al., 2015; Watson, 2014; Wixom et al., 2014). This highlights the importance for a Data Scientist to have effective Communication skills in order to convey the insights discovered to business. Kim and Lee (2016) also discovered in their study, the importance of Communication skills for a Data Scientist. Patil and Davenport (2012) further state the importance of Communication in their statement: "need for Data Scientists to communicate in a language that all their stakeholders understand and to demonstrate the special skills involved in storytelling with data, whether verbally, visually, or ideally both." Communication was a common competency among all Data Scientists interviewed. The importance of a Data Scientist having the Communication competency aligns with literature. Three asterisks have been placed next to the Communication competency as this competency was mentioned as a skills gap in the literature reviewed and also mentioned by DS 1 as a gap between the required and current Communication competency.

c. Business Requirements Elicitation and Specification

In order to make sure that the right requirements have been specified by the business for a problem, the Data Scientist needs to ask the business the right questions in order to get the right requirements to deliver a solution that the business wants (Abbasi et al., 2016; Chatfield et al, 2014; McAfee et al, 2012; Patil & Davenport, 2012; Provost & Fawcett, 2013). The Business Requirements Elicitation and Specification competency is the ability of a Data Scientist to deliver results that match the actual needs of the business which also means being able to understand the actual requirement by asking the right questions (Patil & Davenport, 2012).

Interpreting what business desires is a skill of the Data Scientist in order to make sure that business needs are met. DS 4 states that "A Data Scientist needs to understand the business' needs and develop analytical tools that meet that objective." DS 2, DS 3, DS 4 and DS 6 mentioned having the ability to ask the right questions to business to get the right requirements. DS 6 also mentions "real value comes from delivering the results that match the actual needs of the bank" showing the importance of delivering the right solution. The Business Requirements Elicitation and Specification competency aligns to the literature.

d. Independence

The Independence Competency refers to the way in which the Data Scientist works. They have an independent nature and are very confident in what they do. DS 2 and DS 3 both mentioned that they work well alone while DS 5 mentioned that he could work independently or in a team. The literature spoke more of Data Scientists working well in teams (Patil & Davenport, 2012; Davenport, 2012; Dinter et al., 2015; Kim & Lee, 2016). Currently, not many people can cover all aspects of the vast variety of skills required by Data Science. This encourages companies to invest in creating teams for their Data Science jobs (Kim & Lee, 2016). This may change in the future but currently Data Scientists must work within teams in order to fulfil the requirements of the business (Dinter et al., 2015). The Data Scientist must also be able to work closely with company executives to provide insight on a strategic level (Davenport, 2012). This competency was not aligned to the literature reviewed as the literature spoke more of extroverted Data Scientist who worked in teams and therefore has two asterisks next to it in the competency model. Four out of the six Data Scientists mentioned that they are introverted and prefer to work alone.

e. Adaptability

The Adaptability competency speaks to the ability of a Data Scientist to easily adapt to any situation and also have the flexibility to adjust any solution (Patil & Davenport, 2012). The attributes that fell under the Adaptability competency as gathered from the interviews were a Resilient Personality, Flexibility, Pragmatism and willing to learn new things. DS 1 mentioned several times in the interview the importance of having a resilient personality. He mentioned "you need a resilient personality. Not committed to certain idea or way of doing things. You need to dynamically adjust, and not be overly committed to certain classifications." DS 2 mentions that "pragmatism [find useful solutions, not perfect solutions]" is one of the attributes of a Data Scientist. DS 4 mentioned "I think that I am quite resilient, and enjoy repeatedly stress testing a system in different ways until I get the results I desire" thus showing the importance of a resilient, adaptable and flexible personality. Data Scientists need to be able to adapt to different possibilities of finding a solution. This competency aligned to the literature.

Table 8 shows a summary of the Business Interpersonal competency category and the competencies that were identified with a description. The knowledge, skills and attributes that were identified as part of each competency in the interviews and literature are shown. From the results, the common competencies for the Data Scientist were Curiosity and Creativity, Communication and Business Requirement and Elicitation Specification. Communication has three asterisks indicating that there was a gap in the required and current competencies found in the results. The Independence competency has two asterisks as this competency was mentioned by the Data Scientists interviewed and not found in the literature.

Table 8: Business Interpersonal Competency Category

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Business Interpersonal Competency	**Curiosity and Creativity**	Have the ability to come up with new ideas and is always exploring different possibilities	Inquisitive Creative Introverted but Curious Intuitive Like Challenges Passion working with data Think out of the box	Patil & Davenport (2012) Schoenherr & Speier-Pero (2015)
	***** Communication**	Have the ability to effectively communicate insights to business	Ability to Communicate Presentation Skills Convey business insights	Abbasi et al. (2016) Carter & Sholler (2016) Chatfield et al. (2014) Chen et al. (2012) Davenport (2012) McAfee et al. (2012) Patil & Davenport (2012) Wamba et al. (2015) Watson (2014) Wixom et al. (2014)
	Business Requirements Elicitation and Specification	The ability to gather the correct requirements from business	Ask the right questions Better understand requirements Interpret Queries	Patil & Davenport (2012)
	**** Independence**	Likes to work alone and is confident that goals can be achieved	Work Alone Self Confidence Takes Initiative	Not found in Literature
	Adaptability	Should have the flexibility to adjust to any solution	Resilient Personality Flexibility Pragmatism Willing to learn new things	Patil & Davenport (2012)

* Competency appeared in the Literature and not mentioned by respondents
\# * Specific knowledge or skill appeared in the Literature and not mentioned by respondents
** Competency mentioned by respondents and not in Literature
\# ** Specific knowledge or skill mentioned by respondents and not in Literature
*** Evaluation of Required and Current competencies in respondents

5.5.5 Data Scientist Fundamental Competency category

The Data Scientist Fundamental Competency category includes the critical basic and foundational skills needed to perform their role and underpin their daily activities (Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015). The competencies grouped under the Data Scientist Fundamental Competency category are:

- Data Management
- Mathematics and Statistics
- Data Analytics and Visualisation
- Computer Science and Programming
- Data Exploration
- Critical Thinking
- IT and Information Systems

All the competencies that fall under the Data Scientist Fundamental Competency category as well as the associated knowledge and skills the Data Scientists possess can be seen in Appendix Q. Each of these competencies is now discussed with reference to the knowledge, skills and attributes that have been classified as making up Data Scientist Fundamental Competency category. From the results, the common competencies were Data Management, Mathematics and Statistics, Data Analytics and Visualisation and Computer Science and Programming. These competencies are discussed first followed by the other competencies within this category.

a. <u>Data Management</u>

The Data Management competency is the ability of the Data Scientist to retrieve and work with data from different sources, analyse and transform data to solve problems and gather insight for the organisation (Davis, 2014; Dhar, 2013; Dinter et al., 2015; McAfee et al., 2012; Patil & Davenport, 2012; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Watson, 2014). Data Management was the most common competency for the Data Scientist Fundamental competency category. Kim and Lee (2016) discovered in their study that a Data Scientist needs to understand how to work with data throughout the whole life cycle of data. This statement aligns to the research results.

The main skills a Data Scientist would need to have is data manipulation, data sourcing and data transformation. Five out of the six Data Scientists mentioned these important skills. DS 2 mentioned that his tasks were "Data manipulation, cleaning, data extraction, coding with datasets and SQL." DS 3 states the importance of this competency "I worked in an environment where data needed to be cleaned very well. This taught me how to structure data efficiently. Ultimately this was a big advantage when looking at data and knowing what pitfalls to look out for."

DS 5 states, "I know how to work with different data structures and manipulating them to serve my purpose, as well as how to get data from different sources" which shows the importance of being able to work with data. DS 5 further mentions "Skills that go with Data Science is learning how to manipulate data to suit your needs without loss of generality or data itself." The Data Management competency aligned to the literature. The following knowledge and skills made up the Data Management competency demonstrating that the Data Scientist should be able to work with data throughout the lifecycle of the data:

- Data Manipulation
- Data Sourcing
- Data Transformation
- Database Design

- Data Design
- Data Mining
- Data Modelling
- Data Queries
- Database Management

b. Mathematics and Statistics

The Mathematics and Statistics competency refers to the Data Scientist's ability to use Mathematical and Statistical techniques on data to discover new patterns and trends (Chatfield et al., 2014; Miller, 2014; Carter & Sholler, 2016; Dhar, 2013; Kim & Lee, 2016; Brynjolfsson et al., 2011; McAfee et al., 2012). Mathematics and Statistics are also very important in the literature as Data Scientists have to work with large sets of data to derive meaning as well as creating predictive models to be able to meet future demands and needs of the organisation (Abbasi et al., 2016; Chatfield et al., 2014; Chiang et al., 2012; Davenport, 2012; Davis, 2014; Dinter et al., 2015; Kim & Lee, 2016; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014).

Mathematics and Statistics were common among all Data Scientists as each one had a background in Mathematics and some form of Statistics. DS 3 has an Actuarial background and due to the large amount of Statistics within the course, it has also been grouped within this competency. DS 2 states his skills as having a "Strong understanding of Mathematics and Statistics". DS 2 did a degree in Mathematics and Statistics and an honours and master's degree in Mathematical Statistics and has a strong Mathematics and Statistics background. DS 5 did his degree in Mathematical Statistics as well. The Mathematics and Statistics competency aligned to the literature.

Three asterisks have been placed next to the Mathematics and Statistics competency as the results obtained from the competencies that the Data Scientists identified as required and their current competencies differed. The Data Scientists mentioned the importance of having Statistics as one of the Data Scientist had done a degree in Computational and Applied Mathematics which lacked the Statistics course. The Data Scientist also mentioned the importance of Statistics in their degrees and recommended that anyone wanting to pursue a career in Data Science should be taking a Statistics course in addition to Mathematics.

c. Data Analytics and Visualisation

Data Analytics and Visualisation can be used to support decisions, make discoveries in the data and provide insights to business (Trifu and Ivan, 2014). The Data Analytics and Visualisation competency is Data Scientists' ability to analyse data using Analytic methods and then being able to visualise the data in models to derive insights (Carter & Sholler, 2016; Chatfield et al., 2014; Chen et al., 2012; Chiang et al., 2012; Kim & Lee, 2016). Data visualisation is an extremely important skill for Data Scientists in getting the business to understand the results of their work in a visual manner (Chatfield et al., 2014; Dinter et al., 2015; McAfee et al, 2012; Schoenherr & Speier-Pero, 2015; Wamba et al., 2015). Data Scientists need strong analytic skills and further to this they need to be able to communicate their findings to business in a way they can understand. Data Scientists in this study used Business Intelligence tools to do certain data analysis as well as the visualisations which can be interpreted by business. DS 5 states the following, "I communicate the results of my analysis by using data visualisation which is the fundamental and last part of analysing data, to present certain results to capture business's attention." DS 2 and DS 3 both mentioned data visualisation as part of the skills a Data Scientist should have. The Data Analytics and Visualisation competency aligned to the literature.

Data Analytics is the ability of the Data Scientist to provide the value that an organisation requires through the use of Analytic methods to enable organisations to predict their outcomes for the future (Abbasi et al., 2016; Chatfield et al., 2014; Chiang et al., 2012; Davenport, 2012; Davis, 2014; Dinter et al., 2015; Kim & Lee, 2016; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014). Questions are answered and insights provided through the use of Analytics. DS 2 mentions the importance of using Analytics to create predictive models in his statement, "I help to improve on customer interactions using predictive models to identify customers who may have unmet financial needs or who would be well suited to certain products." This statement is specific to the banking and insurance industry. The importance of meeting customers' needs and the products that meet their needs are crucial in staying ahead in this industry. The importance of this competency gathered in the interviews aligned with the literature.

d. Computer Science and Programming

The Computer Science and Programming competency is the ability to combine a Data Scientist's understanding of software engineering skills along with their numerical and programming skills in order to create solutions that aim to solve business problems (Miller, 2014; Chatfield et al., 2014). All 6 Data Scientists mentioned the need for programming skills to be able to carry out their tasks emphasising how important this skill is. DS 4 had studied Computer Science and states, "I know several computer programming languages and have a major in Computer Science". DS 2 and DS 3, although not having studied Computer Science, both had experience in programming having used SQL, Statistical tools and Java. DS 1 studied Computer Science in his degree and did courses on Machine Learning and Natural process language. Kim and Lee (2016) identified programming as one of the most important skills a Data Scientist should possess. Their study on job advertisements showed more than 80% of organisations requesting Computer Science and Programming as skills that they required in a Data Scientist.

In the literature, the importance of having Programming skills was mentioned (Carter & Sholler, 2016; Chen et al., 2012; Patil & Davenport, 2012; Schoenherr & Speier-Pero, 2015; Wamba et al., 2015; Wixom et al., 2014). According to Provost and Fawcett (2013) it is important to focus on the underlying Data Science principles which is programming instead of focussing on specific languages such as Python and R which could become outdated and replaced, but the competency of being able to program and code will remain important. This competency aligned with the literature.

Three asterisks have been placed next to the Computer Science and Programming competency as the results obtained from the competencies that the Data Scientists identified as required and their current competencies differed. The four Data Scientists mentioned the importance of having Computer Science and Programming skills as they had not done Computer Science in their degrees and recommended that anyone wanting to pursue a career in Data Science should be taking a Computer Science course.

e. Data Exploration

The Data Exploration competency is the ability of the Data Scientist to create new and innovative ways of exploring and optimising data to extract meaning from the data (Patil & Davenport, 2012). Through the exploration of data, insights can be gained (Abbasi et al., 2016; Chatfield et al., 2014). DS 1 mentioned that one of the ways in which he optimises the data he uses to solve problems is to recycle the data and keep it in a database to be able to solve future directed problems.

Patil and Davenport (2012) state that "Data Scientists are the people who understand how to fish out answers to important business questions from today's tsunami of unstructured information". This statement emphasises the importance of data exploration. DS 4 states, "A Data Scientist is someone who extracts value out of data through **exploring** the data." He further states that Data Scientists extract information from data sources for a better understanding of how their business performs showing how important extracting meaning through data exploration is. This competency included skills such as being able to explore data, extract meaning from data and being able to recycle data to predict future outcomes. The Data Exploration competency aligns to the literature.

f. Critical Thinking

The Critical Thinking Competency is the way in which Data Scientists break down problems and create effective solutions (Schoenherr & Speier-Pero, 2015). When DS 2 mentioned the criteria for a Data Scientist, "Critical thinking, analytical ability, a methodical approach to problems, pragmatism [find useful solutions, not perfect solutions], self-critical [asks "what are the potential issues with my solution] it reinforces the notion of how important it is to be able to evaluate a solution critically. DS 5 and DS 6 both mentioned the need to be focused and persistent when working with data. Schoenherr & Speier-Pero (2015) state that critical thinking helps Data Scientists to be able to develop insights from the data. This competency aligns with the literature.

g. IT and Information Systems

The IT and Information Systems competency refers to the Data Scientist's ability to use Information Systems and technology skills to answer business questions as well having the knowledge and skills to use database tools, online processing tools and all software related tools to perform all tasks (Waller & Fawcett, 2013). The Data Scientist also needs to keep up to date with changing technologies to remain competitive and keep pace with changes in the industry (Chiang et al., 2012).

Provost and Fawcett (2013) state the following which highlights the importance of Information Technology:" Information technology can be used to find informative data items from within a large body of data." DS 2, DS3 and DS 4 mentioned the importance of having Information Technology skills which is the ability to use computers to store, analyse and retrieve data. The different knowledge and skills that fell under the IT and Information Systems competency were IT skills, Application Knowledge, Business and Systems Analysis, Informatics, Information Science, Keep ahead with Data Science Techniques and Systems Knowledge. This competency aligned with the literature.

Table 9 shows a summary of the Data Scientist Fundamental competency category and the competencies that were identified with a description. The knowledge, skills and attributes that were identified as part of each competency in the interviews and literature are shown. From the results, the main competencies for the Data Scientist were Data Management, Mathematics and Statistics shown with three asterisks, Data Analytics and Visualisation, and Computer Science and Programming which has three asterisks indicating that there was a gap found in the results.

Table 9: Data Scientist Fundamental Competency Category

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Data Scientist Fundamental Competency	**Data Management**	The management process of data from the sourcing of data, modelling of data and transformation of data	Data Manipulation Data Sourcing Data Transformation Database Design Data Design Data Mining Data Modelling Data Queries Database Management	Carter & Sholler (2016) Chatfield et al. (2014) Chen et al. (2012) Chiang et al. (2012) Dinter et al. (2015) Kim & Lee (2016) McAfee et al. (2012) Miller (2014) Patil & Davenport (2012) Provost & Fawcett (2013) Schoenherr & Speier-Pero (2015) Waller & Fawcett (2013) Wamba et al. (2015) Watson (2014) Wixom et al. (2014)
	***** Mathematics and Statistics**	Have a strong background in Mathematics and Statistics	Mathematics	Carter & Sholler (2016) Chatfield et al. (2014) Dhar (2013) Miller (2014)
			Statistics	Brynjolfsson et al. (2011) Dhar (2013) McAfee et al. (2012) Miller (2014) Kim & Lee (2016)
	Data Analytics and Visualisation	Have the ability to create and maintain predictive models and present data after analysis in a visual way for business to understand	Data Analysis Predictive Models Analytics Prescriptive Models	Abbasi et al. (2016) Chatfield et al. (2014) Chiang et al. (2012) Davenport (2012) Davis (2014) Dhar (2013) Dinter et al. (2015) McAfee et al. (2012) Kim & Lee (2016) Patil & Davenport (2012) Provost & Fawcett (2013) Schoenherr & Speier-Pero (2015) Waller & Fawcett (2013) Wamba et al. (2015) Watson (2014) Wixom et al. (2014)
			Data Visualisation Business Intelligence Business Intelligence Tools	Chatfield et al. (2014) Dinter et al. (2015) McAfee et al. (2012) Schoenherr & Speier-Pero (2015) Wamba et al. (2015)
	***** Computer Science and Programming**	Has a background in Computer Science and programming	Programming	Abbasi et al. (2016) Brynjolfsson et al. (2011) Dhar (2013) Dinter et al. (2015) Kim & Lee (2016) Patil & Davenport (2012) Provost & Fawcett (2013) Watson (2014) Wixom et al. (2014)
			Computer Science	Chatfield et al. (2014) Miller (2014)
	Data Exploration	The exploration of data by extracting meaning and further enhancing processes	Exploring data Extracts meaning from Data Recycle data to predict future outcomes	Abbasi et al. (2016) Chatfield et al. (2014)
	Critical Thinking	Have the ability find best solution to business problems and is critical in the selection of a solution	Focused Persistent Attention to Detail Critical Thinking Self Critical	Schoenherr & Speier-Pero (2015)
	IT and Information Systems	Have a background in information technology to understand the basics	IT skills Application Knowledge Business and Systems Analysis Informatics Information Science Keep ahead with Data Science Techniques Systems Knowledge	Chiang et al. (2012) Dinter et al. (2015) Abbasi et al. (2016) Brynjolfsson et al. (2011) Carter & Sholler (2016) Chiang et al. (2012) Davenport (2012)

5.6 Summary of the Competencies of a Data Scientist

There was a strong alignment between the results gathered in the interviews and literature reviewed implying that there is an alignment in competencies between developed and developing countries in terms of the competencies that a Data Scientist should have. There was also a strong alignment between the competencies that were required and the competencies that the Data Scientists currently have which was gathered from the interviews. The competency model is shown again in Figure 10. Each category was discussed as well as the competencies that fall within each category.

For a competency that was mentioned in the literature and not mentioned by the Data Scientists, one asterisk was placed next to the competency. If a knowledge or skill making up the competency was mentioned in the literature and not mentioned by the Data Scientists, an asterisk with a preceding hash was placed to indicate that a specific skill did not align to literature and not the competency as a whole. In the model, the competency Recommendations to Business is the only competency with one asterisk and a preceding hash. The Data Scientists interviewed did not mention the skill Decision Making which was mentioned in the literature (Abbasi et al., 2016; Miller, 2014; Schoenherr & Speier-Pero, 2015).

A competency that was mentioned by the Data Scientists in this study but not found in the literature was represented with two asterisks next to the competency. If a knowledge or skill making up the competency was mentioned by the Data Scientists but not found in the literature, two asterisks with a preceding hash was placed to indicate that a specific knowledge or skill did not align to literature and not the competency as a whole. Google Analytics which fell under the Big Data Processing competency was mentioned by a respondent but not found in the literature. Mathematica was also not mentioned in the literature reviewed and fell under the Data Processing Competency. Bootstrap, CSS, Django, HTML and JavaScript were all mentioned by one Data Scientist interviewed and fell under the Web Tools competency. This competency was not found in the literature and has two asterisks next to the competency. Global Optimisation was not mentioned in the literature and fell under the competency Process Improvement. This competency had two asterisks with a preceding hash in Figure 10. Data Scientists that were introverted and preferred to work alone was mentioned by the Data Scientists interviewed and fell under the Independence competency. The literature spoke more of extroverted individuals who worked better in teams (Patil & Davenport, 2012; Davenport, 2012; Dinter et al., 2015; Kim & Lee, 2016). For this reason, two asterisks were placed next to this competency.

The gaps identified by Data Scientists between their required and current competencies were highlighted by three asterisks next to the competency. Gaps were identified within the Communication competency which is vital for the transfer of ideas and solutions to business in a way they understand (Chatfield et al., 2014). Computer Science and Programming was also identified as a gap of knowledge and skills from the Data Scientists as four of the Data Scientists interviewed did not do Computer Science as part of their undergraduate degree. Statistics was identified as a gap within the Mathematics and Statistics competency as one Data Scientist had not done Statistics as part of his degree. These gaps are discussed further in the next section.

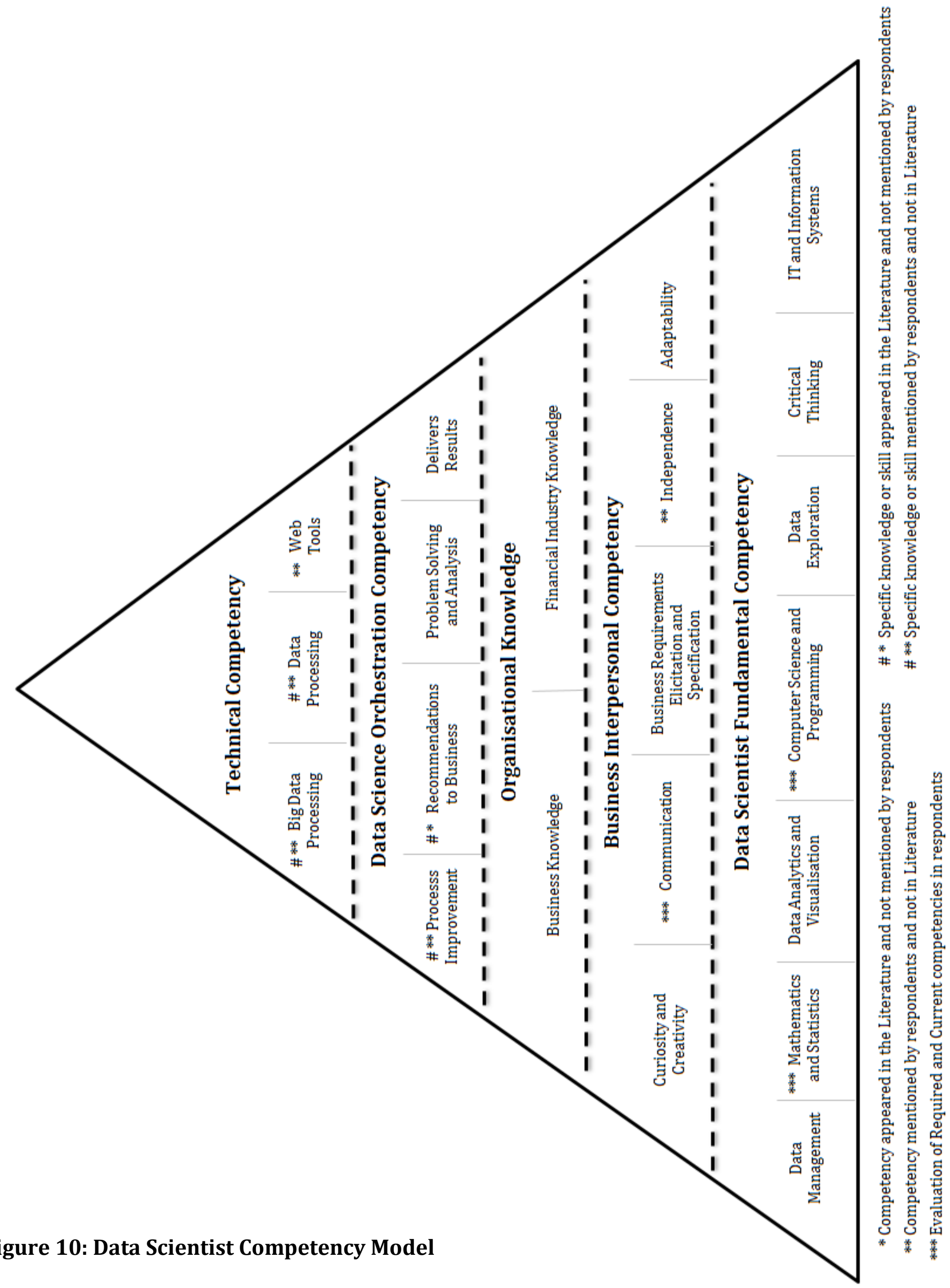


Figure 10: Data Scientist Competency Model

5.7 Gaps in Competencies of a Data Scientist

In order to determine the gaps in the competencies possessed by the Data Scientists, questions were divided into two sections. One section asked about what competencies the Data Scientists considered someone in that role should have, and the second section asked about the current competencies that they have. An overall consensus among the Data Scientists interviewed was that the majority of their competencies aligned to the required competencies they initially described. Three competencies that did not align between the current and required competencies were Communication, Computer Science and Programming, and Mathematics and Statistics. There is a lower emphasis on Statistics as it was mentioned by one Data Scientist but was not indicated as a lack of this skill, merely the need for a more in-depth understanding of Statistics as it was not part of his degree.

5.7.1 Importance of Communication

The gaps that were discussed in the literature review were lack of interpersonal skills, difficulty finding Data Scientists who could effectively communicate and Data Scientists that did not have a business background. One of the gaps, the lack of Communication skills is the ability to effectively translate work into a language that business understands. This study did reveal a similar gap by one of the Data Scientist. DS 1 spoke of the importance of having good Communication skills and when asked if there was anything he wished he could have studied, he mentioned Industrial Psychology which is the study of interpersonal skills and Communication within the business environment. DS 1 also stated that, "Many technological problems can be solved but a problem in many projects is interpersonal human relationships. What is slowing Data Scientists down is the lack of interpersonal skills and how you communicate." DS 5 mentions that "Learning how to **communicate** the results of the analysis is fundamental" which shows how important Communication is. The results obtained from the Data Scientists on the importance of Communication were in line with literature reviewed.

5.7.2 Importance of Computer Science and Programming

As the field is new and emerging, the current Data Scientists have come from various different backgrounds with their current title being a Data Scientist. Due to the emerging nature of this role in South Africa, Universities and colleges have only recently within the past 3 years created some form of Data Science courses aside from the online courses available. DS 3 who is currently 29 years old mentioned he wished he could have studied Data Science and Computer Science sooner as he was previously in an Actuarial role.

Due to the Mathematics background of the Data Scientists interviewed, four of the Data Scientists felt that the gaps were the lack of a Computer Science background. Computer Science enables problems to be solved in an effective manner (Dinter et al., 2015). Even though the Data Scientists know how to program, Computer Science provides one with a background on Software Engineering skills such as Distributed Computing, how to work with different data structures and algorithms so these skills might be lacking (Kim & Lee, 2016).

DS 2 and DS 5 both mentioned they would like to do more Computer Science courses as it was not part of their Mathematical Statistics degree that they had studied. DS 2 stated, "I wish I had done more Computer Science and Programming modules in my undergrad". This statement emphasises the need for Computer Science maybe to be taken as a compulsory subject within the Science degree depending on the career path of a student. DS 5 stated," I think Computer Science would have helped me write efficient code, meaning that the runtime of the program is faster." These statements highlight the need for the Computer Science background for Data Scientists. Two other Data Scientists mentioned that Computer Science should have been part of their degree.

Computer Science and Programming was a required skill for Data Scientists in the literature (Chatfield et al., 2014; Miller, 2014; Carter & Sholler, 2016; Dhar, 2013; Kim & Lee, 2016; Brynjolfsson et al., 2011; McAfee et al., 2012). Patil and Davenport (2012) state "Data Scientists' most basic, universal skill is the ability to write code", highlighting the importance of this skill.

5.7.3 Importance of Statistics

DS 4 had a background in Computational and Applied Mathematics. When asked what he wished he could have studied, he mentioned more Statistics. Other interviewees had a background in Statistics which made the transition from their previous fields into Data Science easier. Data Scientists should have the ability to apply analytic methods on data such as being able to build predictive and prescriptive models to solve problems the organisation has (Chatfield et al., 2014). Through the use of Analytics, discoveries can be made in the data (Abbasi et al., 2016; Chatfield et al., 2014; Chiang et al., 2012; Davenport, 2012; Davis, 2014; Dinter et al., 2015; Kim & Lee, 2016; Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015; Waller & Fawcett, 2013; Wamba et al., 2015; Wixom et al., 2014).

5.8 Recommendations from Data Scientists

All but one of the Data Scientists had migrated from previous roles before taking on the position of a Data Scientist. Data Scientists were asked in the interview to make certain recommendations. Data Scientists were asked the following:

5.8.1 Who should be trained?

Organisations are looking for Data Scientists to take advantage of Big Data opportunities (Chen et al., 2012). It is difficult to find someone with all the competencies that are listed in this study. One of the questions that were posed to the Data Scientist is "who should be trained to migrate into this role?" Most universities are still in the early phases of defining the curriculum on what courses should be studied in Data Science and are not yet producing graduates trained specifically for the role of Data Scientist. Due to this, and the need for experience in certain fields, many Data Scientists have migrated from other backgrounds. The Data Scientists in this study have been in the role for 3 years and less and have migrated from different fields. The question was then asked who they thought could be potential Data Scientists in the future as demand could be expected to grow and the migration route would probably be a popular one for the near future.

The Data Scientists did not mention Business Analysts to migrate into the Data Scientist role. Four of the competencies that were identified in this study concur with a Business Analyst based on the type of skills that are needed for a Data Scientist role which were the Business Requirements Elicitation and Specification competency, Business Knowledge, Financial Industry Knowledge and Communication. The Business Analysts would have the business background which an important skill but would not necessarily have the Technical competencies and the Data Scientist Fundamental Competency which are crucial to being a Data Scientist.

DS 1 felt that anyone could be a Data Scientist as it should not be conformed to a specific job title. DS 2 felt that Business Intelligence Analysts and Developers should be trained. This statement did not align to the study done by Debortoli et al. (2014) who analysed the role of Business Intelligence Analysts and Data Scientists who were referred to as Big Data Analysts in their study. The role of the Big Data Analyst was more specialised when compared to the Business Intelligence Analyst role even though they both required the Analytic, Mathematical and Statistical skills (Gupta et al., 2015; Watson, 2009). The Developers who could migrate into the Data Scientist role already have a background in Programming which was one of the main skills mentioned for a Data Scientist according to Patil and Davenport (2012). The difference would be the Organisational Knowledge, Business Interpersonal Competency and key competencies such as Mathematics and Statistics mentioned within the Data Scientist Fundamental competency category. It is interesting that Developers were viewed as more plausible migrators than Business Analysts, despite neither group having the more mathematical background.

DS 3 identified Business Intelligence Analysts and Developers as well as Statisticians who work with Statistical programming languages such as R. DS 4 mentioned a list of job titles that could be trained to be Data Scientists and included Computer Engineers, Statisticians, Industrial Engineers, Econometric Analysts, Geneticists and Bio Engineers. DS 6 mentioned that Business Intelligence Analysts and Statisticians could be trained to become Data Scientists. This could be a starting point for organisations to consider when looking for this difficult talent even though some of the competencies might be missing, these skills could be picked up through training and development.

Carew (2017) who concurred with studies such as Debortoli et al. (2014) mentioned that even though certain competencies for both Business Intelligence Analysts and Data Scientists overlap, there are key differences between the roles (Carew, 2017). Data Exploration, Data Management, Curiosity and Creativity, and Data Analytics and Visualisation were the competencies that were seen as the major differences between the roles. Business Intelligence Analysts do not have the strong expertise in Analytics when compared to a Data Scientists as well as being able to work with the full data lifecycle from data sourcing, data transformation, data manipulation and other skills related to Data Management, Big Data Processing and the Data Processing competency. As stated by Gartner in Patil & Davenport (2012), which further cements the need for each competency category defined in the Data Scientists Competency model developed in this study:

"It's essential to understand what role this person is going to fulfil and think about hiring a Data Scientist as an investment in the future of your business. Data Scientists are constantly asking 'why'. Great Data Scientists apply new tools and techniques to solve real problems. And they aren't afraid to get their hands dirty – to experiment, fail, recalibrate and try again. You don't actually need mountains of data, but you do need the right people with the right skills to interpret the data you do have. Businesses need Data Scientists with business acumen, individuals who can identify the key data insights needed to take your business to the next level."(Carew, 2017)

5.8.2 What type of Education should they have?

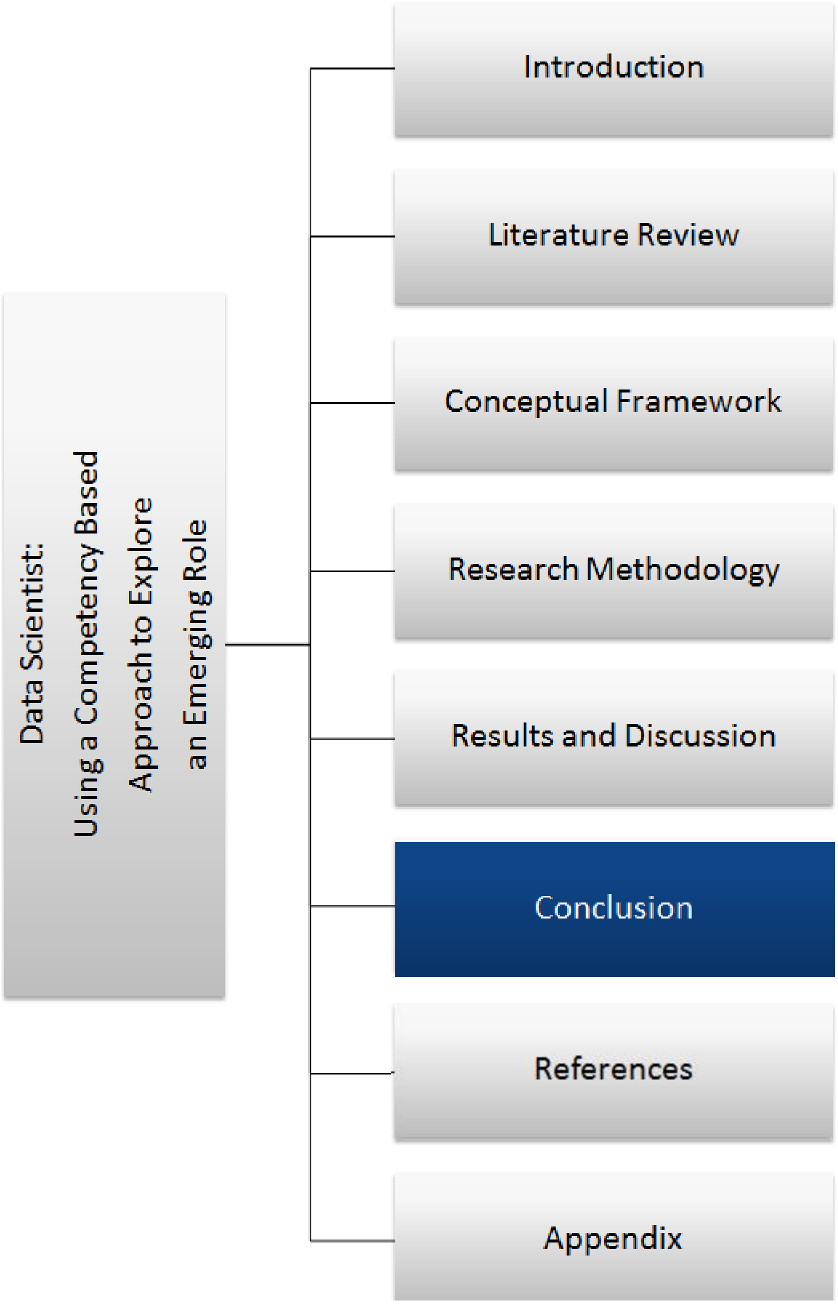
According to the Data Scientists, a Data Scientist needs to a have Mathematics and Statistics degree as well as some background in Computer Science. Patil and Davenport (2012), state that most of the Data Scientists that are in organisations were trained in Mathematics, Statistics or Computer Science. This statement is in line with the statements from the Data Scientists on the type of education they have and also in the statements they made on the education requirements for future Data Scientists.

DS 1 mentioned Mathematics, Statistics and Computer Science. DS 2 mentioned a postgraduate degree in the Science field with strong Mathematics and Statistics as well as Programming components. DS 3 mentioned Programming, Mathematics, Statistics, Data visualisation and design. DS 4 mentioned a postgraduate degree in Mathematics, Computer Science and Statistics while DS 5 and 6 mentioned Computer Science and Mathematical Statistics. DS 5 further mentioned Informatics and Information Science. Overall, Mathematics, Statistics and Computer Science were the subjects that were common in the responses from the Data Scientists.

One knowledge and skill that was mentioned in the literature review was Business Knowledge. The Data Scientists in this study had a good business background gained from experience in the industry with the exception of DS 5 who has only been in the industry for 1 year. One of the gaps in hiring a good Data Scientist is the Business Knowledge which is seen as a rare skill to find in a Data Scientist according to Watson (2014). Business Knowledge is crucial in understanding the problems that business has and whether or not the solutions created by the Data Scientist will be of value (Chiang et al., 2012). Business Knowledge can be gained through integrating subjects from a Commerce degree into a Science degree to enable future Data Scientists to get a background in subjects such as Economics, Business Management and Business Information Systems.

5.9 Summary

This chapter summarised the results of the data that was gathered from the Data Scientists. A role for a Data Scientist was defined within the banking and insurance industry and thereafter, the competencies which are needed to fulfil that role are described. Competencies were grouped according to the categories defined in Sonteya and Seymour's research (2012). A competency model for a Data Scientist was completed and is shown in Figure 10 in this chapter. Gaps in competencies were then discussed followed by recommendations made by the Data Scientists on the type of roles that can migrate into the Data Science field as well as the education needs of those who want to pursue a career in Data Science.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

6 CONCLUSION

Data Science is a new and exciting field that requires individuals possess specific skills to solve problems. The role of Data Scientists is not yet fully defined in literature and practice. Organisations that want to take advantage of opportunities from using Big Data firstly need to employ Data Scientists and the main problem is trying to find Data Scientists with the relevant competencies (Davenport, 2012). It is difficult for organisations to hire the right person which is resulting in a skills shortage. Organisations tend to create a role of a Data Scientist that is specific to their needs (Kim & Lee, 2016). The aim of this study was to explore the role of a Data Scientist within the banking and insurance industry to determine the role that is currently being played by Data Scientists and the competencies Data Scientists should have.

6.1 Research Questions and Findings

Three research questions formed the base of this research.

Research question 1:

- *What is the emerging role of the Data Scientist in South Africa (in the banking and insurance industries)?*

The role of the Data Scientist has been explored in the results section through the thematic analysis of data gathered from a sample of Data Scientists within South Africa. The defined role although in line with what has currently been reviewed in the literature, provides more in depth understanding of the role due to the specific nature of this study which was within the banking and insurance industry. The role for a Data Scientist within the banking and insurance industry can be summarised as:

A Data Scientist is someone who understands business needs and is able to find patterns in the data through the use of data sourcing and tools to analyse data, automate business processes and answer business questions in large data sets to make informed decisions and provide actionable insight. A Data Scientist also creates and maintains predictive models and is able to answer questions that have not been asked which leads to the customisation of products, enhancement of customer behaviour and engagement and an improvement in sales and profitability.

Research question 2:

- *What skills and competencies do Data Scientists currently working in South Africa possess?*

The skills, knowledge and attributes of Data Scientists were explored in this research. The resulting competencies can be seen in Appendix R. The competencies of Data Scientists are as follows grouped according to the competency category:

Within the **Technical Competency** category, the competencies that were identified were Big Data Processing, Data Processing and Web Tools. Each of these competencies focused on the set of knowledge and skills that a Data Scientist should have to perform their role (Kim & Lee, 2016; Song & Zhu, 2016). Within the Big Data Processing competency, Data Scientists should be able to work with tools such as Hadoop, R, Machine Learning, Python, Artificial intelligence, Facebook Marketing, Natural Language Processing and Neural Networks. Within the Data Processing competency, Data Scientists should be able to work with tools such as SQL, Excel, Java, Matlab, Qlikview, Power BI and Tableau. Within the Web Tools competency tools such Bootstrap, CSS, Django, HTML and Javascript were used by one Data Scientists. These knowledge and skills are deemed of a lesser importance than the knowledge and skills identified in other competencies within the Technical Competency category.

Within the **Data Science Orchestration Competency** category, the competencies that were identified were Process Improvement, Recommendations to Business, Problem Solving and Analysis and Delivers Results. Each of these competencies focused on the set of knowledge and skills that a Data Scientist should have in order to provide the actions needed to deliver optimal and accurate results to business (Provost & Fawcett, 2013; Schoenherr & Speier-Pero, 2015). Within the Process Improvement competency, a Data Scientist should be able to Deliver new models, Automate processes, Create new processes, and Dynamically adjust solutions and Optimise processes. Within the Recommendations to Business competency a Data Scientist should Answer Business Questions, provide Actionable Insights, Answer Questions not asked, Drive Results and Change, and Help management understand profitability and Validate Business Assumptions. Within the Problem Solving and Analysis competency, the knowledge and skills a Data Scientist should have are Problem Identification, Problem Solving and Solution Design. Within the Delivers Results competency, the results that a Data Scientist provides should Enhance Customer Behaviour, Enhance Customer Experience, Improve Profitability, and Predict Customer Behaviour and User experience.

Within the **Organisational Knowledge** category, the competencies that were identified were Business Knowledge and Financial Industry Knowledge. These competencies focused on the application of the Data Scientists' extensive and specialised knowledge and skills to accomplish the organisations' objectives effectively (Wamba et al., 2015). Business Knowledge is the ability of a Data Scientist to understand the different processes that occur within the specific organisations context (Wixom et al., 2014). Data Scientists also need to understand the financial processes and procedures as well as having knowledge on all products and services in the industry that they reside in (Waller & Fawcett, 2013).

Within the **Business Interpersonal Competency** category, the competencies that were identified were Curiosity and Creativity, Communication, Business Requirements Elicitation and Specification, Independence and Adaptability. Within the Curiosity and Creativity competency, a Data Scientist needs to be Inquisitive, Creative, Curious, Intuitive, Like Challenges, Passion working with data and Think out of the box. Within the Communication competency, a Data Scientist should have the Ability to Communicate, Presentation Skills and Convey business insights. Within the Business Requirements Elicitation and Specification, a Data Scientist should ask the right questions; Better understand requirements and Interpret Queries. The Independence competency was not mentioned in the literature and might be of a lesser importance than other competencies within this category. Within the Adaptability competency, the Data Scientist should have a Resilient Personality, Flexibility, and Pragmatism and be willing to learn new things.

Within the **Data Scientist Fundamental Competency** category, the competencies that were identified were Data Management, Mathematics and Statistics, Data Analytics and Visualisation, Computer Science and Programming, Data Exploration, Critical Thinking and IT and Information Systems. Within the Data Management competency a Data Scientist should be able to perform the following tasks; Data Manipulation, Data Sourcing, Data Transformation, Database Design, Data Design, Data Mining, Data Modelling, Data Queries and Database Management. Within the Mathematics and Statistics competency were the knowledge and skills Mathematics and Statistics. Within the Data Analytics and Visualisation competency, the knowledge and skills were Data Analysis, Predictive Models, Analytics, Prescriptive Models, Data Visualisation and Business Intelligence. Within the Computer Science and Programming competency, the knowledge and skills were Programming and Computer Science. Within the Data Exploration competency, the knowledge and skills were Exploring data, Extract meaning from Data and Recycle data to predict future outcomes. Within the Critical Thinking competency, the knowledge and skills were Focused, Persistent, Attention to Detail and Critical Thinking. Within the IT and Information Systems competency the knowledge and skills were IT skills, Informatics, Information Science, Keep ahead with Data Science Techniques and Systems Knowledge.

Research question 3:

- *What is the gap (if any) between required and current Data Scientist skills and competencies?*

The gaps identified in this study were firstly the importance of Communication. If Data Scientists could effectively translate their ideas and insights to business better their ideas could be more readily discussed and used by business which could result in an increase in profitability and an improvement in customer behaviour (Kim and Lee, 2016). The second gap was the importance of Computer Science to help Data Scientists in effectively Programming solutions (Chatfield et al., 2014; Miller, 2014; Carter & Sholler, 2016; Dhar, 2013; Kim & Lee, 2016; Brynjolfsson et al., 2011; McAfee et al., 2012). The third gap was the importance of Statistics to help Data Scientists build predictive models and to have the Statistical background needed.

6.2 Migration of Roles

It is difficult to find someone with all the skills and competencies that are listed in this study. One of the questions that were posed to the Data Scientists is who should be trained as most universities are in the early phases of defining the curriculum on what courses should be studied in Data Science. Due to this, and the need for experience in certain fields, many Data Scientists have migrated from other backgrounds. The Data Scientists in this study have only been in the role for 3 years and less and have migrated from different fields. The Data Scientists interviewed did not mention Business Analysts to migrate into the Data Scientist role although four of the competencies that were identified in this study concur with Business Analysts which were Business Requirements Elicitation and Specification competency, Business Knowledge, Financial Industry Knowledge and Communication. The Business Analysts would have the business background which an important skill but would not have the Technical competencies and the Data Scientist Fundamental Competency which are crucial to being a Data Scientist. Courses could be offered which could specifically cater to these competencies such as Big Data Processing, Computer Science and Programming, Critical Thinking, Data Analytics and Visualisation, Data Exploration, Data Management, Data Processing, IT and Information Systems and Mathematics and Statistics.

Data Scientists interviewed felt that code Developers could be trained to become Data Scientists. The developers who could possibly migrate into the Data Scientist role already have a background in Programming which is one of the main skills mentioned for a Data Scientist according to Patil and Davenport (2012). The difference would be the Organisational Knowledge, Business Interpersonal Competency and key competencies such as Mathematics and Statistics mentioned within the Data Scientist Fundamental competency category. Courses could be offered which could address this gap with competencies such as Adaptability, Business Knowledge, Business Requirements Elicitation and Specification, Communication, Critical Thinking, Curiosity and Creativity, Data Analytics and Visualisation, Delivers Results, Financial Industry Knowledge, Independence, Mathematics and Statistics, Problem Solving and Analysis, Process Improvement and Recommendations to Business.

Data Scientists interviewed also felt that Business Intelligence Analysts could be trained to become Data Scientists. Carew (2017) who concurred with studies such as Debortoli et al. (2014) mentioned that even though certain competencies for both Business Intelligence Analysts and Data Scientists overlap, there are key differences between both roles (Carew, 2017). The skills that overlap are the specific skills such as Mathematics, Statistics and the use of data visualisation tools. Data Exploration, Data Management, Curiosity and Creativity, and Data Analytics were the competencies that were seen as the major differences between both roles. Business Intelligence Analysts don't have strong expertise in Analytics when compared to a Data Scientists as well as being able to work with the full data lifecycle from data sourcing, data transformation, data manipulation and other skills related to Data Management and Big Data Processing. Courses could be created to cater for this gap in knowledge. The roles mentioned could be a starting point for organisations to consider when looking for this difficult talent even though some of the competencies might be missing, these skills could be developed through training and education.

6.3 Contributions of this Study

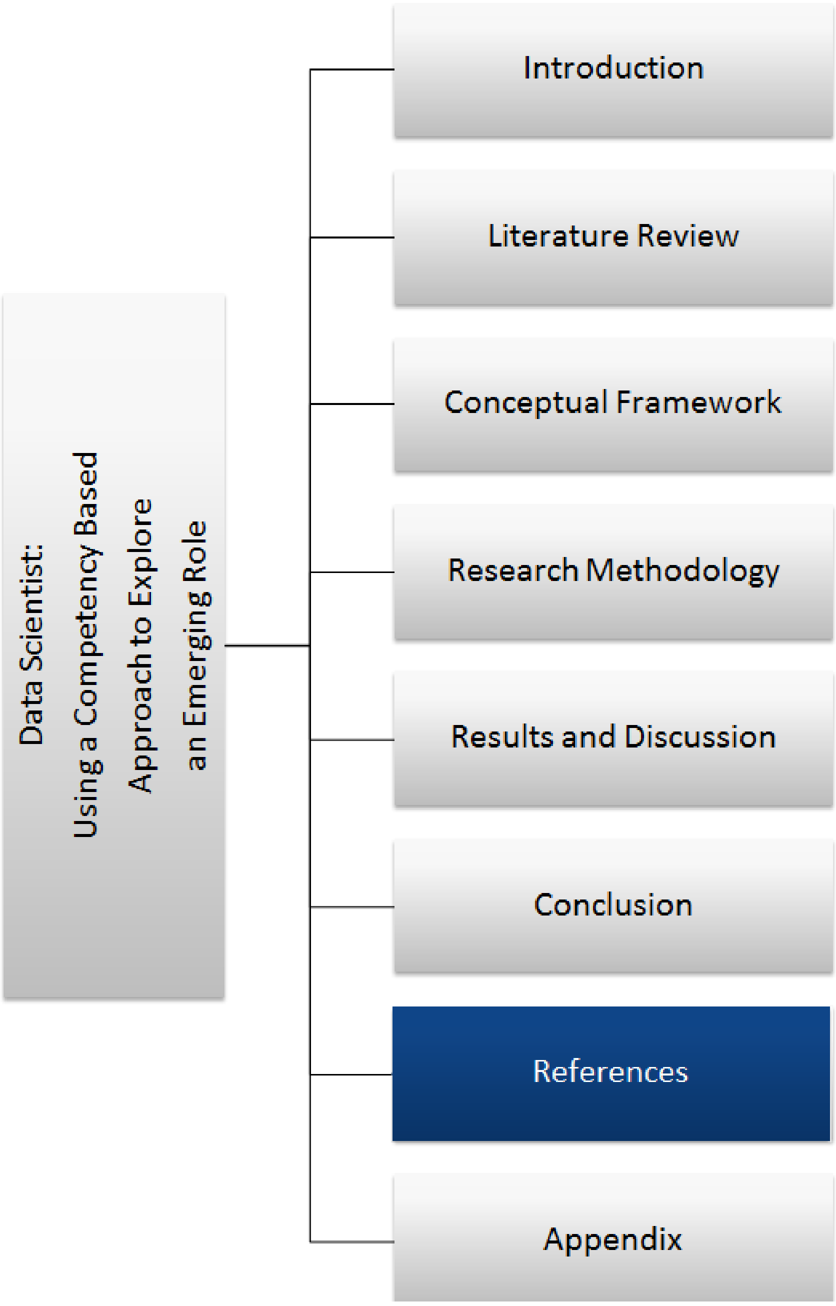
The roles and competencies examined in this research add to the emerging Data Science body of knowledge in South Africa. The findings of this research will further benefit organisations when attempting to hire Data Scientists due to the difficulties faced to acquire certain skills. This could potentially lead to an increase in adoption of Big Data technologies in South Africa. The Data Scientist competencies identified in this research can also provide guidance in the development of skills and in education for students who wish to pursue career in Data Science. Educational institutions can cater to these needs by offering degrees based on the role and competencies examined in this research which could potentially decrease the skill shortages of Data Scientists.

6.4 Future Research

Future research can use larger sample sizes to get a broader view of skills as a sample of only six Data Scientists was used for this research. This research only focused on Data Scientists within the Banking and insurance industries. Data Scientists could be researched within different industries to determine skills transferability of Data Scientists.

6.5 Summary

This research aimed to address the challenges faced by organisations when hiring Data Scientists by answering the research questions on the role and competencies of Data Scientists in South Africa. This qualitative study also highlighted gaps in research through an exploratory study of existing literature which defined a role of a Data Scientist as well as assessing the required skills and level of education and the current skills and level of education. The research will add to the emerging Data Science body of knowledge in South Africa by having identified the role and competencies of Data Scientists. The Data Scientist competencies examined in this research could also provide guidance for those who want to pursue a career in Data Science by highlighting education and skill needs. Current roles were also discussed which could migrate into the Data Scientist role. Educational institutions could specifically cater to these needs by offering degrees based on the skills and competencies examined in this research which could potentially decrease the skill shortages which is faced in South Africa. And lastly, the competencies identified in this research strongly align to competencies that were created in developed countries.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

7 REFERENCES

1. Abbasi, A., Sarker, S. and Chiang, R.H., 2016. Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17(2), p.3 – 34
2. ABSA. 2015. Absa Life pioneers predictive medical underwriting solution using Big Data. [ONLINE] Available at: https://www.absa.co.za/media-centre/press-statements/absa-life-pioneers. [Accessed 1 November 2017].
3. Bassellier, G. and Benbasat, I., 2004. Business Competence of Information Technology Professionals: Conceptual Development and Influence on IT-Business Partnerships, *MIS Quarterly* (28:4), pp. 673-694.
4. Bhattacherjee, A., 2012. Social science research: principles, methods, and practices.
5. Boyatzis, R.E., 1998. *Transforming qualitative information: Thematic analysis and code development*. sage.
6. Brynjolfsson, E., Hammerbacher, J. and Stevens, B., 2011. Competing through data: Three experts offer their game plans. *McKinsey Quarterly*, 4, pp.36-47.
7. Cai, L. and Zhu, Y., 2015. The challenges of data quality and data quality assessment in the Big Data era. *Data Science Journal*, *14*, p.2.
8. Campion, M.A., Fink, A.A., Ruggeberg, B.J., Carr, L., Phillips, G.M. and Odman, R.B., 2011. Doing competencies well: Best practices in competency modeling. *Personnel Psychology*, *64*(1), pp.225-262.
9. Carew, J., 2017. King Shaka: The earliest Data Scientist. [ONLINE] Available at: http://www.brainstormmag.co.za/features/14183-king-shaka-the-earliest-data-scientist. [Accessed 22 February 2018].
10. Carter, D. and Sholler, D., 2016. Data science on the ground: Hype, criticism, and everyday work. *Journal of the Association for Information Science and Technology*, *67*(10), pp.2309-2319.
11. Chatfield, A.T., Shlemoon, V.N., Redublado, W. and Rahman, F., 2014. Data Scientists as game changers in Big Data environments.
12. Chen, H., Chiang, R.H. and Storey, V.C., 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, *36*(4), pp.1165-1188.
13. Chiang, R.H., Goes, P. and Stohr, E.A., 2012. Business intelligence and Analytics education, and program development: a unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems* (TMIS),3(3), p.12.
14. Collis, J. and Hussey, R., 2003. *Business research: A practical guide for undergraduate and postgraduate students*. Palgrave Macmillan.
15. Creswell, J.W., 2013. *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications.

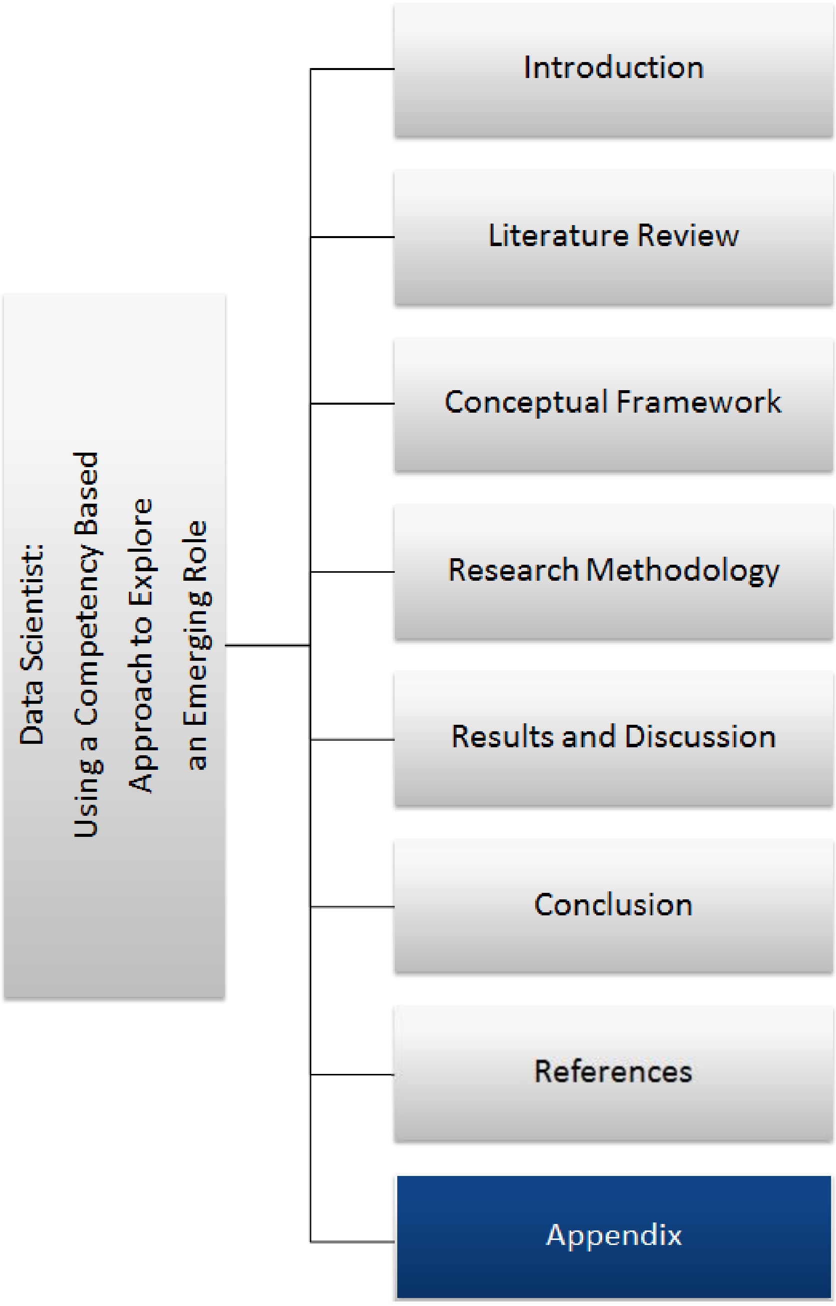
16. Davenport, T.H., 2012. The human side of Big Data and high-performance analytics. *International Institute for Analytics*, pp.1-13.
17. Davenport, T.H. and Dyché, J., 2013. Big Data in big companies. *International Institute for Analytics, May*.
18. Davis, C.K. 2014. , Beyond data and analysis. *Communications of the ACM,* 57, 6 (June 2014), 39-41. DOI=http://dx.doi.org/10.1145/2602326
19. De Mauro, A., Greco, M. and Grimaldi, M., 2015, February. What is Big Data? A consensual definition and a review of key research topics. In*AIP Conference Proceedings* (Vol. 1644, No. 1, pp. 97-104).
20. Debortoli, S., Müller, O. and vom Brocke, J., 2014. Comparing business intelligence and Big Data skills. *Business & Information Systems Engineering*, *6*(5), pp.289-300.
21. Demchenko, Y., Belloum, A., Los, W., Wiktorski, T., Manieri, A., Brocks, H., Becker, J., Heutelbeck, D., Hemmje, M. and Brewer, S., 2016, December. EDISON data science framework: a foundation for building data science profession for research and industry. *In Cloud Computing Technology and Science (CloudCom), 2016 IEEE International Conference* on (pp. 620-626). IEEE.
22. Dhar, V., 2013. Data science and prediction. *Communications of the ACM*, *56*(12), pp.64-73.
23. Dinter, B., Douglas, D., Chiang, R.H., Mari, F., Ram, S. and Schoder, D., 2015. Big Data Panel at SIGDSS Pre-ICIS Conference 2013: A Swiss-Army Knife? The Profile of a Data Scientist. In *Reshaping Society through Analytics, Collaboration, and Decision Support* (pp. 7-11). Springer, Cham.
24. Draganidis, F. and Mentzas, G., 2006. Competency based management: a review of systems and approaches. *Information management & computer security*, 14(1), pp.51-64.
25. Gioia, D.A. and Pitre, E., 1990. Multiparadigm perspectives on theory building. *Academy of management review*, 15(4), pp.584-602.
26. Gupta, B., Goul, M. and Dinter, B., 2015. Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs. *Communications of the Association for Information Systems*, 36(1), p.23.
27. Halaweh, M. and Massry, A.E., 2015. Conceptual Model for Successful Implementation of Big Data in Organizations. *Journal of International Technology and Information Management, 24(*2), p.2.
28. Harrington, S.J., 1996. The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. *MIS quarterly*, pp.257-278.
29. Harvard University, 2015, Harvard University Competency Dictionary, [ONLINE] Available at: https://apps2.campusservices.harvard.edu/cas/empldocs/cs/harvard_competency_dictionary_complete.pdf. [Accessed 12 December 2017]

30. Hawk, S., Kaiser, K.M., Goles, T., Bullen, C.V., Simon, J.C., Beath, C.M., Gallagher, K.P. and Frampton, K., 2012. The information technology workforce: A comparison of critical skills of clients and service providers. *Information Systems Management*, 29(1), pp.2-12.
31. Hu, H., Wen, Y., Chua, T.S. and Li, X., 2014. Toward scalable systems for Big Data Analytics: A technology tutorial. *IEEE Access*, *2*, pp.652-687.
32. Hussey, J., & Hussey, R. (1997). *Business Research*. Palgrave: New York.
33. Isik, Ö., Jones, M.C. and Sidorova, A., 2013. Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, 50(1), pp.13-23.
34. Nestler, S., Levis, J. and Klimack, B, 2012, Certified Analytics Professional. [ONLINE] Available at: https://www.informs.org/ORMS-Today/Public-Articles/October-Volume-39-Number-5/Certified-Analytics-Professional. [Accessed 15 November 2017]
35. Kamioka, T. and Tapanainen, T., 2014. Organizational Use of Big Data and Competitive Advantage-Exploration of Antecedents. In PACIS (p. 372).
36. Kim, J.Y. and Lee, C.K., 2016. An empirical analysis of requirements for Data Scientists using online job postings. *International Journal of Software Engineering and Its Applications,* 10(4), pp.161-172.
37. Kim, M., Zimmermann, T., DeLine, R. and Begel, A., 2016, May. The emerging role of Data Scientists on software development team s. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 96-107). ACM.
38. Lee, S.M. and Lee, C.K., 2006. IT managers' requisite skills. *Communications of the ACM*, 49(4), pp.111-114.
39. Lewis, N, 2017, Mind the Data Science Skills Gap. [ONLINE] Available at: http://capgemini.ft.com/closer-look/mind-the-data-science-skills-gap_a-48-2500.html. [Accessed 20 May 2017].
40. Malinga, S. 2016. Barclays Pronto helps identify clients' sentiments. [ONLINE] Available at: http://www.itweb.co.za/index.php?option=com_content&view=article&id=150658. [Accessed 16 August 2017].
41. Mamabolo, M. 2017. Shortage of data science skills offers Africa an opportunity. [ONLINE] Available at: http://www.itwebafrica.com/business-intelligence/507-africa/237359-shortage-of-data-science-skills-offers-africa-an-opportunity. [Accessed 12 May 2017].
42. Manieri, A., Brewer, S., Riestra, R., Demchenko, Y., Hemmje, M., Wiktorski, T., Ferrari, T. and Frey, J., 2015, November. Data Science Professional uncovered: How the EDISON Project will contribute to a widely accepted profile for Data Scientists. In *Cloud Computing Technology and Science (CloudCom), 2015 IEEE 7th International Conference on* (pp. 588-593). IEEE.

43. Massachusetts Institute of Technology, 2015, Data Scientist Position Description, [ONLINE] Available at: https://ist.mit.edu/sites/default/files/about/org/roles/Data_Scientist_Position_Description_v4.pdf [Accessed 20 May 2017]
44. Matook, S. and Maruping, L.M., 2014. A competency model for customer representatives in agile software development projects. *MIS Quarterly Executive*, 13(2).
45. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D., 2012. Big Data. *The management revolution. Harvard Bus Rev*, *90*(10), pp.61-67.
46. McClelland, D.C., 1973. Testing for competence rather than for" intelligence." *American psychologist*, *28*(1), p.1.
47. Miller, S., 2014. Collaborative Approaches Needed to Close the Big Data Skills Gap. *Journal of Organization Design,* 3(1), pp.26-30.
48. Moyo, A. 2015. *Data Scientists in short supply*. [ONLINE] Available at: http://www.itweb.co.za/index.php?option=com_content&view=article&id=145199. [Accessed 10 April 2017].
49. Ozkose, H., Ari, E.S. and Gencer, C., 2015. Yesterday, Today and Tomorrow of Big Data.*Procedia-Social and Behavioral Sciences*, *195*, pp.1042-1050.
50. Patil, D.J. and Davenport, T. H., 2012. Data Scientist: the sexiest job of the 21st century. *Harvard Business Review.* 90(10), pp.70-76
51. Poleto, T., de Carvalho, V.D.H. and Costa, A.P.C.S., 2015, May. The Roles of Big Data in the Decision-Support Process: An Empirical Investigation. In *International Conference on Decision Support System Technology* (pp. 10-21). Springer International Publishing.
52. Power, D.J., 2014. Using 'Big Data' for Analytics and decision support. *Journal of Decision Systems*, *23*(2), pp.222-228.
53. Provost, F. and Fawcett, T., 2013. Data science and its relationship to Big Data and data-driven decision making. *Big Data*, *1*(1), pp.51-59.
54. Reddy, V., Bhorat, H., Powell, M., Visser, M.M. and Arends, F., 2016. Skills supply and demand in South Africa.
55. Russom, P., 2011. Big Data Analytics. *TDWI best practices report, fourth quarter*, *19*, p.40.
56. Sagiroglu, S. and Sinanc, D., 2013, May. Big Data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE.
57. Saldana, J., 2016. *The coding manual for qualitative researchers*. Sage.
58. Saltz, J.S., 2015, October. The need for new processes, methodologies and tools to support Big Data teams and improve Big Data project effectiveness. In *Big Data (Big Data), 2015 IEEE International Conference on* (pp. 2066-2071). IEEE.
59. Saunders, M., Lewis, P. & Thornhill, A. (2009*). Research Methods for Business Students*. Prentice Hall: London.

60. Schoenherr, T. and Speier-Pero, C., 2015. Data science, Predictive Analytics, and Big Data in supply chain management: Current state and future potential. *Journal of Business Logistics*, *36*(1), pp.120-132.
61. Sekaran, U. and Bougie, R. (2013) Research Methods for Business: A Skill-Building Approach.
62. Shah, D. 2017. Data Scientist Skills for Success. [ONLINE] Available at: https://upside.tdwi.org/Articles/2017/02/14/Data-Scientist-Skills-for-Success.aspx?Page=1. [Accessed 16 May 2017].
63. Silverman, D., 2015. *Interpreting qualitative data*. Sage.
64. Song, I.Y. and Zhu, Y., 2016. Big Data and data science: what should we teach?. *Expert Systems*, *33*(4), pp.364-373.
65. Sonteya, T. and Seymour, L.F., 2012. Towards an understanding of the business process analyst: An analysis of competencies. *Journal of Information Technology Education: Research*, 11, pp.43-63.
66. Strengell, T., 2017. Competitiveness from data and Analytics: required competency in organization.
67. Todd, P.A., McKeen, J.D. and Gallupe, R.B., 1995. The evolution of IS job skills: a content analysis of IS job advertisements from 1970 to 1990. *MIS quarterly*, pp.1-27.
68. Trifu, M.R. and Ivan, M.L., 2014. Big Data: present and future. *Database Systems Journal*, 5(1), pp.32-41.
69. Van der Aalst, W.M., 2014. Data Scientist: The engineer of the future. In *Enterprise Interoperability VI* (pp. 13-26). Springer, Cham.
70. Vahn, G.Y., 2014. Business Analytics in the age of Big Data. Business Strategy Review, 25(3), pp.8-9.
71. van Zyl, G, Fin24. 2014. SA's finance sector banking more on Big Data. [ONLINE] Available at: http://www.fin24.com/Tech/Companies/SAs-finance-sector-banking-more-on-big-data-20141207. [Accessed 20 March 2017]
72. Viitala, R., 2005. Perceived development needs of managers compared to an integrated management competency model. *Journal of workplace learning*, 17(7), pp.436-451.
73. Waller, M.A. and Fawcett, S.E., 2013. Data science, Predictive Analytics, and Big Data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), pp.77-84.
74. Walsham, G., 1995. Interpretive case studies in IS research: nature and method. *European Journal of information systems*, 4(2), pp.74-81.
75. Wamba, S.F., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D., 2015. How 'Big Data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, pp.234-246.

76. Watson, H.J. (2009). Business Intelligence – Past, Present, and Future. *Communications of the AIS*, 25(1), 487-510.
77. Watson, H.J., 2014. Tutorial: Big Data Analytics: Concepts, technologies, and applications. *Communications of the AIS*, *34*, p.65.
78. Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., Kulkarni, U., Mooney, J.G., Phillips-Wren, G. and Turetken, O., 2014. The current state of business intelligence in academia: The arrival of Big Data. *Communications of the Association for Information Systems*, *34*(1), p.1.
79. World Economic Forum, 2015. *New vision for education: Unlocking the potential of technology*. British Columbia Teachers' Federation.
80. Yin, S. and Kaynak, O., 2015. Big Data for Modern Industry: Challenges and Trends [Point of View]. *Proceedings of the IEEE*, *103*(2), pp.143-146.
81. Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J. and Corrigan, D., 2012. *Harness the power of Big Data: The IBM Big Data platform*. McGraw Hill Professional.
82. Ziora, A.C.L., 2015. The Role of Big Data Solutions in the Management of Organizations. Review of Selected Practical Examples. *Procedia Computer Science*, *65*, pp.1006-1012.


Data Scientist:
Using a Competency Based
Approach to Explore
an Emerging Role
Introduction
Literature Review
Conceptual Framework
Research Methodology
Results and Discussion
Conclusion
References
Appendix

8 APPENDIX

8.1 APPENDIX A: LITERATURE REVIEW 103
SEARCH STRINGS AND QUALITY CRITERIA 106
8.2 APPENDIX B: SUMMARY OF THEMES FOR DATA SCIENTIST ROLE FROM THE LITERATURE 108
8.3 APPENDIX C: LITERATURE REVIEW MAPPED TO RESEARCH MODEL 109
8.4 APPENDIX D: ETHICAL CLEARANCE CERTIFICATE 113
8.5 APPENDIX E: PARTICIPATION LETTER (SPL) 114
8.6 APPENDIX F: PARTICIPATION CONSENT FORM (SPCF) 115
8.7 APPENDIX G: CONSENT FORM FOR INTERVIEW RECORDING (CIR) 116
8.8 APPENDIX H: INTERVIEW QUESTIONS 117
8.9 APPENDIX I: EXAMPLE OF INTERVIEW DATA 120
8.10 APPENDIX J: A VIEW OF THE TRANSCRIBED DATA WITH CODES DONE IN EXCEL 122
8.11 APPENDIX K: SUMMARY OF THEMES FOR DATA SCIENTIST ROLE FROM THE RESULTS 123
8.12 APPENDIX L: COMPETENCY PROFILES FROM INTERVIEWS 124
8.13 APPENDIX M: TECHNICAL COMPETENCY CATEGORY 130
8.14 APPENDIX N: DATA SCIENCE ORCHESTRATION COMPETENCY CATEGORY 133
8.15 APPENDIX O: ORGANISATIONAL KNOWLEDGE CATEGORY 135
8.16 APPENDIX P: BUSINESS INTERPERSONAL COMPETENCY CATEGORY 136
8.17 APPENDIX Q: DATA SCIENTIST FUNDAMENTAL COMPETENCY CATEGORY 138
8.18 APPENDIX R: DATA SCIENTIST COMPETENCY MODEL TABLE 141
8.19 APPENDIX R: TURNITIN PAGE 144

8.1 Appendix A: Literature Review

No	Source	Purpose	Role of a Data Scientist	Skills and Competencies of a Data Scientist
1	Abbasi, Sarker & Chiang (2016)	Big Data is seen as a disruption to the value chain. Look at values and opportunities and put forward Big Data value chain	People who know how to get answers to important questions from a large amount of unstructured information	Machine Learning Hadoop Derive knowledge from internal and external data sources Work closely with analyst in the decision making phase
2	Brynjolfsson, Hammerbacher & Stevens (2011)	New research about the relationships between the data, Analytics, productivity and profitability	Working with very larger data sets, sampling methodologies, designing experiments	Skills of social scientist, programming skills, ability to do aggressive prioritisation, knowledge Statistics and Machine Learning
3	Carter & Sholler (2016)	Compares the discourse around data science to the workplace	Gathering and formatting data, constructing models, communicating analysis to members of their organisation and customers	Mathematician, product development and detective.
4	Chatfield, Shlemoon, Redublado & Rahman (2014).	To understand the emerging role of the Data Scientist and the skills needed	Working with data to answer questions that are used to make informed business decisions	Computer Science Communication Mathematics Data visualisation
5	Chen, Chiang & Storey (2012)	Challenges and opportunities within Business Intelligence and Analytics is analysed	Turn data and information into meaningful and actionable insights for an organisation.	Technical skills and system implementation, business or domain knowledge, communication skills
6	Chiang, Goes & Stohr(2012)	To develop the skills needed to interpret Big Data and take advantage of the opportunities derived from Big Data	Skill required to interpret Big Data	Analytical skills Information technology skills Business knowledge and communication skills
7	Davis (2014)	Looks at Big Data and Analytics from past to present	To implement and diffuse analytic methodologies throughout an organisation	
8	Davenport (2012)	To learn more about Data Scientist and activities, interviews were done.	Data Scientists understand Analytics and are skilled in exploring and using information technology	Technical, business, analytical and relationship skills and have advanced degrees in computer science.
9	Dhar (2013)	Discusses the meaning of Data science and the skills needed for a Data Scientist	Extraction of knowledge from data	Integrated Skill set for Mathematics,

				Machine Learning, artificial intelligence, Statistics, databases, and optimisation, problem formulation skills
10	Dinter et al. (2015)	Provide a forum for exchanging ideas in the data science and Big Data field to create a profile for a Data Scientist	Ability to take data and understand, process, extracts value, visualise it and communicate it.	Analytics skills, IT and programming, business and domain knowledge and interpersonal skills such as communicating.
11	Kim & Lee (2016)	Explore the definition of a Data Scientist by looking at 1240 jobs for Data Scientist	Defined as an expert that makes discoveries in large quantities of data to solve problems and provide insights for organisations.	Machine Learning Analytical skills Statistical skills Data mining Programming Business knowledge
12	McAfee et al (2012)	Looks at the challenges of becoming Big Data enabled	Skilled at working large quantities of data	Statistics Data mining Visualisation Communicate to business leaders
13	Miller (2014)	Universities and Industry that are in partnership need to ensure graduates have the required skills	Data Scientist role as an Information strategist, information system professional Data governance ethics	Mathematics Statistics Machine Learning Predictive Analytics Decision management Computer science Data ethics Information law Data mining
14	Patil & Davenport (2012)	To address the shortage of Data Scientists	A highly ranked professional having the necessary training and curiosity to make discoveries in Big Data	Hybrid of a data hacker, analyst, communicator and trusted adviser. Basic skills should be coding
15	Provost & Fawcett (2013)	Define principles underlining data science	View a business problem from the data side	Skills to access data, prepare it, process it using data tools
16	Schoenherr & Speier-Pero (2015)	How are Data Scientists trained in university and the skills needed for successful Data Scientist	Knowledge of data processes to create insights	Enterprise business processes, decision making, Data management, analytical modelling tools
17	Waller & Fawcett		Deep domain knowledge and broad analytic skill set	Deep domain knowledge and broad analytic skill

	(2013)			set
18	Wamba et al (2015)		Using new and unusual sources of data advanced technologies and visualisation, predictive Analytics and rare combinations of user skills.	Technical, analytical, business and governance skills, networking skills, communication skills, operationalising skill and implementation of analysis
19	Watson (2014)	Provides a tutorial on Big Data analytic concepts and technologies	To discover patterns and relationships no one has seen or thought about and turn the discovery into insights that can be actioned and create value for the organisation	Understand different types of data and its storage, write code, access data and analyse it Communicate findings to management through reports.
20	Wixom et al (2014)	Study based on a survey to determine the market needs for students with skills to analyse Big Data	Using new and unusual sources of data advanced technologies and visualisation, predictive Analytics and rare combinations of user skills.	Communication Programming Basic Analytics Data management Business knowledge

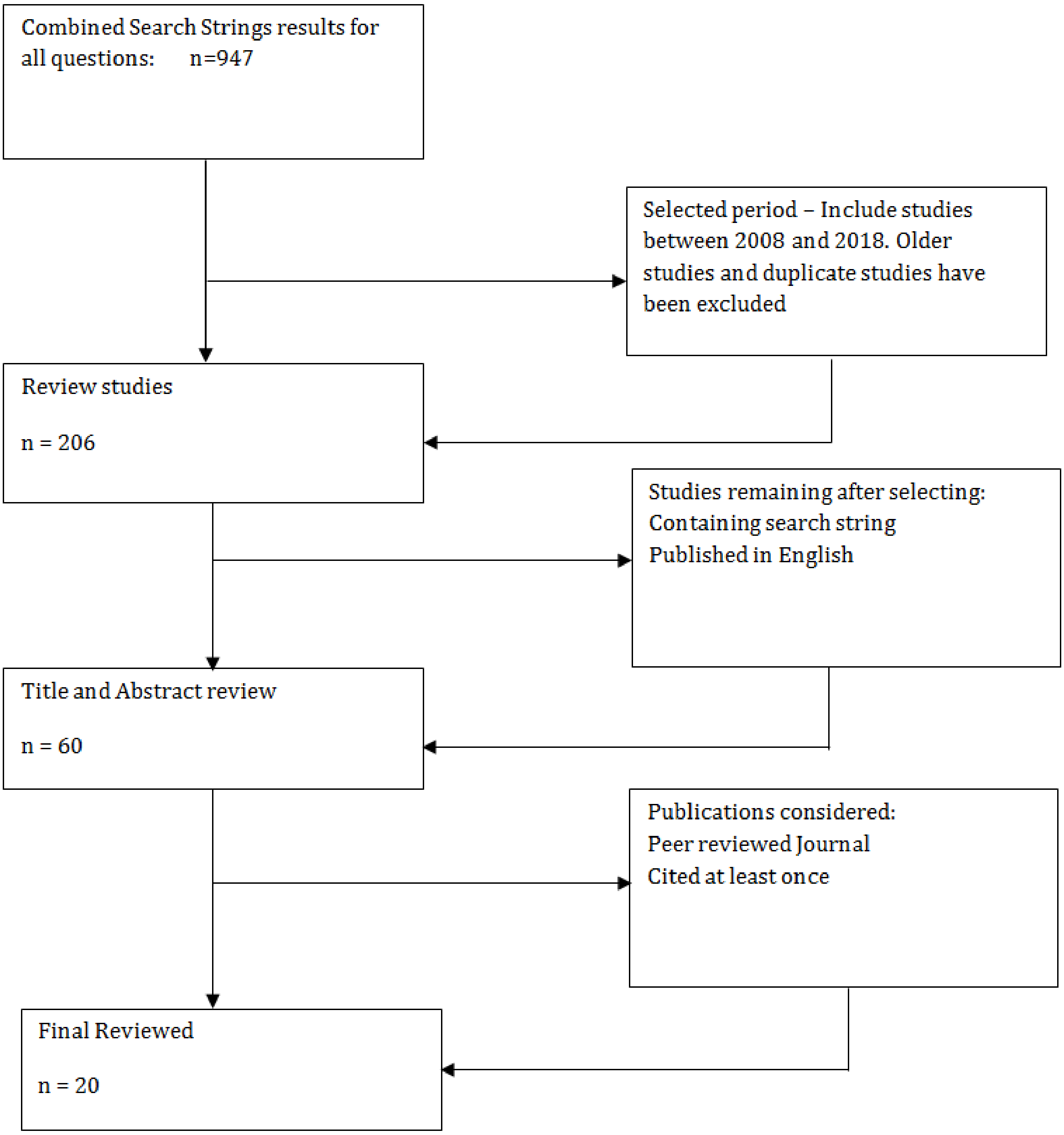
Search Strings and Quality Criteria

The following databases and search engines were used to obtain the relevant literature. The databases were chosen as they comprise of information on peer-reviewed journal articles and conference proceedings. The data selected for this study is based on journals and articles published between 2008 and 2018. Publications were sourced with keywords such as, Data Scientist and Data Scientist skills.

Source Name / Search String	Uniform Resource Locator (URL)	Link to Search	Results
Web of science "Data Scientist" OR "Data Scientist skills"	http://0-apps.webofknowledge.com.innopac.wits.ac.za/UA_GeneralSearch_input.do?product=UA&search_mode=GeneralSearch&SID=Y2dbhM2swCArijME7Zv&preferencesSaved=	http://0-apps.webofknowledge.com.innopac.wits.ac.za/Search.do?product=WOS&SID=C2cxWDON3MuohFMYcGK&search_mode=GeneralSearch&prID=6967280e-25db-4b9a-b721-7a7fd74bdc4d	52
IEEE Transactions "Data Scientist" OR "Data Scientist skills"	http://0-ieeexplore.ieee.org.innopac.wits.ac.za/xpl/RecentIssue.jsp?punumber=6245516	http://0-ieeexplore.ieee.org.innopac.wits.ac.za/search/searchresult.jsp?newsearch=true&queryText=.QT.data%20scientist.QT.	52
Google scholar "Data Scientist" OR "Data Scientist skills"	https://scholar.google.co.za	https://scholar.google.co.za/scholar?as_q=&as_epq=data+scientist&as_oq=&as_eq=&as_occt=title&as_sauthors=&as_publication=&as_ylo=&as_yhi=&btnG=&hl=en&as_sdt=0%2C5 https://scholar.google.co.za/scholar?hl=en&as_sdt=0%2C5&q=allintitle%3A+%22data+scientist+skills%22&btnG=	218 2
Wiley online library "Data Scientist" OR "Data Scientist skills"	http://0-onlinelibrary.wiley.com.innopac.wits.ac.za/	https://0-onlinelibrary-wiley-com.innopac.wits.ac.za/action/doSearch?AllField=%22data+scientist%22 https://0-onlinelibrary-wiley-com.innopac.wits.ac.za/action/doSearch?AllField=%22data+scientist+skills%22	437 3
Science Direct "Data Scientist" OR "Data Scientist skills"	https://0-www-sciencedirect-com.innopac.wits.ac.za/	https://0-www-sciencedirect-com.innopac.wits.ac.za/search?qs=%22data%20scientist%22&show=25&sortBy=relevance&publicationTitles=271653%2C271625%2C273567%2C280203%2C312594%2C314942%2C271521&lastSelectedFacet=publicationTitles	183

Quality Criteria

The quality criteria aim to increase the validity and reliability of the results. The following illustration shows the selection criteria. After a vigorous exercise of filtering and ensuring that the inclusion and exclusion criteria has been met 20 articles were chosen for the final of this study.



8.2 Appendix B: Summary of Themes for Data Scientist Role from the Literature

Theme From Literature	Source	No. in Appendix A
Answers questions in large datasets to make informed decisions	Abbasi, Sarker & Chiang (2016) Chatfield et al(2014) McAfee et al (2012) Provost & Fawcett (2013)	1, 4, 12, 15
Designing experiments and sampling methodologies in large data sets	Brynjolfsson et al (2011) Davis (2014)	2, 7,
Turn data and information into meaningful and actionable insights	Chen et al. (2012) Chiang, Goes & Stohr (2012) Dhar (2013)	5, 6, 9
Turn data and information into meaningful and actionable insights and communicate results	Carter et al (2015) Dinter et al (2015)	3, 10
Skilled in using and exploring Information Technology	Davenport (2012)	8
Turn data and information from large data sets into meaningful and actionable insights	Kim & Lee (2016) Miller (2014) Waller & Fawcett (2013) Watson (2014)	11, 13, 17, 19
Necessary training and curiosity to make discoveries in Big Data	Patil & Davenport (2012)	14
Using advanced technologies and unusual data sources with predictive Analytics to gain insights	Wixom et al (2014)	20

8.3 Appendix C: Literature Review Mapped to Research Model

No	Source	Role of a Data Scientist	Themes for Role of a Data Scientist	Skills and Competencies of a Data Scientist	Themes for Competencies of a Data Scientist	High Level Competency from Research Model
1	Abbasi, Sarker & Chiang (2016)	People who know how to get answers to important questions from a large amount of unstructured information	Answers questions in large datasets to make informed decisions	Machine Learning Hadoop Derive knowledge from internal and external data sources Work closely with analyst in the decision making phase	Technical System Answer Questions Systems Analytical Programming Communication Decision Making	Technical Competency Data Scientist Fundamental Competency Data Science Orchestration Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Business Interpersonal Competency Data Science Orchestration Competency
2	Brynjolfsson, Hammerbacher & Stevens (2011)	Working with very larger data sets, sampling methodologies, designing experiments	Designing experiments and sampling methodologies in large data sets	Skills of social scientist, programming skills, ability to do aggressive prioritisation, knowledge Statistics and Machine Learning	Programming Statistics Technical System	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Technical Competency Data Scientist Fundamental Competency
3	Carter et al (2015)	Gathering and formatting data, constructing models, communicating analysis to members of their organisation and customers	Turn data and information into meaningful and actionable insights and communicate results	Mathematician, product development and detective.	Data Transformation Mathematics Communication Technical System Answer Questions	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Business Interpersonal Competency Technical Competency Data Scientist Fundamental Competency Data Science Orchestration Competency
4	Chatfield, Shlemoon, Redublado & Rahman (2014).	Working with data to answer questions that are used to make informed business decisions	Answers questions in large datasets to make informed decisions	Computer Science Communication Mathematics Data visualisation	Data Transformation Computer Science Communication Mathematics Data visualisation Analytics Communication Answer Questions Decision Making	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Business Interpersonal Competency Data Science Orchestration Competency Data Science Orchestration Competency
5	Chen, Chiang & Storey (2012)	Turn data and information into meaningful and actionable insights for an organisation.	Turn data and information into meaningful and actionable insights	Technical skills and system implementation, business or domain knowledge, communication skills	Data Transformation Technical Systems Business Communication	Data Scientist Fundamental Competency Technical Competency Data Scientist Fundamental Competency Organisational knowledge Business Interpersonal Competency

6	Chiang, Goes & Stohr (2012)	Skill required to interpret Big Data	Turn data and information into meaningful and actionable insights	Analytical skills Information technology skills Business knowledge and communication skills	Data Transformation Analytical skills Information Technology System Business knowledge Communication Actionable Insights	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Technical Competency Data Scientist Fundamental Competency Organisational knowledge Business Interpersonal Competency Data Science Orchestration Competency
7	Davis (2014)	To implement and diffuse analytic methodologies throughout an organisation	Designing experiments and sampling methodologies in large data sets		Analytic Skills Data Analysis	Data Scientist Fundamental Competency Data Scientist Fundamental Competency
8	Davenport (2012)	Data Scientists understand Analytics and are skilled in exploring and using information technology	Skilled in using and exploring Information Technology	Technical, business, analytical and relationship skills and have advanced degrees in computer science.	Communication Technical System Business	Business Interpersonal Competency Technical Competency Data Scientist Fundamental Competency Organisational knowledge
9	Dhar (2013)	Extraction of knowledge from data	Turn data and information into meaningful and actionable insights	Integrated Skill set for mathematics, Machine Learning, artificial intelligence, Statistics, databases, and optimisation, problem formulation skills	Mathematics Technical (Machine Learning) System Programming Statistics Data Analysis Problem Solving Actionable Insights	Data Scientist Fundamental Competency Technical Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency Data Science Orchestration Competency
10	Dinter, Douglas, Chiang, Mari, Ram & Schoder (2015)	Ability to take data and understand, process, extracts value, visualise it and communicate it.	Turn data and information into meaningful and actionable insights and communicate results	Analytics skills, IT and programming, business and domain knowledge and interpersonal skills such as communicating.	Analytics Information Technology Programming Business knowledge Communication Data Mining Data Analysis Data Visualisation Actionable Insights	Data Scientist Fundamental Competency Technical Competency Data Scientist Fundamental Competency Organisational knowledge Business Interpersonal Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency
11	Kim & Lee (2016)	Defined as an expert that makes discoveries in large quantities of data to solve problems and provide insights for organisations.	Turn data and information from large data sets into meaningful and actionable insights	Machine Learning Analytical skills Statistical skills Data mining Programming	Machine Learning Analytical skills Statistical skills Data mining Data Transformation	Technical Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency

				Business knowledge	Programming Business knowledge Problem Solving Actionable Insights	Data Scientist Fundamental Competency Organisational knowledge Data Science Orchestration Competency Data Science Orchestration Competency
12	McAfee et al (2012)	Skilled at working large quantities of data	Answers questions in large datasets to make informed decisions	Statistics Data mining Visualisation Communicate to business leaders	Statistics Data mining Visualisation Communication Data Analysis Answer Questions	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Business Interpersonal Competency Data Scientist Fundamental Competency Data Science Orchestration Competency
13	Miller (2014)	Data Scientist role as an Information strategist, information system professional Data governance ethics	Turn data and information from large data sets into meaningful and actionable insights	Mathematics Statistics Machine Learning Predictive Analytics Decision management Computer science Data ethics Information law Data mining	Mathematics Statistics Machine Learning Predictive Analytics Decision management Computer science Data ethics Information law Data mining Actionable Insights	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Technical Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency
14	Patil and Davenport (2012)	A highly ranked professional having the necessary training and curiosity to make discoveries in Big Data	Necessary training and curiosity to make discoveries in Big Data	Hybrid of a data hacker, analyst, communicator and trusted adviser. Basic skills should be coding	Technical Programming Communication Data Mining Data Analysis	Technical Competency Data Scientist Fundamental Competency Business Interpersonal Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency
15	Provost & Fawcett (2013)	View a business problem from the data side to make decisions	Answers questions in large datasets to make informed decisions	Skills to access data, prepare it, process it using data tools	Data Mining Data Analysis Programming Analytics Answer Questions Decision Making Problem Solving	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency Data Science Orchestration Competency Data Science Orchestration Competency
16	Schoenherr & Speier-Pero (2015)	Knowledge of data processes to create insights	Turn data and information into meaningful and actionable insights	Enterprise business processes, decision making, Data management, analytical modelling tools	Technical Analytics Data Mining Data Analysis Actionable Insights Decision Making	Technical Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency Data Science Orchestration Competency

17	Waller & Fawcett (2013)	Deep domain knowledge and broad analytic skill set	Turn data and information from large data sets into meaningful and actionable insights	Deep domain knowledge and broad analytic skill set	Business knowledge Analytics Data Mining Data Analysis Actionable Insights	Organisational knowledge Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency
18	Wamba et al (2015)	Using new and unusual sources of data advanced technologies and visualisation, predictive Analytics and rare combinations of user skills.	Using advanced technologies and unusual data sources with predictive Analytics to gain insights	Technical, analytical, business and governance skills, networking skills, communication skills, operationalising skill and implementation of analysis	Technical Analytics Business Communication Predictive Model Data Mining Data Analysis Actionable Insights	Technical Competency Data Scientist Fundamental Competency Organisational knowledge Business Interpersonal Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Science Orchestration Competency
19	Watson (2014)	To discover patterns and relationships no one has seen or thought about and turn the discovery into insights that can be actioned and create value for the organisation	Turn data and information from large data sets into meaningful and actionable insights	Understand different types of data and its storage, write code, access data and analyse it Communicate findings to management through reports.	Data Mining Data Analysis Programming Communication Actionable Insights Answer Questions not Asked	Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Business Interpersonal Competency Data Science Orchestration Competency Data Science Orchestration Competency
20	Wixom et al (2014)	Using new and unusual sources of data advanced technologies and visualisation, predictive Analytics and rare combinations of user skills.	Using advanced technologies and unusual data sources with predictive Analytics to gain insights	Communication Programming Basic Analytics Data management Business knowledge	Technical Programming Analytics Data management Business knowledge Communication Predictive Model Actionable Insights	Technical Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Data Scientist Fundamental Competency Organisational knowledge Business Interpersonal Competency Data Scientist Fundamental Competency Data Science Orchestration Competency

8.4 Appendix D: Ethical Clearance Certificate

Faculty of Commerce, Law and Management
University of the Witwatersrand, Johannesburg

School of Economic and Business Sciences
Private Bag X3, WITS, 2050, South Africa • Telephone: +27 11 717 8004 •
email: Siyabonga.Molaba@wits.ac.za

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

CLEARANCE CERTIFICATE **PROTOCOL NUMBER: CINFO/1170**

PROJECT: DATA SCIENTIST: USING A COMPETENCY BASED APPROACH TO EXPLORE AN EMERGING ROLE

INVESTIGATOR: Naseema Banu Nosarka

STUDENT NUMBER: 1558925

SCHOOL: SEBS

DATE CONSIDERED: 20 October 2017

DECISION OF THE ETHICS COMMITTEE: Approved

NOTE

Unless otherwise specified this ethics clearance is valid for 1 year and may be renewed upon application. Please remember to include the protocol number above to your participation letter.

DATE: 03/11/2017 **CHAIRPERSON:** Jean-Marie Bancilhon

cc: Supervisor:

Susan Benvenuti

SCHOOL OF ECONOMIC & BUSINESS SCIENCES

8.5 Appendix E: Participation Letter (SPL)

Date: 16 October 2017

Dear Participant

My name is Naseema Nosarka. I am currently registered as a Masters Student in the Information Systems Division at the University of the Witwatersrand, Johannesburg. I am conducting research on the role and desired competencies of a Data Scientist which could help benefit organisations in their quest to hire Data Scientists which could lead to an increase in adoption of Big Data technologies.

As a Data Scientist, you are invited to participate in this exploratory research study. The purpose of this interview is to explore the role of a Data Scientists as experienced by Data Scientists or other professionals undertaking the role.

Your response is important and there are no right or wrong answers. Your responses will be kept confidential and anonymous. Anonymity and confidentiality are guaranteed by not needing identifying information such as your name, identification number, phone number, or any details that may be used to identify you. Your participation in this research is completely voluntary and involves no risk, penalty, or loss of benefits. Please note that you are free to not answer particular questions or to withdraw from the research at any time. If you decide to withdraw, you can stop the interview at any point. You can also choose to withdraw at the end of the interview. In either case, any information obtained from you prior to that point will be destroyed. The survey was approved by the SEBS Ethics Committee (Non-Medical), Protocol Number: CINFO/1170.

The first part of the interview captures demographic data. The second part of the semi structured interview consists of 12 questions exploring the role and associated competencies of a Data Scientist. The entire interview should take one hour to complete.

Thank you for taking the time to participate in this interview. Should you have any questions, or should you wish to obtain a copy of the results of the study, please contact me at 1558925@students.wits.ac.za or 083 950 0919. My supervisor's name and email are: Susan Benvenuti Susan.Benvenuti@wits.ac.za
(Researcher's Signature)

Kind regards
Naseema Nosarka
Master of Commerce Student (Information Systems)
School of Economic and Business Sciences
University of Witwatersrand

8.6 Appendix F: Participation Consent Form (SPCF)

Title of research project: Data Scientist: Using a Competency Based Approach to Explore an Emerging Role
Name/s of principal researcher/s: Naseema Nosarka

Department/research group address: Information Systems.

Telephone: 083 9500 919.

Email: 1558925@students.wits.ac.za

Nature of the research:
The aim of this research is to explore the role and desired competencies of a Data Scientist which could help benefit organisations in their quest to hire Data Scientists. This could lead to an increase in adoption of Big Data technologies.

Participant's involvement:

To provide responses to the interview administered.

What's involved?

Risks: None

Benefits: Gain an understanding of the skills Data Scientists should have in an emerging context. This research can add insight by highlighting the role that Data Scientists are currently portraying by providing information on the specific skills and competencies that are required through the assessment of responses.
I acknowledge the following:

- I agree to participate in this research project.
- I have read this consent form and the information it contains and had the opportunity to ask questions about them.
- I agree to my responses being used for education and research on condition that my privacy is respected, subject to the following:
  - I understand that my personal details will not / may be included in the research / will be used in aggregate form only, so that I will not be personally identifiable (delete as applicable.)
  - I understand that I am under no obligation to take part in this project.
  - I understand I have the right to withdraw from this project at any stage.

☐ I hereby consent to participate in this interview.

☐ I do not consent to participate in this interview.

Signature of Participant:
Signature of person who sought consent:
Name of person who sought consent: Naseema Nosarka
Date: 19/10/2017

8.7 Appendix G: Consent Form for Interview Recording (CIR)

Title of research project: Data Scientist: Using a Competency Based Approach to Explore an Emerging Role

Name/s of principal researcher/s: Naseema Nosarka
Department/research group address: Information Systems
University of the Witwatersrand, Johannesburg
Telephone: 083 9500 919
Email: 1558925@students.wits.ac.za

Name of participant:____

Nature of the research: The aim of this research is to explore the role and desired competencies of a Data Scientist which could help benefit organisations in their quest to hire Data Scientists. This could lead to an increase in adoption of Big Data technologies.

Participant's involvement: Provide his/her experience on the current and required role of a Data Scientist

What's involved? An interview to explore the role of a Data Scientist
Risks: None
Benefits: Gain an understanding of the skills Data Scientists should have in an emerging context. This research can add insight by highlighting the role that Data Scientists are currently portraying by providing information on the specific skills and competencies that are required through the assessment of responses.

I acknowledge the following:

- I agree to participate in this research project.
- I agree that the interview will be recorded
- I have read this consent form and the information it contains and had the opportunity to ask questions about them.
- I agree to my responses being used for education and research on condition that my privacy is respected, subject to the following:
  - I understand that my personal details will not be included in the research and that I will not be personally identifiable.
  - I understand that I am under no obligation to take part in this project.
  - I understand I have the right to withdraw from this project at any stage.

Signature of Participant:_________________________________
Name of Participant: ____
Signature of person who sought consent: _____________________________________
Name of person who sought consent: Naseema Nosarka_____________________________
Date:____

8.8 Appendix H: Interview Questions

Section A: Demographics

Age (Optional):
Gender (Optional):

Section B: Interview Questions

If you were in the role of management and could define the ideal role of the Data Scientist in its truest form:

REQUIRED - ROLE OF A DATA SCIENTIST:

1. What would you describe the current role of a Data Scientist in the banking and insurance industry as being?

REQUIRED - COMPETENCIES OF A DATA SCIENTIST:

Data Science Orchestration Competency

2. What value do you think a Data Scientist can add to an organisation, and how?

Organisational knowledge, Technical Competency, Business Interpersonal Competency

3. In your experience, what knowledge, skills, and attributes does a Data Scientist need to be effective in the role? Can we look at each of those individually?
    a. If we define Knowledge as: factual/conceptual information that a person knows and that is needed for a specific job … what ideally would a Data Scientist need?
    b. If we define Skills as: an ability that has been acquired through training and education that enables a person to consistently perform a complex task efficiently… what ideally would a Data Scientist need?
    c. If we define Attributes as: including personal characteristics, traits, motives and values or ways of thinking that affect an individual's behaviour or attitude … what ideally would a Data Scientist need?

Data Scientist Fundamental Competency

4. What are your thoughts on developing or training people to work in the Big Data Arena?
    a. Who should we be training to move into those roles? (E.g. Developers, BI Analysts, Statisticians, etc.)
    b. What type of education ideally should a typical Data Scientist have?

c. What type of experience would you value in a Data Scientist?
d. In your opinion, which of the following areas are important for a Data Scientist to be educated/trained in?

CURRENT - ROLE AS DATA SCIENTIST (SPECIFIC TO RESPONDENT):

5. May I ask you a few questions relating to your current job and role in your organisation?
    a. What is your current role and job title in the organisation?
    b. How would you describe your role and responsibilities (what is your job description and expectations)?
    c. To what extent would you say your responsibilities align with how you defined the role of a Data Scientist for me earlier?
    d. How long have you been in your current role?

CURRENT - COMPETENCIES OF THE DATA SCIENTIST:

Data Science Orchestration Competency

6. What value do you add to the organisation, and how?
7. What do you believe would make your job more effective in terms of the value you could deliver to the organisation?
8. What kind of data do you generally work with (structured or unstructured)? (define the terms if needed)

Organisational knowledge, Technical Competency, Business Interpersonal Competency

9. What knowledge, skills, and attributes do you possess as a Data Scientist to be effective in your role? Can we look at each of those individually?
    a. If we define Knowledge as: factual/conceptual information that a person knows and that is needed for a specific job … what do you currently possess?
    b. If we define Skills as: an ability that has been acquired through training and education that enables a person to consistently perform a complex task efficiently… what do you currently possess?
    c. If we define Attributes as: including personal characteristics, traits, motives and values or ways of thinking that affect an individual's behaviour or attitude … what do you currently possess?
    d. Are you part of a broader team? What does the rest of the team look like in terms of knowledge, skills and responsibilities?

Data Scientist Fundamental Competency

10. The role of the Data Scientist is very new and would not have been something that you originally trained for. Would you mind taking me through your career path to-date so that I can understand how your formal training and experience have led to you being able to take on the role of Data Scientist?

11. Then as a follow-up (depending on what their answer is): Is there any specific part of your prior education, training, courses or experience that you think is particularly valuable in your current role as Data Scientist? (Skills required to perform job)

12. Is there anything you wish you had studied in the past that might be useful to you now?

8.9 Appendix I: Example of Interview Data

Interview Number	Section	Sub Section	Question	Answer
3	**Demographics**		Age	29
			Gender	Male
	Required Role		Describe role of a Data Scientist	Find patterns in the data that can be used to make a positive change in the business.
	Required Competencies	**Data Science Orchestration Competency**	Value a Data Scientist can add	Data can be overlooked due to the other pressures of running a business. The insights that come from analysis and AI can be used to notice issues before they become problems. Data Scientists are equipped to bring these issues to light sooner. It's sometimes small changes in how a business operates that can have big results. Data scientists can use their expertise to find the best levers to drive change.
		Organisational knowledge, Technical Competency, Business Interpersonal Competency	What knowledge should a Data Scientist have	Business Knowledge – how the specific business operates Systems Knowledge – The systems the business uses to warehouse data
			What skills should a Data Scientist have	Applications Knowledge – knowledge of the tools (BI tools, Programming languages, SQL etc) used to run the analysis
			What attributes should a Data Scientist have	Takes initiative (it's not up to the employer to tell you where to look), Problem solver (given a business problem, the data scientist must translate the data insights to formulate a solution), Self Confidence (the data may be contrary to the status quo and the data scientist must be able to stand his/her ground), ability to work on their own, presentation skills, ability to ask questions from business to understand the business.
		Data Scientist Fundamental Competency	Which people to train in Data Science	Developers, BI Analysts, Statisticians, etc.) All three of these people can benefit from the knowledge held by the others
			Who should be trained	Developers, BI Analysts, Statisticians
			Type of Education	Programming, basic maths and stats, data visualisation/design
			Type of Experience	Prior track record in a variety of projects in different parts of the organisation. Prior projects dealing with variety of stakeholders (eg management, business owners, customers)
			Important Areas for Data Scientists to be trained in	IT Skills, statistical/mathematical Design/user experience/presentation skills
	Current Role		Job Title	Data Scientist
			Role and Responsibilities	My role is to find marketing levers to improve profitability.
			Does current role align with the role defined earlier	I'd say they line up well

			Time in Current role	2 years
	Current Competencies	**Data Science Orchestration Competency**	Value you add as a Data Scientist	I help management better understand the source of profitability and how this can be changed
			What would make job more effective	More face time with the business owners after each piece of work has been submitted. This could allow me to better understand their requirements
			What data do you work with	Structured
		Organisational knowledge, Technical Competency, Business Interpersonal Competency	What knowledge do you have	Actuarial knowledge, Business Knowledge, Systems knowledge
			What skills do you have	Data manipulation, Qlikview, Google Analytics, understanding of Facebook Marketing, Excel, SQL and Java, Programming
			What attributes do you have	Take initiative, Work well alone, presentation skills, problem solver
			Part of a bigger team? Makeup of team?	Yes. The rest of the team is business focused i.e. marketing, operations, sales.
		Data Scientist Fundamental Competency	Career path and former training	Actuarial student -> BI analyst -> Software Developer -> Consulting in BI -> Consulting in Data Science
			Prior education, training, experience that is valuable	I worked in an environment where data needed to be cleaned very well. This taught me how to structure data efficiently. Ultimately this was a big advantage when looking at data and knowing what pitfalls to look out for.
			Anything you wish you had studied	Only that I wish I started sooner and studied part time over my career.

8.10 Appendix J: A View of the transcribed data with Codes done in Excel

Section	Sub Section	Question	Interview 2	Interview 3	
Demographics		Age	27	29	
		Gender	Male	Male	
Required Role		Describe role of a Data Scientist	**Difficult Role** **Very Broad** **Answer Questions not asked** D ve a broad range of problems in bu a solution can be found by looking at e problem. The range of problems can be broad and often relate to predicting individual customer's behaviour, or finding levers to use to improve sales, but the method of finding the solution is consistently to look at data and use the necessary tools (be it statistical or otherwise), to find the solution.	**Problem Resolution** **Predict Customer Behaviour** **Improve Sales** **Data Analysis** Find patterns in the data that can be used to make a positive change in the business.	**Find Patterns in Data** **Positive Outcomes for Business** R operational reporting. Proj classification outcomes in an
	Data Science Orchestration Competency	Value a Data Scientist can add	**Answer Questions not asked** **New processes** D ssumptions made in board-room m of business. They can also help in op proving on customer interactions us entify customers who may have unmet financial needs or who would be well suited to certain products. data from customer traffic through portals can be used to improve automated call-centre experience or website experience via optimisation. In short, most problems an insurance or bank experiences that are tangential to its core business (ie. the work of actuaries), can be tackled be data scientists, provided the problem generates some data that can be analysed.	**Validate Assumptions made by Business** **Process Optimisation** D o the other pressures of running a business. The in sis and AI can be used to notice issues before they become problems. Data Scientists are equipped to bring these issues to light sooner. It's sometimes small changes in how a business operates that can have big results. Data scientists can use their expertise to find the best levers to drive change.	**Problem Identification** **Drive Results and Change**
		What knowledge should a Data Scientist have	**Broad Business Knowledge** **Machine Learning** **Big data** A ics and mathematics. k\Knowledge of ues and how to handle them. Basic ur roblem. Some knowledge of nu python/R/etc). An understanding of common statistical traps and pitfalls (eg Anscombe's quartet, reverse causation, spurious correlation, nonstationarity).	Business Knowledge – how the specific business operates Systems Knowledge – The systems the business uses to warehouse data	**Business Knowledge** **Systems Knowledge** **Application Knowledge**

8.11 Appendix K: Summary of Themes for Data Scientist Role from the Results

No	Job Title	Required/ Current	Role of a Data Scientist	Themes for Role of a Data Scientist
DS 1	Data Scientist	Required	A Data Scientist can help organisations answer questions they have not asked.	Difficult Role Very Broad Answer Questions not asked
		Current	Source, transform data to answer business questions. Recycle data to answer future directed questions.	Data Sourcing Data Analysis Answer Business Questions Recycle data to predict future outcomes
DS 2	Data Scientist	Required	Solve a broad range of problems in business. Predicting individual customer's behaviour. Look at data and use the necessary tools to find the solution.	Problem Resolution Predict Customer Behaviour Improve Sales Data Analysis Data Tools Answers questions in large datasets to make informed decisions
		Current	Create and maintain ongoing predictive models for customer behaviour. Perform ad-hoc data-analysis. Quickly find answers to questions being asked in business	Create and Maintain Predictive Models Enhance Customer Behaviour Data Analysis Answer Questions from Business
DS 3	Data Scientist	Required	Find patterns in the data that can be used to make a positive change in the business.	Find patterns in data Improve Profitability Answer Questions not asked
		Current	Find marketing levers to improve profitability.	Improve Profitability Data Analysis Improve Sales Answer Questions from Business
DS 4	Data Scientist	Required	Answers to Business Predictive Models Understands Business Needs Develop Analytic tools Enhance Customer Engagement Automate Business Processes Delivering results that meet the bank's needs (Answers Questions)	Answer Business Questions Predictive Models Understands Business Needs Develop Analytic tools Enhance Customer Engagement Automate Business Processes Answers questions in large datasets to make informed decisions
		Current	Provide answers to business regarding bonuses Predictive modelling	Deliver new model Answers questions in large datasets to make informed decisions
DS 5	Data Scientist	Required	Data Analysis Actionable Insights Customise Products communications Stakeholder engagement Automate Tasks	Data Analysis Actionable Insights Customise Products communications Stakeholder engagement Automate Business Processes
		Current	Deliver new model Business Intelligence Business and Systems Analysis Solution Design	Deliver new model Business Intelligence Business and Systems Analysis Solution Design
DS 6	Data Scientist	Required	Collect and report on data, as well as develop and maintain statistical models, perform strategic analysis to determine what they mean and recommend ways to apply the models and data aligned to business strategies	Data Sourcing Data Analysis Answer Business Questions
		Current	Responsible for building & supporting new analytical initiatives as well as operational reporting.	Deliver new models

8.12 Appendix L: Competency Profiles from Interviews

DATA SCIENTIST 1

This table shows all Data Scientist 1 competencies that were identified

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 1	Business Interpersonal Competency	Adaptability	Flexibility
		Adaptability	Resilient Personality
		Communication	Ability to communicate
		Curiosity and Creativity	Passion working with data
	Data Science Orchestration Competency	Process Improvement	Create new processes
		Process Improvement	Dynamically Adjust Solutions
		Recommendations to Business	Answer Business Questions
		Recommendations to Business	Answer Questions not asked
	Data Scientist Fundamental Competency	Computer Science and Programming	Computer Science
		Computer Science and Programming	Programming
		Critical Thinking	Attention to Detail
		Data Analytics and Visualisation	Analytics
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Predictive Models
		Data Exploration	Recycle data to predict future outcomes
		Data Management	Data Mining
		Data Management	Data Modelling
		Data Management	Data Sourcing
		Data Management	Data Transformation
		IT and Information Systems	Keep ahead with Data Science Techniques
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
	Organisational Knowledge	Business Knowledge	Broad Business Knowledge
	Technical Competency	Big Data Processing	Algorithms
		Big Data Processing	Big data
		Big Data Processing	Hadoop
		Big Data Processing	Machine Learning
		Big Data Processing	Natural Language Processing
		Data Processing	SQL

DATA SCIENTIST 2

This table shows all Data Scientist 2 competencies that were identified

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 2	Business Interpersonal Competency	Adaptability	Pragmatism
		Business Requirements Elicitation and Specification	Interpret Queries
		Communication	Convey business insights
		Independence	Work Alone
	Data Science Orchestration Competency	Delivers Results	Enhance Customer Behaviour
		Delivers Results	Enhance Customer Experience
		Problem Solving and Analysis	Problem Solving
		Process Improvement	Process Optimisation
		Recommendations to Business	Actionable Insights
		Recommendations to Business	Answer Business Questions
		Recommendations to Business	Validate Business Assumptions
	Data Scientist Fundamental Competency	Computer Science and Programming	Programming
		Critical Thinking	Critical Thinking
		Critical Thinking	Self Critical
		Data Analytics and Visualisation	Analytics
		Data Analytics and Visualisation	Predictive Models
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Management	Data Manipulation
		Data Management	Data Queries
		Data Management	Data Sourcing
		Data Management	Data Transformation
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
	Organisational Knowledge	Business knowledge	Business knowledge
	Technical Competency	Big Data Processing	Python
		Big Data Processing	Hadoop
		Big Data Processing	R
		Data Processing	SQL

DATA SCIENTIST 3

This table shows all Data Scientist 3 competencies that were identified

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 3	Business Interpersonal Competency	Business Requirements Elicitation and Specification	Better understand requirements
		Communication	Ability to Communicate
		Communication	Presentation Skills
		Independence	Work alone
		Independence	Self Confidence
		Independence	Takes Initiative
	Data Science Orchestration Competency	Delivers Results	Improve Profitability
		Delivers Results	User experience
		Problem Solving and Analysis	Problem Identification
		Problem Solving and Analysis	Problem Solving
		Recommendations to Business	Drive Results and Change
		Recommendations to Business	Help management better understand profitabilty
	Data Scientist Fundamental Competency	Computer Science and Programming	Programming
		Data Analytics and Visualisation	Business Intelligence Tools
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Management	Data Design
		Data Management	Data Manipulation
		IT and Information Systems	Application Knowledge
		IT and Information Systems	IT skills
		IT and Information Systems	Systems Knowledge
		Mathematics and Statistics	Actuarial
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
	Organisational Knowledge	Business Knowledge	Business Knowledge
	Technical Competency	Big Data Processing	Google Analytics
		Big Data Processing	Facebook Marketing
		Big Data Processing	Python
		Big Data Processing	R
		Data Processing	Excel
		Data Processing	Java
		Data Processing	Qlikview
		Data Processing	SQL

DATA SCIENTIST 4

This table shows all Data Scientist 4 competencies that were identified

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 4	Business Interpersonal Competency	Adaptability	Resilient Personality
		Business Requirements Elicitation and Specification	Ask the right questions
		Communication	Ability to Communicate
		Curiosity and Creativity	Creative
		Curiosity and Creativity	Inquisitive
		Curiosity and Creativity	Introverted but Curious
		Curiosity and Creativity	Intuitive
		Curiosity and Creativity	Think out of the box
	Data Science Orchestration Competency	Problem Solving and Analysis	Problem solving
		Process Improvement	Automate processes
		Process Improvement	Deliver new model
		Process Improvement	Global Optimisation
	Data Scientist Fundamental Competency	Computer Science and Programming	Programming
		Computer Science and Programming	Computer Science
		Data Analytics and Visualisation	Analytics
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Analytics and Visualisation	Analytics
		Data Exploration	Exploring data
		Data Exploration	Extracts meaning from Data
		Data Management	Data Manipulation
		Data Management	Data Sourcing
		Data Management	Data Transformation
		IT and Information Systems	IT skills
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
		Mathematics and Statistics	Mathematics
	Organisational Knowledge	Financial Industry Knowledge	Financial Industry Knowledge
	Technical Competency	Big Data Processing	Artificial Intelligence
		Big Data Processing	Machine Learning
		Big Data Processing	Neural Networks
		Big Data Processing	Hadoop
		Big Data Processing	R
		Data Processing	Mathematica
		Data Processing	Programming on Matlab
		Data Processing	SQL

DATA SCIENTIST 5

This table shows all Data Scientist 5 competencies that were identified

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 5	Business Interpersonal Competency	Adaptability	Willing to learn new things
		Communication	Ability to communicate
		Curiosity and Creativity	Inquisitive
		Curiosity and Creativity	Inquisitive
		Curiosity and Creativity	Like Challenges
		Independence	Work alone
	Data Science Orchestration Competency	Problem Solving and Analysis	Solution Design
		Process Improvement	Deliver new model
	Data Scientist Fundamental Competency	Computer Science and Programming	Computer Science
		Computer Science and Programming	Programming
		Critical Thinking	Focused
		Critical Thinking	Persistent
		Data Analytics and Visualisation	Business Intelligence
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Analytics and Visualisation	Predictive Models
		Data Analytics and Visualisation	Prescriptive Models
		Data Management	Data Manipulation
		Data Management	Data Sourcing
		Data Management	Data Transformation
		Data Management	Database Design
		Data Management	Database Management
		IT and Information Systems	Business and Systems Analysis
		IT and Information Systems	Informatics
		IT and Information Systems	Information Science
		IT and Information Systems	IT skills
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
	Organisational Knowledge	Financial Industry Knowledge	Financial
	Technical Competency	Big Data Processing	Hadoop
		Big Data Processing	Machine learning
		Big Data Processing	Python
		Big Data Processing	R
		Data Processing	Power BI
		Data Processing	Tableau
		Web Tools	Bootstrap
		Web Tools	CSS
		Web Tools	Django
		Web Tools	HTML
		Web Tools	Javascript

DATA SCIENTIST 6

This table shows all Data Scientist 6 competencies that were identified

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 6	Business Interpersonal Competency	Business Requirements Elicitation and Specification	Ask the right questions
		Communication	Presentation Skills
		Curiosity and Creativity	Exploring data
		Curiosity and Creativity	Inquisitive
		Independence	Work Alone
	Data Science Orchestration Competency	Delivers Results	Predict Customer Behaviour
		Process Improvement	Automate processes
		Process Improvement	Deliver new model
	Data Scientist Fundamental Competency	Computer Science and Programming	Computer Science
		Computer Science and Programming	Programming
		Critical Thinking	Focused
		Critical Thinking	Persistent
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Analytics and Visualisation	Predictive Models
		Data Management	Data Manipulation
		Data Management	Data Sourcing
		Data Management	Data Transformation
		Data Management	Database Design
		Mathematics and Statistics	Economics
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
	Organisational Knowledge	Business Knowledge	Organisational Knowledge
		Financial Industry Knowledge	Financial
	Technical Competency	Big Data Processing	Machine learning
		Big Data Processing	Hadoop
		Big Data Processing	Python
		Big Data Processing	R
		Data Processing	Qlikview
		Data Processing	SQL
		Data Processing	Power BI
		Data Processing	Tableau

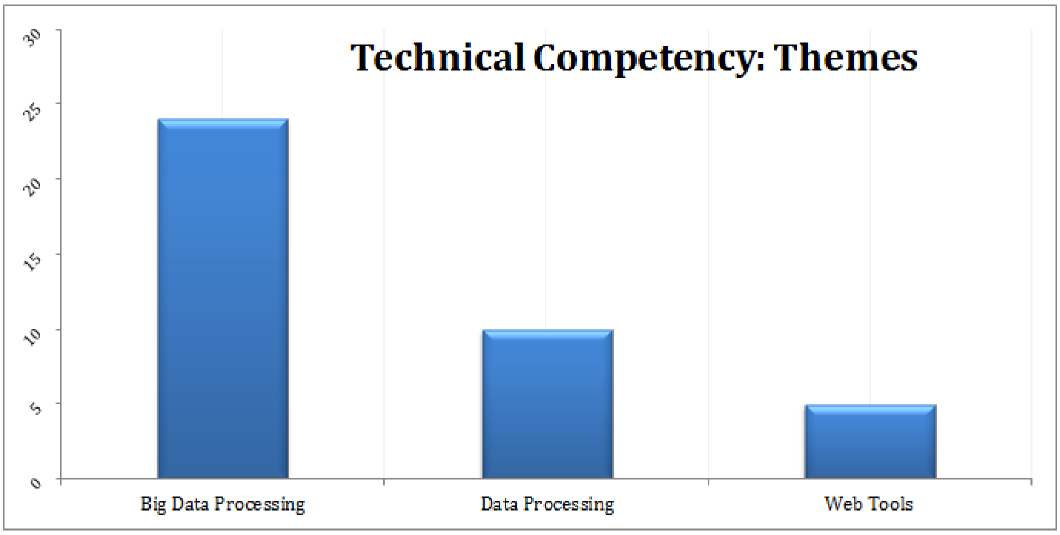
8.13 Appendix M: Technical Competency Category

This table shows all Data Scientist skills and competencies that fall within the Technical Competency category.

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 1	Technical Competency	Big Data Processing	Algorithms
		Big Data Processing	Big data
		Big Data Processing	Hadoop
		Big Data Processing	Machine Learning
		Big Data Processing	Natural Language Processing
		Data Processing	SQL
DS 2	Technical Competency	Big Data Processing	Python
		Big Data Processing	Hadoop
		Big Data Processing	R
		Data Processing	SQL
DS 3	Technical Competency	Big Data Processing	Google Analytics
		Big Data Processing	Facebook Marketing
		Big Data Processing	Python
		Big Data Processing	R
		Data Processing	Excel
		Data Processing	Java
		Data Processing	Qlikview
		Data Processing	SQL
DS 4	Technical Competency	Big Data Processing	Artificial Intelligence
		Big Data Processing	Machine Learning
		Big Data Processing	Neural Networks
		Big Data Processing	Hadoop
		Big Data Processing	R
		Data Processing	Mathematica
		Data Processing	Programming on Matlab
		Data Processing	SQL
DS 5	Technical Competency	Big Data Processing	Hadoop
		Big Data Processing	Machine learning
		Big Data Processing	Python
		Big Data Processing	R
		Data Processing	Power BI
		Data Processing	Tableau
		Web Tools	Bootstrap
		Web Tools	CSS
		Web Tools	Django
		Web Tools	HTML
		Web Tools	Javascript
DS 6	Technical Competency	Big Data Processing	Machine learning
		Big Data Processing	Hadoop
		Big Data Processing	Python
		Big Data Processing	R
		Data Processing	Qlikview
		Data Processing	SQL
		Data Processing	Power BI
		Data Processing	Tableau

Competencies identified from Interview Data

Counting all skills was done to gain an idea of how to work with the data and not necessarily to generalise as the sample contained only 6 Data Scientists.

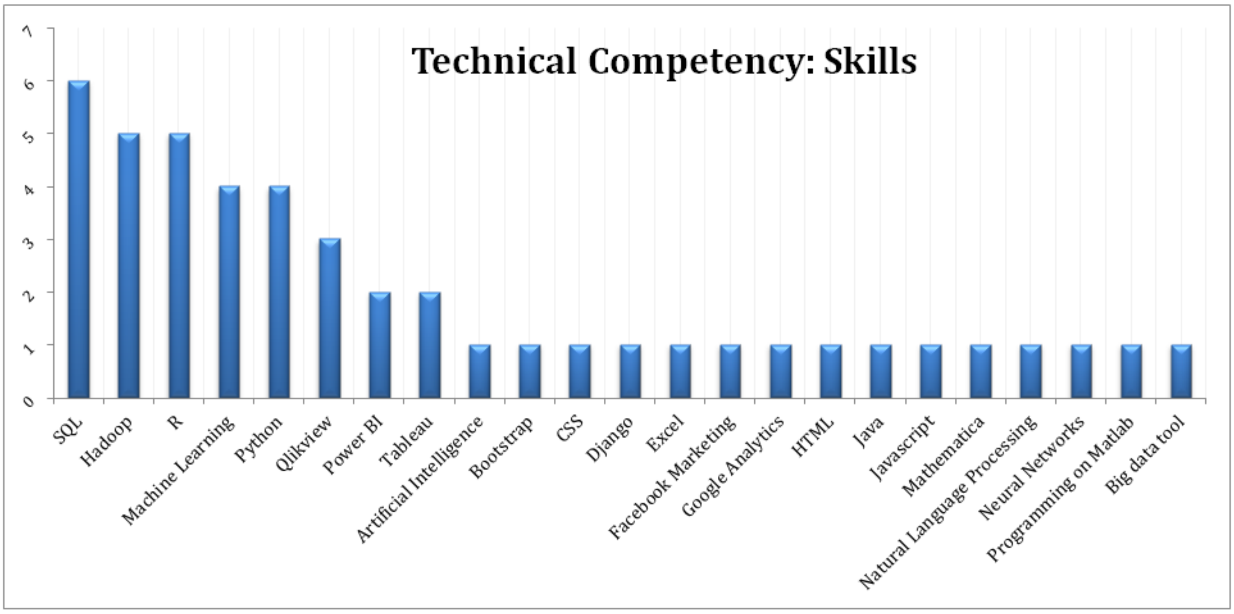


Skills identified from Interview Data

This table was done to show how common each knowledge and skill was within each competency within the Technical Competency category.

Big Data Processing	
Hadoop	5
R	5
Machine Learning	4
Python	4
Artificial Intelligence	1
Big data tool	1
Facebook Marketing	1
Google Analytics	1
Natural Language Processing	1
Neural Networks	1
Data Processing	
SQL	6
Qlikview	3
Power BI	2
Tableau	2
Excel	1
Java	1
Mathematica	1
Programming on Matlab	1
Web Tools	
Bootstrap	1
CSS	1
Django	1
HTML	1
Javascript	1

The knowledge and skills which was the data collected from the interviews skills are counted to give a full view to identify which are the most common skills to a Data Scientist within the sample under the Technical competency category.



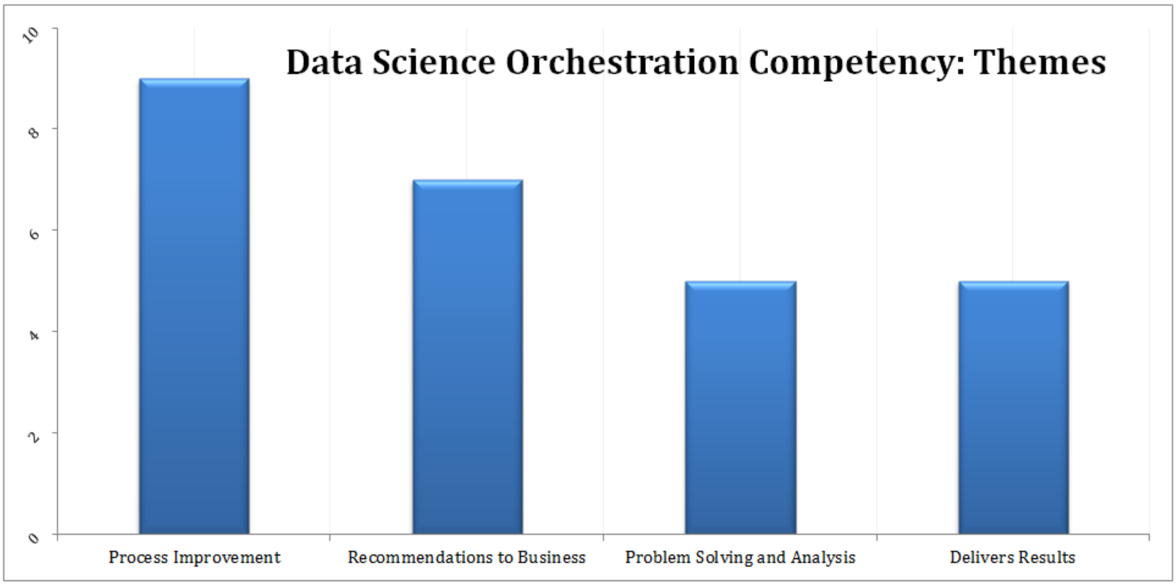
8.14 Appendix N: Data Science Orchestration Competency Category

This table shows all Data Scientist skills and competencies that fall within the Data Science Orchestration Competency category.

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 1	Data Science Orchestration Competency	Process Improvement Process Improvement Recommendations to Business Recommendations to Business	Create new processes Dynamically Adjust Solutions Answer Business Questions Answer Questions not asked
DS 2	Data Science Orchestration Competency	Delivers Results Delivers Results Problem Solving and Analysis Process Improvement Recommendations to Business Recommendations to Business Recommendations to Business	Enhance Customer Behaviour Enhance Customer Experience Problem Solving Process Optimisation Actionable Insights Answer Business Questions Validate Business Assumptions
DS 3	Data Science Orchestration Competency	Delivers Results Delivers Results Problem Solving and Analysis Problem Solving and Analysis Recommendations to Business Recommendations to Business	Improve Profitability User experience Problem Identification Problem Solving Drive Results and Change Help management better understand profitabilty
DS 4	Data Science Orchestration Competency	Problem Solving and Analysis Process Improvement Process Improvement Process Improvement	Problem solving Automate processes Deliver new model Global Optimisation
DS 5	Data Science Orchestration Competency	Problem Solving and Analysis Process Improvement	Solution Design Deliver new model
DS 6	Data Science Orchestration Competency	Delivers Results Process Improvement Process Improvement	Predict Customer Behaviour Automate processes Deliver new model

Competencies identified from Interview Data

Counting all skills was done to gain an idea of how to work with the data and not necessarily to generalise as the sample contained only 6 Data Scientists.



Skills identified from Interview Data

This table was done to show each knowledge and skill within each competency within the Data Science Orchestration Competency. The knowledge and skills was obtained from the data collected from the interviews.

Process Improvement
Deliver new model
Automate processes
Create new processes
Dynamically Adjust Solutions
Global Optimisation
Process Optimisation
Recommendations to Business
Answer Business Questions
Actionable Insights
Answer Questions not asked
Drive Results and Change
Help management better understand profitabilty
Validate Business Assumptions
Problem Solving and Analysis
Problem Solving
Problem Identification
Solution Design
Delivers Results
Enhance Customer Behaviour
Enhance Customer Experience
Improve Profitability
Predict Customer Behaviour
User experience

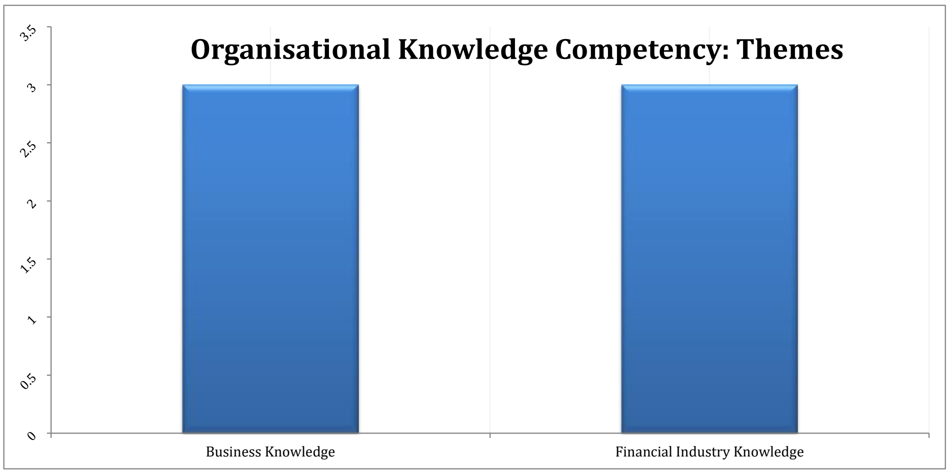
8.15 Appendix O: Organisational Knowledge Category

This table shows all Data Scientist skills and competencies that fall within the Organisational Knowledge category.

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 1	Organisational Knowledge	Business Knowledge	Broad Business Knowledge
DS 2	Organisational Knowledge	Business knowledge	Business Knowledge
DS 3	Organisational Knowledge	Business Knowledge	Business Knowledge
DS 4	Organisational Knowledge	Financial Industry Knowledge	Financial Industry Knowledge
DS 5	Organisational Knowledge	Financial Industry Knowledge	Financial
DS 6	Organisational Knowledge	Business Knowledge Financial Industry Knowledge	Organisational Knowledge Financial

Competencies identified from Interview Data

Counting all skills was done to gain an idea of how to work with the data and not necessarily to generalise as the sample contained only 6 Data Scientists.



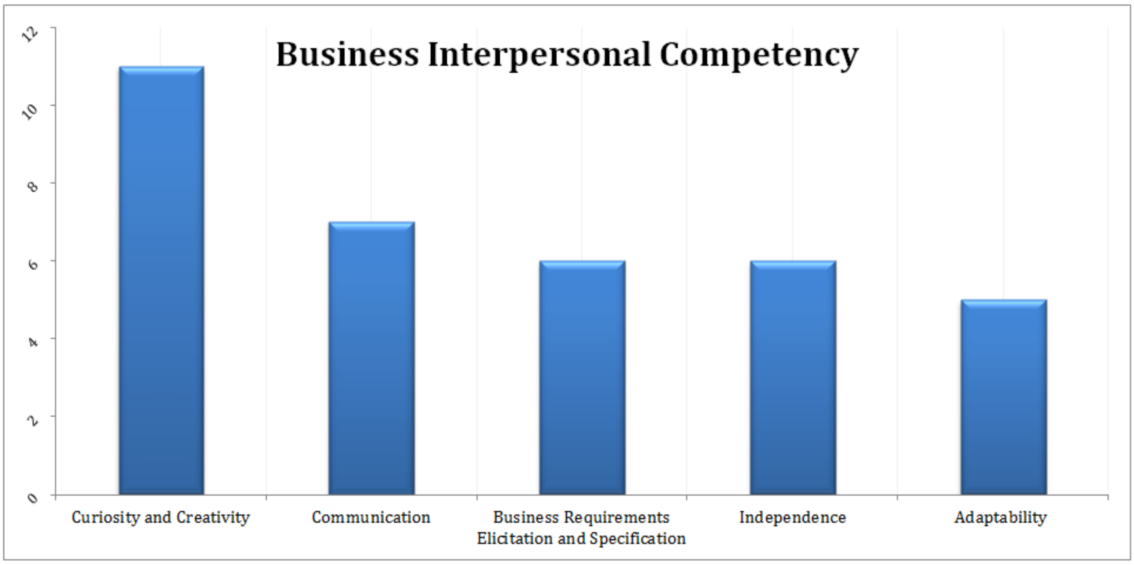
8.16 Appendix P: Business Interpersonal Competency Category

This table shows all Data Scientist skills and competencies that fall within the Business Interpersonal Competency category.

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 1	Business Interpersonal Competency	Adaptability Adaptability Communication Curiosity and Creativity	Flexibility Resilient Personality Ability to communicate Passion working with data
DS 2	Business Interpersonal Competency	Adaptability Business Requirements Elicitation and Specification Communication Independence	Pragmatism Interpret Queries Convey business insights Work Alone
DS 3	Business Interpersonal Competency	Business Requirements Elicitation and Specification Communication Communication Independence Independence Independence	Better understand requirements Ability to Communicate Presentation Skills Work alone Self Confidence Takes Initiative
DS 4	Business Interpersonal Competency	Adaptability Business Requirements Elicitation and Specification Communication Curiosity and Creativity Curiosity and Creativity Curiosity and Creativity Curiosity and Creativity Curiosity and Creativity	Resilient Personality Ask the right questions Ability to Communicate Creative Inquisitive Introverted but Curious Intuitive Think out of the box
DS 5	Business Interpersonal Competency	Adaptability Communication Curiosity and Creativity Curiosity and Creativity Curiosity and Creativity Independence	Willing to learn new things Ability to communicate Inquisitive Inquisitive Like Challenges Work alone
DS 6	Business Interpersonal Competency	Business Requirements Elicitation and Specification Communication Curiosity and Creativity Curiosity and Creativity Independence	Ask the right questions Presentation Skills Exploring data Inquisitive Work Alone

Competencies identified from Interview Data

Counting all skills was done to gain an idea of how to work with the data and not necessarily to generalise as the sample contained only 6 Data Scientists



Skills identified from Interview Data

This table was done to show each knowledge and skill within each competency within the Business Interpersonal Competency category. The knowledge and skills was obtained from the data collected from the interviews.

Curiosity and Creativity
Inquisitive
Creative
Exploring data
Introverted but Curious
Intuitive
Like Challenges
Passion working with data
Think out of the box
Communication
Ability to Communicate
Presentation Skills
Convey business insights
Business Requirements Elicitation and Specification
Ask the right questions
Better understand requirements
Interpret Queries
Independence
Work Alone
Self Confidence
Takes Initiative
Adaptability
Resilient Personality
Flexibility
Pragmatism
Willing to learn new things

8.17 Appendix Q: Data Scientist Fundamental Competency Category

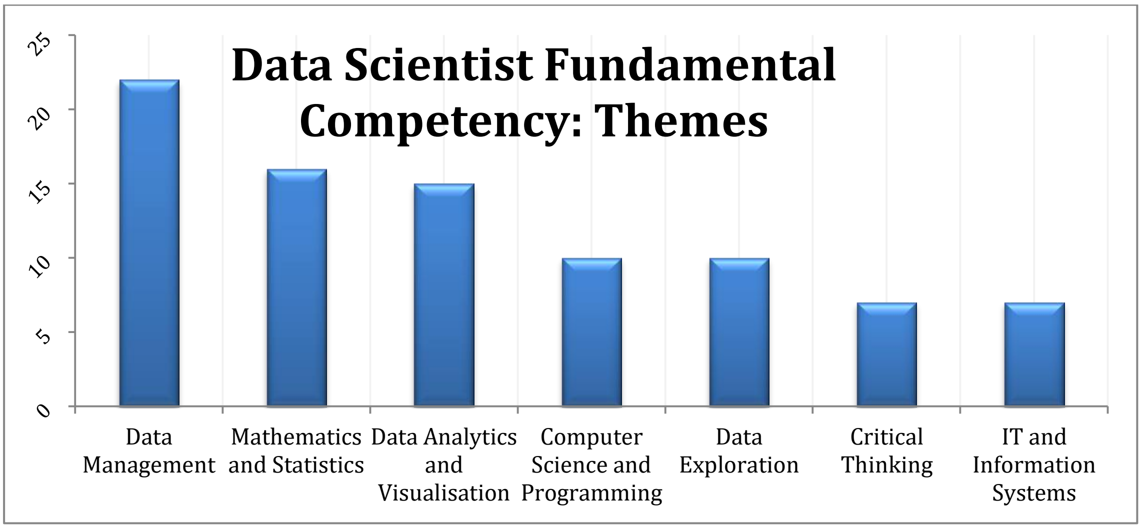
This table shows all Data Scientist skills and competencies that fall within the Data Scientist Fundamental Competency category.

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 1	Data Scientist Fundamental Competency	Computer Science and Programming	Computer Science
		Computer Science and Programming	Programming
		Critical Thinking	Attention to Detail
		Data Analytics and Visualisation	Analytics
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Predictive Models
		Data Exploration	Recycle data to predict future outcomes
		Data Management	Data Mining
DS 2	Data Scientist Fundamental Competency	Computer Science and Programming	Programming
		Critical Thinking	Critical Thinking
		Critical Thinking	Self Critical
		Data Analytics and Visualisation	Analytics
		Data Analytics and Visualisation	Predictive Models
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Management	Data Manipulation
		Data Management	Data Queries
		Data Management	Data Sourcing
		Data Management	Data Transformation
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
DS 3	Data Scientist Fundamental Competency	Computer Science and Programming	Programming
		Data Analytics and Visualisation	Business Intelligence Tools
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Management	Data Design
		Data Management	Data Manipulation
		IT and Information Systems	Application Knowledge
		IT and Information Systems	IT skills
		IT and Information Systems	Systems Knowledge
		Mathematics and Statistics	Actuarial
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics

No	Competency Category from Research Model	Competencies of a Data Scientist	Knowledge, Skills, Attributes of a Data Scientist
DS 4	Data Scientist Fundamental Competency	Computer Science and Programming	Programming
		Computer Science and Programming	Computer Science
		Data Analytics and Visualisation	Analytics
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Analytics and Visualisation	Analytics
		Data Exploration	Exploring data
		Data Exploration	Extracts meaning from Data
		Data Management	Data Manipulation
		Data Management	Data Sourcing
		Data Management	Data Transformation
		IT and Information Systems	IT skills
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
		Mathematics and Statistics	Mathematics
DS 5	Data Scientist Fundamental Competency	Computer Science and Programming	Computer Science
		Computer Science and Programming	Programming
		Critical Thinking	Focused
		Critical Thinking	Persistent
		Data Analytics and Visualisation	Business Intelligence
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Analytics and Visualisation	Predictive Models
		Data Analytics and Visualisation	Prescriptive Models
		Data Management	Data Manipulation
		Data Management	Data Sourcing
		Data Management	Data Transformation
		Data Management	Database Design
		Data Management	Database Management
		IT and Information Systems	Business and Systems Analysis
		IT and Information Systems	Informatics
		IT and Information Systems	Information Science
		IT and Information Systems	IT skills
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics
DS 6	Data Scientist Fundamental Competency	Computer Science and Programming	Computer Science
		Computer Science and Programming	Programming
		Critical Thinking	Focused
		Critical Thinking	Persistent
		Data Analytics and Visualisation	Data Analysis
		Data Analytics and Visualisation	Data Visualisation
		Data Analytics and Visualisation	Predictive Models
		Data Management	Data Manipulation
		Data Management	Data Sourcing
		Data Management	Data Transformation
		Data Management	Database Design
		Mathematics and Statistics	Economics
		Mathematics and Statistics	Mathematics
		Mathematics and Statistics	Statistics

Competencies identified from Interview Data

Counting all skills was done to gain an idea of how to work with the data and not necessarily to generalise as the sample contained only 6 Data Scientists.



Skills identified from Interview Data

This table was done to show each knowledge and skill within each competency within the Data Scientist Fundamental Competency category. The knowledge and skills was obtained from the data collected from the interviews.

Data Management
Data Manipulation
Data Sourcing
Data Transformation
Database Design
Data Design
Data Mining
Data Modelling
Data Queries
Database Management
Mathematics and Statistics
Mathematics
Statistics
Actuarial
Applied Mathematics
Economics
Computer Science and Programming
Programming
Computer Science
Data Exploration
Exploring data
Extracts meaning from Data
Recycle data to predict future outcomes

Data Analytics and Visualisation
Data Analysis
Data Visualisation
Business Intelligence
Business Intelligence Tools
Predictive Models
Analytics
Prescriptive Models
Critical Thinking
Focused
Persistent
Attention to Detail
Critical Thinking
Self Critical
IT and Information Systems
IT skills
Application Knowledge
Business and Systems Analysis
Informatics
Information Science
Keep ahead with Data Science Techniques
Systems Knowledge

8.18 Appendix R: Data Scientist Competency Model Table

Category	Competency	Description	Skills, Knowledge, Attributes
Technical Competency	**# ** Big Data Processing**	The ability to use skills and tools to work on data to answer business questions	Hadoop
			R
			Machine Learning
			Python
			Artificial Intelligence
			Facebook Marketing
			# ** Google Analytics
			Natural Language Processing
			Neural Networks
			Technical programming skills
	# ** Data Processing	The ability to use data tools for analysis,transfromation and visualisation of data	SQL
			Excel
			Java
			# ** Mathematica
			Programming on Matlab
			Qlikview
			Power BI
			Tableau
	**** Web Tools**	The ability to build web applications using certain tools and applications	Bootstrap
			CSS
			Django
			HTML
			Javascript

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Organisational Knowledge	**Business Knowledge**	Specialised business knowledge	Business Knowledge	Chen et al. (2012)
	Financial Industry Knowledge	Applies Financial knowledge and skills and specialised knowledge of the Industry	Financial Knowledge	Davenport (2012) Kim & Lee (2016) Waller & Fawcett (2013) Wamba et al. (2015) Wixom et al. (2014)
Business Interpersonal Competency	**Curiosity and Creativity**	Have the ability to come up with new ideas and is always exploring different possibilities	Inquisitive Creative Introverted but Curious Intuitive Like Challenges Passion working with data Think out of the box	Patil & Davenport (2012) Schoenherr & Speier-Pero (2015)
	***** Communication**	Have the ability to effectively communicate insights to business	Ability to Communicate Presentation Skills Convey business insights	Abbasi et al. (2016) Carter & Sholler (2016) Chatfield et al. (2014) Chen et al. (2012) Davenport (2012) McAfee et al. (2012) Patil & Davenport (2012) Wamba et al. (2015) Watson (2014) Wixom et al. (2014)
	Business Requirements Elicitation and Specification	The ability to gather the correct requirements from business	Ask the right questions Better understand requirements Interpret Queries	Patil & Davenport (2012)
	**** Independence**	Likes to work alone and is confident that goals can be achieved	Work Alone Self Confidence Takes Initiative	Not found in Literature
	Adaptability	Should have the flexibility to adjust to any solution	Resilient Personality Flexibility Pragmatism Willing to learn new things	Patil & Davenport (2012)

Category	Competency	Description	Skills, Knowledge, Attributes	Source in Literature
Data Scientist Fundamental Competency	**Data Management**	The management process of data from the sourcing of data, modelling of data and transformation of data	Data Manipulation Data Sourcing Data Transformation Database Design Data Design Data Mining Data Modelling Data Queries Database Management	Carter & Sholler (2016) Chatfield et al. (2014) Chen et al. (2012) Chiang et al. (2012) Dinter et al. (2015) Kim & Lee (2016) McAfee et al. (2012) Miller (2014) Patil & Davenport (2012) Provost & Fawcett (2013) Schoenherr & Speier-Pero (2015) Waller & Fawcett (2013) Wamba et al. (2015) Watson (2014) Wixom et al. (2014)
	***** Mathematics and Statistics**	Have a strong background in Mathematics and Statistics	Mathematics	Carter & Sholler (2016) Chatfield et al. (2014) Dhar (2013) Miller (2014)
			Statistics	Brynjolfsson et al. (2011) Dhar (2013) McAfee et al. (2012) Miller (2014) Kim & Lee (2016)
	Data Analytics and Visualisation	Have the ability to create and maintain predictive models and present data after analysis in a visual way for business to understand	Data Analysis Predictive Models Analytics Prescriptive Models	Abbasi et al. (2016) Chatfield et al. (2014) Chiang et al. (2012) Davenport (2012) Davis (2014) Dhar (2013) Dinter et al. (2015) McAfee et al. (2012) Kim & Lee (2016) Patil & Davenport (2012) Provost & Fawcett (2013) Schoenherr & Speier-Pero (2015) Waller & Fawcett (2013) Wamba et al. (2015) Watson (2014) Wixom et al. (2014)
			Data Visualisation Business Intelligence Business Intelligence Tools	Chatfield et al. (2014) Dinter et al. (2015) McAfee et al. (2012) Schoenherr & Speier-Pero (2015) Wamba et al. (2015)
	***** Computer Science and Programming**	Has a background in Computer Science and programming	Programming	Abbasi et al. (2016) Brynjolfsson et al. (2011) Dhar (2013) Dinter et al. (2015) Kim & Lee (2016) Patil & Davenport (2012) Provost & Fawcett (2013) Watson (2014) Wixom et al. (2014)
			Computer Science	Chatfield et al. (2014) Miller (2014)
	Data Exploration	The exploration of data by extracting meaning and further enhancing processes	Exploring data Extracts meaning from Data Recycle data to predict future outcomes	Abbasi et al. (2016) Chatfield et al. (2014)
	Critical Thinking	Have the ability find best solution to business problems and is critical in the selection of a solution	Focused Persistent Attention to Detail Critical Thinking Self Critical	Schoenherr & Speier-Pero (2015)
	IT and Information Systems	Have a background in information technology to understand the basics	IT skills Application Knowledge Business and Systems Analysis Informatics Information Science Keep ahead with Data Science Techniques Systems Knowledge	Chiang et al. (2012) Dinter et al. (2015) Abbasi et al. (2016) Brynjolfsson et al. (2011) Carter & Sholler (2016) Chiang et al. (2012) Davenport (2012)

8.19 Appendix R: Turnitin Page

1558925:1558925_-_Final_Research.docx

ORIGINALITY REPORT

2%	3%	1%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.tudelft.nl Internet Source	1%
2	www.authorstream.com Internet Source	1%
3	edhs1.gsfc.nasa.gov Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		