general contractors for installation in an apartment complex or development. The size of a typical order, of course, varies dramatically between these two served markets. Large builders and contractors buy in carload (even trainload) lots with little variation in the features, functions, colors of models, etc.

It is important that you define the served market for each business carefully, since our analyses have shown that the *share* of its market achieved by a business is one of the most important single determinants of profitability. Generally speaking, the more *narrowly* the market is defined, the better. Of course, this means that it will often be necessary to make *estimates* of total market size based on appropriate allocations of published figures to specific regions, products, or types of customers.

APPENDIX XI

Aberrant clusters

Aberrant clusters

Since the cluster analysis was essentially measuring the grouping of the data according to one particular variable the aberrant clusters can be described in terms of that single variable, in this case, ICU the customer concentration variable. The table below gives the cluster number and the mean value of ICU.

Cluster Number	Number of Observations in Cluster	Mean Value of ICU
1	85	10,18
2	1	300,00
3	1	414,00
4	1	3182,00
5	15	153,33

Clearly the 3 aberrant clusters represent discrete high values of the variable "customer concentration". A closer look at these SBUs revealed that there was nothing unique about them other than that they have an unusually large value of ICU (the number of customers that make up 50% of sales ,the definition of ICU).

APPENDIX XII

Fast clustering procedure

Fast clustering procedure : Extract from SAS Manual

The FASTCLUS procedure combines an effective method for finding initial clusters with a standard iterative algorithm for minimizing the sum of squared distances from the cluster means. The result is an efficient procedure for disjoint clustering of large data sets. FASTCLUS was directly inspired by Hartigan's *leader* algorithm (1975) and MacQueen's *k-means* algorithm (1967)

PROC FASTCLUS uses a method that Anderberg (1973) calls *nearest centroid sorting.* A set of points called *cluster seeds* is selected as a first guess of the means of the clusters. Each observation is assigned to the nearest seed to form temporary clusters. The seeds are then replaced by the means of the temporary clusters and the process is repeated until no further changes occur in the clusters. Similar techniques are described in most references on clustering (Anderberg 1973; Hartigan 1975; Everitt 1989; Spath 1980).

The FASTCLUS procedure differs from other nearest centroid sorting methods in the way the initial cluster seeds are selected. The initial method of FASTCLUS guarantees that if there exist clusters such that all distances between observations in the same cluster are less than all distances between observations in different clusters, and if you tell FASTCLUS the correct number of clusters to find, then it always finds such a clustering without iterating. Even with clusters that are not as well separated, FASTCLUS usually finds sufficiently good initial seeds that few iterations are required. The importance of initial seed selection is demonstrated by Milligan (1980).

References

Anderberg, M.R. (1973), **Cluster Analysis for Applications**, New York : Academic Press.

Everitt, B.S. (1980), **Cluster Analysis**, Second Edition, London: Heineman Educational Books Ltd.

Hartigan, J.A. (1975), **Clustering Algorithms**, New York: John Wiley & Sons.

MacQueen, J. B. (1967), "Some Methods for Classification and Analysis of Multivariate Observations," **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**, 1, 281-297.

Milligan, G.W. (1980), "An Examination of the Effect of Size Types of Error Perturbation of Fifteen Clustering Algorithms," **Psychometrika**, 45, 325-342.

SAS User's Guide : Statistics Version 5 SAS Institute Cary, North Carolina, 1985 ISBN 0-917382-66-8 p378

INDEX

25X