# Project 1

## Task 1

## Task 2

I created a new dataframe ‘data\_att’, where I isolated the attitudinal variables of the original dataset. I considered attitudinal variables to be the six variables ‘variety\_of\_choice’, ‘electronics’, ‘furniture’, ‘quality\_of\_service’, ‘low\_prices’, and ‘return\_policy’.

I normalised these variables to create a new dataframe, ‘data\_att\_norm’, to contain the normalised versions of these values. I used z-score standardisation as the method of normalisation for all variables.

The ‘electronics’ variable had both the smallest minimum value and the largest maximum value across all six normalized variables, at -1.77534 and 2.85598 respectively. Other minimums and maximums can be observed in Figure X.

## Task 3

To perform hierarchical clustering on the normalised attitudinal variables, I first calculated the Euclidean distances between observations.

Chart

Description automatically generated with medium confidenceUsing these distances, I then performed hierarchical clustering with the appropriate library and the method “ward.D2”. The dendogram of the algorithm’s resulting clusters can be seen in Figure X.

Figure X: Dendogram for hierarchical clustering algorithm.

## Task 4

Observation numbers for each cluster of a six-cluster solution can be seen in Table X. 59 observations were assigned to the largest cluster, 1, and only 8 observations were assigned to the smallest cluster, 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
| 59 | 8 | 52 | 17 | 35 | 29 |

Table X: Number of observations assigned to each cluster for a six-cluster solution.

## Task 5

Chart, bar chart

Description automatically generatedThe segment profile plot generated by the ‘flexclust’ package can be observed in Figure X.

Figure X: Segment profile plots for each cluster in a six-cluster solution.

The table of cluster memberships can be observed in Table X. The results indicate that ‘hclust’ and ‘as.kcca’ procedures are in full agreement, as no cluster memberships changed between the two procedures.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0 | 59 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 8 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 52 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 17 |
| 5 | 35 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 29 | 0 |

Table X: Cluster assignments for ‘hclust’ (vertical) and ‘as.kcca’ (horizontal) procedures.

## Task 6

## Task 7

Cluster 2 only contains eight observations, which makes it a significantly smaller cluster when compared to all other clusters. This indicates that the cluster membership is very specific, which are less useful when observing customer behaviour as we are attempting to generate a generalised solution. Hence, it would be wiser to observe the results of a five or less cluster solution, where the cluster could be absorbed by another to create a larger cluster that describes a larger customer type.

## Task 8

## Task 9

## Task 10

## Task 11

## Task 12

## Task 13

# Project 2

## Task 1

## Task 2

## Task 3

## Task 4

## Task 5

## Task 6

## Task 7

## Task 8

## Task 9

## Task 10

## Task 11

## Task 12

## Task 13

## Task 14

## Task 15

## Task 16

## Task 17

## Task 18

## Task 19

# Project 3

## Task 1

## Task 2

## Task 3

## Task 4

## Task 5

## Task 6

## Task 7

## Task 8

## Task 9

## Task 10

## Task 11

## Task 12

## Task 13

## Task 14

## Task 15

## Task 16

## Task 17

## Task 18

## Task 19

## Task 20