Calculation Assignment 3 – Cluster Analysis

Class Name: Golden Gate University MSBA 324 Web and Social Analytics

Name: Akshay Danthi

Agenda

Introduction

A market survey using a 7-point Likert scale to identify clusters among customers who purchase Dog food has been conducted. Each respondent has been asked to state their agreement on 6 statements from 1 (strongly disagree) to 4 (neutral) to 7 (strongly agree). The statements are:

S1: It is important for me to buy dog food that prevents canine cavities.

S2: I like dog food that gives my dog a shiny coat.

S3: Dog food should strengthen gums.

S4: Dog food should make my dog's breath fresher.

S5: It is not a priority for me that dog food prevents tooth decay or cavities (reverse coded).

S6: When I buy dog food, I look for food that gives my dog shiny teeth.

The aim is to identify the needs of different clusters of customers to better serve them.

Research

The conclusion I have arrived from this case study is in line with the research published by (**Boya, Dotson, Hyatt, 2015**) which states that there are 3 segments of Customers in the Dog Food market, namely (i) Dog People (who highly anthropomorphize (self-identify) with their dogs and prioritize the health of their Dogs), (ii) Dog Parents (who view their dogs as family members but not as much as the Dog people), (iii) Pet Owners who do not anthropomorphize with their dogs.

This is in line with the conclusion of this case study where 3 clusters were found based on the survey options. The 3 clusters found were (i) Customers who prioritized with the Oral health of their dogs the most (paralleling with Dog People who anthropomorphized most with their dogs and thus cared about their dogs oral health instead of aesthetic characteristics such as shiny coat or shiny teeth), (ii) Customers who prioritized with the Appearance of their dogs the most (paralleling with Dog Parents who were not as attached as Cluster 1 respondents) and (iii) Customers who did not have strong opinions on any of the opinions on the survey, paralleling with the Pet Owners who least anthropomorphized with their dogs.

Software

```
# Read the CSV file into a R dataframe
dogfood_need <- read.csv("MSBA324_Week13_CalculationAssignmentCluster.csv")
dogfood_need

#Calculate the distance matrix
dist_mat <- dist(dogfood_need, method="euclidean")

# Perform hierarchical agglomerative clustering with Ward's revised method
clusters <- hclust(dist_mat, method="ward.D2")
clusters

# Plot the cluster dendogram
plot(clusters)

# *From the plot, by intersecting at the tallest vertical lines, we find that there are 3 clusters here. Hence number of clusters (k) has been chosen to be 3 below. *

# Identify each of the 3 clusters using a blue rectangle
rect.hclust(clusters, k=3, border="blue")

# Divide the data into 3 clusters
clusterid <- cutree(clusters, k=3)
clusterid

# Create separate variables for each cluster data

cluster_1 <- subset(dogfood_need, clusterid==1)
cluster_2 <- subset(dogfood_need, clusterid==2)
cluster_3 <- subset(dogfood_need, clusterid==3)

# Make it an iterable list to traverse through them using a loop
clusters <- list(cluster_1, cluster_2, cluster_3)

# Initialize a matrix for 3 clusters and 6 S columns
cluster_means_matrix <- matrix(nrow = 3, ncol = 6)


# Loop over the clusters
for(i in 1:length(clusters))
{
  # Loop over the columns S1 to S6
  for (j in 1:6)
  {
  column_name <- paste("S", j, sep = "")
  mean_value <- mean(clusters[[i]][[column_name]], na.rm = TRUE)
  # Store mean value directly in the matrix
  cluster_means_matrix[i, j] <- mean_value
  }
}
```

```
# Set the column and row names for the matrix
colnames(cluster_means_matrix) <- paste("S", 1:6, sep = "")
rownames(cluster_means_matrix) <- paste("Cluster", 1:3, sep = "")

# View the matrix
cluster_means_matrix
```

Screenshots:

```
R version 4.3.1 (2023-06-16 ucrt) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> setwd("C:/Users/axays/Downloads/MSBA_Assignment_3")
> # Read the CSV file into a R dataframe
> dogfood_need <- read.csv("MSBA324_Week13_CalculationAssignmentCluster.csv")
> dogfood_need
   Respondent S1 S2 S3 S4 S5 S6
1           1  7  3  6  4  2  4
2           2  1  3  2  4  5  4
3           3  6  2  7  4  1  3
4           4  4  5  4  6  2  5
5           5  1  2  2  3  6  2
6           6  6  3  6  4  2  4
7           7  5  3  6  3  4  3
8           8  6  4  7  4  1  4
9           9  3  4  2  3  6  3
10         10  2  6  2  6  7  6
11         11  6  4  7  3  2  3
12         12  2  3  1  4  5  4
13         13  7  2  6  4  1  3
14         14  4  6  4  5  3  6
15         15  1  3  2  2  6  4
16         16  6  4  6  3  3  4
17         17  5  3  6  3  3  4
18         18  7  3  7  4  1  4
```

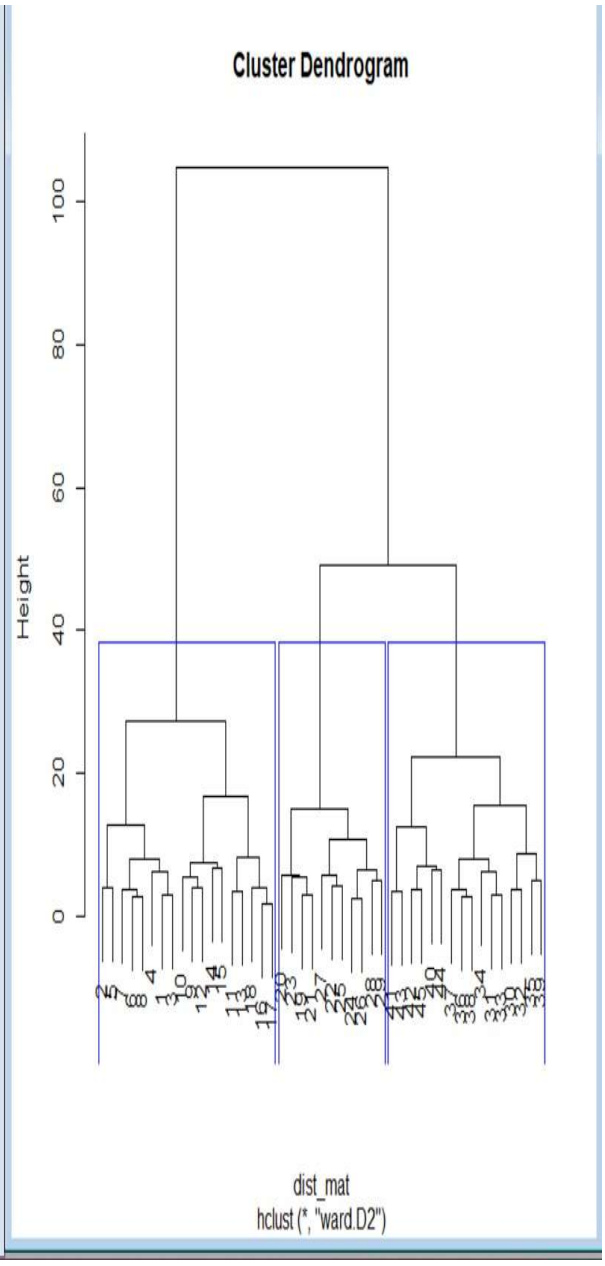
```
19          19  2  4  3  3  6  3
20          20  3  5  3  6  4  6
21          21  1  3  2  3  5  3
22          22  5  4  5  4  2  4
23          23  2  2  1  5  4  4
24          24  4  6  4  6  4  7
25          25  6  5  4  2  1  4
26          26  3  5  4  6  4  7
27          27  4  4  7  2  2  5
28          28  3  7  2  6  4  3
29          29  4  6  3  7  2  7
30          30  2  3  2  4  7  2
31          31  7  3  6  4  2  4
32          32  1  3  2  4  5  4
33          33  6  2  7  4  1  3
34          34  4  5  4  6  2  5
35          35  1  2  2  3  6  2
36          36  6  3  6  4  2  4
37          37  5  3  6  3  4  3
38          38  6  4  7  4  1  4
39          39  3  4  2  3  6  3
40          40  2  6  2  6  7  6
41          41  6  4  7  3  2  3
42          42  2  3  1  4  5  4
43          43  7  2  6  4  1  3
44          44  4  6  4  5  3  6
45          45  1  3  2  3  6  4
>
> #Calculate the distance matrix
> dist_mat <- dist(dogfood_need, method="euclidean")
>
> # Perform hierarchical agglomerative clustering with Ward's revised method
> clusters <- hclust(dist_mat, method="ward.D2")
> clusters

Call:
hclust(d = dist_mat, method = "ward.D2")

Cluster method   : ward.D2
Distance         : euclidean
Number of objects: 45

>
> # Plot the cluster dendogram
```

```
> # Plot the cluster dendogram
> plot(clusters)
>
> # *From the plot, by intersecting at the tallest vertical lines, we find that there are 3 clusters he$
>
> # Identify each of the 3 clusters using a blue rectangle
> rect.hclust(clusters, k=3, border="blue")
>
> # Divide the data into 3 clusters
> clusterid <- cutree(clusters, k=3)
> clusterid
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3
[39] 3 3 3 3 3 3 3
>
> # Create separate variables for each cluster data
> cluster_1 <- subset(dogfood_need, clusterid==1)
> cluster_2 <- subset(dogfood_need, clusterid==2)
> cluster_3 <- subset(dogfood_need, clusterid==3)
>
> # Make it an iterable list to traverse through them using a loop
> clusters <- list(cluster_1, cluster_2, cluster_3)
>
> # Initialize a matrix for 3 clusters and 6 S columns
> cluster_means_matrix <- matrix(nrow = 3, ncol = 6)
>
> # Loop over the clusters
> for(i in 1:length(clusters)) {
+
+   # Loop over the columns S1 to S6
+   for (j in 1:6) {
+     column_name <- paste("S", j, sep = "")
+     mean_value <- mean(clusters[[i]][[column_name]], na.rm = TRUE)
+
+     # Store mean value directly in the matrix
+     cluster_means_matrix[i, j] <- mean_value
+   }
+ }
>
> # Set the column and row names for the matrix
> colnames(cluster_means_matrix) <- paste("S", 1:6, sep = "")
> rownames(cluster_means_matrix) <- paste("Cluster", 1:3, sep = "")
```



```
>
> # Set the column and row names for the matrix
> colnames(cluster_means_matrix) <- paste("S", 1:6, sep = "")
> rownames(cluster_means_matrix) <- paste("Cluster", 1:3, sep = "")
>
> # View the matrix
> cluster_means_matrix
                S1       S2       S3       S4       S5       S6
Cluster1 4.388889 3.500000 4.611111 3.833333 3.333333 3.888889
Cluster2 3.363636 4.636364 3.454545 4.545455 3.454545 4.818182
Cluster3 3.937500 3.500000 4.125000 4.000000 3.750000 3.750000
>
```

Commentary:

We see that for Cluster 1, S1 and S3 opinions carry a disproportionately higher weight.

We see that for Cluster 2, S2 and S6 opinions carry a disproportionately higher weight.

We see that for Cluster 3, the means are relatively even and not disproportionately higher.

We can see that Cluster 1 respondents while having a strong preference to Strengthen gums (S3), also prefer to buy food that prevents cavities (S1). This conclusion is also further strengthened by their lower score to the reverse-coded question about Cavities (S5).

Similarly, respondents of Cluster 2 care more about the appearance of their Dog (S2->Shiny coat and S6-> Shiny teeth).

However, Cluster 3 respondents have a relatively even score for all the opinions and do not have any disproportionately high Mean for any opinion. This means that Either A. They do not have strong opinions on Dog food they purchase OR B. Their interests are not represented in this survey.

This is in line with the research paper published by the (**Boya, Dotson, Hyatt, 2015**) as referenced in the Research section of this paper.

References

1. Boya, U. O., Dotson, M. J., & Hyatt, E. M. (2015). A comparison of dog food choice criteria across dog owner segments: an exploratory study. International Journal of Consumer Studies, 39(1), 74–82. https://doi.org/10.1111/ijcs.12145