**Cluster Analysis for Boston Boats**

Young Doek Jung, Nikki, Seong Shin, Tyron Greg Sabatta (student ID:2019730615)

**1) Import the data**

First, create a new R code script file, and Save this script file to your designated course folder. Name it as “Group assignment 7variables.R”

#Step 1: Import the Data

***setwd("C:/Users/user/Desktop/Dataset")***

***DT=read.csv("GroupProject\_Boats\_7variables.csv")***

**2) Check regularities of the data**

***head(DT,5) dim(DT)***

***summary(DT)***

***summary(DT$Q1.25****) # to double check.*

**3) Scale the data if necessary:** Not applicable

**4) Select segmentation variables**

Note this was done first with 29 Variables, and then also with 3 segments so we could eliminate similar variables. We then condensed the variables to 7 questions we felt covered all areas and had no overlap.

***Segment = DT[,c("Q1.1","Q1.2","Q1.4","Q1.11","Q1.13","Q1.19","Q1.25")]***

***#View(Segment)***

***#Segment = DT[,2:8]***

***#View(Segment)***

**5) Define similarity measure**

***EuclideanD <- dist(Segment[1:2813, 1:7])***

***EuclideanD <- dist(Segment[1:10, 1:7])***

***EuclideanD = as.matrix(EuclideanD)***

***View( round(EuclideanD, 1) )***

**6) Clustering methods**

***EuclideanD = dist(Segment[, 1:7])***

***Hierarchical\_Cluster = hclust(EuclideanD)***

***plot(Hierarchical\_Cluster)***

#Heatmap

***heatmap(as.matrix(Segment[, 1:7]))***

#Membership

***HC\_boat= as.vector(cutree(Hierarchical\_Cluster, k = 4))***

***YD\_HC1 = cbind(DT, HC\_boat)***

***head(YD\_HC1)***

**7) Profile and interpret the segments**

***prop.table(table(YD\_HC1$HC\_boat))***

***Mean\_by\_Group <- function(data, groups) { aggregate(data, list(groups), function(x) mean(as.numeric(x))) }***

***Group\_Character\_HC= Mean\_by\_Group(YD\_HC1, YD\_HC1$HC\_boat)***

***View(Group\_Character\_HC)***

***write.csv(Group\_Character\_HC, "Group\_Character\_HC7-1.csv", row.names = FALSE)***

**8) Business insights marketing decisions**

Based on the data we were able to segment it into 4 main segments. We then analyzed these four segments further and selected two of the segments we felt gave us the best opportunity and coverage in the American market. (*this data can be seen in the Appendix).*

We then created ***two personas***, for each of the segments and from selected the best boat ***Montauk 170 & Vantage 270*** that satisfied each persona. We then comprised the two profiles below:

**Montauk 170**

**Persona:** Family/ Single/married, Male, Salary $50,000, price sensitive consumers, Enjoy fishing, cruising, entertaining/socializing.

**STP:**

*For family men looking for an affordable enjoyable fishing boat, the Montauk 170 is simply the best for the everyday boating family, providing enhanced comfort, performance and versatility unlike any other boating brand, so that they can make a positive difference in their family and friends lives. The brand will exceed your expectations in every way with its superior storage, affordability, is easy to use and economical to run*.

**Vantage 270**

**Persona:** Married, Male, Salary $90,000, outgoing, have knowledge about Boats (Advanced), usually boat with others, like cruising, swimming, Entertaining, Boating more than 50% of Boating season’s weekends

**STP:**

*For family men who never compromise and never settle for less with knowledge about boat specs and functionality and are looking for outgoing cruising and entertainment, Vantage 270 would give unmatched ride quality with every activity dreamed up. Optional Summer Kitchen with luxury touches, including a refrigerator, electric grill will satisfy the need of leisure cruising. Unlike others, Armrest seats and solid performance ensure the quality level experience and safety first.*

**Appendix: ## 7Variables#**

**Segment Variables: ("Q1.1","Q1.2","Q1.4","Q1.11","Q1.13","Q1.19","Q1.25")**

**Profiling Variables: Q7.1, Q8, Q9.1 – Q9.5, Q10 and Q11 – Q14**

*Variables corrected: Q13, Q14 and Q10.*

Note: For *Q13,* the average is 2 for widowed, however this is inaccurate because 70% of the recipients are married in Q13. Therefore, average reflects only widowed.

For Q14, where the average can be seen to be self-employed however 60% of the recipients in Q14 are paid employees. This Q14 is important to note for Persona, as persona should include paid employees**.**

Note:

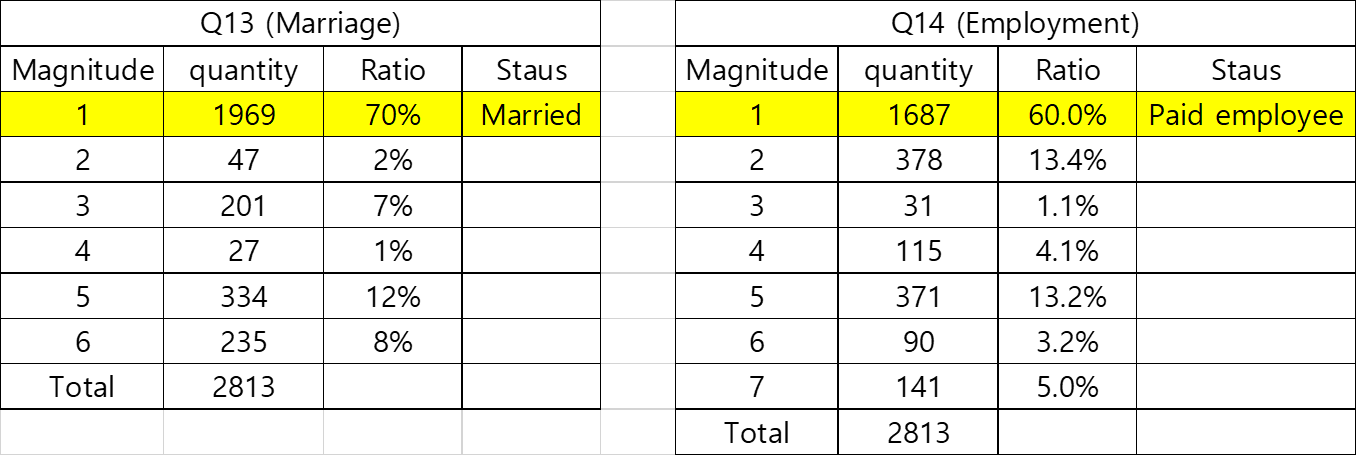
1 = married and majority of Q13 is 1. So, all we need to do is how many is #1 = married and others are not important. If married is 1 then others are 0, with that proportion the average is very close to 2.1.

and Q14 with that proportion the average is very close to 2.4. Here, both do not make any sense.

i.e: (1X1969)+(2X47)+(3X201)+(4X27)+(5X334)+(6X235) = 5854

5854 / 2813

= 2.08, that is why the average is 2.1.



We added an additional variable for Q10 which reflects the days of boating per a persona**.** *The equation is = 44days (Boating seasons Sat and Sun) X Avg(Q9.1~9.5) X 5/7*

5/7 is a magnitude

1 - 0%

2 - 20%

3 - 40%

4- 60%

5- 80%

Note: we can say that usual costumers boating days are about 50% days of boating seasons' weekends.

Appendix: We selected Group 1 & 2

*Segment:*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | *Elastic,Considerable* | *Elastic, cheapest* | *Inelastic, Brandish* | *Moderate, expert, Self repair* | *Inelastic, Show off* | *Moderate, outgoing* | *Moderate, outgoing, Expert* | *Boat Frequency* | *Level* | *Fishing* |
|  | **ResponderID** | **Q1.1** | **Q1.2** | **Q1.4** | **Q1.11** | **Q1.13** | **Q1.19** | **Q1.25** | **Q7.1** | **Q8** | **Q9.1** |
| Vantage | 191787.465 | 4.1 | 2.7 | 4.1 | 3.7 | 3.2 | 4.3 | 4.0 | 2.3 | 2.4 | 3.6 |
| Montage | 207968.958 | 3.9 | 3.8 | 3.3 | 3.1 | 2.5 | 4.2 | 3.6 | 2.2 | 2.0 | 3.6 |
| 3 | 243409.514 | 3.8 | 2.4 | 4.0 | 1.8 | 2.5 | 4.1 | 2.6 | 1.6 | 2.0 | 3.2 |
| 4 | 253246.364 | 2.8 | 2.7 | 2.6 | 2.7 | 2.2 | 2.7 | 1.8 | 2.2 | 2.0 | 3.3 |

*Behavior & Demographics:*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Swimming | Crusing | Water Sports | Entertaining/  Socializing | Boating  Days | Gender | Household  income | Marriage | Employment |  |
|  | **ResponderID** | **Q9.2** | **Q9.3** | **Q9.4** | **Q9.5** | **Q10** | **Q11** | **Q12** | **Q13** | **Q14** | **HC\_boat** |
| Group1 | 191787.5 | 3.5 | 3.7 | 3.2 | 3.5 | 22.2 | 1.4 | 13.6 | 2.1 | 2.2 | 1 |
| Group2 | 207969 | 3.4 | 3.7 | 3.1 | 3.4 | 21.6 | 1.5 | 12.7 | 2.2 | 2.4 | 2 |
| 3 | 243409.5 | 3.2 | 3.7 | 3.0 | 3.5 | 22.0 | 1.8 | 14.3 | 1.9 | 2.6 | 3 |
| 4 | 253246.4 | 2.8 | 3.2 | 3.0 | 3.0 | 19.1 | 1.5 | 12.8 | 1.7 | 2.6 | 4 |