**SDM-FINAL**

**Vinay ratnala**

**U12127232**

**Introduction:**

The issue bargains upon the retail analytics and we are inquired to unravel them utilizing RFMD analysis. The issue comes up with three sorts of campaigns sending coupons to each family units in case they are appearing any impact on the family’s acquiring behavior. Identify the components that might help in decreasing the recency of every family. Our concern is to analyze if there's any impact of the campaign and report it to the store manager.

We can observe customer behavior and learn from it. We have a variety of independent factors such as age, marital status, and income. We can predict their consequences and plan accordingly.

Customer loyalty can be established because we prefer to focus and market based on the information obtained from customers.

Promotions can be tailored to the preferences of customers. We can classify households and plan successful marketing efforts.

**Adding a 'D' variable to the equation can help us learn more about household spending trends.**

**We can, for example, look at how many weeks each household has spent for all the weeks.**

**Part 2. Exploring K-type of Heterogeneity**

**Elbow Plots:**

Chart, line chart, scatter chart

Description automatically generated

The value of k at the "elbow," or the point where distortion/inertia starts to fall linearly, must be determined to establish the appropriate number of clusters. As a result, we find that 4 clusters are better when compared to 3 clusters.

K\_means segmentation with three clusters

Chart

Description automatically generated

K\_means segmentation with 4 clusters

Chart

Description automatically generated

**Cluster analysis:**

**Summarizing values in a table for each category in each segment**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Avg\_recency | Avg\_Frequency | Avg\_Monetary | Avg\_log(D) |
| 3Clusters-Segment1 | 0.698 | 2.88 | 62.1 | 0.901 |
| 3 Clusters -Segment2 | 0.647 | 3.00 | 91.6 | 1.02 |
| 3 Clusters -Segment3 | 0.722 | 2.90 | 51.7 | 0.837 |
| 4 Clusters -Segment1 | 0.672 | 2.96 | 80.5 | 0.982 |
| 4 Clusters -Segment2 | 0.704 | 2.88 | 61.9 | 0.909 |
| 4 Clusters -Segment3 | 0.533 | 3.05 | 97.0 | 0.913 |
| 4 Clusters -Segment4 | 0.722 | 2.90 | 51.7 | 0.837 |

3Clusters-Segment 3 has least d and monetary values.

3Clusters -Segment 1 has highest recency values.

3Clusters -Segment2 has highest frequency, D and monetary values.

4Clusters- Segment1 has highest D value.

4Clusters - Segment3 has highest frequency and monetary value.

4Clusters -Segment4 has highest recency value

**Applying Fixed effects to all the 3 segments cluster:**

f\_cluster1=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d+as.factor(hhsize),data=cluster1)

f\_cluster2=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d+as.factor(hhsize),data=cluster2)

f\_cluster3=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d+as.factor(hhsize),data=cluster3)

Stargazer output for all the three models:

**Table

Description automatically generated**

**Applying Fixed effects to all the 3 segments cluster:**

**f\_cluster\_4\_1=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d+as.factor(hhsize),data=clusters\_4\_1)**

**f\_cluster\_4\_2=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d+as.factor(hhsize),data=clusters\_4\_2)**

**f\_cluster\_4\_3=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d+as.factor(hhsize),data=clusters\_4\_3)**

**f\_cluster\_4\_4=lm(weeklyspend ~ recency+frequency+spendingavg+as.factor(NumA\_Ever\_cat)+as.factor(NumB\_Ever\_cat)+as.factor(NumC\_Ever\_cat)+log\_d +as.factor(hhsize),data=clusters\_4\_4)**

**StarGazer output for all the four models:**

**Table

Description automatically generated**

**Part 3-Please explain which response model you would prefer as a data analyst interested in designing marketing campaigns: Fixed Effect or K-means+Fixed Effects?. Discuss the pros and cons of segmentation from both a managerial and IT perspective.**

**Text

Description automatically generated**

**From the AIC values we can say that custer\_4\_2 which is 16355.2 having the Least AIC value.**

I prefer the Kmeans + Fixed Effects model since it allows us to categorize and examine the effects based on AIC values. From a commercial standpoint, deciding promotions and designing marketing campaigns based on segmentation will be simple. People who are specifically targeted would be more particular.

However, segmentation has a few drawbacks. If we mindlessly follow the results, we may miss a group of households. Household habits may alter in the actual world, and the data we have may not be completely accurate. So, to some extent, we should use these segmented models and make business decisions based on domain knowledge.

**Section 3: Conclusions, directions for further investigation, and comments to the instructor about the exam.**

To figure out which campaigns are more successful. I utilized glm regressions OLS models, but because the data is multi-leveled and the independence test is frequently broken, our interpretation is skewed. We employed multilayer modeling for these reasons.By looking at the m7 output. We can get greater weekly spending if we run more A campaigns for clients.