MARKET ANALYSIS OF THE BATH SOAP INDUSTRY

A FINAL PROJECT REPORT WAS SUBMITTED

TO

KENT STATE UNIVERSITY

MASTER OF SCIENCE

IN

BUSINESS ANALYTICS

SUBMITTED BY

MEGHANA UDIGA (811251566)

MUDIGA@KENT.EDU

INSTRUCTOR

Dr. CHAOJIANG(CJ) WU

PROFESSOR

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

KENT STATE UNIVERSITY

CONTENTS

INTRODUCTION

In the past, ABC has divided markets based on the characteristics of its customers. About 50,000 household panels, or 80% of the Indian urban market, have been assembled by ABC in 105 Indian cities and towns. Utilizing stratified sampling, the households are selectively chosen. About 30 product categories (such as detergents, for example) and 60 brands are tracked by ABC for each category**.** ABC keeps the following data for its household data, which consists of transaction data (each row represents a transaction) characteristics of the households. -Ownership of long-lasting items (car, washing machine, etc.; updated annually). -Product category and brand purchase statistics (updated monthly**).**

PROBLEM STATEMENT

The ABC Company is interested in learning how the business is doing based on consumer feedback and which of the elements has an impact on the company's profit. After a quick data analysis, I learned how to answer the above issue for the project and how to use k means clustering to solve the problem. The primary focus of the project is on how well we can analyze data and segment it to make suggestions for how to improve services where the company lags.

DATA DESCRIPTION

The dataset was taken from the Kaggle website which consists of 600 rows and 47 variables.

Data cleaning:

For our project, some of the columns are not required so I removed some of the columns. Converting all character variable values into numeric values. Later I checked for null values in the dataset. I found that there are no null values in the dataset.

The data was considered in the dataset as follows.

1. SEX has 1 = male, 2 = female
2. EDU has 1 to 9 Levels
3. CS has 1 = available, 2 = unavailable
4. CHILD has 1 to 4 Levels.

ANALYSIS & DISCUSSION

K means Clustering:

K-means separates the collection of data items into distinct subgroups (clusters), where each data item refers to a single subset.

For finding the optimum number of clusters I used WSS and silhouette methods.

Silhouette Method:

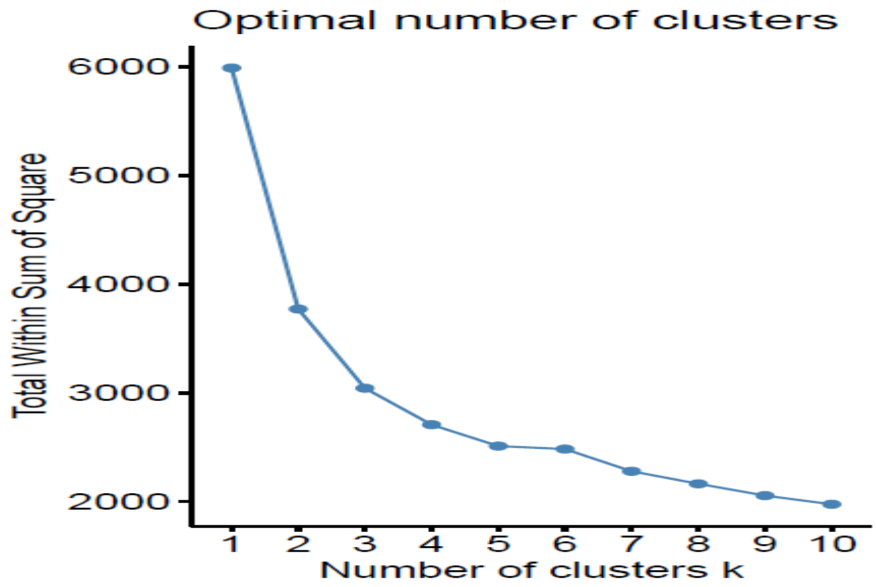


Fig: Silhouette Method Graph

From the above Silhouette Method Graph, we can say that the optimum number of clusters is 3.

WSS Method:

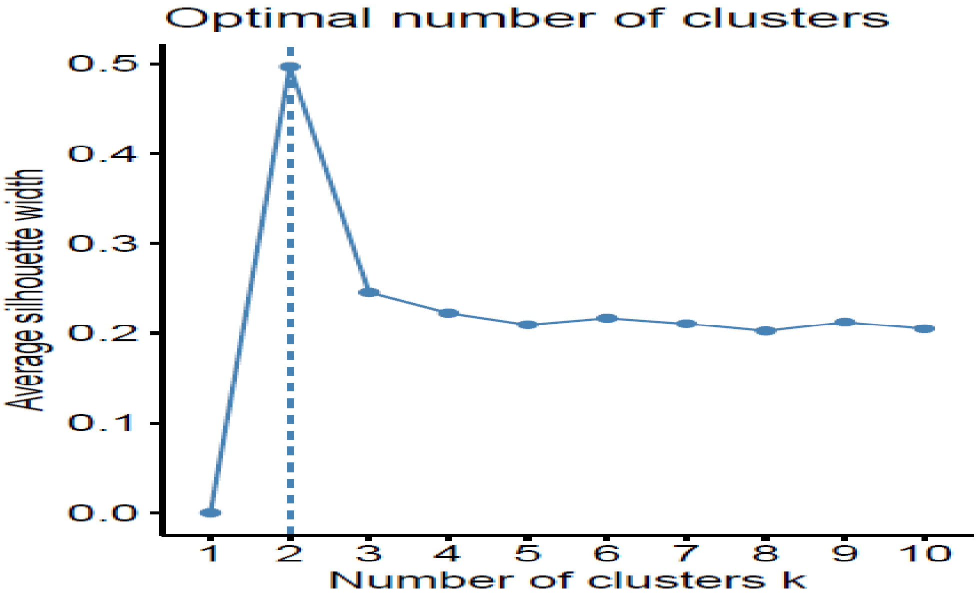


Fig: WSS Method Graph

From the above WSS Method Graph, we can say that the optimum number of clusters is 2.

Based on the WSS method I divided data into 2 clusters.

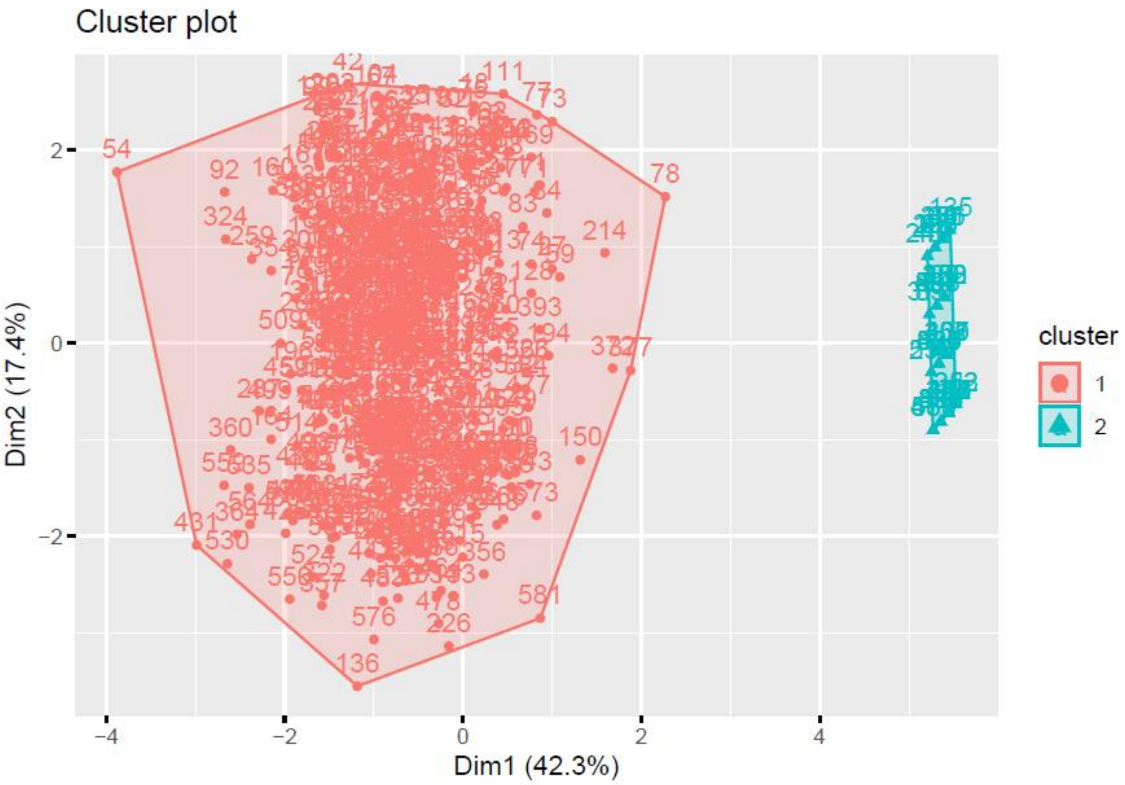


Fig: Clusters Formation Based on WSS Method

According to the abovementioned graph, cluster 1 has excellent customer ratings, which indicates that the region has a higher percentage of dedicated consumers and a very high level of customer satisfaction.

We need to improve services in cluster 2 factors because Cluster 2 has few customer reviews of the industry.

Based on the Silhouette Method I divided data into 3 clusters:

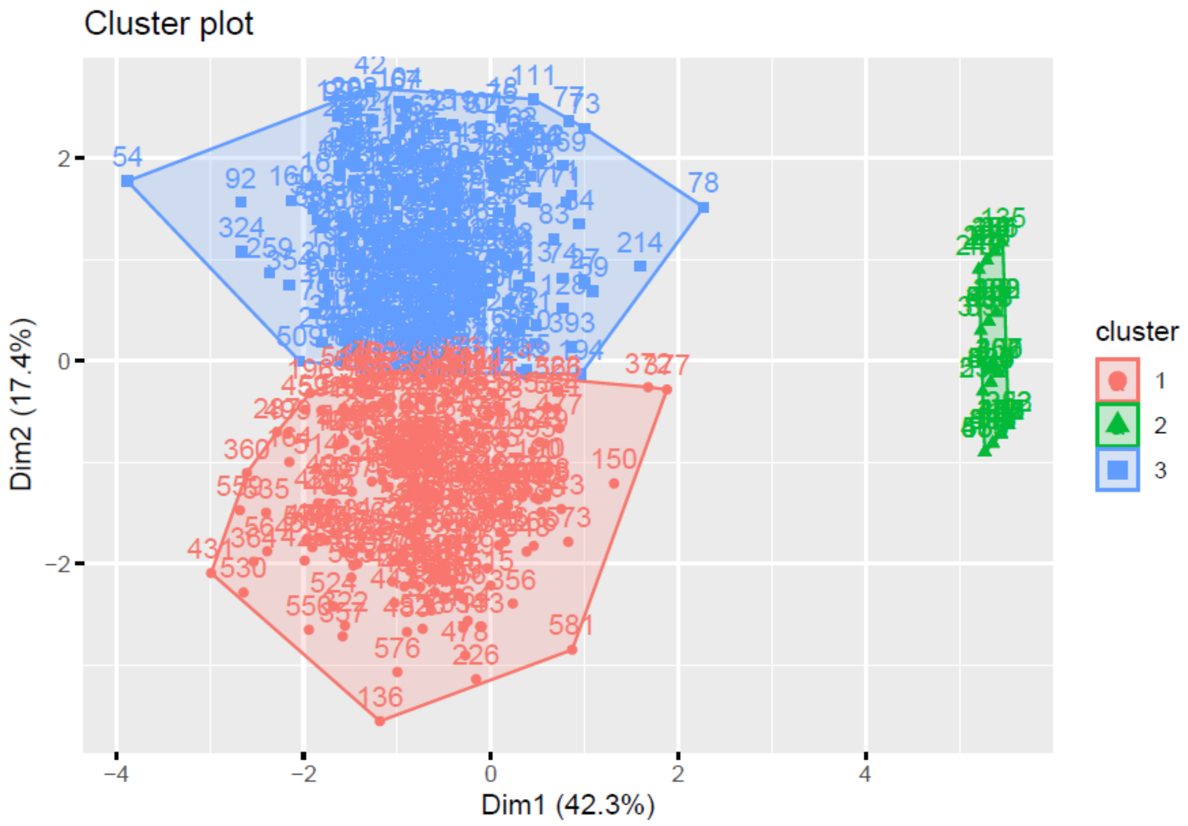


Fig: Clusters Formation Based on Silhouette Method

We may infer from the preceding graph that cluster 3 has positive customer ratings, which indicates that there are more devoted customers there.

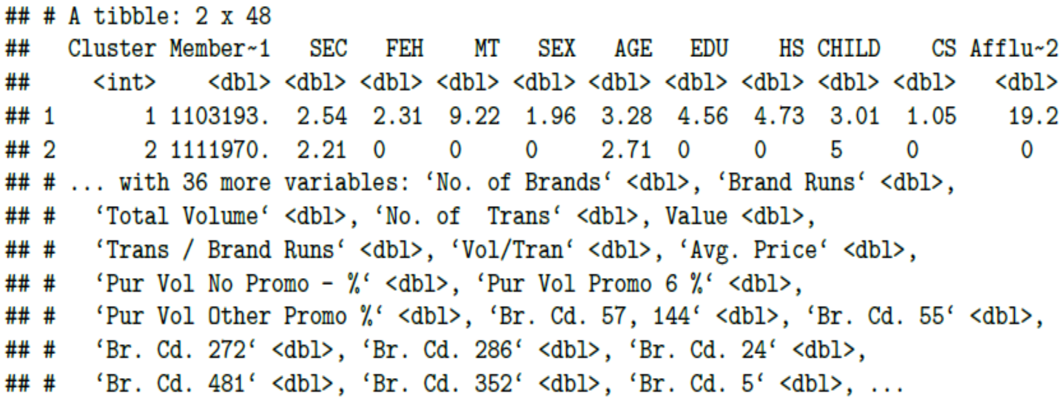
The consumer feedback for cluster 2 is mixed, and our services need to be improved.

There are hardly any customer reviews of the industry for cluster 1. We must prioritize serving improvements.

The clusters divide the data uniformly and each cluster contains the power plants which use all fuel types such as coal, Petroleum, natural gas, Petroleum Coke, etc.

The main agenda is to find which cluster consists of low fuel costs and Low emission of pollutants. So that we can predict which power plants generate power with lower operating costs in the US.

WSS Method clustering data:

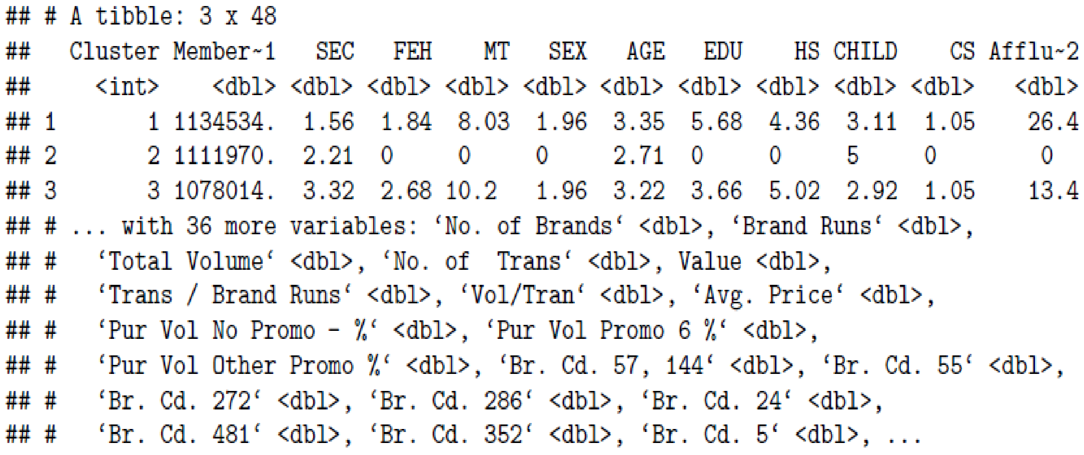


```
## # A tibble: 2 x 48
##   Cluster Member~1   SEC   FEH    MT   SEX   AGE   EDU    HS CHILD    CS Afflu~2
##     <int>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1       1 1103193.  2.54  2.31  9.22  1.96  3.28  4.56  4.73  3.01  1.05    19.2
## 2       2 1111970.  2.21  0     0     0     2.71  0     0     5     0        0
## # ... with 36 more variables: ‘No. of Brands‘ <dbl>, ‘Brand Runs‘ <dbl>,
## #   ‘Total Volume‘ <dbl>, ‘No. of  Trans‘ <dbl>, Value <dbl>,
## #   ‘Trans / Brand Runs‘ <dbl>, ‘Vol/Tran‘ <dbl>, ‘Avg. Price‘ <dbl>,
## #   ‘Pur Vol No Promo - %‘ <dbl>, ‘Pur Vol Promo 6 %‘ <dbl>,
## #   ‘Pur Vol Other Promo %‘ <dbl>, ‘Br. Cd. 57, 144‘ <dbl>, ‘Br. Cd. 55‘ <dbl>,
## #   ‘Br. Cd. 272‘ <dbl>, ‘Br. Cd. 286‘ <dbl>, ‘Br. Cd. 24‘ <dbl>,
## #   ‘Br. Cd. 481‘ <dbl>, ‘Br. Cd. 352‘ <dbl>, ‘Br. Cd. 5‘ <dbl>, ...
```

Fig: WSS Method clustered data

According to the above table, Cluster 1 has mean values for characteristics like SEC, FEH, MT, SEX, AGE, and HS that are higher than those in Cluster 2 in comparison. When compared to cluster 1, the youngster spends more money in cluster 2 since its mean values are higher**.**

Silhouette Method clustering data:



```
## # A tibble: 3 x 48
##   Cluster Member~1   SEC   FEH    MT   SEX   AGE   EDU    HS CHILD    CS Afflu~2
##     <int>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1       1 1134534.  1.56  1.84  8.03  1.96  3.35  5.68  4.36  3.11  1.05    26.4
## 2       2 1111970.  2.21  0     0     0     2.71  0     0     5     0        0
## 3       3 1078014.  3.32  2.68 10.2   1.96  3.22  3.66  5.02  2.92  1.05    13.4
## # ... with 36 more variables: ‘No. of Brands‘ <dbl>, ‘Brand Runs‘ <dbl>,
## #   ‘Total Volume‘ <dbl>, ‘No. of  Trans‘ <dbl>, Value <dbl>,
## #   ‘Trans / Brand Runs‘ <dbl>, ‘Vol/Tran‘ <dbl>, ‘Avg. Price‘ <dbl>,
## #   ‘Pur Vol No Promo - %‘ <dbl>, ‘Pur Vol Promo 6 %‘ <dbl>,
## #   ‘Pur Vol Other Promo %‘ <dbl>, ‘Br. Cd. 57, 144‘ <dbl>, ‘Br. Cd. 55‘ <dbl>,
## #   ‘Br. Cd. 272‘ <dbl>, ‘Br. Cd. 286‘ <dbl>, ‘Br. Cd. 24‘ <dbl>,
## #   ‘Br. Cd. 481‘ <dbl>, ‘Br. Cd. 352‘ <dbl>, ‘Br. Cd. 5‘ <dbl>, ...
```

Fig: Silhouette Method clustered data

According to the above table, Cluster 1 has higher mean values than Clusters 2 and 3, which indicates that Cluster 1 consumers make more purchases than customers in Clusters 2 and 3.

Compared to the other clusters, Cluster 2 has the lowest consumer weight. When compared to the other two clusters, Cluster 3 has moderate client purchases.

CONCLUSION

In concluding that, based on the WSS method based on k means clustering. Cluster 1 has excellent customer ratings, which indicates that the region has a higher percentage of dedicated consumers and a high level of customer satisfaction. We need to improve services in cluster 2 factors because Cluster 2 has few customer reviews of the industry.

Comparatively speaking, Cluster 1 has mean values for traits that are higher than Cluster 2. Since cluster 2's mean values are greater than cluster 1's, the child spends more money there.

In the table above, Cluster 1 has mean values that are greater than Clusters 2 and 3, indicating that Cluster 1 consumers spend more money than consumers in Clusters 2 and 3. Cluster 2 has the lowest consumer weight when compared to the other clusters. Cluster 3 has a moderate level of customer purchases as compared to the other two clusters.

From these two methodologies, Silhouette Method clustering gives optimum results in my thoughts because the splitting of the data is uniform and clear.

EXECUTIVE SUMMARY

In this project, I could observe the below results

The ABC company’s growth is mainly depending on customer satisfaction. The impacting factors of customer satisfaction are analyzed using K means clustering.

For the solution we have done two methods one is WSS-based and the other one is the Silhouette method. In Comparative of both the methods, the Silhouette method gave better results because the whole data is slitted not 3 clusters. Cluster 1 has more loyal customers and cluster 2 has more child-related reviews but the results are not favorable. So, we need to improve the services in that area. Cluster 3 has a moderate level of purchase review, so we need to improve the services as a high priority.

I believe that ABC company will implement new strategies and services to get exceptional customer satisfaction and reviews.

References:

1. **https://www.kaggle.com/search?q=DATASETS**
2. **https://www.researchgate.net/publication/341236998_MARKET_ANALYSIS_INSTRUMENTS_IN_THE_DEVELOPMENT_OF_THE_STARTUP_MARKETING_STRATEGY**
3. **https://www.researchgate.net/publication/271616608_A_Clustering_Method_Based_on_K-Means_Algorithm**