# Customer Segmentation based on Transactional Data

## A Executive Summary

The behavioural data which is captured in the form of transactional data is used for customer segmentation. It categorizes the customers based on spending and consumption habits/behaviour. The customers in a group or cluster shares similar characteristics and hence it can be used to identify important groups for the business and to develop customized marketing campaigns for individual clusters. To begin the segmentation process, first, the datasets were explored and processed to remove any discrepancies. Then, 25 features were selected from the existing datasets and 10 features were generated describing customers' spending behaviour. As the features exhibited skewed distribution, power transformation was applied so that they followed the gaussian-like distribution. The features were then standardized to ensure that each variable contributed equally. Following this, Principal Component Analysis was applied to reduce the number of features, 8 Principal components were selected, which captured 70% of the total variance of data. After the dimensionality reduction, K-means clustering was applied to divide the customers into clusters which best describe the customer groups. The number of clusters was chosen to be 5 based on the highest silhouette score. The clusters are then visualised on a 2-D scatter plot. Finally, from the 5 segments, these groups are created: Occasional Buyers, Convenience-seeker Buyers, Premium Buyers, Regular Buyers and Budget Buyers. Based on specific marketing strategy, the business can target each of these buyer segments.

## B Data Preparation

There are 4 datasets provided, all of which contain neither duplicate rows nor missing values.

- **baskets_sample** - There are 195,547 rows and 5 columns. Each row represents a single visit or purchase by a customer.
- **category_spends_sample** - There are 3000 rows and 21 columns. Each row represents a customer's spend across 20 product categories.
- **customers_sample** - There are 3000 rows and 6 columns. Each row represents a customer's overall spending habits.
- **lineitems_sample** - There are 1,461,315 rows and 6 columns. Each row represents purchase of individual product category items.

The data cleaning process is required to ensure that the data is in the correct format for analysis. For the analysis, (i) the first step towards data preparation included removing of pound logo "£" and comma "," from the spend values. (ii) The next step was to convert those values from string data type to float data type. On further investigation, it was observed that there were some discrepancies found in the (i) "customers_sample" ("baskets" column had incorrect values) and (ii) "category_spends_sample" ("bakery" column had all zero values) tables. Hence, to ensure the correctness of the data, all the features were re-generated from the raw data of "lineitems_sample" table. Moreover, it was also used to generate additional features that were used for exploratory analysis and subsequently for customer segmentation.

## C Feature Description

The features are generated and selected from "lineitems_sample" table such that they can help to distinguish and describe the customers of the clusters. There are total **35** features which are described below. **A data dictionary has been provided in Appendix-A which explains the description of each feature.**

- **Feature Selection**: All 20 features from "category_spends" table are selected, which shows the amount spent by the customer on each category of products. The product categories include Bakery, Cashpoint, Confectionary, ... Soft_Drinks, Tobacco, World_Foods. These features can help to identify customer preferences for the products and to identify which product categories are most popular among customers. Additionally, all 5 features from "customers" table are selected. These features provide information about the customer's purchase behavior, such as the number of purchases made, the total quantity of items purchased, the average quantity of items purchased per transaction, the total spend and the average spend per transaction. These features are important for understanding customer's general buying habits and identifying the most valuable customers for marketing.
- **Feature Generation**: There are 10 additional features generated to get more insights about consumer spending behaviour, such as the average spend per product category, customer's average spend per item, spend on food and non-food items, the recency of the last purchase, and the average interval between purchases. The features describing spending during weekdays and weekends and spending during daytime and nighttime tells when customers are more likely to make purchases.

## D Customer Base Summary

The data is a transactional dataset generated over 6 months period at the retailer's store. There are 3000 customers who have shopped during this period at the store. Following are the key highlights from the exploratory analysis of the data summarizing the company's market. (Please refer to Appendix-B containing the visual description of the highlights)

- **Visits** - There are more than 1000 visits per day on an average at the store. On an average, a customer visits the store 65 times in 6 months period.
- **Revenue generated** - The item category that generated the highest revenue is "Tobacco" having revenue generated of £277,642 and lowest is "Discount Bakery" with revenue generation of £5,444. (Figure B1)
- **Quantities sold** - the item category with the highest sold quantities is "Dairy" with total quantities sold equal to 224,304 and lowest is "Practical Items" with total quantities sold equal to 3,103. (Figure B2)
- **Busy time** - Most busy time of the day is Noon, with total visits equal to 73,069 and least busy time is Night with total visits equal to 6,819. (Figure B3)
- **Busy day** - Most busy day of the week is Friday, with total visits equal to 32,552 and least busy day is Sunday with total visits equal to 14,730. Also, most revenue is generated on Friday and least on Sunday (Figure B4)

## E Feature Engineering

The feature engineering is done to normalize the data and reduce the number of features, as it will help to ensure that clustering algorithm performs well and produces meaningful clusters.

### E.1 Data transformation

It has been observed that most of the features have skewed distribution, hence a transformation needs to be applied to make the distribution more gaussian-like.

- The data contained some negative values under "LOTTERY" feature, hence log transformation could not be applied. The power transformation using the "Yeo-Johnson" method was applied which could handle both positive and negative values. Additionally, the power transformation also standardizes the data to zero mean and unit variance after eliminating the skewness in the distribution.
- Standardization is important so that all features are on the same scale and hence each feature is equally treated. For instance, "total_spend" feature had values ranging from 7.28 to 6588.65 whereas "recency" feature had values ranging from 0 to 164. If features are not standardized, it will lead to over-emphasizing some features having higher magnitudes.

### E.2 Dimensionality reduction

The distance based clustering algorithms are affected by curse of dimensionality, hence we need to minimize the number of features. To do that, a dimensionality reduction technique called **PCA (Principal Component Analysis)** is used. It rotates the dataset in such a way that the rotated features are uncorrelated. It can be used to reduce the number of features while preserving as much information as possible. The PCA produces principal components equal to the number of original features. To reduce the dimensionality of the data, we need to specify the number of components for PCA. Following are the methods which can be used to select the number of principal components:

- Cumulative Explained Variance: The number of principal components are selected such that all the components together captures atleast 70% of the total variance in the data.

- Scree Plot: A scree plot is a graphical representation of eigenvalues of principal components when presented in decreasing order. The number of principal components to retain can be determined by looking for the "elbow" point on the scree plot.

- Kaiser's rule: According to Kaiser's rule, all the principal components having eigenvalues greater than 1 should be selected.

For the analysis, **cumulative explained variance** is used to select 8 principal components. These components capture at least 70% of the total variance, as can be seen from Figure C1 (Appendix C).

## F Segmentation Methodology

A popular unsupervised algorithm called **K-means clustering** is useful to segment customers for exploring the customer base and marketing strategies. It partitions the customers into k clusters based on the similarity of the features. It creates clusters by minimizing the sum of the squared distances between the data points and their assigned cluster centres. The main advantages of this algorithm are that we can specify the exact number of clusters that we require and also it is a computationally light clustering algorithm compared to others. It also assumes that the data points are represented as continuous variables. Therefore, to begin the clustering, first, let us determine the number of appropriate clusters. This can be found using the following methods: (i) Silhouette Score - The average silhouette score is calculated for 5, 6 and 7 clusters. Here, **the number of clusters resulting in the highest average silhouette score is 5. (Refer to Figure D1, Appendix D)** (ii) Elbow Method - The number of clusters are selected by taking the elbow point on the graph of "No. of clusters" vs "WCSS" (Within-Cluster-Sum-of-Squares).

## G Results and Analysis

A good clustering is said to be done if the distance between datapoints in a cluster is very low compared to the distance between the clusters. The distance measure used in the algorithm is "Eucledian Distance" by default. This section presents the statistical summary of 5 clusters which have been formed using K-means algorithm. Here, the tables for each segment depict various statistics and highlight the major features which differentiate between the various segments. Out of 3000 customers, the distribution of customers in each segment is also given. Along with the statistical summary, the pen portraits have also been described as follows. The pen portraits are created by analysing cluster centroids and by comparing the mean and median values of all features of each cluster. (Ref Figure E2, Appendix E)

### Occaional Buyers

The customers of this segment have made very few purchases as their average visit interval is high among all the clusters. In other words, their last purchase was also quite a time ago. They spend moderately during their purchases buying not many different categories of products. They have made the least purchase across all the product categories with almost zero spendings on "Lottery" and "Practical items". Hence, they are also the least spenders among all segments.

Table 1: **Statistical Summary for Occasional Buyers having 442 customers**

| **Feature** | **Mean** | **Std** | **Min** | **25%** | **50%** | **75%** | **Max** |
|---|---|---|---|---|---|---|---|
| no_of_purchases | 20.04 | 9.99 | 1.00 | 12.00 | 20.00 | 27.00 | 58.00 |
| total_quantity | 171.49 | 76.26 | 6.00 | 112.00 | 168.00 | 227.75 | 443.00 |
| total_spend | 224.46 | 94.75 | 7.28 | 156.36 | 225.41 | 281.89 | 689.80 |
| avg_visit_interval | 9.11 | 5.25 | 0.00 | 5.77 | 7.70 | 10.97 | 38.00 |
| recency | 28.10 | 39.71 | 0.00 | 3.00 | 10.00 | 35.75 | 164.00 |
| DRINKS | 22.79 | 41.93 | 0.00 | 0.00 | 3.57 | 26.31 | 280.53 |
| SOFT_DRINKS | 6.20 | 8.28 | 0.00 | 0.90 | 3.45 | 7.56 | 50.64 |

### Convenience-seeker Buyers

This segment consists of customers who make a lot of purchases and frequently visit to the store, almost every other day. They tend not to buy too many different items per visit and hence spend less during each visit. Most of their shopping is conducted during day time. They spend considerably high on "Lottery", "Cashpoints", "Drinks", "Newspapers_Magazines" and "Tobacco" products. This implies they spend enough on non-food items compared to other segments.

### Premium Buyers

They visit the store once a week on average. Compared to other clusters, their average basket quantity and category is the least, but on the other hand, they have the highest spend per item. Additionally, they spend relatively more on "Tobbaco" and "Cashpoint" as compared to other products. This implies they also spend a relatively good amount on non-food items.

### Regular Buyers

The customers in this segment buy many products spread across a number of different categories during each visit. They visit the store every 2-3 days and spend high on each visit making them the highest spenders among all the segments. They tend to spend heavily on all food items as compared to non-food

Table 2: **Statistical Summary for Convenience-seeker Buyers having 660 customers**

| **Feature** | **Mean** | **Std** | **Min** | **25%** | **50%** | **75%** | **Max** |
|---|---|---|---|---|---|---|---|
| no_of_purchases | 114.63 | 48.69 | 38.00 | 81.75 | 104.00 | 138.25 | 374.00 |
| total_quantity | 753.81 | 226.07 | 358.00 | 600.75 | 712.50 | 872.25 | 2080.00 |
| avg_basket_quantity | 7.16 | 2.22 | 2.73 | 5.59 | 6.96 | 8.53 | 15.38 |
| avg_visit_interval | 1.81 | 0.66 | 0.49 | 1.29 | 1.73 | 2.18 | 4.51 |
| CASHPOINT | 101.23 | 189.44 | 0.00 | 0.00 | 21.00 | 130.00 | 1751.90 |
| DRINKS | 95.59 | 164.23 | 0.00 | 1.76 | 19.56 | 109.27 | 1005.33 |
| LOTTERY | 32.52 | 72.07 | -10.00 | 0.00 | 3.00 | 30.00 | 591.00 |
| NEWSPAPERS_MAGAZINES | 29.20 | 34.68 | 0.00 | 6.82 | 17.96 | 37.17 | 265.10 |
| TOBACCO | 193.47 | 281.12 | 0.00 | 0.15 | 56.76 | 287.15 | 2264.31 |
| total_spend_nonfood | 431.45 | 372.74 | 28.32 | 158.71 | 302.86 | 594.73 | 2415.72 |
| total_spend_weekend | 250.64 | 152.28 | 0.00 | 148.68 | 220.99 | 321.02 | 1421.39 |

Table 3: **Statistical Summary for Premium Buyers, having 670 customers**

| **Feature** | **Mean** | **Std** | **Min** | **25%** | **50%** | **75%** | **Max** |
|---|---|---|---|---|---|---|---|
| avg_basket_quantity | 5.73 | 1.90 | 1.37 | 4.41 | 5.64 | 6.90 | 14.04 |
| avg_basket_spend | 8.80 | 4.81 | 2.07 | 5.90 | 7.70 | 10.65 | 57.32 |
| avg_basket_category | 4.78 | 1.54 | 1.22 | 3.69 | 4.72 | 5.79 | 10.44 |
| avg_spend_per_item | 1.60 | 0.81 | 0.56 | 1.08 | 1.39 | 1.84 | 7.92 |
| CASHPOINT | 58.33 | 137.92 | 0.00 | 0.00 | 8.50 | 60.00 | 2137.01 |
| TOBACCO | 105.73 | 167.52 | 0.00 | 0.00 | 18.11 | 157.23 | 1272.29 |
| total_spend_nonfood | 224.46 | 221.80 | 1.77 | 70.83 | 142.74 | 317.11 | 2164.03 |

items like “Cashpoint”, “Tobacco” and “Lottery”. They also spend high on “World Foods”, “grocery health pets” and “seasonal gifting”. They mostly buy products on weekdays and are not time bound, i.e. they visit the store at day or night more frequently as compared to other segments.

Table 4: **Statistical Summary for Regular Buyers having 375 customers**

| **Feature** | **Mean** | **Std** | **Min** | **25%** | **50%** | **75%** | **Max** |
|---|---|---|---|---|---|---|---|
| no_of_purchases | 86.68 | 53.39 | 12.00 | 50.00 | 73.00 | 109.50 | 354.00 |
| total_quantity | 1310.69 | 490.52 | 606.00 | 969.00 | 1206.00 | 1544.00 | 4949.00 |
| avg_basket_quantity | 19.66 | 12.72 | 5.67 | 11.43 | 16.05 | 23.11 | 90.75 |
| total_spend | 1691.75 | 693.36 | 855.66 | 1226.76 | 1511.49 | 1982.28 | 6588.65 |
| avg_basket_spend | 25.41 | 17.31 | 6.81 | 14.86 | 20.61 | 29.61 | 152.62 |
| avg_basket_category | 15.91 | 9.97 | 4.97 | 9.26 | 13.03 | 18.97 | 72.42 |
| BAKERY | 84.89 | 55.88 | 6.52 | 49.39 | 71.68 | 102.71 | 444.37 |
| FRUIT_VEG | 169.90 | 113.58 | 13.02 | 94.77 | 145.74 | 213.49 | 1262.97 |
| MEAT | 149.27 | 115.80 | 0.00 | 72.18 | 120.25 | 190.76 | 713.28 |
| PRACTICAL_ITEMS | 5.57 | 7.56 | 0.00 | 0.00 | 3.22 | 7.45 | 65.06 |

### Budget Buyers

They are less frequent buyers, however, their last purchase is quite recent. When they shop, they buy more and purchase a variety of products. Whereas, they have the least average spend per item indicating they only buy required products. For instance, they spend more on “Dairy”, “Fruit Veg” and "Grocery Food". They have the least percent of spending on non-food items, which is just 14% of their total spending. They also have the least percent of spending in evening and night. Their spending on "Tobbaco", "Cashpoint" and "Practical items" is also one of the least among all segments.

Table 5: **Statistical Summary for Budget Buyers having 853 customers**

| **Feature** | **Mean** | **Std** | **Min** | 25% | 50% | 75% | Max |
|---|---|---|---|---|---|---|---|
| avg_basket_quantity | 15.61 | 8.30 | 5.51 | 10.21 | 13.15 | 18.12 | 83.25 |
| avg_basket_spend | 18.97 | 11.08 | 5.37 | 11.85 | 15.68 | 23.24 | 100.48 |
| avg_basket_category | 12.86 | 6.59 | 4.67 | 8.57 | 10.92 | 15.00 | 55.00 |
| avg_spend_per_item | 1.21 | 0.29 | 0.61 | 1.03 | 1.18 | 1.34 | 2.88 |
| avg_visit_interval | 5.32 | 2.89 | 1.32 | 3.43 | 4.71 | 6.44 | 35.75 |
| total_spend_food | 518.03 | 169.40 | 213.50 | 393.05 | 483.25 | 625.80 | 1143.19 |
| total_spend_daytime | 483.68 | 233.19 | 15.86 | 329.05 | 454.92 | 609.36 | 1450.00 |
| total_spend_weekday | 473.95 | 223.01 | 0.00 | 325.74 | 445.23 | 598.74 | 1450.00 |

## H Summary

The k-means clustering analysis was conducted on transactional data to segment customers into 5 clusters named Occasional Buyers, Convenience-seeker Buyers, Premium Buyers, Regular Buyers and Budget Buyers, based on their purchase behaviour.

- Business Case: The clustering provides valuable insights into customer behaviour that can be leveraged by the company to better understand their customer base and tailor marketing campaigns to specific segments. By focusing on the most important segments, the company can achieve a better return on investment and increase revenue.

- Recommendations: Based on the analysis, the two segments that are most important for the company to focus on are Premium Buyers and Regular Buyers. Premium Buyers are customers who are willing to spend more on high-end products, while Regular Buyers are the ones who are loyal customers and helps to generate maximum revenue for the company. These segments offer the highest potential for revenue growth and can be targeted with personalized marketing campaigns that highlight relevant products and promotions.

- Key Take-Home Points and Marketing Recommendations: (i) Convenience buyers are busy individuals who lead fast-paced lifestyles and often prioritize their convenience. This category may respond well to promotions such as click-and-collect options or home delivery services. (ii) Budget buyers may be willing to shop at multiple stores to find the best deals and may prefer to buy in bulk to save money. From a marketing perspective, retailers can appeal to them by highlighting affordable prices, offering discounts and promotions on value products. (iii) Occasional buyers can be busy persons or maybe open to trying new products. Retailers can use targeted promotions and discounts to encourage these buyers to make more frequent purchases. (iv) Regular Buyers can be targeted with loyalty programs to encourage repeat purchases. (v) Premium Buyers should be targeted with high-end products and personalized promotions that highlight exclusivity and quality.

- Further Analysis: Company can take customer surveys and connect with focus groups to gain a deeper understanding of their specific needs and preferences and additional data analysis to identify patterns and trends. This information can be used to further optimize marketing campaigns and increase customer engagement.

# Appendix A

### Data Dictionary

Table 6: Data Dictionary

| **Feature** | **Description** | **Data Type** |
|---|---|---|
| customer_number | Unique identifier for each customer | Integer |
| no_of_purchases | Total number of purchases made by the customer | Float |
| total_quantity | Total quantity of items purchased by the customer | Float |
| avg_basket_quantity | Average number of items in each purchase | Float |
| total_spend | Total amount spent by the customer | Float |
| avg_basket_spend | Average amount spent in each purchase | Float |
| avg_basket_category | Average spend in each product category | Float |
| recency | Number of days since the customer's last purchase | Integer |
| avg_visit_interval | Average number of days between purchases | Float |
| avg_spend_per_item | Average spend per item purchased | Float |
| BAKERY | Amount spent on bakery products | Float |
| CASHPOINT | Amount spent on cashpoint services | Float |
| CONFECTIONARY | Amount spent on confectionary products | Float |
| DAIRY | Amount spent on dairy products | Float |
| DELI | Amount spent on deli products | Float |
| DISCOUNT_BAKERY | Amount spent on discounted bakery products | Float |
| DRINKS | Amount spent on drinks | Float |
| FROZEN | Amount spent on frozen products | Float |
| FRUIT_VEG | Amount spent on fruits and vegetables | Float |
| GROCERY_FOOD | Amount spent on grocery food items | Float |
| GROCERY_HEALTH_PETS | Amount spent on health and pet products | Float |
| LOTTERY | Amount spent on lottery | Float |
| MEAT | Amount spent on meat products | Float |
| NEWSPAPERS_MAGAZINES | Amount spent on newspapers and magazines | Float |
| PRACTICAL_ITEMS | Amount spent on practical items | Float |
| PREPARED_MEALS | Amount spent on prepared meals | Float |
| SEASONAL_GIFTING | Amount spent on seasonal gifting | Float |
| SOFT_DRINKS | Amount spent on soft drinks | Float |
| TOBACCO | Amount spent on tobacco products | Float |
| WORLD_FOODS | Amount spent on world foods | Float |
| total_spend_food | Total amount spent on food items | Float |
| total_spend_nonfood | Total amount spent on non-food items | Float |
| total_spend_weekend | Total amount spent on weekends | Float |
| total_spend_weekday | Total amount spent on weekdays | Float |
| total_spend_daytime | Total amount spent during the daytime | Float |
| total_spend_night | Total amount spent during the night | Float |

# Appendix B

## Visualisation for customer base summary

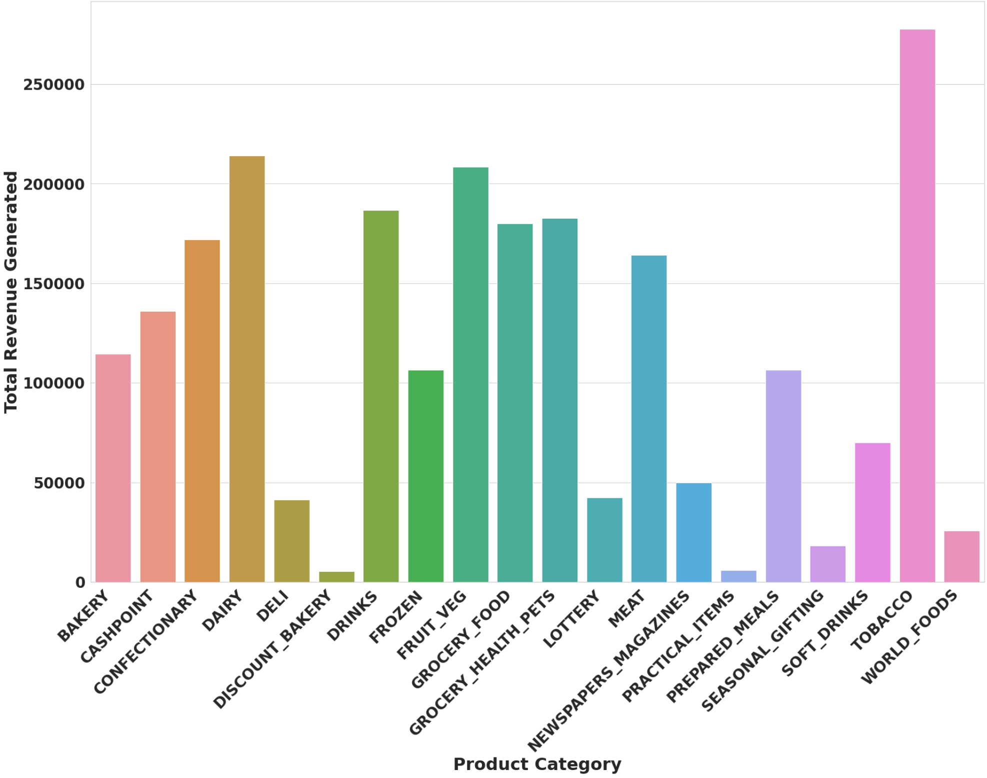


Figure B1: Total revenue generated by product category

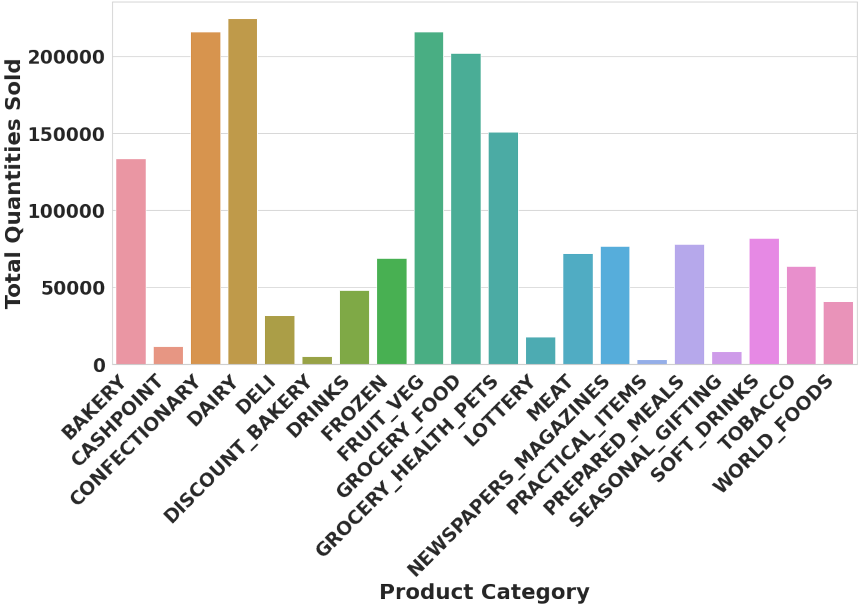


Figure B2: Total quantities sold by product category

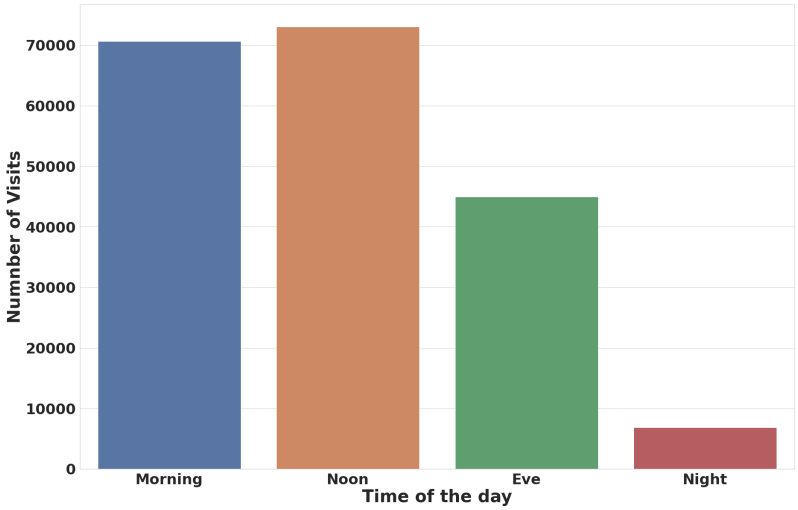


Figure B3: Number of purchases by the time of day

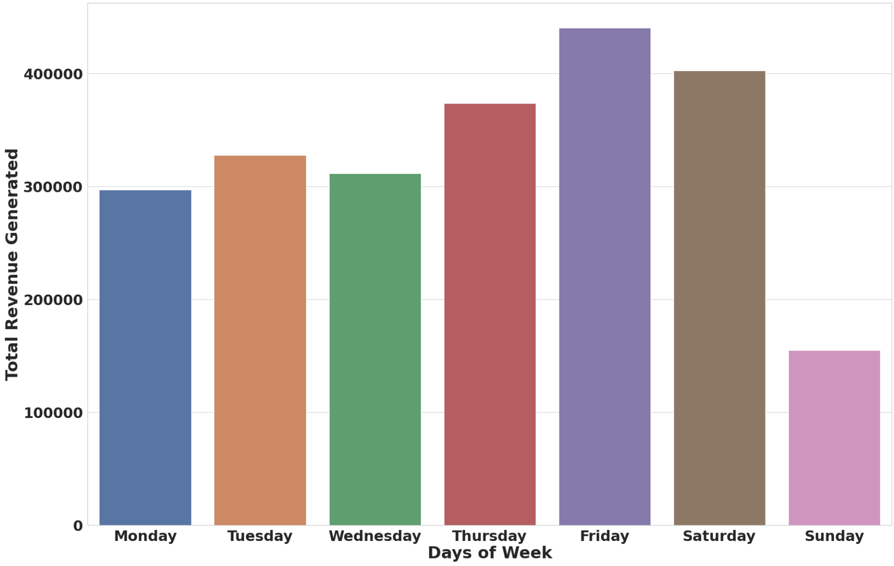


Figure B4: Total revenue generated by the day of week

# Appendix C

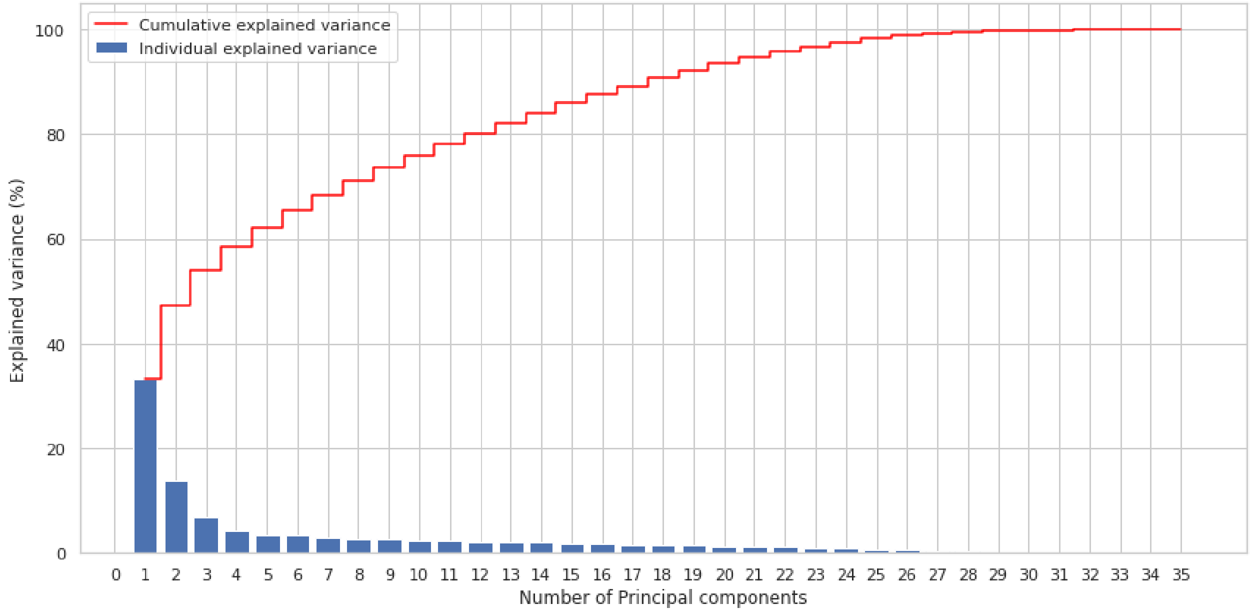


Figure C1: Plot showing individual and cumulative explained variance captured by the principal components

# Appendix D

The figures describe the thickness of the silhouette plot on the left and the cluster visualisation on the right. The silhouette score for the number of clusters equal to 5,6,7 is 0.1547,0.1497, and 0.1434, respectively. It can be observed that for 5 clusters, the thickness of the silhouette plot and also the respective cluster size does not vary much unlike for clusters 6 and 7. Hence, we chose number of clusters to be equal to 5.

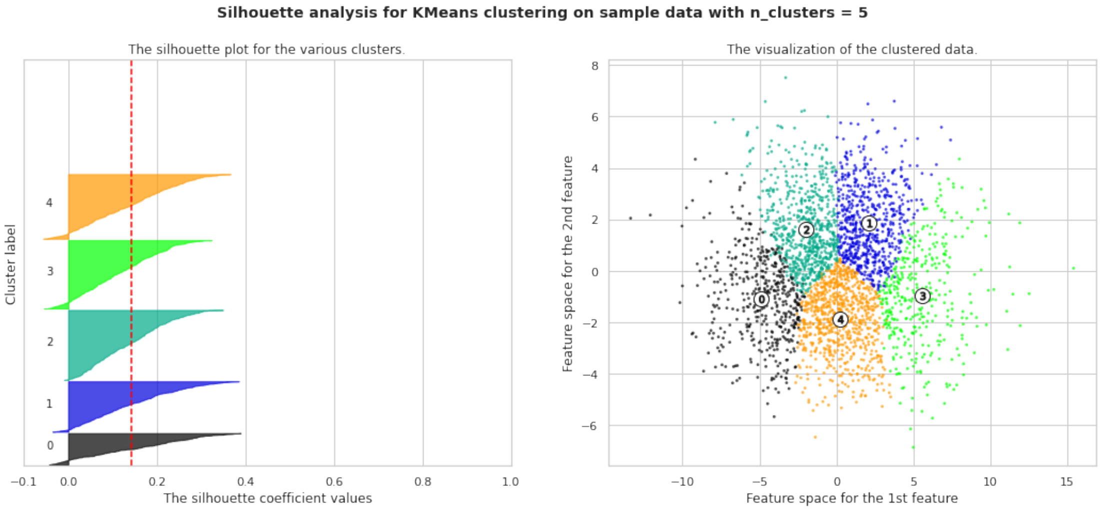


Figure D1: Number of Clusters =5

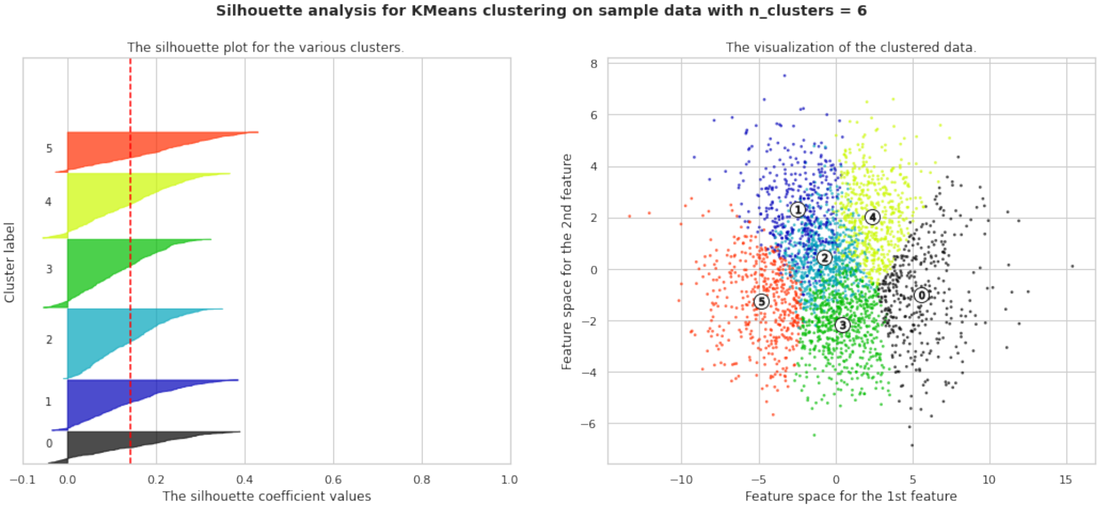


Figure D2: Number of Clusters =6

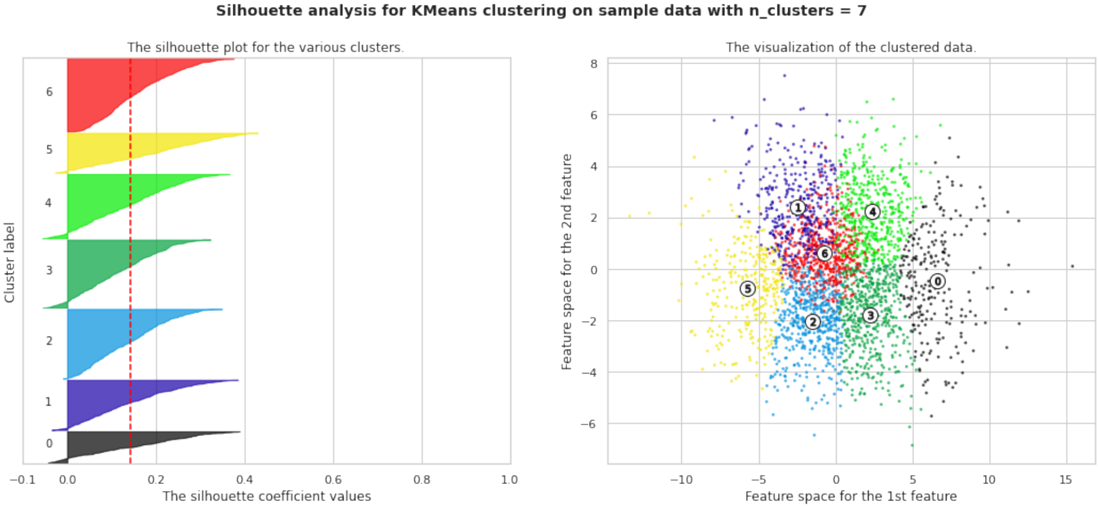


Figure D3: Number of Clusters =7

# Appendix E

This figure shows the cluster formation when K-means is applied.

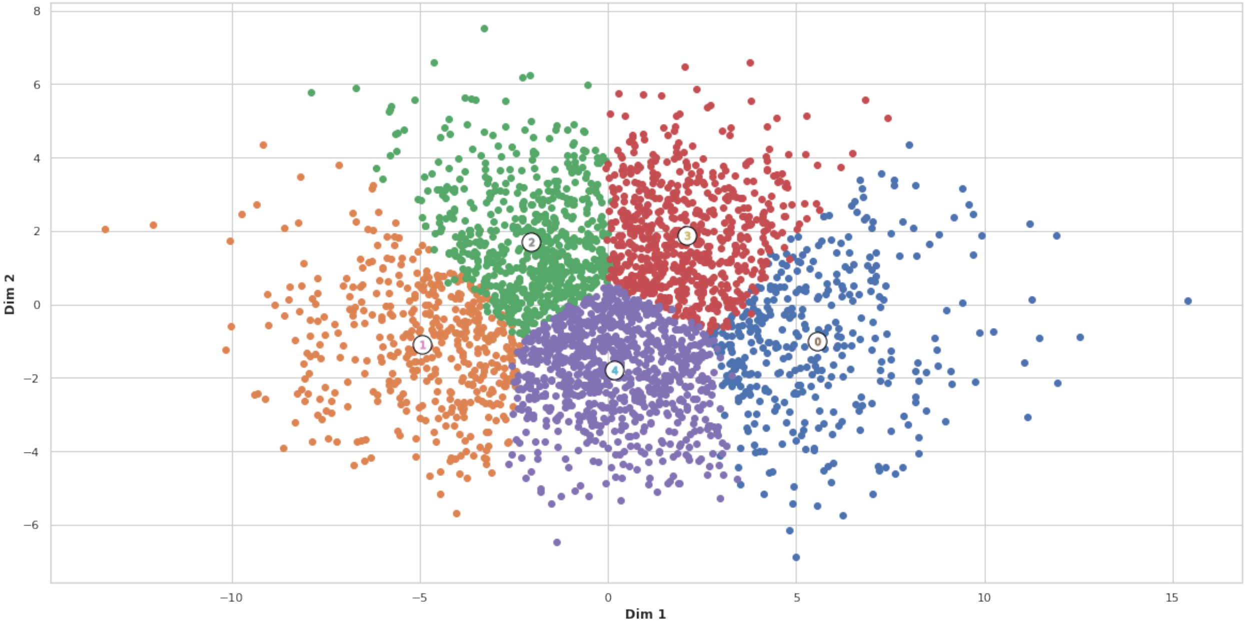
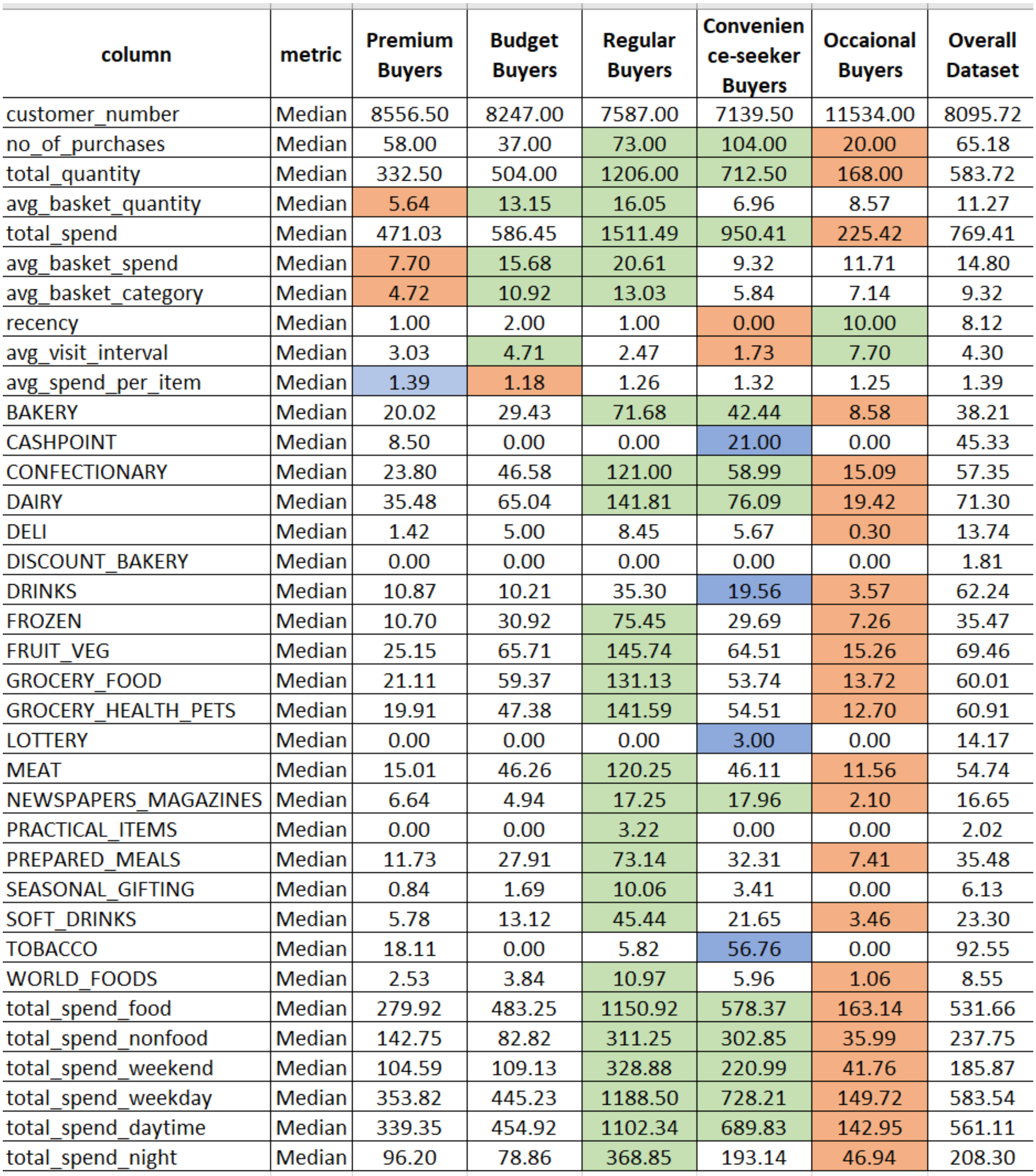


Figure E1: Cluster Visualisation

The figure E2 shows the profiling method using medians of all features in all the clusters.



| column | metric | Premium Buyers | Budget Buyers | Regular Buyers | Convenience-seeker Buyers | Occaional Buyers | Overall Dataset |
|---|---|---|---|---|---|---|---|
| customer_number | Median | 8556.50 | 8247.00 | 7587.00 | 7139.50 | 11534.00 | 8095.72 |
| no_of_purchases | Median | 58.00 | 37.00 | 73.00 | 104.00 | 20.00 | 65.18 |
| total_quantity | Median | 332.50 | 504.00 | 1206.00 | 712.50 | 168.00 | 583.72 |
| avg_basket_quantity | Median | 5.64 | 13.15 | 16.05 | 6.96 | 8.57 | 11.27 |
| total_spend | Median | 471.03 | 586.45 | 1511.49 | 950.41 | 225.42 | 769.41 |
| avg_basket_spend | Median | 7.70 | 15.68 | 20.61 | 9.32 | 11.71 | 14.80 |
| avg_basket_category | Median | 4.72 | 10.92 | 13.03 | 5.84 | 7.14 | 9.32 |
| recency | Median | 1.00 | 2.00 | 1.00 | 0.00 | 10.00 | 8.12 |
| avg_visit_interval | Median | 3.03 | 4.71 | 2.47 | 1.73 | 7.70 | 4.30 |
| avg_spend_per_item | Median | 1.39 | 1.18 | 1.26 | 1.32 | 1.25 | 1.39 |
| BAKERY | Median | 20.02 | 29.43 | 71.68 | 42.44 | 8.58 | 38.21 |
| CASHPOINT | Median | 8.50 | 0.00 | 0.00 | 21.00 | 0.00 | 45.33 |
| CONFECTIONARY | Median | 23.80 | 46.58 | 121.00 | 58.99 | 15.09 | 57.35 |
| DAIRY | Median | 35.48 | 65.04 | 141.81 | 76.09 | 19.42 | 71.30 |
| DELI | Median | 1.42 | 5.00 | 8.45 | 5.67 | 0.30 | 13.74 |
| DISCOUNT_BAKERY | Median | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.81 |
| DRINKS | Median | 10.87 | 10.21 | 35.30 | 19.56 | 3.57 | 62.24 |
| FROZEN | Median | 10.70 | 30.92 | 75.45 | 29.69 | 7.26 | 35.47 |
| FRUIT_VEG | Median | 25.15 | 65.71 | 145.74 | 64.51 | 15.26 | 69.46 |
| GROCERY_FOOD | Median | 21.11 | 59.37 | 131.13 | 53.74 | 13.72 | 60.01 |
| GROCERY_HEALTH_PETS | Median | 19.91 | 47.38 | 141.59 | 54.51 | 12.70 | 60.91 |
| LOTTERY | Median | 0.00 | 0.00 | 0.00 | 3.00 | 0.00 | 14.17 |
| MEAT | Median | 15.01 | 46.26 | 120.25 | 46.11 | 11.56 | 54.74 |
| NEWSPAPERS_MAGAZINES | Median | 6.64 | 4.94 | 17.25 | 17.96 | 2.10 | 16.65 |
| PRACTICAL_ITEMS | Median | 0.00 | 0.00 | 3.22 | 0.00 | 0.00 | 2.02 |
| PREPARED_MEALS | Median | 11.73 | 27.91 | 73.14 | 32.31 | 7.41 | 35.48 |
| SEASONAL_GIFTING | Median | 0.84 | 1.69 | 10.06 | 3.41 | 0.00 | 6.13 |
| SOFT_DRINKS | Median | 5.78 | 13.12 | 45.44 | 21.65 | 3.46 | 23.30 |
| TOBACCO | Median | 18.11 | 0.00 | 5.82 | 56.76 | 0.00 | 92.55 |
| WORLD_FOODS | Median | 2.53 | 3.84 | 10.97 | 5.96 | 1.06 | 8.55 |
| total_spend_food | Median | 279.92 | 483.25 | 1150.92 | 578.37 | 163.14 | 531.66 |
| total_spend_nonfood | Median | 142.75 | 82.82 | 311.25 | 302.85 | 35.99 | 237.75 |
| total_spend_weekend | Median | 104.59 | 109.13 | 328.88 | 220.99 | 41.76 | 185.87 |
| total_spend_weekday | Median | 353.82 | 445.23 | 1188.50 | 728.21 | 149.72 | 583.54 |
| total_spend_daytime | Median | 339.35 | 454.92 | 1102.34 | 689.83 | 142.95 | 561.11 |
| total_spend_night | Median | 96.20 | 78.86 | 368.85 | 193.14 | 46.94 | 208.30 |

Figure E2: Cluster Analysis using Median values