Segmentation & Profiling Project

Merrimack College

Data Exploration

Amol Gote

(Data science student)

July 3rd, 2020

Executive Summary

The objective of this report is to provide customer segmentation for customers of a telecommunication company. This will help the organization in making effective economical customer retention strategies. Based on the data analytics exercise customers have been categorized into four segments Platinum, Gold, Silver, and Bronze. These four segments have been primarily driven by the revenue generated by the customers over the tenure of their association with the telecommunications company. The key value factor is revenue generated over the tenure and average monthly revenue generated. Apart from the value generation factors, other attributes related to demographics like age, gender, education, employment length, marital status, and household size have been factored in. Financial features like household income, debt to income ratio, total debt have been considered. Homeownership, how long the customer in months has been associated with the company have also been taken into consideration. Below is the summarized view of the Distribution of various customer segments in Figure 1.

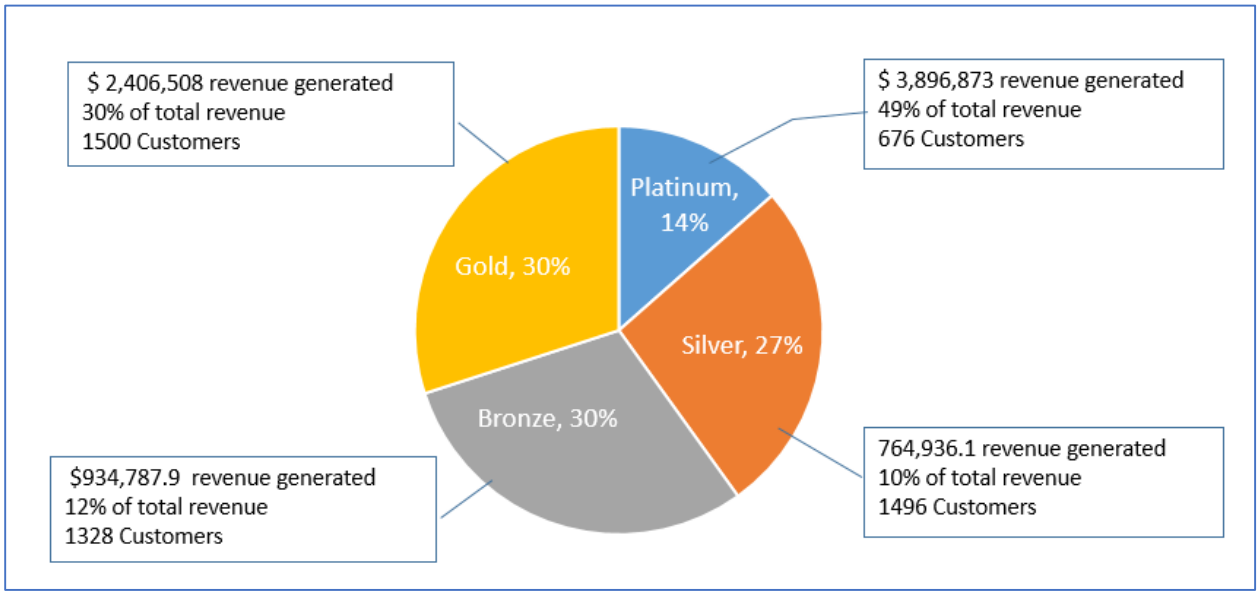


Figure 1: Customer segmentation distribution

Platinum customers who account for only 14% of the total customer base generate almost 50% of the total revenue, followed by gold customers with 30% of total revenue, silver and bronze generate very minimal revenue with overall percentage hovering around 10-12%. The average monthly revenue generated by platinum customers is $103, by gold customers is $27.50, for silver and bronze, it is almost the same which is $24. (*Refer the Appendix figure 12 for distribution of average monthly revenue across four customer segments*). Platinum customers are the high-value customers as they account for nearly 49% of the revenue generated, these customers are consuming all 3 services voice, data, and equipment heavily. Gold customers generate nearly 30% of total revenue and are significant consumers of voice related products and services compared to data and equipment. Silver customers are in between low and medium value customers and generate 12% of the revenue. Bronze customers only generate 10% of total revenue and are the low-value customers. Silver and bronze customers have a somewhat similar pattern of usage of different services like voice, data, and equipment.

As far as demographics are concerned, for Platinum customers, it is dominated by customers aged in between 40 to 60, with an average age of 53, gold customers are dominated by the customers with age greater than 60, with an average age of 63, while silver and bronze are dominated by customers aged between 18 to 40 years of age, with an average age of 35 and 37 respectively. Customers below the age of 18 years are either in silver or bronze category. Gender distribution is equal across all customer segments, there is no bias. For platinum marital status is not of significance as there are nearly equal married and unmarried customers, with married ones on the higher side a bit. Silver customer segment comprises of the married customers and with very less unmarried customers. The Bronze customer segment hardly contains any married customers. Median household size for platinum and gold customers is 2, while that for silver is 4, for bronze household size hovers around 1 which is self-evident from the fact that this customer segment is dominated by unmarried customers. Across all 4 segments Professional, Sales and service are three top job categories (Refer Appendix Figure 13).

Most Platinum customers close to 95% are multiline, for Gold customers multiline drops to 65%, and for Silver and Bronze customer percentage drop to 25%. Platinum and gold customers have been associated with the company with an average timespan of 56 months, followed by silver for 28 months and bronze by 20 months.

For household income average for platinum customers is $90,000, for gold customers, it is $65,000, for bronze and silver the average household income is around $40,000. The average monthly revenue gradually increases as household income increases across 3 segments Platinum, Gold and Bronze, for Silver it drops slightly (Refer Appendix figure 10). Debt to income ratio for platinum, gold, and silver customers is around 10% while that for bronze customers it shoots up to 37%. The Loan default rate for the Platinum customers is higher around 18% compared to that of Gold Customers which is a mere 6%, for Silver customers, it is 32% and for Bronze customers, it is 35%. For car ownership, there is a uniform trend across all four segments, with most of the customers preferring to buy rather than leasing it. Also, the number of cars owned across all 4 segments is the same, on average every customer owns 2 cars in all segments. For credit cards, leading credit card type is Discover across all segments, then followed by Master Card, Visa, and then American Express (Refer Appendix figure 14). Credit card spend is the same across all 4 segments around $1000, Silver customers segment has a bit higher credit card spend compared to the other 3 segments (Refer Appendix figure 9).

Across all 3 services voice, equipment, and data average revenue for voice has increased as the tenure for company association for customer grows, this is evident in all 4 customer segments. The average revenue for data and equipment for Gold, Silver, and Bronze customers has remained flat as tenure has increased, with platinum customers this has linear growth (Refer Appendix figure 11).

Gold and platinum customers are having long term relationships with the company so the telecommunication company must give discounts on new products and help them port to new offerings, this way it ensures that they continue their existing relationships. Platinum members use all three services voice, equipment, and data, gold members on the contrary use the voice offering extensively, so to strengthen and to continue their existing relationships, data, and equipment related offerings must be targeted to these segments. Silver and bronze members are primarily young adults who in general are heavy consumers of data and tend to churn a lot, so to retain them targeted campaigns can be made to sell more data related products bundled with

voice and equipment. Silver customers have an average household size of greater than 2, with 60% population not using multiline and the majority of the customers married, it provides a good opportunity to target these customers with family deals which can result in more acquisition as well as retention of existing customers. Across all segments Discover card is heavily used followed by Visa and Master card, so promotion campaigns can be launched in conjunction with the credit card company to offer discounts on equipment related services which can help in customer retention and satisfaction. Across all segments loyalty programs can be introduced to give perks based on higher usage of certain services like equipment or data, this will result in customer loyalty and increased business.

Segmentation solution

There were 2 types of clustering techniques that were utilized to identify the customer segments firstly K-mean and secondly Hierarchical Clustering. For both clustering techniques following variables were used Gender, Age, Education Years, Employment Length, Household Income, Debt to Income Ratio, Marital Status, Household Size, Homeowner, customer tenure (association of customer with the company in months) and Multiline. For statistical analysis added following variables, over the tenure total value and average monthly revenue, these variables could help in finding high-value customers segment.

K-Mean

Firstly K-mean technique was performed, k-mean only accepts numerical variables so all categorical variables were converted to numeric variables, for the k-mean clustering technique the number of clusters (k) must be set before starting the algorithm. So, the k-mean analysis was performed with varying cluster sizes from 2 to 9. Determining the exact cluster size is a challenge and this is where graphical methods of Elbow method and Silhoutte method came to rescue. Elbow method indicated the bend at cluster number 4. Based on Appendix Figure 6 & 7 it is evident that cluster sizes of 2, 3, and 4 give clear differentiating customer segments, other cluster sizes there is a considerable amount of overlap especially with higher cluster sizes of 6, 7, 8, and 9. Figure 3 below which is an output of the elbows method shows bend at the cluster size of 4, 5, and 6, if you compare the cluster distribution in Appendix Figure 6, then cluster with a size of 4 is more prominent with less overlap. For cluster size 5 and 6 all segments are not obviously visible. *So, based on the analysis of the 4 segments* ***Cluster 1 is Platinum, Cluster 3 is Gold, Cluster 4 is silver and finally, Cluster 2 is Bronze***

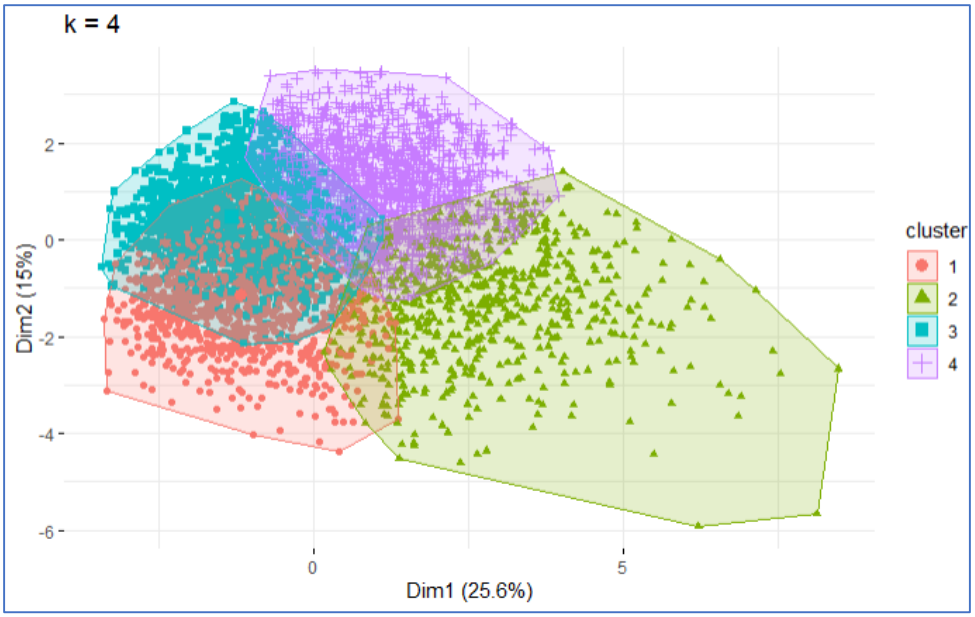


Figure 2: K-Mean clustering with 4 segments

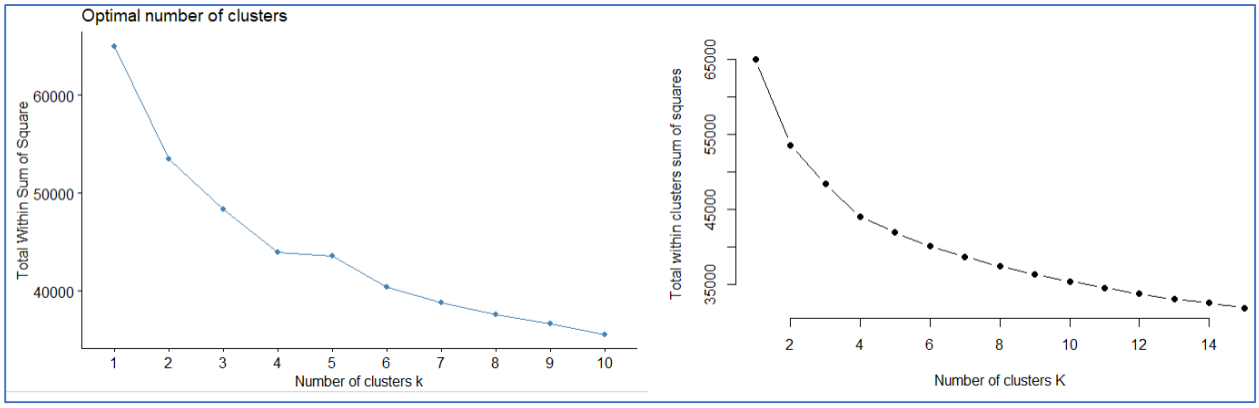


Figure 3: Elbow method output

Hierarchical Clustering

In this analysis cluster sizes ranging from 2 to 9 were tried to find customer segments, refer Appendix figure 8. One major distinguishing factor between K-mean and hierarchical is that one of the clusters in hierarchical is consistently large (cluster with red color in Figure 4) and there is no even distribution of points across the clusters. The elbow method indicated that the optimal cluster size is 4 but based on the points distribution analysis there are a smaller number of points in clusters 3 and 4.

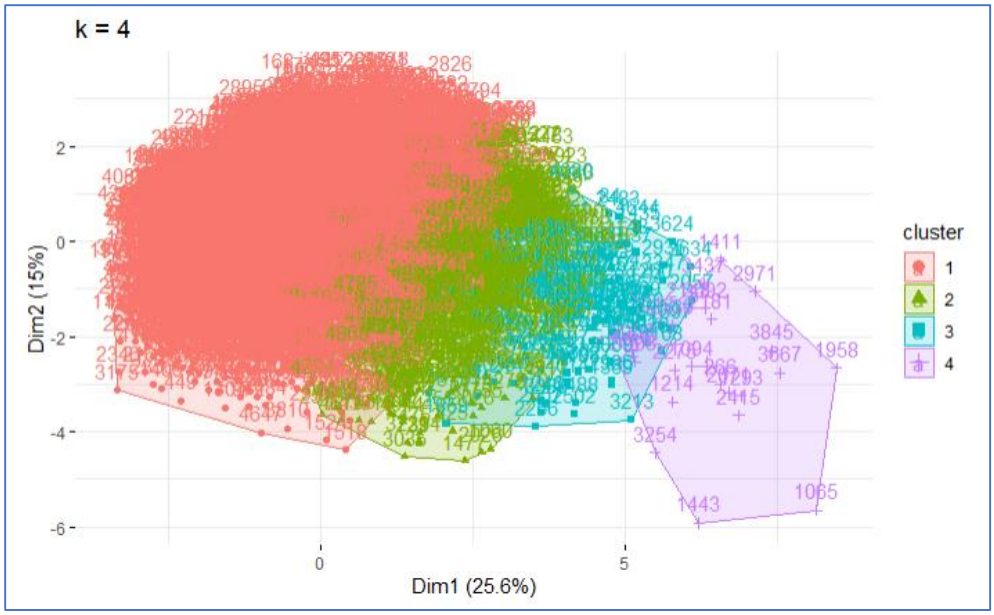


Figure 4: Hierarchical clustering with 4 segments

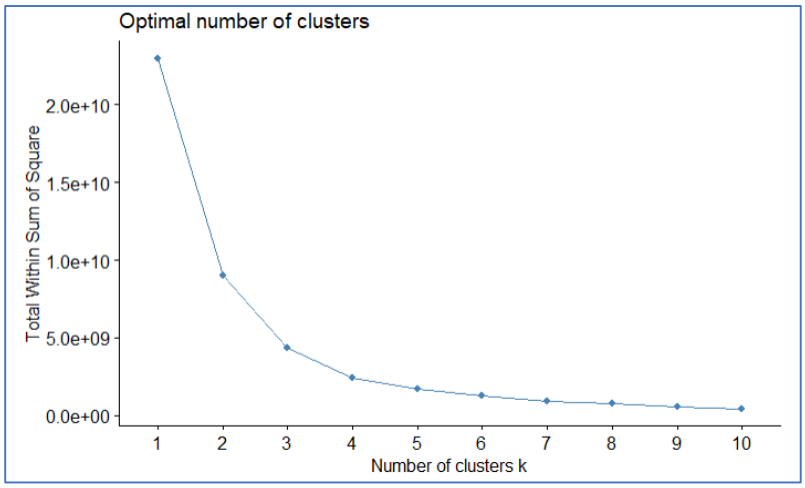
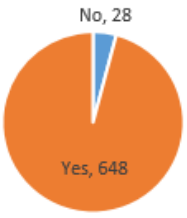
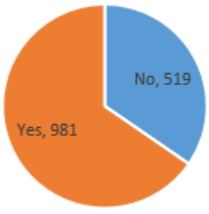
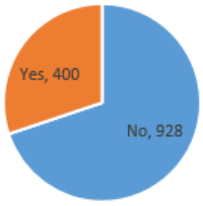
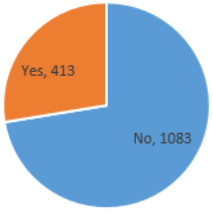
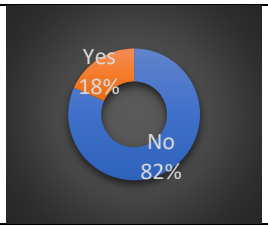
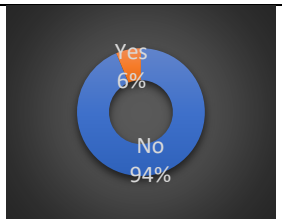
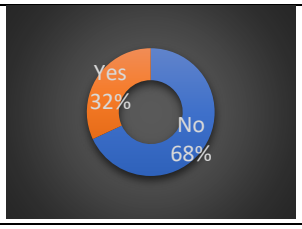
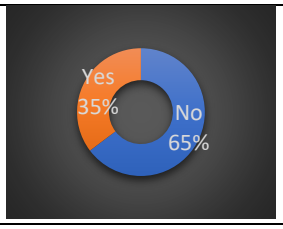
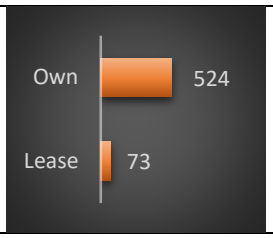
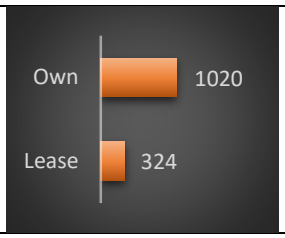
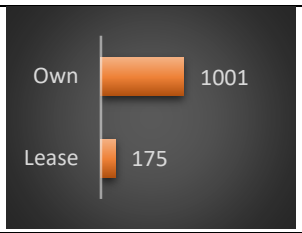
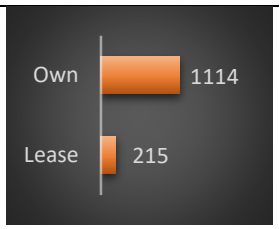
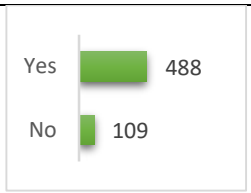
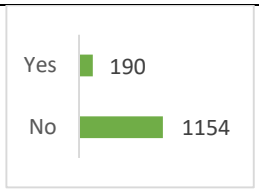
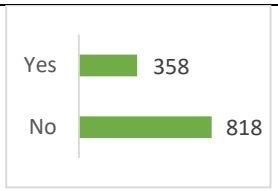
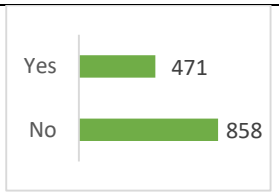
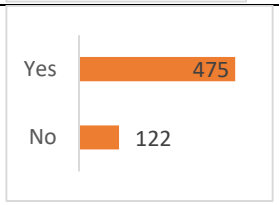
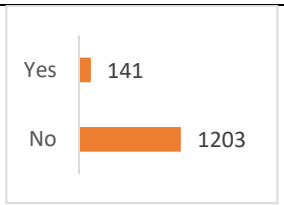
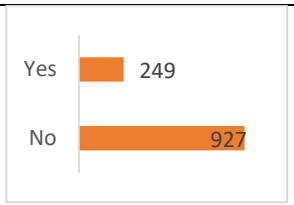
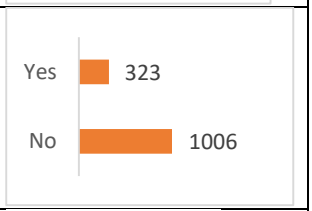
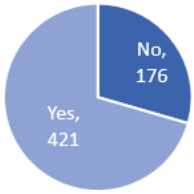
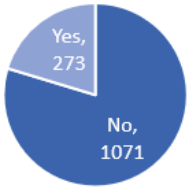
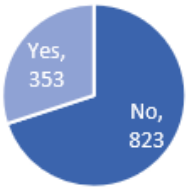
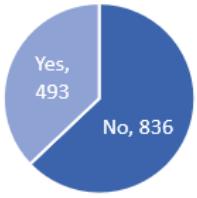
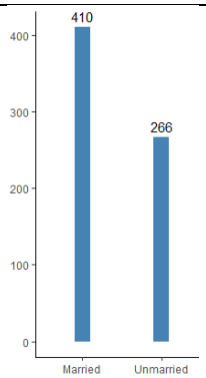
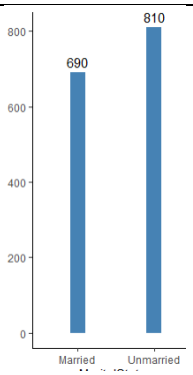
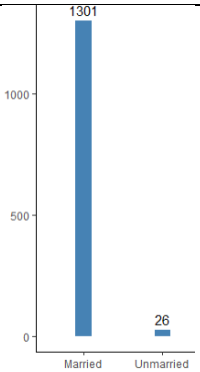
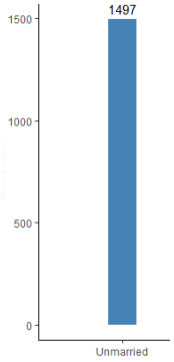


Figure 5: Elbow method output.

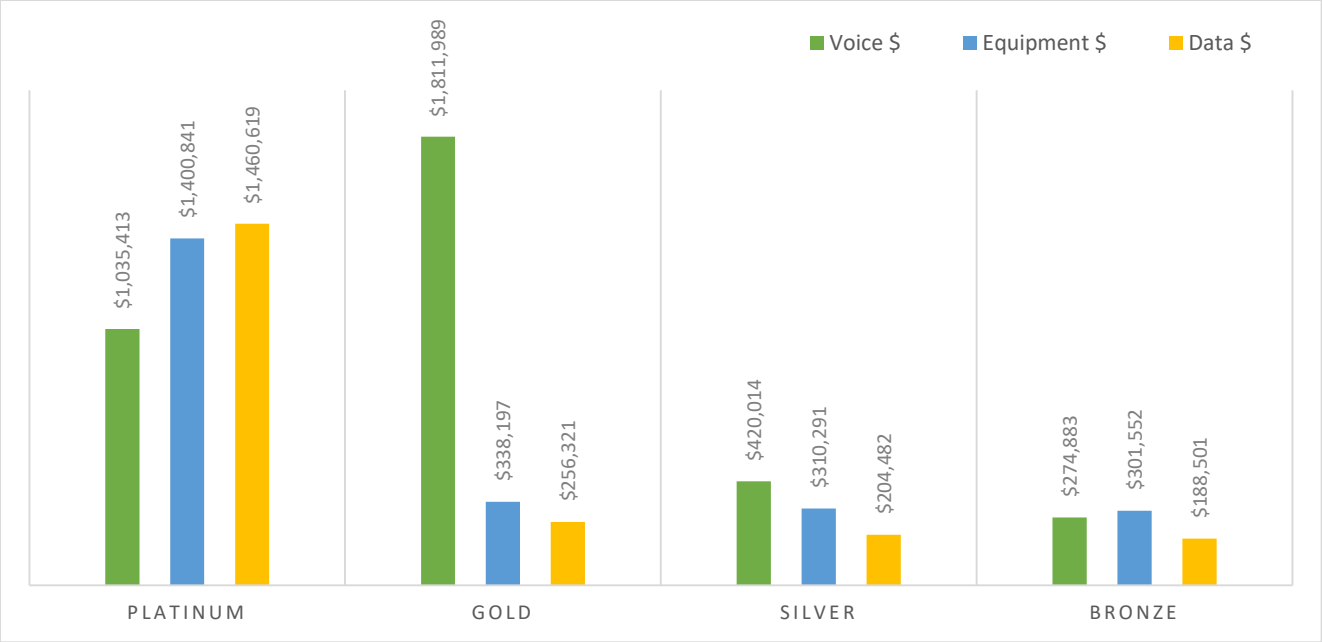
Elbow method shows the bend at 4 clusters, so segmentation with a cluster size of 4 was chosen, but the distribution is highly biased with one segment having 4245 customers and remaining 3 having 609, 121, and 25 customers, respectively. In contrast, the k-mean had even distribution of the number of customers across 4 clusters. So, considering all this analysis K-mean clustering technique was the preferred choice for segmenting the customer of the telecommunication company.

Customer Segments compared

	PLATINUM	GOLD	SILVER	BRONZE
# of Customers	676 14%	1500 30%	1496 12%	1328 10%
$ Revenue	$3,896,873 49%	$2,406,508 30%	$934,788 12%	$764,936 10%
$ Revenue (Average Monthly)	$103	$28	$25	$24
Household Income (Avg)	$90,000	$65,000	$41,000	$40,000
Household size (Avg)	2	2	4	1
Education (Avg)	17 Years	13 Years	14 Years	14 years
Age (Avg)	54 years	65 years	35 years	37 years
Association with the company (Avg)	56 months	57 months	28 months	20 months
Debt to income ratio	10%	10.5%	9.5%	37% ↑
Gender	Male: 355 Female: 321	Male: 752 Female: 748	Male: 577 Female: 751	Male: 805 Female: 691
Multiline				
Loan Default				

Car Ownership				
Cars Owned	1424	3206	2802	3221
Voice Revenue $	$1,035,413	$1,811,989	$420,014	$ 274,883
Equipment Revenue $	$1,400,841	$338,197	$310,291	$301,552
Data Revenue $	$1,460,619	$256,321	$204,482	$188,501
Equipment				
Data				
E-Billing				
Marital Status				

Revenue distribution across 3 services (Voice, Equipment, and Data)



Platinum customers generate almost equal equipment and data revenue close to $1.4 million and close to $1 million for voice. Gold customers are heavy consumers of voice products and services and generate close $1.8 million through voice services. For Gold customers, data and equipment revenue is almost near to that of customers in the Silver segment. For Silver customers revenue for all services are somewhat similar, voice products leading, followed by equipment and then data. For Bronze customers similar revenue generated across 3 services, with equipment leading followed by voice and then data.

Value proposition for various segments

Platinum, Gold, and Silver segments must be given precedence for any retention effort or campaigns, Bronze customers are with high churn rate, and the payment due risks associated with this segment is relatively higher than other segments.

Platinum customers

These are high-value customers and need to give them one-on-one highly personalized attention. New products and offerings must be cross sold to these customers so that they maintain existing relationships and help them migrate from older products to the latest offerings. There are 16% of the Platinum customers who do not use the equipment product offering, so these customers can be provided with equipment related special offers so that they can be retained and continue their relationship. Equipment products usually make the customers stay with the telecommunication company until the equipment cost is repaid. 18% of platinum customers do not use any data related products, these customers can be targeted with offerings for data for strengthening their existing client relationship. For this customer segment as the customer stays longer with

company, average equipment and data revenue has increased linearly which is a clear indicator that these customers lean towards new products and offerings easily (Refer Appendix figure 11), so this segment should be the first primary target for any new products.

Gold Customers

This is a medium value customer segment. This segment is a heavy user of voice products and services, so these customers must be provided with equipment and data related offerings to reinforce customer relationships and increased revenue. From all customers in this segment, 76% do not have the equipment products and 80% do not have the data products and services, so appropriate campaigns can be targeted towards these customers. 30% of the customer are not multiline, there are 189 married customers with no multiline, these can be targeted with special offerings for group deals to maintain and bring in more business and customers.

Silver Customers

This is a low to medium value customer segment, it has an average household size of 4 and 87% of the population is married, there are close to 800 customers which are married and do not have multiline. Since the average household size is greater over here, customized group deals or family plans can be targeted to these customers, which can result in long-lasting customer relationships. Household income for these customers is at low around $40000, so these customers can be offered less expensive equipment services as part of an upgrade.

Bronze Customers

This is the lowest value customer segment. The debt to income ratio is highest for this customer segment. There is a higher churn rate in this customer segment. Considering the customer profile in this segment, there is a high probability of customers defaulting on payments, to avoid that customer can be targeted with new products in equipment and data offerings with discounts provided if they set up E-Billing which minimizes the risk of payment defaults. This strengthens the existing relationship and avoids churning. In case if the budget needs to be allocated across these 4 segments then this segment should be given the least preference as the ROI (Return on investment) cannot be guaranteed.

Appendix

K-mean distribution

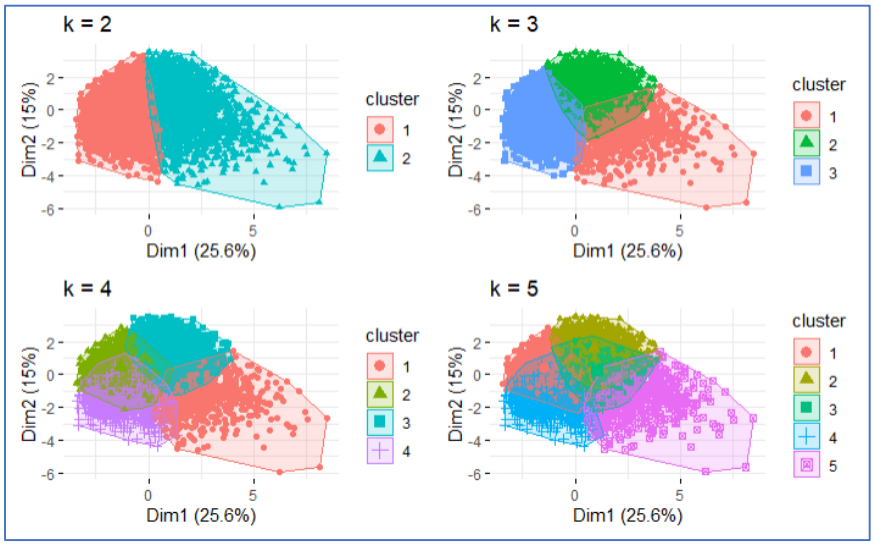


Figure 6: K-mean distribution with a cluster size of 2, 3, 4 and 5

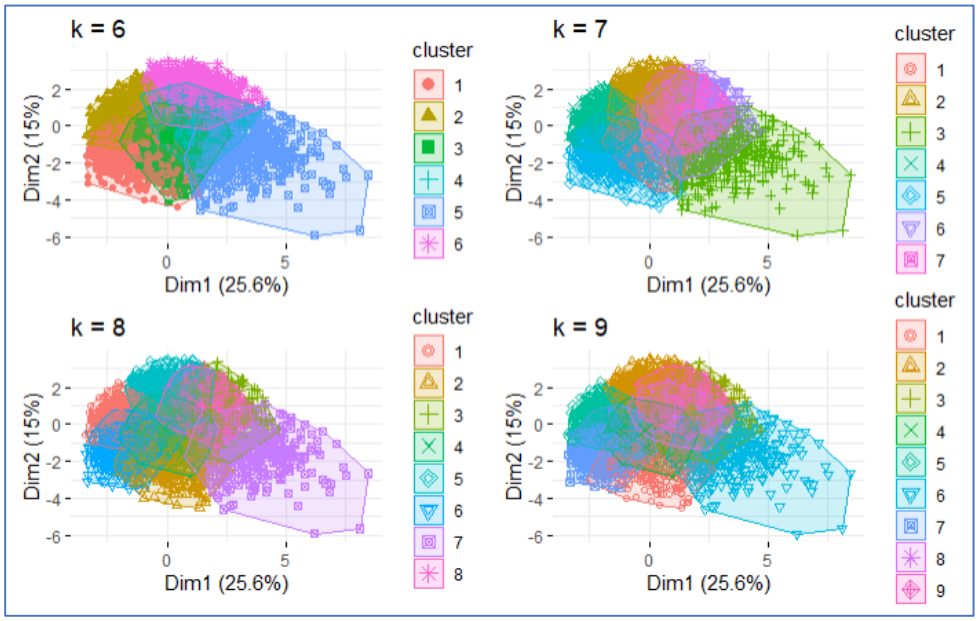


Figure 7: K-mean distribution with a cluster size of 6, 7, 8 and 9

Hierarchical Clustering

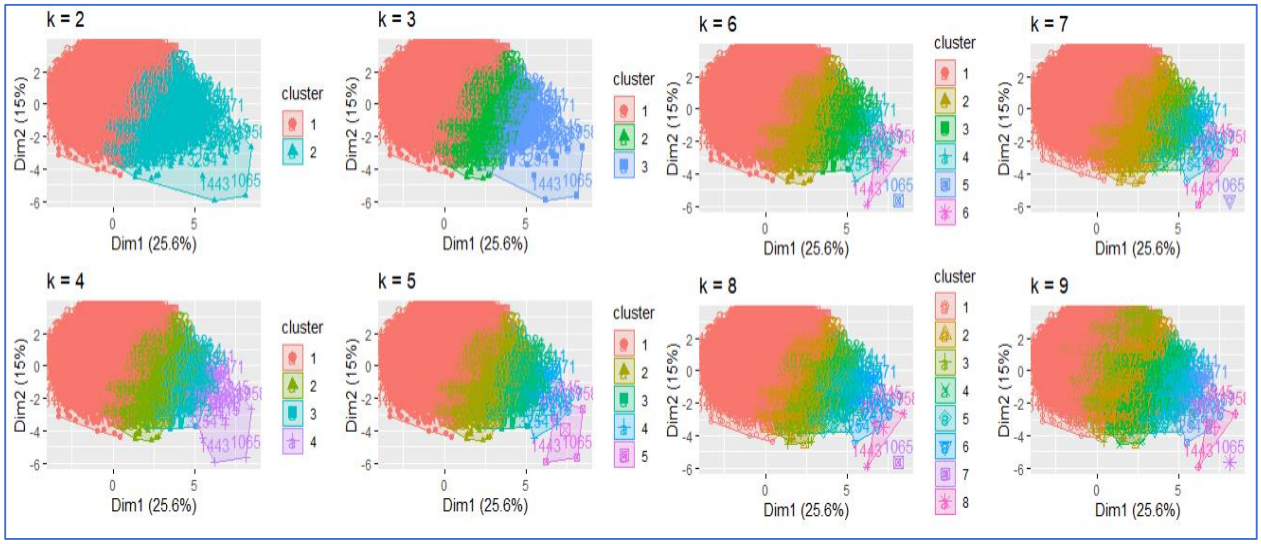


Figure 8: Hierarchical Clustering with cluster size from 2 to 9

Credit card spend across customer segments

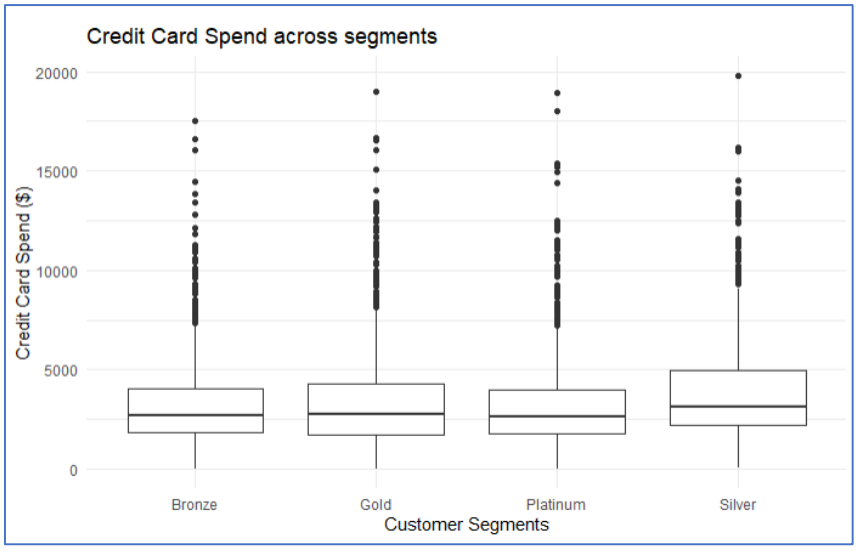


Figure 9: Credit card spend across customer segments

Average Monthly revenue generated to Household income

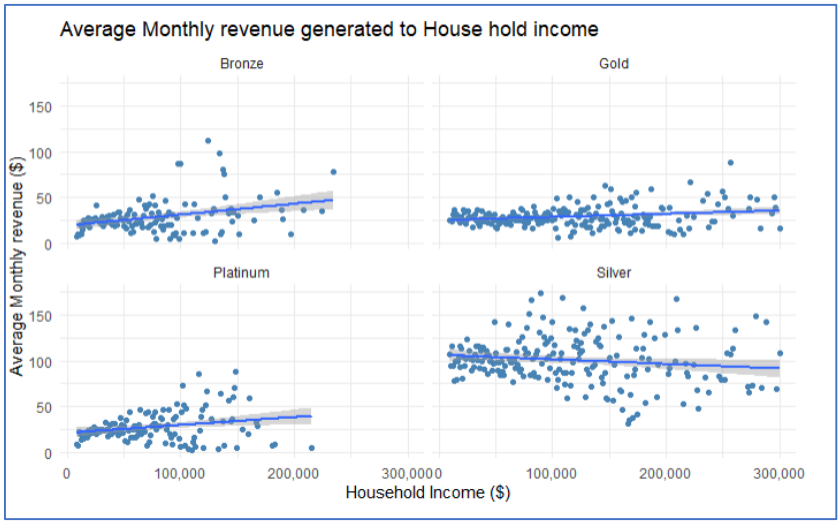


Figure 10:Average Monthly revenue generated to Household income

Average equipment and data revenue by tenure

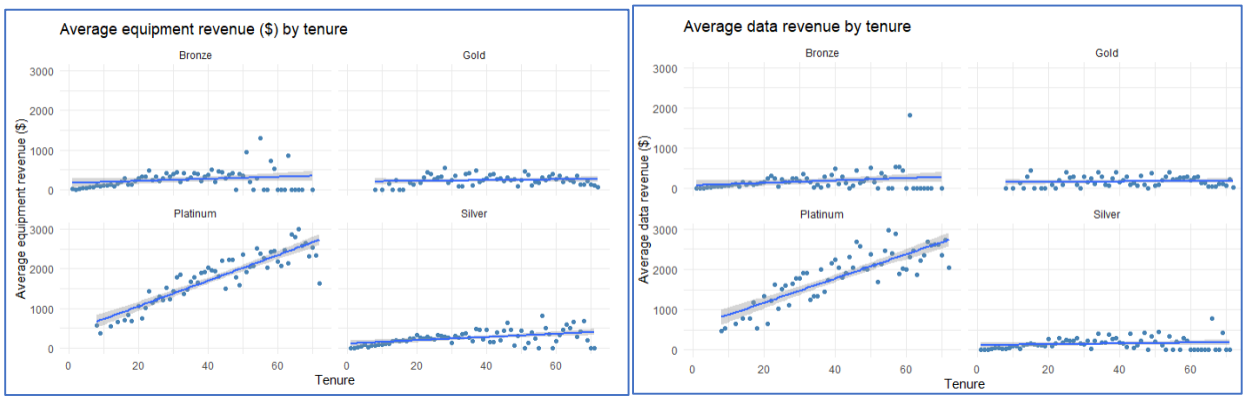


Figure 11: Average equipment and data revenue by tenure

Average monthly distribution across segments

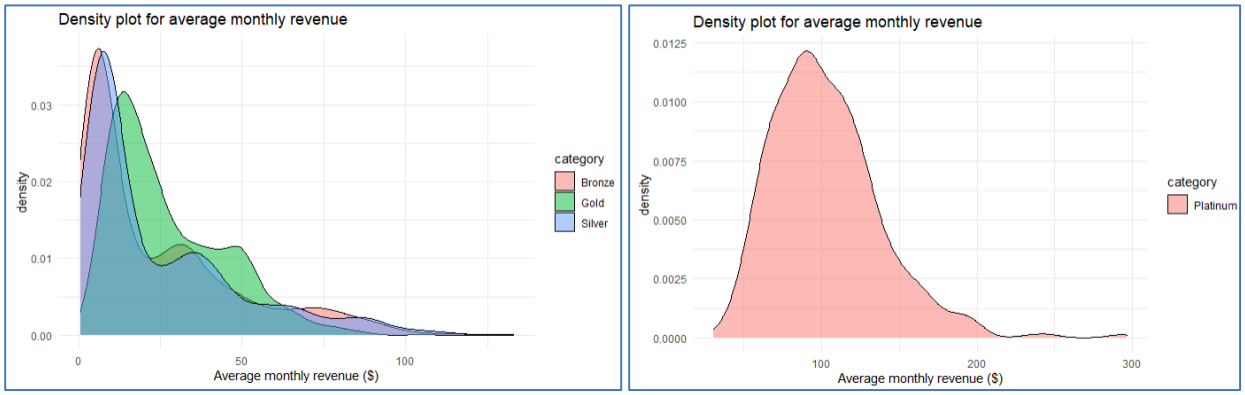


Figure 12: Average monthly distribution across segments

Job Category Distribution

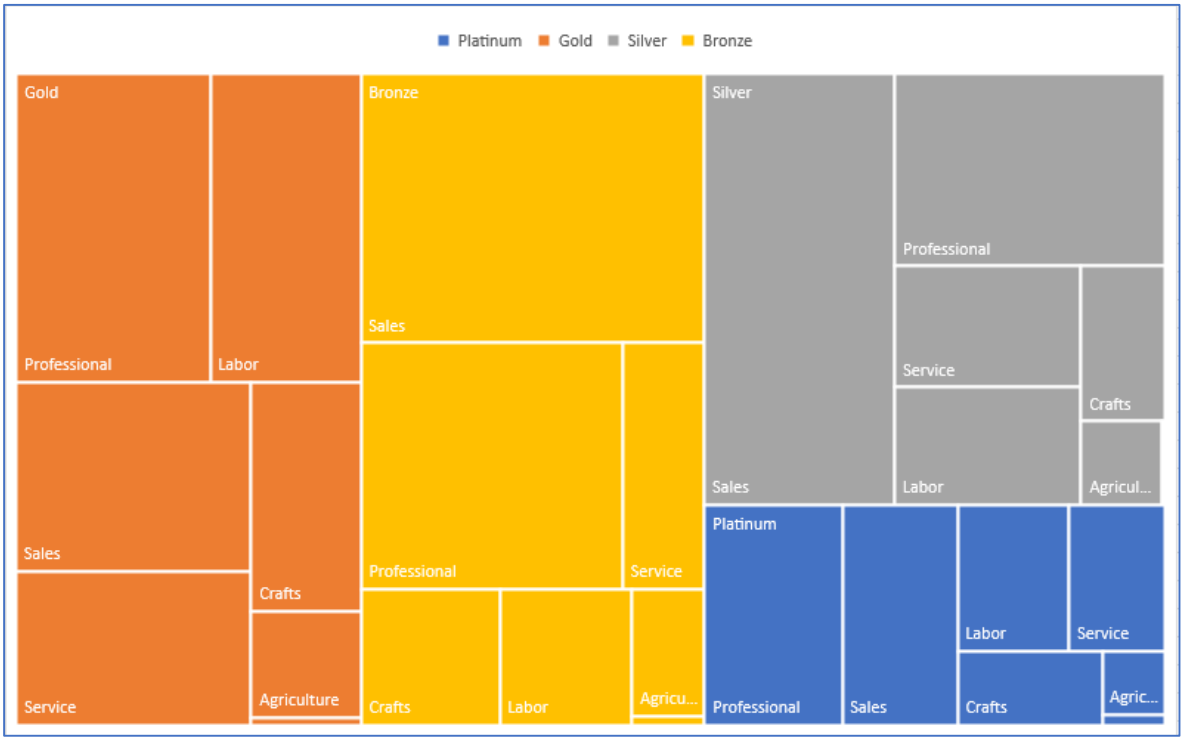


Figure 13: Job Category Distribution

Credit Card type usage

Figure 14: Credit Card type usage

Bibliography

K-means Cluster Analysis. Retrieved from University of Cincinnati Business Analytics R Programming Guide: https://uc-r.github.io/kmeans_clustering

Hierarchical Cluster Analysis. Retrieved from University of Cincinnati Business Analytics R Programming Guide: https://uc-r.github.io/hc_clustering