Alisha C. Souza

DSA5400

Segmentation and Profiling Project

Executive Summary

Background

Continuing the role of an incoming marketing analytics manager for a telecommunication company the use of customer analytics through evidence-based management is of high importance. In this analysis customer retention is the goal and we will be looking at different methods of segmentation to be able to understand our customer base to be able to support effective efforts in meeting our customer's needs and overall strengthening customer retention. Prior to this, the data was cleaned and variables were set to different categories (Customer Profile, Customer Education, Customer Profession, Services and Devices, Income and Monetary Utilization and Customer Loyalty) for a general understanding of the customer base.

Processes/Methods

The segmentation methods chosen for this process were rules-based segmentation and unsupervised segmentation. The Rule Based segmentation was chosen because I had specific values in mind that I was interested in segmenting by. Having had some experience with this data set in the due diligence project I felt this method would be a great addition to the knowledge I have already obtained about the customers of this company. In this method input variables were selected, and separated into bins of high medium and low. Additionally, rules were defined and 6 different segments were created (Segment 1: low phone company tenure customers that are married, Segment 2: medium phone company tenure customers that are married, Segment 3: high phone company tenure customers that are married, Segment 4: low phone company tenure customers that are unmarried, Segment 5: medium phone company tenure customers that are unmarried, Segment 6: high phone company tenure customers that are unmarried.)

The second method unsupervised segmentation was chosen as a way of comparing and contrasting efficient segmentation using the k means algorithm for clustering like segments. This method of segmentation starts by preparing a numerical dataset for k-means clustering. After standardizing the data, the k-means algorithm assigns customers to clusters, and the number of segments is evaluated for effectiveness. Following this, segment analysis and profiling similar to the rule-based approach can be conducted using the determined optimal number of clusters.

Results/Recommendations

The objective of the segmentation analysis was to enhance customer insight for the telecom company with a focus on retention, using two distinct segmentation methods. While both methods offered customer insights, the rule-based approach provided more detailed information on specific variables. This suggests the potential benefits of continuing with a rule-based strategy for further analysis or reconsidering efforts based on the results. This approach not only revealed customer value but also unveiled life stage and demographic segments such as age and marital status.

The findings align and add to the due diligence project, indicating that the most valuable customers often belong to crafts and labor job categories, are married, and reside in the southeast and northeast regions. It's recommended to prioritize maintaining products, services, and customer support for this group to sustain loyalty. Conversely, focusing on unmarried customers in Sales and professional careers is advised, as they exhibit lower tenure. A key observation is that the largest customer group in the company, comprising sales and professional job categories, demonstrates lower loyalty. Establishing strong customer relationships within these categories is essential. Overall, unmarried individuals and the southwest region show reduced loyalty, emphasizing the need to engage customers in these categories to enhance company tenure.

Technical Report

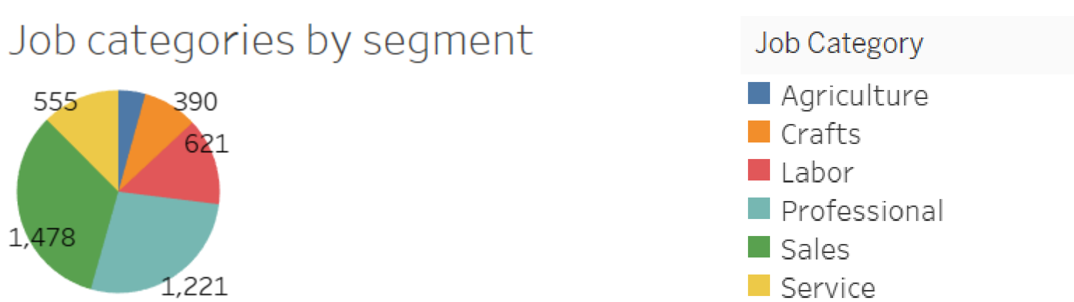
Rule Based: Value Segmentation

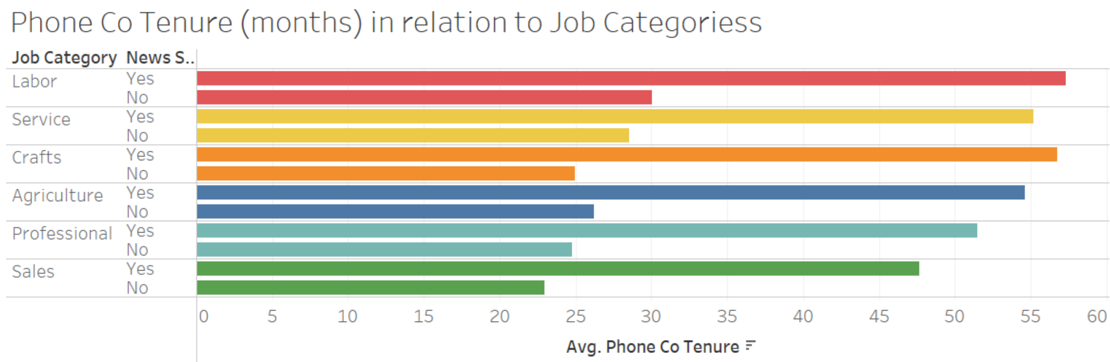
Having gained a good foundation of knowledge of the customer base from the due diligence project I began further analysis using Rule based segmentation. Initially I chose this method for its ability to be customized. Since I had already set my variables into groups I knew that there were specific elements that I wanted to focus on while keeping customer retention in mind. With this being said I also chose this method because I felt it was the best suited for customized marketing campaigns in the future which could help meet the company's goals. Rule based segmentation allows us to identify distinctive customer segments that are differentiated by values. The data prepared is defined into different bucket values that correspond with particular rules that are determined. Additional rules are added to then define "low" "medium" and "high" customer segments. From here customer segment profiles are created and analyzed or adjusted to make sense.

With this customer data and Rule-based method I first began by creating categorical values from these numerical values. These values were split into 3 equal buckets (1-33, 33-66,66-100) that were then classified into values low medium and high. From here rules were made to create segments. I chose to focus on the data in relation to phone company tenure as I was curious to see which customers are already loyal and see what attributes these customers have. The segments created are as follows: Segment 1: low phone company tenure customers that are married, Segment 2: medium phone company tenure customers that are married, Segment 3: high phone company tenure customers that are married, Segment 4: low phone company tenure customers that are unmarried, Segment 5: medium phone company tenure customers that are unmarried, Segment 6: high phone company tenure customers that are unmarried. Each customer in the data set was then classified into each of these 6 segments. A summary statistic of these segments was ran and further analysis was done through visualizations.

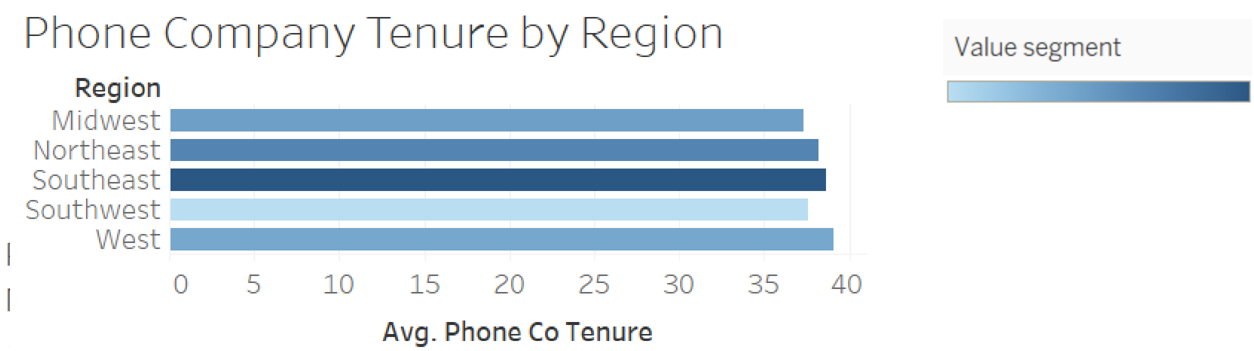
Looking at the summary statistics (see appendix) of our Rule-based segmentation we can gather insights on the classifications set by our self-determined rules. From this we can already see an interesting pattern in the count of customers in relation to each segment. Segments 3 and 6 have two of the lowest count values which corresponds to the customers with high value of phone company tenure of both married and unmarried couples. Having these be the lowest of the segments reassures us that there is room for extreme growth with retention in the company. On the contrary, segments 1 and 4 are of the lowest tenure with the phone company for both married and unmarried customers and have significantly higher counts with segment 4 (unmarried) having the largest count of all segments. Looking at the age of these customers within the segments we do see an increase in average age from segments 1-3 (35yrs, 49 yrs and 63yrs) and the same from segments 4-6 (37yrs, 54yrs and 65yrs) as expected since tenure will increase as age of the customer increases. Similar to this trend is the card tenure of customers within each segment that mirrors the age in regards to tenure. We also see similarities among segments in the customers education years, debt to income ratio, card items monthly and TV watching hours. So far, our segmentation has shown that there is plenty of room for growth in tenure loyalty amongst both married and unmarried customers.

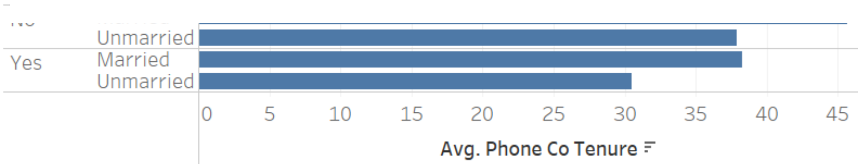
Taking a deeper dive into the visualizations of segmented data we can get a better understanding of each segment and how it compares and contrasts to different variables. Looking at the job categories by segment we can see that majority of the customer base is in sales and professional job categories. This is an interesting component as we can see that the average phone company tenure in comparison to these job segments is lower than all other job categories. Once again this indicates the potential for company growth seeing as there is a greater number of customers in these areas. On the other hand, we can see that the Labor and crafts job categories show stronger tenure. Another key aspect indicated in the graphs below is the value of customers within each region. We can see that the highest value customers based on phone company tenure are seen in the southeast and the lowest value customers are in the south west. Looking at activity level we can see that customers who have shown the most loyalty in the past are married and also not active however being married and non-active also showed more loyalty than unmarried customers.





Average of Phone Co Tenure for each News Subscriber broken down by Job Category. Color shows details about Job Category.





Unsupervised Segmentation

The process for unsupervised based method starts off slightly different and then uses some of the same method seen I the rule-based method. Overall, it is a good idea to start off by making a copy of the customer data set with just columns of numerical values of interest for k-means clustering since this can only be done with numerical values. The data is then scaled so variables are all weighted equally. Next, the k-means clustering is run and customers are assigned to their cluster. An analysis is also run to visualize the effectiveness of the number of clusters used and once the number of segments is determined based on interpretation of k-means cluster further analysis and profiling of segments can be conducted in the same way as the rule-based segmentation method. One of the reasons why I choose this method is because of the general knowledge base I already had with the customer data set. Having already understood many of these variables I was curious to see how models such as the k-means segmented the data in opposed to how I chose to segment with the rule-based model.

The main goal with this method is to identify different clusters and choosing the number of clusters fit for further analysis. I began by converting several column variables to numerical values. This process also included recoding variables such as marital status to either 1 or 0 to represent married or unmarried. Once all of my data was numerical it was then scaled and k-means clustering was conducted. Analysis from the k-means clustering graph shows 2-12 different clusters. We can see a strong change in this curve around the 5-6 cluster area. For further analysis I decided to look at 6 clusters and ran a summarized result of the 6 segments. Averages and counts of each segment were calculated along with looking at overall patterns and correlations of segments (see appendix).

Looking at the summary statistics output from the unsupervised method we can begin to take a deeper look at the customers within each segment. Originally, I choose to use 6 segments in the k-means segmentation however I found to much similarity and decided to go with 4. Once thing that really stuck out to me initially was the number low number of counts in the first segment (206) and I n segments 2-4 we can see more similar counts. Customers in segment 1 also suck out because they high higher debt overall including credit debt and other debt showing an overall higher debt to income ratio. Customers in segments 2 and 4 showed much higher relations to card tenure and phone company tenure but also had a higher age as expected. The ratio of gender, and marital status was close to even across the board in all segments.

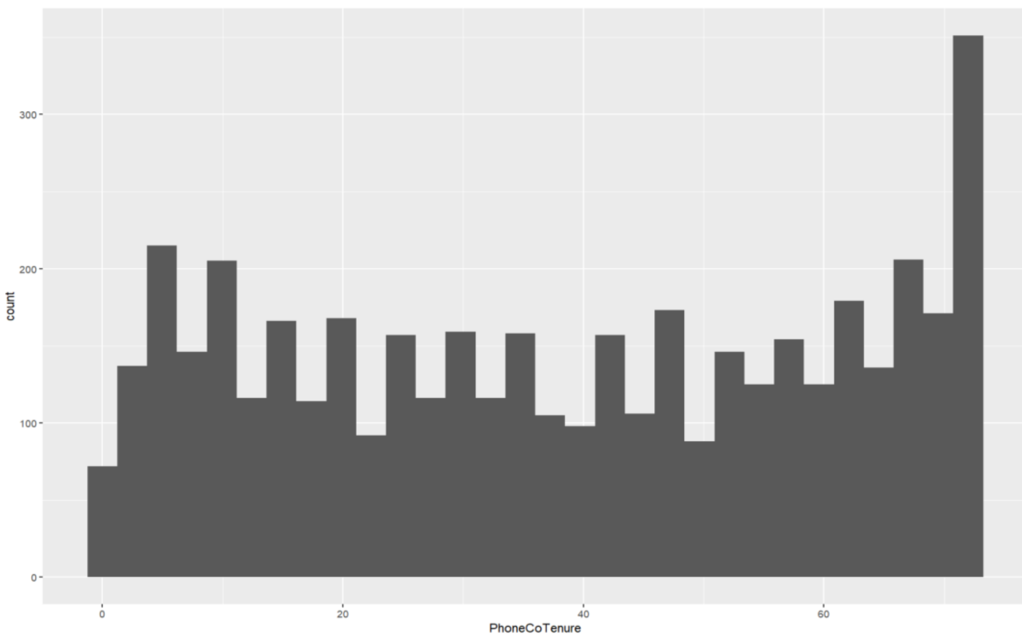
Recommendation

The goal of this segmentation analysis was to gain more insight on the customers within this telecommunication company keeping customer retention in mind while using two different segmentation methods. Each of the segmentation's methods provided information on the customer base however the use of rule-based segmentation gave more depth in regards to specific variables. From this the continuation of a rules-based strategy may be beneficial for additional analysis or to re evaluate any efforts made from the results of this analysis. Using the rule-based segmentation analysis we have gained information on the types of customers that are of highest value to the company. In addition to this rules-based segmentation not only gave us the value our customers have to the company but provided life stage and demographic segmentation of age and marital status for example.

Similar to what was seen in the results of the due diligence project we can see that the highest value customers tend to be in the crafts and labor job categories and are married and in the southeast and northeast regions. As a company the continuation of products, services and customer service for this group is important for continuation of loyalty. In the future learning about how this group of customers interacts with the company may lead to further findings of customer loyalty. On the other hand, I would recommend putting emphasis on customers in Sales and professional careers who are unmarried as they have shown the lowest tenure. Another important note as stated previously is that the largest number of customers in the company are in the sales and professional job categories of which show the lowest loyalty. Targeting strong relationships with customers in these categories would be a great start. Overall, unmarried individuals showed a decrease in loyalty as opposed to married individuals. The southwest region also stuck out as an area with low tenure. Figuring out how to gain the interest and trust of customers that call within these categories is sure to increase overall company tenure.

Appendix

Histogram of Phone Company Tenure used to visualize variable in order to split the numerical variables into “low”, “medium” and “high” categorical variables for rule setting



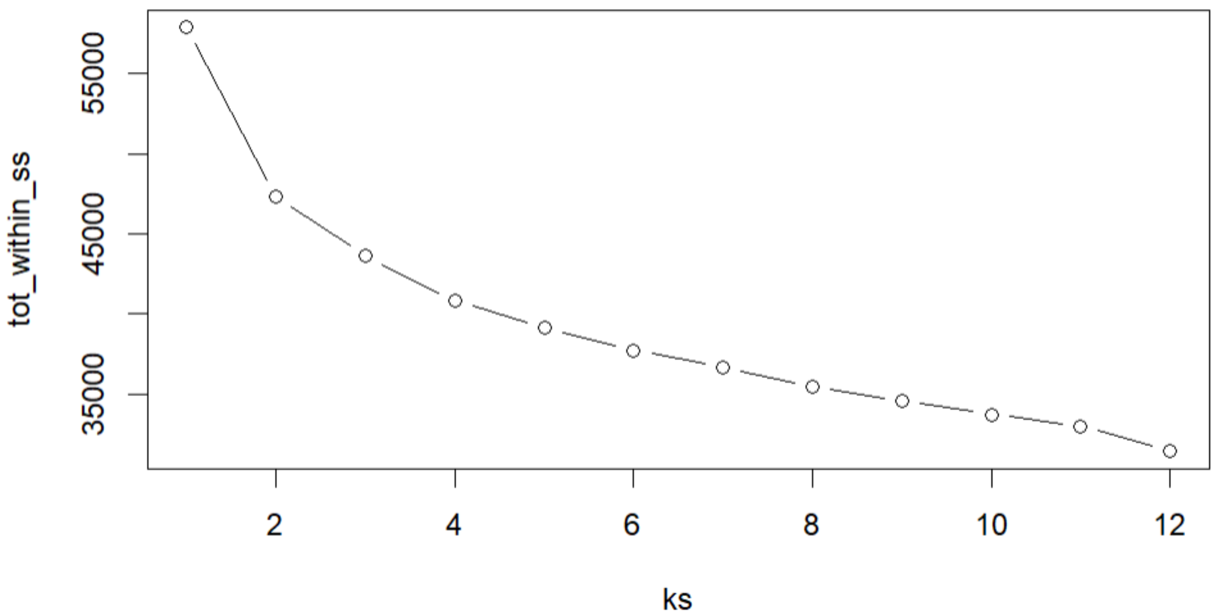

```r
#cerate a rule using the cut function
```{r}
customer_data$PhoneCoTenureGroup<-cut(customer_data$PhoneCoTenure,
breaks=c(-1,33,66,100), labels=c("low", "medium", "high"))

summary(customer_data)

```


#segmenting data
```{r}
customer_data<-customer_data %>%mutate(segment=case_when(
  PhoneCoTenureGroup == "low" & MaritalStatus == "Married" ~1,
  PhoneCoTenureGroup == "medium" & MaritalStatus == "Married" ~2,
  PhoneCoTenureGroup == "high" & MaritalStatus == "Married" ~3,
  PhoneCoTenureGroup == "low" & MaritalStatus == "Unmarried" ~4,
  PhoneCoTenureGroup == "medium" & MaritalStatus == "Unmarried" ~5,
  PhoneCoTenureGroup == "high" & MaritalStatus == "Unmarried" ~6
))
```
```

Summarized results of segments from Rules-based segmentation method

segment <dbl>	n <int>	Age <dbl>	EducationYears <dbl>	EmploymentLength <dbl>	DebtToIncomeRatio <dbl>	CardTenure <dbl>	CardItemsMonthly <dbl>	PhoneCoTenure <dbl>	TVWatchingHours <dbl>
1	780	35.35385	14.66154	4.074359	9.435000	6.617949	10.08718	17.19103	19.72692
2	960	49.86667	14.55729	9.926042	9.836250	20.753125	10.36667	51.09063	19.93437
3	395	63.89873	13.90633	20.015190	9.821772	35.508861	10.02278	70.32658	19.55696
4	1199	37.68974	14.91910	4.984987	9.981068	6.276063	10.08090	15.20183	19.22686
5	859	54.20256	14.53900	12.498254	10.531665	21.164144	10.20838	50.22119	19.77183
6	264	65.56818	13.49242	22.757576	9.820076	35.700758	10.10985	70.08712	19.57197

Analysis of clusters for k-means cluster



Summarized results of segments from unsupervised segmentation method

kmeans_segment <fctr>	n <int>	TVWatchingHours <dbl>	TownSize <dbl>	Age <dbl>	EducationYears <dbl>	EmploymentLength <dbl>	CarsOwned <dbl>	CommuteTi... <dbl>	CardTenure <dbl>	CardItemsMonthly <dbl>
1	206	19.66505	2.553398	56.59709	15.58738	17.927184	2.441748	25.48544	25.334951	10.04854
2	1456	19.61126	1.683379	37.15522	15.03091	4.425824	2.359203	21.97047	8.548077	10.07280
3	1589	19.55821	2.600378	63.29201	13.79295	17.679673	2.376337	25.10699	29.093770	10.07992
4	1206	19.70978	4.039801	36.00332	14.78441	4.232172	2.340796	29.82338	8.578773	10.40713

PhoneCoTenure <dbl>	TVWatchingHoursGroup <dbl>	DebtToIncomeRatio <dbl>	CreditDebt <dbl>	OtherDebt <dbl>	Female <dbl>	Married <dbl>
52.64563	1.839806	20.332039	11.314993	19.934977	0.5000000	0.5000000
24.03915	1.843407	9.630082	1.352191	2.721952	0.4320055	0.4320055
59.77281	1.833858	9.266457	1.614682	3.314564	0.5191945	0.5191945
24.27861	1.833333	9.414594	1.254465	2.552785	0.4792703	0.4792703

Scatter plot of k-means segments in relation to age of customer and phone company tenure