Telecommunication Customer Database

Segmentation & Profiling Project

Cullen Nauck

12/20/2019

Merrimack College

DSEG5113

Contents

[Table of Figures 3](#_Toc27752699)

[Executive Summary 4](#_Toc27752700)

[Overview 4](#_Toc27752701)

[Business Goals 4](#_Toc27752702)

[Data Description 4](#_Toc27752703)

[Methodological Summary 4](#_Toc27752704)

[Results 4](#_Toc27752705)

[Recommendations 4](#_Toc27752706)

[Core Report 5](#_Toc27752707)

[Overview 5](#_Toc27752708)

[Business Goals 5](#_Toc27752709)

[Data Description 6](#_Toc27752710)

[Data Quality 6](#_Toc27752711)

[Data Types 6](#_Toc27752712)

[Methodological Summary 7](#_Toc27752713)

[Segmentation Solution Logic 7](#_Toc27752714)

[Classification Model 8](#_Toc27752715)

[Results 9](#_Toc27752716)

[Segmentation 9](#_Toc27752717)

[Customer Segment Profiles 9](#_Toc27752718)

[Business Segment Profiles 13](#_Toc27752719)

[Classification Model 13](#_Toc27752720)

[Conclusions and Recommendations 13](#_Toc27752721)

[Appendix 14](#_Toc27752722)

# Table of Figures

[Figure 1: K-means Inertia Comparison 7](#_Toc27754877)

[Figure 2: K-means Segment Scatter Plots 8](#_Toc27754878)

[Figure 3: Segment Populations 9](#_Toc27754879)

[Figure 4: Segment Age Bin Percentages 10](#_Toc27754880)

[Figure 5: Segment Average Age 10](#_Toc27754881)

[Figure 6: Segment Region Percentages 10](#_Toc27754882)

[Figure 7: Segment Town Size Percentages 10](#_Toc27754883)

[Figure 8: Segment Town Size Average 10](#_Toc27754884)

[Figure 9: Segment Household Income Percentages 11](#_Toc27754885)

[Figure 10: Segment Household Income Average 11](#_Toc27754886)

[Figure 11: Segment Loan Default Percentages 12](#_Toc27754887)

[Figure 12: Segment Debt To Income Ratio 12](#_Toc27754888)

[Figure 13: Segment Phone Company Tenure 12](#_Toc27754889)

[Figure 14: Segment Customer Profiles 13](#_Toc27754890)

[Figure 15: Segment PC 13](#_Toc27754891)

[Figure 16: Segment Game System 13](#_Toc27754892)

[Figure 17: Segment Fax 13](#_Toc27754893)

[Figure 18: Segment News 13](#_Toc27754894)

[Figure 19: Segment Internet 13](#_Toc27754895)

[Figure 20: Segment Mobile Device 14](#_Toc27754896)

[Figure 21: Segment Wireless Data 14](#_Toc27754897)

[Figure 22: Segment Data Per Month 14](#_Toc27754898)

[Figure 23: Segment Voice Per Month 15](#_Toc27754899)

[Figure 24: Segment Calling Card 15](#_Toc27754900)

[Figure 25: Segment Multiline 15](#_Toc27754901)

[Figure 24: Segment Calling Card 17](#_Toc27754902)

[Figure 25: Segment Multiline 17](#_Toc27754903)

# Executive Summary

## Overview

## Business Goals

## Data Description

## Methodological Summary

## Results

## Recommendations

# Core Report

## Overview

For this project, a 5,000 customer database was given with various consumer information and telecommunications spending information. The goal for this project was to develop a customer segmentation that can support effective and economically sound customer retention efforts. The business goals developed for this project was to determine what segments would be in the market for high-speed internet, unlimited mobile data and/or unlimited voice packages (or a bundle of all three). By segmenting the database into distinct clusters, we can more efficiently target customers based on their previous history.

To start with, some feature engineering was required to recode and reformat variables into data sets that could be used for the analysis. Some of this refactoring included cleaning the null values and reformatting the monetary values to be floating point numbers. Also, some variables were converted from total spending over the customers tenure into *per month* spending. This was done to get a better feel of how much each customer was spending on a month-to-month basis, so customers of varying tenures could be compared and analyzed against each other. For instance, for voice per month, the voice over tenure spending values were divided by the phone company tenure for each customer, giving a *per month* voice spending value.

Next, some of the continuous variables went through a process of binning into similar groups (primarily based on the quartile ranges). This was done to be able to view the cross-segment differences more clearly and effectively. Finally, some of the nominal categorical data was coded numerically for use in the k-means clustering method.

Before running the clustering and segmenting the data, the variables were split into various groups. The first group was the clustering variables, which included general overview variables, debt behavior variables and finally communications behavior variables. These variables were chosen for the basis of the segmentation analysis. The second group was the business profile variables, which included internet and TV behavior, mobile data behavior and voice behavior variables. These variables would be used to compare the segments to determine high and low value segments.

Once the data was cleaned and refactored, the k-means clustering approach was used to segment the data into 5 segments. This process included scaling the clustering variable data and then running the k-means clustering technique (using the scikit-learn package in Python). After performing this analysis, the segmentation solution was evaluated statistically and practically. The 5 segments were compared using crosstabs and pivot tables both on the clustering variable data and the business profile variables. This gave an in-depth overview of the segment profiles and cross-segment differences in order to arrive at the objective of determining which segments would be in the market for various packages.

Finally, using the segments determined for each customer, a classification model was built using a Random Forest. This was done so that this model can be taken to the remaining customers outside this small dataset to target specific customers for packages. The random forest was evaluated using training and test data from the 5,000 customer database. The accuracy of the model was evaluated by running the model on the test data.

## Business Goals

The goal for this project is to determine segments that can be targeted for packages including high-speed internet, unlimited mobile data and/or unlimited voice packages (or a bundle of all three). To do this, spending and various telecom behavior was gathered for each customer to determine their previous history. By determining which segments are more inclined to internet, mobile phones and/or voice, we can better know which upgrades and packages will have a higher probability of retaining customers. Since we have a broad customer range over various markets, it is important that we recognize the needs of each segment separately rather than assume all customers want the same thing.

## Data Description

The 5,000 customer database was provided in a .csv format with various data types. Variables included general information such as age, gender and region as well as spending and debt behavior. The 59 variables provided were truncated and reformatted into two smaller groups: demographic clustering variables and business profiling variables. In order to ensure the data was in the correct format for analysis, some feature engineering was done to clean data, bin data and code data into numerical values.

### Data Quality

Since the raw data was in various formats, some work was required to correct the data for analysis. The null (missing) values in analysis variables were replaced in different ways according to their variable type. The TownSize variable had two missing values, so it was assumed sufficient to simply replace these with the mode of TownSize. The missing Gender and JobCategory values were replaced with “Unknown”. Finally, the Card Spend, Voice and Data over tenure missing values were replaced with 0 because it was determined that a missing value simply meant the customer had not spent money in that category. Also, these monetary values were feature engineered to remove the ‘$’ and ‘,’ to format them into floating point numbers. The Internet variable contained 3 numeric values as well as ‘yes’ and ‘no’. It was assumed that these three values corresponded to ‘yes’ on Internet, and they were recoded in this way.

To more thoroughly capture spending trends over the customer database, the voice and data over tenure spending values were converted to a *per month* value by dividing by the total tenure with the phone company. This was done to get a better feel of how much each customer was spending on a month-to-month basis, so customers of varying tenures could be compared and analyzed against each other. For instance, for voice per month, the voice over tenure spending values were divided by the phone company tenure for each customer, giving a *per month* voice spending value.

Next, some of the continuous variable data went through a process of binning to gather the data into similar groups. To do this, statistical analysis was done on the continuous variable to determine quartile locations. These were then used to determine bin group regions across the range of the variable. The variables binned using this process were Age, Data spending, HHIncome, EducationYears, CardSpendMonth, DebtToIncomeRatio, PhoneCoTenure, TVWatchingHours, and VoicePerMonth. These were binned to more clearly distinguish between the various segments.

Finally, for use in the k-means clustering approach, some of the categorical nominal variables had to be numerically coded. These included Gender, JobCategory and LoanDefault. This was done by simply assigning a particular category to a specific integer for each of these variables.

### Data Types

After determining the relevant variables for this analysis, these remaining variables were split into two groups: demographic clustering variables and business profiling variables. The first group was the clustering variables, which included general overview variables, debt behavior variables and finally one communications behavior variable. These variables were chosen for the basis of the segmentation analysis. The second group was the business profile variables, which included internet and TV behavior, mobile data behavior and voice behavior variables. These variables would be used to compare the segments to determine high and low value segments.

Within the demographic variables used for clustering were the general overview variables. These provided general demographic information for each customer in the database. They included age, town size, region, gender, job category, household income and education years. These variables were selected because it was expected that these could thoroughly capture the various demographic differences between segments of the customer set. Next, the debt behavior variables included debt to income ratio, card expenditures per month and loan default. These were included to capture some of the financial factors that can play into telecommunication usage. Finally, phone company tenure was included as a clustering variable so that we can get clarity into the differences between customers of various tenures with our company.

The business profiling variables were split into three groups: internet/streaming variables, mobile data variables and voice usage variables. The internet and streaming variables included PC, Game System and Fax ownership, as well as news subscriptions, TV watching information and finally internet access. Next, the mobile data variables included wireless data access, data spending per month and mobile phone ownership. Finally, the voice usage variables included voice spending per month, calling card, three way calling, multiline and voice mail usage.

## Methodological Summary

After the feature engineering and variable recoding was completed, the selected clustering variables could be used to segment the data set into groups. To do this segmentation a k-means clustering approach was used. The KMeans algorithm from the scikit-learn package in Python was used to perform this analysis. After the segmentation was completed, a classification model for the data set was built using a Random Forest.

### Segmentation Solution Logic

Although it has various drawbacks since it is a Euclidian distance spaced model (which required the recoding of categorical values), k-means was determined to be sufficient for this level of analysis. K-means separates samples into a selected number of groups of equal variance by minimizing inertia. To perform this analysis, first the variables selected for clustering were scaled by applying a standard scaler to them. This was done to normalized the variables so that differences in orders of magnitude between the variables would skew the results. Next, an inertia plot was created to determine how the inertia values were affected by the number of clusters:

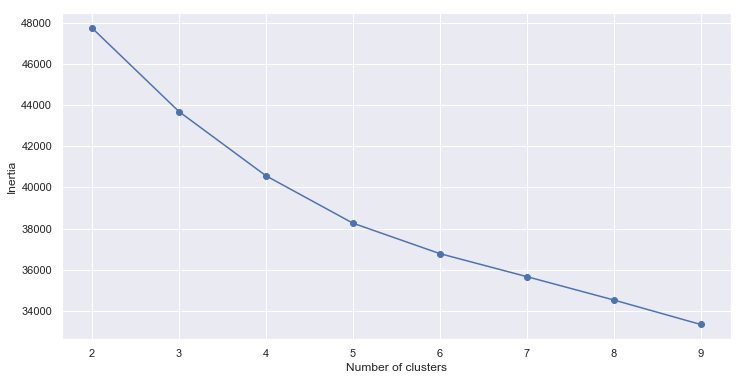


Figure 1: K-means Inertia Comparison

Although in this case the Inertia plot does not show a leveling off after a certain number of clusters, it was determined for this level of analysis that 5 clusters would be sufficient. The k-means function was run for 5 clusters based on the clustering variables previously discussed. After this was completed, the clusters were reviewed to evaluate the statistical and practical efficacy of the solution. The inertia for this 5 segment k-means clustering was determined to be 38260. The segments ranged in size from 1291 customers to 383 customers. Some of the cluster centers were compared against each other for the k-means variables to view the scatter of each cluster across the variable space.

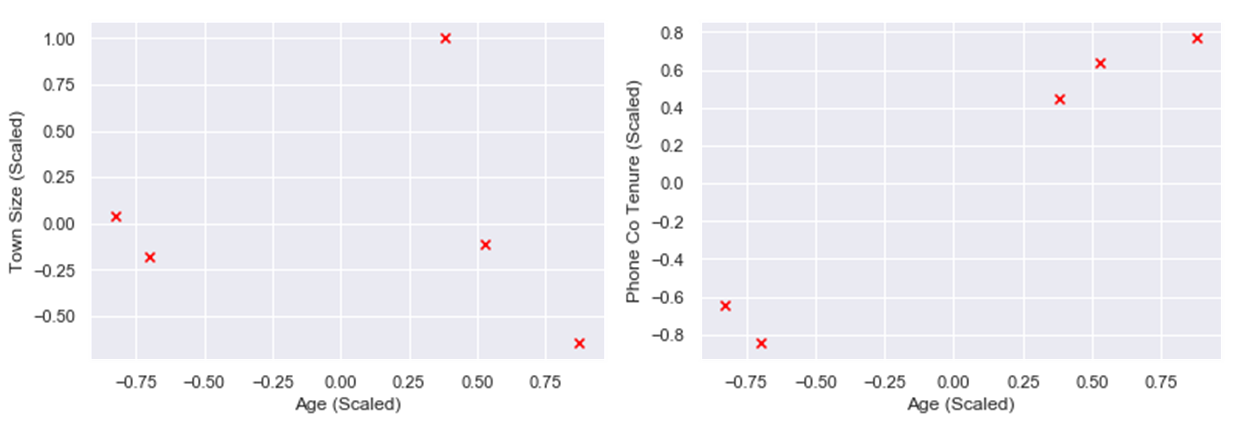


Figure 2: K-means Segment Scatter Plots

Though these are just two of many scatter plots that can be used to visualize the clusters, these show how each of the clusters fit into various segments of the market. Although some of the segments are relatively close to each other these figures show how the clusters are dispersed across the scaled variable space (in this case for age vs. phone co. tenure and town size). The segmentation results were then viewed using crosstabs to determine the cross-segment differences between each of the clusters (see Results section).

### Classification Model

Once the segmentation is completed using k-means clustering, a classification model was built using a Random Forest approach. This Random Forest algorithm used was also from the scikit-learn package for Python. To develop such a model with the created segments for all 5,000 customers in the dataset, the data was split into two groups: training and testing. To do so the train\_test\_split function was imported from scikit-learn, which randomly splits the x variables (the clustering variables) and the y variables (the segment for each customer) into 3,750 training points and 1,250 testing points. This split was necessary so that the classification model can be evaluated on data points not used to train the model.

The training data was then input into the random forest classifier from scikit-learn with the number of estimators (number of trees in the ‘forest) set to 100. Once this model was trained, the test x values could be used to predict the segments for each data point (customer). The accuracy of the model was determined to be 93% accurate, so the Random Forest classification model does a relatively good job of predicting the segment for each of the customers. With this model in hand, we will be able to determine which segments other customers (outside of this 5,000 customer database) fit into and efficiently target specific groups for programs and upgrades.

## Results

### Segmentation

To determine cross-sectional differences between the segments both for the clustering variables (demographic information) and for the business profile variables, crosstabs and pivot tables were created for the relevant information. This information is split into the two sets of results, with the first describing the customer demographics for each segment and the second describing the telecom spending behaviors for each segment. The 5 segments were relatively evenly split, with one exception. Segment 2 had significantly fewer customers than the other 4. Note that the crosstab tables have been normalized across each index (segment) to show the percentage of each segment in each bin. This was done to more clearly see how each segment is comprised and then could be compared to how other segments are comprised. Since segment 2 had significantly fewer customers, crosstabs normalized by the columns showed a low percentage of customers in each category for that segment, making it more difficult to determine the actual composition of that segment.

|  |  |  |
| --- | --- | --- |
| Segment | Population | % of Database |
| 0 | 1211 | 24.22 |
| 1 | 1047 | 20.94 |
| 2 | **383** | **7.66** |
| 3 | 1291 | 25.82 |
| 4 | 1068 | 21.36 |

Figure 3: Segment Populations

### Customer Segment Profiles

The first four variables reviewed were general demographic information including age, region, town size and gender. I have included crosstab information for age, region and town size. As shown in the tables below, segments 1 and 5 tend to be in the younger age group bins and this is also shown by the average age of the segments as well. Segments 2 and 3 are slightly above average age, with the majority being middle-aged. Finally segment 4 is the oldest with an average age of 63 years old. Most of the segments included customers across the 5 regions, however segment 4 had an above average number of customers from region 1 and segment 2 had over half of its customers from region 5. A majority of segment 2’s customers were from cities of size 4 or 5 while segment 4 had a majority of its customers from towns of size 1. The other 3 segments mostly held to the overall customer database distribution for town size. Finally, gender was evenly distributed for each of the segments, however segment 3 was slightly more male and segment 4 slightly more female (see appendix).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Age | 18-31 | >31-47 | >47-62 | >62 | | 1 | **45.0** | **40.8** | 12.7 | 1.5 | | 2 | 7.0 | 28.8 | 31.5 | 32.7 | | 3 | 2.9 | 16.4 | 50.4 | 30.3 | | 4 | 0.7 | 11.9 | 31.1 | 56.3 | | 5 | **57.6** | 29.0 | 10.1 | 3.3 | | All | 25.1 | 24.8 | 23.7 | 26.4 |   Figure 4: Segment Age Bin Percentages | |  |  | | --- | --- | |  | Age | | 1 | 34.6 | | 2 | 53.7 | | 3 | 56.4 | | 4 | 62.6 | | 5 | 32.4 | | All | 47.0 |   Figure 5: Segment Average Age |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Region | 1 | 2 | 3 | 4 | 5 |
| Segment | 1 | 21.6 | 23.5 | 22.7 | 20.4 | 11.8 |
| 2 | 0.6 | 3.9 | 13.8 | 28.6 | **53.1** |
| 3 | 16.2 | 17.0 | 24.8 | 20.1 | 21.9 |
| 4 | **36.4** | 28.7 | 21.8 | 9.8 | 3.4 |
| 5 | 19.6 | 22.0 | 19.4 | 20.3 | 18.7 |
| All | 20.2 | 19.9 | 20.1 | 19.3 | 20.5 |

Figure 6: Segment Region Percentages

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | Town Size | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | | Segment | 1 | 30.4 | 25.4 | 21.7 | 16.1 | 6.4 | | 2 | 0.4 | 6.2 | 18.7 | **31.3** | **43.4** | | 3 | 29.8 | 24.8 | 19.6 | 14.6 | 11.2 | | 4 | **51.2** | 28.4 | 13.3 | 6.3 | 0.9 | | 5 | 27.2 | 20.0 | 18.8 | 18.4 | 15.4 | | All | 28.8 | 21.0 | 18.1 | 17.1 | 15.0 |   Figure 7: Segment Town Size Percentages | |  |  | | --- | --- | |  | Town Size | | 1 | 2.4 | | 2 | 4.1 | | 3 | 2.5 | | 4 | 1.8 | | 5 | 2.7 | | All | 2.7 |   Figure 8: Segment Town Size Average |

The next three general demographic variables reviewed were job category, household Income and education years. Household income is shown below while job category and education information can be found in the appendix. From this information it was determined that segment 3 has above average distribution of agriculture, crafts and labor, while segment 4 has above average labor jobs and segment 5 has above average sales jobs. As shown in the figure below, segment 3 has extremely high household income compared to the other 4 segments. The other 4 averages fall below the overall average of the data set. Segments 4 and 5 have above average numbers of customers in the lowest quartile of household incomes. Finally, the education results show that customers in segments 1 and 2 typically have some college education, in segment 5 most have college education and segment 3 typically has higher levels of education than all the other segments. Segment 4 has the lowest education, typically only high school education.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Household Income | <$24K | $24K-38K | $38K-67K | >$67K | | 1 | **29.0** | **33.2** | 25.8 | **12.1** | | 2 | 19.1 | 24.2 | **30.9** | 25.8 | | 3 | 0.3 | 1.0 | 7.6 | **91.1** | | 4 | **32.5** | 18.3 | 25.4 | 23.8 | | 5 | **33.5** | 27.8 | 24.4 | 14.2 | | All | 23.8 | 26.6 | 24.5 | 25.1 |   Figure 9: Segment Household Income Percentages | |  |  | | --- | --- | |  | Household Income | | 1 | $39813 | | 2 | $51719 | | 3 | **$174052** | | 4 | $46660 | | 5 | $41699 | | All | $54760 |   Figure 10: Segment Household Income Average |

The next three variables analyzed across the segments were the debt behavior information including debt to income ratio, card spending per month and loan default. The loan default and debt to income ratio crosstab information is shown below while the card spending per month is included in the appendix. As shown below, loan default was a key driver in 4 of the 5 segments, with segments 1,2 and 4 having very few loan defaults. However, all segment 5 customers have a loan default, showing this is a key attribute for this segment. Segment 3 had around the average loan default distribution for the whole dataset. The debt to income ratio table shows that segment 1 has more customers with little to no debt while 42% of customers in segment 5 have a debt to income ratio of greater than 13.6. The card spending per month showed that segment 3 has significantly higher card spending per month compared to other segments, which corresponds to their higher income.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Loan Default | No | Yes | | 1 | **100.0** | **0.0** | | 2 | **99.0** | **1.0** | | 3 | 79.1 | 20.9 | | 4 | **99.0** | **1.0** | | 5 | **0.0** | **100.0** | | All | 76.6 | 23.4 |   Figure 11: Segment Loan Default Percentages | |  |  |  |  |  | | --- | --- | --- | --- | --- | | Debt to Income Ratio | <5.1 | 5.1-8.8 | 8.8-13.6 | >13.6 | | 1 | **36.8** | 28.7 | 20.9 | 13.5 | | 2 | 27.9 | 28.2 | 25.6 | 18.3 | | 3 | 28.2 | 21.7 | 25.1 | 25.1 | | 4 | 21.0 | 26.7 | 26.1 | 26.2 | | 5 | 12.5 | 18.3 | 27.2 | **42.0** | | All | 24.8 | 24.9 | 25.0 | 25.3 |   Figure 12: Segment Debt To Income Ratio |

The final clustering variable analyzed was the phone company tenure. In this table it is shown that segments 1 and 5 are significantly newer customers to the company, while customers in 3 and 4 have typically been at the company several years.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phone Co. Tenure | <18 | 18-38 | 38-59 | >59 |
| 1 | **55.7** | 34.0 | 9.7 | 0.7 |
| 2 | 6.4 | 23.3 | **35.8** | **34.5** |
| 3 | 5.2 | 17.8 | 29.5 | **47.5** |
| 4 | 1.5 | 13.6 | 35.6 | **49.3** |
| 5 | **49.0** | 29.3 | 16.4 | 5.3 |
| All | 26.1 | 24.9 | 24.3 | 24.8 |

Figure 13: Segment Phone Company Tenure

A summary of the customer segment profiles is found below. Segment 1 is typically younger, with lower income and lower debt to income ratio. They typically have some college education and many of them live in region 5. They have been customers at the company for a short time. Segment 2 is middle aged, typically living in larger cities, with middle income and some college education. They have a longer tenure as customers. Segment 3 is primarily defined as higher income customers and are typically middle aged and have higher education. They tend to have been customers longer. Segment 4 is the oldest and least educated segment, typically living in small towns with a significant population in region 3. They also have been customer for longer time periods. Finally, segment 5 is younger customers who all have had loan defaults. They have some college education and have a significant population in Sales. They have the highest debt to income ratio and typically low incomes. They have been customers for shorter tenures.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Segment Customer Profiles | | | | | | | | | | | |
| Segment | **% Data Set** | **Age** | **Significant Region** | **Town Size** | **Job** | **HH Income** | **Education** | **Debt to Income** | **Card Spend** | **Loan Default** | **Phone Co Tenure** |
| 1 | 24.2 | Young | 5 | Average |  | Low | College | 7.6 | Average | None | Short |
| 2 | 20.9 | Middle Aged |  | City |  | Middle | College | 8.9 | Average | 1% | Long |
| 3 | 7.7 | Middle Aged |  | Average |  | Wealthy | Higher | 9.8 | High | 20% | Long |
| 4 | 25.8 | Elderly | 3 | Town | Labor | Middle | High School | 10.3 | Average | 1% | Long |
| 5 | 21.4 | Young |  | Average | Sales | Low | College | 13.3 | Average | ALL | Short |

Figure 14: Segment Customer Profiles

### Business Segment Profiles

In a similar fashion to the segment customer profiles, the business profiles for each segment were created using crosstab tables for each business profile variable. The first set of variables reviewed was the internet and streaming variables. Segments 3 and 5 show slightly higher ownership of PCs while segment 4 has significantly lower PC ownership. Segments 1 and 5 are more inclined to gaming while segment 4 has less interest. Segment 3 has much higher fax ownership than the remaining segments. Segments 2,3 and 4 have higher news subscription rates than 1 and 5. Finally, segments 3 and 5 have slightly above average internet usage while segment 4 has below average internet use. TV watching was analyzed but showed similar results for each segment, it is included in the appendix.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | PC | No | Yes | | 1 | 32.1 | 67.9 | | 2 | 37.4 | 62.6 | | 3 | 29.5 | **70.5** | | 4 | **47.4** | 52.6 | | 5 | 30.9 | **69.1** | | All | 36.7 | 63.3 |   Figure 15: Segment PC | |  |  |  | | --- | --- | --- | | Game System | No | Yes | | 1 | 43.8 | **56.2** | | 2 | 56.9 | 43.1 | | 3 | 51.4 | 48.6 | | 4 | **66.9** | 33.1 | | 5 | 41.1 | **58.9** | | All | 52.5 | 47.5 |   Figure 16: Segment Game System |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Fax | No | Yes | | 1 | 82.3 | 17.7 | | 2 | 83.1 | 16.9 | | 3 | 68.4 | **31.6** | | 4 | 86.6 | 13.4 | | 5 | 80.4 | 19.6 | | All | 82.1 | 17.9 |   Figure 17: Segment Fax | |  |  |  | | --- | --- | --- | | News Subscriber | No | Yes | | 1 | **77.8** | 22.2 | | 2 | 39.8 | **60.2** | | 3 | 34.2 | **65.8** | | 4 | 29.7 | **70.3** | | 5 | **71.5** | 28.5 | | All | 52.7 | 47.3 |   Figure 18: Segment News | |  |  |  | | --- | --- | --- | | Internet | No | Yes | | 1 | 46.3 | 53.7 | | 2 | 52.2 | 47.8 | | 3 | 40.5 | **59.5** | | 4 | **59.6** | 40.4 | | 5 | 43.6 | **56.4** | | All | 50.0 | 50.0 |   Figure 19: Segment Internet |

The next set of variables evaluated was mobile data information. Segments 1 and 5 show higher usage of mobile devices while segment 4 shows lower usage. Segment 3 shows the highest wireless data subscriptions (and data spending per month) while segment 4 shows the lowest wireless data subscriptions and data spending per month. The three remaining segments showed similar results to one another.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Mobile Device | No | Yes | | 1 | 44.3 | **55.7** | | 2 | 57.4 | 42.6 | | 3 | 50.4 | 49.6 | | 4 | **64.5** | 35.5 | | 5 | 41.3 | **58.7** | | All | 52.1 | 47.9 |   Figure 20: Segment Mobile Device | |  |  |  | | --- | --- | --- | | Wireless Data | No | Yes | | 1 | 70.9 | 29.1 | | 2 | 75.9 | 24.1 | | 3 | 57.2 | **42.8** | | 4 | **79.9** | 20.1 | | 5 | 70.4 | 29.6 | | All | 73.1 | 26.9 |   Figure 21: Segment Wireless Data |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Spending per Month | 0 | 7-26 | 26-35 | 35-46 | >46 |
| 1 | 70.9 | 12.7 | 6.9 | 5.5 | 4.0 |
| 2 | 75.9 | 3.2 | 5.7 | 6.5 | 8.6 |
| 3 | 57.2 | 3.9 | 9.7 | 13.1 | **16.2** |
| 4 | **79.9** | 1.8 | 5.3 | 6.4 | 6.5 |
| 5 | 70.4 | 10.6 | 8.7 | 5.5 | 4.8 |
| All | 73.1 | 6.9 | 6.5 | 6.8 | 6.7 |

Figure 22: Segment Data Per Month

The last group of variables analyzed was the voice information. This included voice spending per month, calling card usage, three way calling, multiline and voicemail. Segments 1 and 5 show lower voice spending per month while 3 and 4 show high levels of voice spending. Similarly, segments 1 and 5 have lower rates of usage for calling cards and multiline than other segments. Segments 2,3 and 4 all use calling cards significantly while 3 and 4 also subscribe to multiline more than average. Segment 3 showed the highest rate of usage for three way calling and voicemail, significantly higher than all other segments (see appendix).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Voice per Month | <5 | 5-10 | 10-17 | >17 |
| 1 | **41.0** | **42.9** | 14.7 | 1.5 |
| 2 | 6.1 | **28.8** | **32.5** | 32.6 |
| 3 | 5.2 | 23.2 | 26.6 | **44.9** |
| 4 | 3.6 | 21.0 | 30.5 | **44.9** |
| 5 | **34.3** | **39.1** | 18.9 | 7.8 |
| All | 24.3 | 19.8 | 23.9 | 32.0 |

Figure 23: Segment Voice Per Month

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Calling Card | No | Yes | | 1 | **47.1** | 52.9 | | 2 | 19.4 | **80.6** | | 3 | 10.7 | **89.3** | | 4 | 12.3 | **87.7** | | 5 | **41.8** | 58.2 | | All | 28.4 | 71.6 |   Figure 24: Segment Calling Card | |  |  |  | | --- | --- | --- | | Multiline | No | Yes | | 1 | **69.3** | 30.7 | | 2 | 41.8 | 58.2 | | 3 | 28.7 | **71.3** | | 4 | 38.7 | **61.3** | | 5 | **62.8** | 37.2 | | All | 51.2 | 48.8 |   Figure 25: Segment Multiline |

A summary of the business profiling for each segment is shown below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Segment Business Profiles | | | | | | | | | | | | | | |
| Segment | **% Data Set** | **Mobile Device** | **PC** | **Game System** | **Fax** | **News** | **Internet** | **VM** | **Data per Month** | **Wireless Data** | **Voice Per Month** | **Calling Card** | **Three Way Calling** | **Multiline** |
| 1 | 24.2 | High |  | High |  | Low |  |  |  |  | Low | Low |  | Low |
| 2 | 20.9 |  |  |  |  | High |  |  |  |  | High | High |  |  |
| 3 | 7.7 |  | High |  | High | High | High | High | High | High | Highest | High | High | High |
| 4 | 25.8 | Low | Low | Low |  | High | Low |  |  | Low | High | High |  | High |
| 5 | 21.4 | High | High | High |  | Low | High |  |  |  | Low | Low |  | Low |

### Classification Model

## Conclusions and Recommendations

# Appendix

|  |  |  |  |
| --- | --- | --- | --- |
| Gender | Female | Male | Unknown |
| 1 | 49.5 | 49.9 | 0.6 |
| 2 | 47.9 | 51.0 | 1.1 |
| 3 | **43.3** | **56.4** | 0.3 |
| 4 | **53.4** | **46.2** | 0.4 |
| 5 | 50.2 | 49.0 | 0.8 |
| All | 49.9 | 49.5 | 0.7 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| JobCategory | Agriculture | Crafts | Labor | Professional | Sales | Service | Unknown |
| 1 | 3.7 | 9.3 | 11.0 | 28.7 | 36.5 | 10.5 | 0.2 |
| 2 | 4.3 | 8.2 | 14.4 | 26.4 | 31.7 | 15.0 | 0.0 |
| 3 | **7.3** | **14.9** | **22.2** | 22.2 | 13.8 | 19.3 | 0.3 |
| 4 | 5.0 | 9.5 | **17.0** | 27.0 | 29.5 | 11.5 | 0.6 |
| 5 | 2.8 | 6.9 | 9.2 | 30.2 | **40.0** | 10.6 | 0.3 |
| All | 4.2 | 9.0 | 13.7 | 27.6 | 32.7 | 12.4 | 0.3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Education Years | <12 | 12-14 | 14-17 | >17 |
| 1 | 24.4 | 23.5 | 31.3 | 20.7 |
| 2 | 28.0 | 21.9 | 30.5 | 19.7 |
| 3 | 14.6 | 16.7 | **33.9** | **34.7** |
| 4 | **43.4** | 19.9 | 23.7 | 13.0 |
| 5 | 19.5 | 24.8 | **32.0** | **23.7** |
| All | 29.5 | 22.0 | 28.3 | 20.2 |

|  |  |
| --- | --- |
|  | EducationYears |
| 1 | 14.8 |
| 2 | 14.5 |
| 3 | 16.0 |
| 4 | 13.5 |
| 5 | 15.1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Card Spending per Month | <1800 | 1800-2800 | 2800-4200 | >4200 |
| 1 | 23.8 | 29.4 | 26.7 | 20.1 |
| 2 | 25.3 | 25.1 | 27.3 | 22.3 |
| 3 | 1.0 | 5.0 | 12.3 | **81.7** |
| 4 | **29.3** | **30.4** | 23.1 | **17.2** |
| 5 | 25.3 | 28.7 | 24.7 | 21.3 |
| All | 24.1 | 26.7 | 24.4 | 24.8 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TV Watching Hours | <17 | 17-20 | 20-23 | >23 |
| 1 | 29.6 | 29.0 | 21.7 | 19.7 |
| 2 | 30.6 | 27.6 | 21.7 | 20.2 |
| 3 | 31.1 | 24.0 | 22.7 | 22.2 |
| 4 | 31.6 | 23.5 | 23.3 | 21.6 |
| 5 | 30.9 | 25.7 | 21.9 | 21.5 |
| All | 30.7 | 26.2 | 22.2 | 20.9 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | ThreeWayCalling | No | Yes | | 1 | 53.2 | 46.8 | | 2 | 51.9 | 48.1 | | 3 | 35.2 | **64.8** | | 4 | 55.0 | 45.0 | | 5 | 54.1 | 45.9 | | All | 52.2 | 47.8 |   Figure 24: Segment Calling Card | |  |  |  | | --- | --- | --- | | VM | No | Yes | | 1 | 69.3 | 30.7 | | 2 | 70.8 | 29.2 | | 3 | 59.0 | **41.0** | | 4 | **76.1** | 23.9 | | 5 | 65.3 | 34.7 | | All | 69.7 | 30.3 |   Figure 25: Segment Multiline |