WGU D211 Report

Andrew Shrestha

Introduction:

In today's extremely competitive world of Telecom services, it is vital for companies to consistently obtain and analyze data relating to customers with great significance to customer loyalty and retention. The ability of a company to keep customers is as important, if not more important, as its ability to gain new customers. We often see customer turnover or in this case "Churn" to be a vital metric brought up time and time again in any service industry as a Key Performance Metric of Interest (**Kenza M. 2020)**

Purpose:

The purpose of this dashboard will be to utilize tableau's visualization and interactive properties as well as several SQL supporting commands to display clear patterns and statistics relevant to our function to reduce customer churn. This will be done through the use of analyzing our primary dataset which consists of customer churn along with several other independent variables of varying significance playing a role in the churn outcome. This data set is the primary since it strongly correlates to our business need of analyzing and predicting customer churn.

The secondary Dataset will consist customer churn specifically in the region of California. This provides quite abit of further insights as California is the region with the highest data usage. This data set also provides a more in-depth analysis into customer churn reasoning, which can be extrapolated to cover all customers in general which is very useful. Integrating a secondary data set with regards to churn reasonings. In combination, these two will grant us a strong insight into the mindset of the customer characteristics and aid in predicting churn as well as avoiding churn results from being the outcome.

Business Intelligence Tool:

The intelligence tool utilized as previously mentioned is tableau. The reasoning behind this was because Tableau has extremely comfortable implementation of visualizations, ease of handling large amounts of data, and able to incorporate scripting language via python with SQL. Tableau is also heavily favored for its user experience, where coding is not necessarily a prerequisite to understand its outputs. This makes it highly attractive for business meetings and discussions where the findings are easy to comprehend. **(McLean H. 2020)**

Steps used to clean the data:

1) Utilized the “churn” database given in the WGU pgadmin database

2) Imported data sets in jyupter and renamed customer table columns/variables to more appropriate labels (Item1 to TimelyResponse)

3) Took an initial look at each of the tables to get a grasp of the columns, rows, and variables. Viewed structure and data types as well

4) Obtained summary statistics get a good idea of the dataset

5) Removed outliers and imputed missing data with significant values via “mean” or retain data integrity.

6) Conducted both univariate and bivariate visualizations

7) Prepared dataset and extracted under Churn_Clean.csv

8) Ensured that the Churn_Clean.csv dataset was still similar to that of the WGU pgadmin database given

9) Did the same procedure of checking for missing variables and imputing variables with the secondary Telco dataset obtained online.

10) Ensured that referential integrity was maintained between the tables with primary key and foreign keys correspondingly. Structural coherency being uncorrupted was taken into account while cleaning the datasets. **(IBM. 2021)**

Summary of Steps Used To Create Dashboard:

1) Top left hand side of the dashboard consists of the churn ratio by state. This is color coded by a light and dark shade of blue so that it is easy to interpretate visually by all audiences including color blindness. This portion was chosen since there could be locational characteristics influencing the churn rate that maybe interesting for a more detailed analysis.

2) Top middle of the dashboard display the Key Performance Indicators consisting of : Average Bandwidth Usage, Average Income, Average Monthly Charge, and Average Tenure all with regards to churn. This will be very useful as we see strong positive correlations with Bandwidth Usage and inverse correlations with Tenure and Monthly Charge when it comes to Churn. This is also color coded by orange and blue for ease of interpretation from all audiences including those with color blindness.

3) Bottom left side of the dashboard consists of the Monthly Revenue by state. This is displayed as a single color and bar chart due to this display only focusing on a univariate analysis of

revenue. It is assumed that revenue plays a significant part in churn, however this graph with the culmination of the others reenforce this idea.

4) Bottom middle of the dashboard display churn ratio by gender. This utilizes a stacked bar chart to avoid unnecessary clutter and keep the binary bivariate analysis simple.

5) on the right side of the dashboard is one of the most important displays, that which is reasoning for churn. This covers any bases missed by the other dashboard parts as it takes actual surveyed customers and explains the specific reason why customers choose to churn without ambiguity. This is displayed via string response output.

Results of Data Analysis and Impact on Executive Decisions:

- Three most significant and notable conclusions from the visualizations are that as Average Bandwidth GB Year usage increases (above 2,000) then there seems to be a correlation with customers deciding to stay and not to churn. Inversely, Any Average Bandwidth GB Year lower than 2,000 by the customer generally leads to churning
- Also, as Tenure increases, there is a sharp decrease in customer churn. This could be due to the tenure having an impact on loyalty.
- Finally, Monthly Charge of customers that do not churn seem to be lower than customers who churn. This seems to suggest that customers who churn use more services and pay more for the short term they are with the telecom company before deciding to change.

The two different data representations that Executives can use to support their business decision would be that of the Key Performance Indicators and Churn reasoning. The Key Performance Indicators consist of : Average Bandwidth Usage, Average Income, Average Monthly Charge, and Average Tenure all with regards to churn. This will be very useful as we see strong positive correlations with Bandwidth Usage and inverse correlations with Tenure and Monthly Charge when it comes to Churn. Executives can use this information of averages to get a better picture of the average variable value deciding between a customer churn or retention. An example of this is that we see retention at around $163 where churn happens at $199. This can support an idea of a discount once a customer's monthly charge gets close to the $199 per month mark.

Churn Reasoning via churn count data representation is also extremely useful as executives can pinpoint exactly the top reasons why customers churned and thus have clearly defined opportunities to fix these issues. Top reason for churn in this case is customer service attitude, which is totally trainable. This can support the idea of Executives to put more resources into training and hiring the right people.

Limitations:

The one obvious limitation that would have hindered the data analysis process would be that of the limited number of observations. With a field as important to corporate interests as “churn” we would want to ensure that we are able to collect as much data as possible from our sources. The issue was that the churn dataset we were given was limited to 10,000 observations while our IBM sourced Telco churn dataset was limited to just over 7,000. In order to have more certainty in the results we have found, it would be better to have these numbers be multiplied by several fold so that we can safely correlate the results to the sample population. If this was a real world example, I would definitely go to the data warehouse/ company and request more data in order to increase the accuracy and hopefully remove any biases/outliers effects from our results.

Acknowledged Sources and References:

Kenza M. (2020). How to Calculated Customer Churn Rate.
https://www.zendesk.com/blog/customer-churn-rate/

McLean H. (2020). 10 Advantages of Tableau for Data Visualization.

https://www.newhorizons-jax.com/blog/10-advantages-of-tableau-for-data-visualization

IBM. (2021). Referential Integrity.
https://www.ibm.com/docs/en/informix-servers/14.10?topic=integrity-referential