

**ASSIGNMENT COVER SHEET**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PROGRAMME | : | MASTER OF BUSINESS ANALYTICS | | |
| SUBJECT CODE AND TITLE | : | BAA5083 ADVANCED BUSINESS ANALYTICS & VISUALIZATION | | |
| ASSIGNMENT TITLE | : | TABLEAU PROJECT | | |
|  |  |  | | |
| LECTURER | : | DR TANG TIONG YEW | ASSIGNMENT DUE DATE: | 14/6/2024 |

STUDENT’S DECLARATION

1. I hereby declare that this assignment is based on my own work except where acknowledgement of sources is made.
2. I also declare that this work has not been previously submitted or concurrently submitted for any other courses in Sunway University/College or other institutions.

[ Submit “Turn-it-in” report (please tick √): Yes \_\_√\_\_ No \_\_\_\_\_ ]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NO. | NAME | STUDENT ID NO. | SIGNATURE | DATE |
| 1. | AHMED SHIYAAN IBRAHIM | 23118482 | *Shiyaan* | 13/6/2024 |
| 2. | RAHKESH A/L KUMAR | 19032754 | *Rahkesh* | 13/6/2024 |
| 3. | HARRESH A/L RAGUNATHAN | 19076090 | *Harresh* | 13/6/2024 |
| 4. | ABDUL HAKIM BIN KAMALUR RAHMAN | 24015257 | *Hakim* | 13/6/2024 |
| 5. |  |  |  |  |
| 6. |  |  |  |  |

E-mail Address / Addresses (according to the order of names above):

|  |  |
| --- | --- |
| 1. 23118482@imail.sunway.edu.my | 4. 24015257@imail.sunway.edu.my |
| 2. 19032754@imail.sunway.edu.my | 5. |
| 3. 19076090@imail.sunway.edu.my | 6. |

APPROVAL FOR LATE SUBMISSION OF ASSIGNMENT (If applicable)

IF extension is granted, what is the revised due date? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of Lecturer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |
| --- |
| Marker’s Comments: |

Marks and / or Grade Awarded: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ADDENDUM**

**USE OF ARTIFICAL INTELLIGENCE (A.I.) DECLARATION**

Students are allowed to use AI to support completion of assessments. However, students are reminded to do so ethically and transparently. This is so that (a) submissions can be fairly and accurately marked; and (b) feedback can be provided on the content that reflects student ability, in order to help with future submissions. Students are also reminded that in accordance with the University’s Academic Malpractice Policy, Item 4.11.2, “*… the representation of work: written, visual, practical or otherwise, of any other person, including another student or* ***anonymous web-based material*** *[emphasis added], or any institution, as the candidate’s own*” is considered malpractice.

**Declaration**

[ √ ] I / We used the following A.I. tools to produce content in this submission:

|  |  |  |  |
| --- | --- | --- | --- |
| **Tool** | **Purpose** | **Prompts** | **Sections where AI output was used / Outcome(s) in the submission** |
| *e.g. ChatGPT* | *e.g. Generating points for the essay*  *Structuring the essay* | *e.g.* “*Give me 5 key talking points for an essay on…”*  *“Show me a structure for an essay on…”* | *e.g. The main point for Section 1.2 and 1.3 were generated by AI, but the discussion was not.*  *The organization / structure of the essay was suggested by AI* |
| *e.g. Grammarly* | *e.g. Correcting grammar and spelling, improving sentence structure* | *N/A* | *e.g. Grammarly suggestions were used for all sections of the essay* |
|  |  |  |  |
|  |  |  |  |

*Note: Add additional rows if necessary.*

**OR**

[ ] I / We did not use any A.I. tools to produce any of the content in this submission.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NO. | NAME | STUDENT ID NO. | SIGNATURE | DATE |
| 1. |  |  |  |  |
| 2. |  |  |  |  |
| 3. |  |  |  |  |
| 4. |  |  |  |  |
| 5. |  |  |  |  |
| 6. |  |  |  |  |

E-mail Address / Addresses (according to the order of names above):

|  |  |
| --- | --- |
| 1. | 4. |
| 2. | 5. |
| 3. | 6. |

**TABLEAU PROJECT: REPORT WRITE-UP**

**Contents**

[**Section 1: Introduction** 3](#_Toc169217817)

[**Dataset 1** 4](#_Toc169217818)

[**Dataset 2** 6](#_Toc169217819)

[**Detail of the observations** 8](#_Toc169217820)

[I. **Independent Variable** 8](#_Toc169217821)

[II. **Demographic Variables** 9](#_Toc169217822)

[III. **Service Features Variables** 9](#_Toc169217823)

[IV. **Tenure and Charges Variables** 9](#_Toc169217824)

[V. **Outcome Variable** 9](#_Toc169217825)

[**Research Questions** 9](#_Toc169217826)

[**Research Objectives** 10](#_Toc169217827)

[**Section 2: Analysis, Storytelling and Visualization** 12](#_Toc169217828)

[**Exploratory Data Analysis (EDA)** 12](#_Toc169217829)

[**Univariate Analysis** 12](#_Toc169217830)

[**Bivariate Analysis** 13](#_Toc169217831)

[**Multivariate Analysis** 20](#_Toc169217832)

[**Correlation Matrix** 21](#_Toc169217833)

[**Location against Churn** 22](#_Toc169217834)

[**Services Subscribed against Churn** 23](#_Toc169217835)

[**Section 3: Discussion** 24](#_Toc169217836)

[**Section 4: Limitations** 27](#_Toc169217837)

[**Section 5: Conclusion** 28](#_Toc169217838)

[**Section 6: Additional Work** 29](#_Toc169217839)

[**References** 32](#_Toc169217840)

[**Tableau Project: Proposal** 33](#_Toc169217841)

# **Section 1: Introduction**

Much like the phrase ‘customers are always right’, businesses are learning that customers are also always important. Especially in a competitive and saturated market like telecommunication businesses are keen to understand the needs and wants of their customers in order to retain them. Retaining existing customers is a key approach for these businesses.

Telecommunication industry over the past decade has been on a constant upward trend when it comes to invention and innovation. From landlines to fibre optic cable to satellite internet and mobile connection improving its technology by one generation every year. Only way to survive such a harsh and competitive industry is by identifying the needs of the customers and catering for their best needs. Businesses need this strategy to hold their market share and remain competitive in the industry.

For this project, our objective is to determine the factors affecting churn of customers and provide solutions to retain them. To analyse these factors in the industry, we have gathered quarterly churn update data for Q3 of a California based telecommunication company. The initial dataset includes 7043 unique customers with 21 variables.

## **Dataset 1**

|  |  |
| --- | --- |
| **Observations** | **Descriptions** |
| 1. CustomerID | A unique ID that identifies each customer. |
| 1. Gender | The customer’s gender: Male, Female |
| 1. Senior Citizen | Indicates if the customer is 65 or older: Yes, No |
| 1. Partner | Indicate if the customer has a partner: Yes, No |
| 1. Dependents: | Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc. |
| 1. Tenure Months | Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above. |
| 1. Phone Service | Indicates if the customer subscribes to home phone service with the company: Yes, No |
| 1. Multiple Lines | Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No |
| 1. Internet Service | Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable. |
| 1. Online Security | Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No |
| 1. Online Backup | Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No |
| 1. Device Protection | Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No |
| 1. Tech Support: | Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No |
| 1. Streaming TV | Indicates if the customer uses their Internet service to stream television programming from a third-party provider: Yes, No. The company does not charge an additional fee for this service. |
| 1. Streaming Movies | Indicates if the customer uses their Internet service to stream movies from a third-party provider: Yes, No. The company does not charge an additional fee for this service. |
| 1. Contract | Indicates the customer’s current contract type: Month-to-Month, One Year, Two Year. |
| 1. Paperless Billing | Indicates if the customer has chosen paperless billing: Yes, No |
| 1. Payment Method | Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check |
| 1. Monthly Charge | Indicates the customer’s current total monthly charge for all their services from the company. |
| 1. Total Charges | Indicates the customer’s total charges, calculated to the end of the quarter specified above |
| 1. Churn Label | Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value. |

In relevance to the initial data obtained, and to enhance the result of our analysis we collected further information on the customers through our second dataset. The second set of data further elaborate on analysing the demographic factors such as location of the customers. Dataset also provides customers' opinions on why they decided to leave the company.

## **Dataset 2**

|  |  |
| --- | --- |
| **Observations** | **Descriptions** |
| 1. CustomerID | A unique ID that identifies each customer. |
| 1. Count | A value used in reporting/dashboarding to sum up the number of customers in a filtered set. |
| 1. Country | The country of the customer’s primary residence. |
| 1. State | The state of the customer’s primary residence. |
| 1. City | The city of the customer’s primary residence. |
| 1. Zip Code | The zip code of the customer’s primary residence. |
| 1. Lat Long | The combined latitude and longitude of the customer’s primary residence. |
| 1. Latitude | The latitude of the customer’s primary residence. |
| 1. Longitude | The longitude of the customer’s primary residence. |
| 1. Gender | The customer’s gender: Male, Female |
| 1. Senior Citizen | Indicates if the customer is 65 or older: Yes, No |
| 1. Partner | Indicate if the customer has a partner: Yes, No |
| 1. Dependents: | Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc. |
| 1. Tenure Months | Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above. |
| 1. Phone Service | Indicates if the customer subscribes to home phone service with the company: Yes, No |
| 1. Multiple Lines | Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No |
| 1. Internet Service | Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable. |
| 1. Online Security | Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No |
| 1. Online Backup | Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No |
| 1. Device Protection | Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No |
| 1. Tech Support: | Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No |
| 1. Streaming TV | Indicates if the customer uses their Internet service to stream television programming from a third-party provider: Yes, No. The company does not charge an additional fee for this service. |
| 1. Streaming Movies | Indicates if the customer uses their Internet service to stream movies from a third-party provider: Yes, No. The company does not charge an additional fee for this service. |
| 1. Contract | Indicates the customer’s current contract type: Month-to-Month, One Year, Two Year. |
| 1. Paperless Billing | Indicates if the customer has chosen paperless billing: Yes, No |
| 1. Payment Method | Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check |
| 1. Monthly Charge | Indicates the customer’s current total monthly charge for all their services from the company. |
| 1. Total Charges | Indicates the customer’s total charges, calculated to the end of the quarter specified above |
| 1. Churn Label | Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value. |
| 1. Churn Value: | 1 = the customer left the company this quarter. 0 = the customer remained with the company. Directly related to Churn Label. |
| 1. Churn Score | A value from 0-100 that is calculated using the predictive tool IBM SPSS Modeler. The model incorporates multiple factors known to cause churn. The higher the score, the more likely the customer will churn. |
| 1. CLTV | Customer Lifetime Value. A predicted CLTV is calculated using corporate formulas and existing data. The higher the value, the more valuable the customer. High value customers should be monitored for churn. |
| 1. Churn Reason | A customer’s specific reason for leaving the company. Directly related to Churn Category. |

For the analysis, both datasets will be combined using left join. The independent variable is Customer ID to merge the datasets. Using the left join technique will provide a table that contains all values from the left table and corresponding matches from the right table. When a value in the left table doesn't have a corresponding match in the right table a null value will be created.

## **Detail of the observations**

1. **Independent Variable**

The observations are based on 7043 unique customer information in the state of California for the Q3 of 2023.

1. **Demographic Variables**

The customers are categorized based on the gender, age (senior citizen or not), marital status (dependents or with partners) and location. We believe the demographic information would bring structure to our analysis in determining the factors affecting churn and provide detailed solutions. In the dataset, there are a total of 3555 Male and 3488 Female. Out of which 1142 are senior citizens (above the age of 65). The customers are scattered across 1129 cities of California.

1. **Service Features Variables**

The company provides 9 unique services for its customers including Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, and Streaming Movies. Our objective is to find out how these services affect churn both individually and collectively.

1. **Tenure and Charges Variables**

This is to analyse how the length of service tenure affects the churn and what kind of contracts the company must sell to retain the customers. We will also be analysing the impact of both monthly and total charges on the churn rate.

1. **Outcome Variable**

The outcome variable for this analysis will be the churn status. Related variables include studying the churn reasons, scores, and customer lifetime value in the dataset to determine factors resulting in customer churn and identifying solutions to retain customers.

Studying and understanding the data is a key importance of this project. This includes pre-processing of the data and exploratory data analysis. For data preprocessing we plan to use python coding. The process will involve identifying structure, data types, missing and duplicate values of the datasets.

## **Research Questions**

A company will have a sense of how much is really being lost because of the customer churn and the scale of the efforts that would be appropriate for a retention campaign (Teemu Mutanen, n.d.). Businesses often wonder what precisely those factors are. However, it is rather not simple to pinpoint where to focus.

Upon studying the data for this report, provided by a telecommunication company based in California. We have categorized the customer churn analysis into three main categories. These categories are made based on the variables in the case data, questioning how they impact the company's churn rate. The categories involve demographic variables such as genders, locations, senior citizens, dependants, and partners. Services features such as phone service, multiple lines, online security, online backup, device protection, tech support, streaming TV, and streaming movies. Length of services which is determined from variables such as tenure, contract period, and total charges. The analysis in this report aims to answer how all of these variables impact customer churn rate.

## **Research Objectives**

The first objective of this study is to identify demographic segments that are more likely to leave the telco company. Demographic factors such as gender, age, dependent, partners and location would significantly influence customer behavior and satisfaction. Identifying the churn rate across the demographic segments will help to pinpoint the specific groups that are at higher risk to leave. This would be crucial information for tailoring retention effects, and it also could help in creating effective marketing campaigns that can resonate with the needs and preference of these segments (Oghojafor et al., 2012).

The second objective is to measure the impact of subscribing to additional services on the likelihood of churn. In this study, the telco company that is being analyzed offers a variety of services including internet, phone, and television packages. It is important to understand how the adoption of these services may affect churn and it can provide valuable information into customer satisfaction and utilization. For example, customers who subscribed to bundle services may have different churn behaviors compared to the customers who use the standalone services. Analyzing this pattern can identify the opportunities to enhance service offerings and promote service packages that can increase customer loyalty.

The third objective is to identify if longer-tenured customers are more or less likely to churn and identify the points in their tenure where churn risk increases or decreases. In churn analysis, customer tenure or the length of time a customer has been with the company is a critical factor (Huang et al., 2012). Customers who have stayed long in the subscriptions are less likely to churn as he has established loyalty and satisfaction. However, there might be a period in the customer’s lifecycle that the risk of churn spikes. Analyzing the churn rate over different tenure periods may help to develop strategies to address these critical points and increase customer retention.

As the consultant service company that is assisting the Telco company to analyze their data to minimize their churning rate, we have set that our main priority is to solve the created research questions to obtain an in-depth insight regarding the relationship of customer churning. The research questions will provide a clear direction for the researchers to focus on, thus efficiently and effectively understanding which variables majorly impacts the rate of churning and allowing the Telco company to be aware of it. The 3 research questions and its objectives are listed below:

RQ 1: How do demographic factors (age, gender, geographic location, marital status) influence customer churn rates?

RO 1: To identify which demographic segments are more likely to leave the company and provide recommendations for targeted retention strategies

RQ 2: How do additional service features (Online Security, Online Backup, Device Protection, Tech Support) affect customer retention?

RO 2: To evaluate the impact of subscribing to additional services on the likelihood of churn and suggest enhancements or marketing strategies to promote these services.

RQ 3: What is the relationship between the length of service tenure and customer churn?

RO 3: To determine if longer-tenured customers are more or less likely to churn and at what point in their tenure churn risk increases or decreases.

Before the analysis, we are going to explain all the business-related data and the main response variable that is in the dataset, however only specific variables will be chosen to be applied in the analysis as some variable does not have any significant impact on the response variable. The dataset is collected from the Telco’s company records which consists of details of 7043 customers from California. The dataset is a structured arrangement of qualitative and quantitative variables. Following that, the response variable that is kept an eye on is the churn label/value variable which explains whether the customer stays or leaves.

Since the main aim for Telco approaching us is to understand the relationship between their variables and the response variable, therefore the best analysis approach to solve this issue is by executing the Exploratory Data Analysis (EDA), as this analysis will help us to understand the patterns and relationships in the data. Before the EDA, we have to clean the data to ensure there are no missing or duplicate variables and values. In this dataset, the churn reason variable has some missing values, but it would not impact the analysis as it will not be applied in the EDA. After cleaning, the dataset is approved for the analysis.

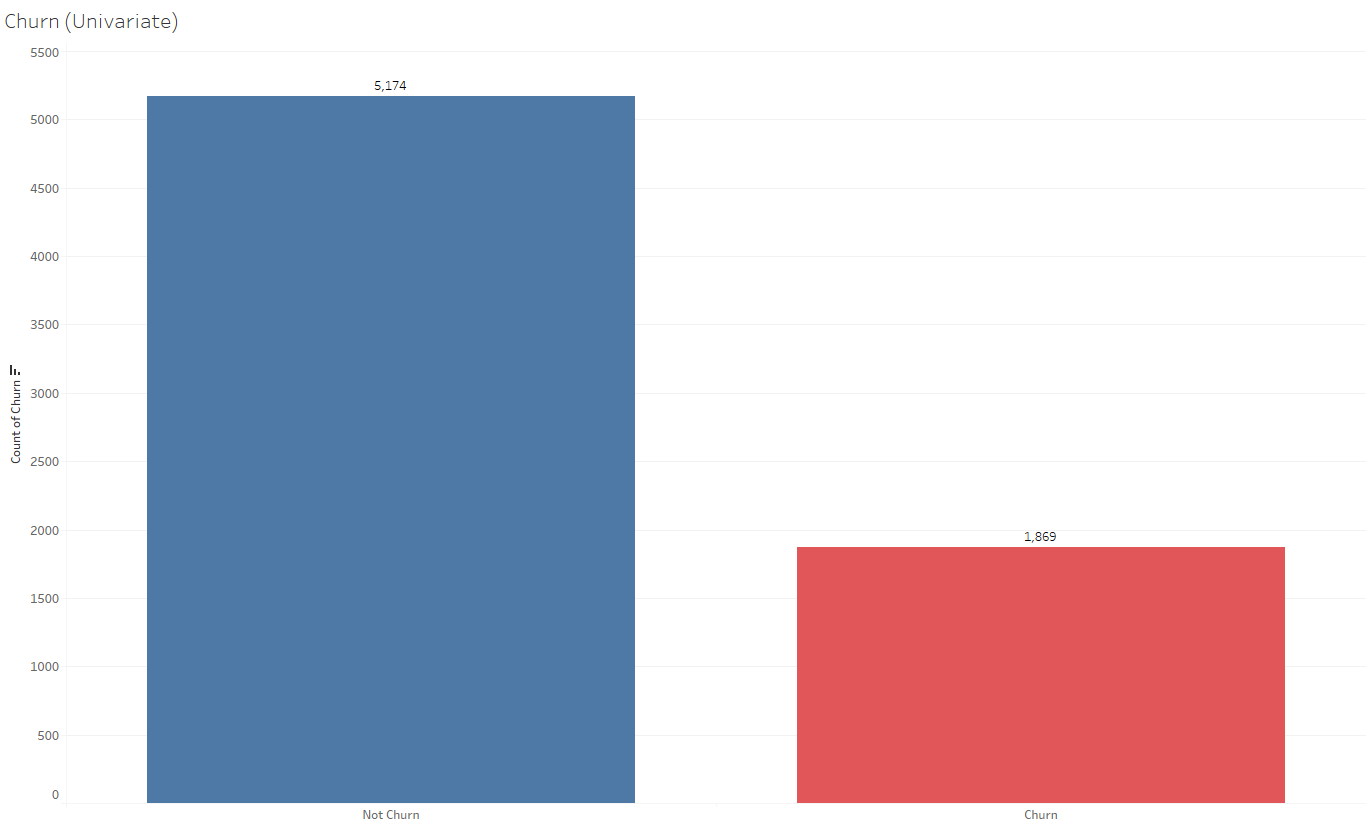
# **Section 2: Analysis, Storytelling and Visualization**

**Exploratory Data Analysis (EDA)**

The project's exploratory data analysis (EDA) is about identifying the relationship between other variables and the outcome variable (Churn). This process will help delve deeper into the data to see what it seems to say. EDA will include correlation matrix for the numeric data types and separate graphs for each variable.

### **Univariate Analysis**

Univariate analysis is a useful tool to visualize an initial summary of a given variable’s response pattern. In this case, Churn is the main variable that is being analyzed, as the overall objective is to understand the churn rate of the Telco company.

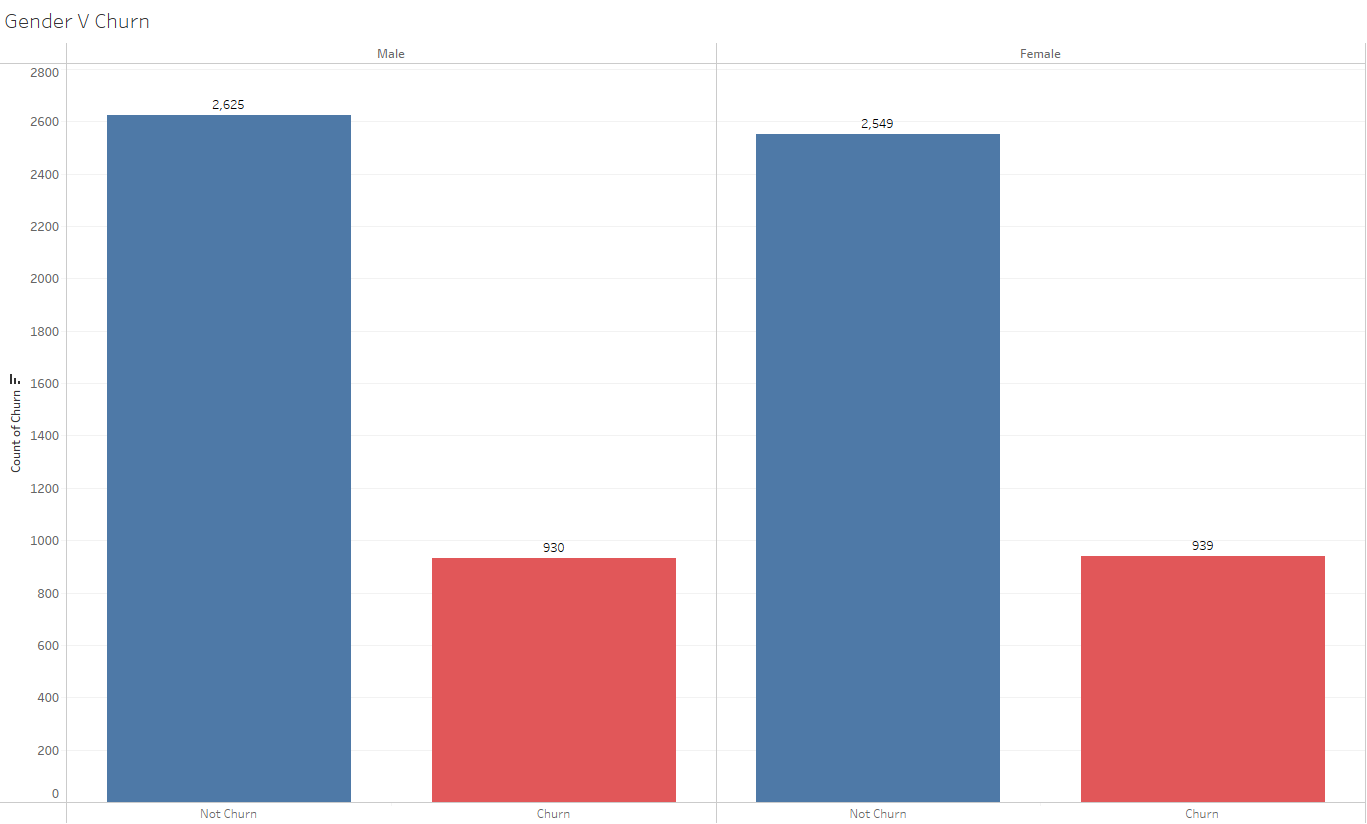


**Figure 1 – Customer Churn**

Figure 1 is a bar chart that indicates whether customers have churned or not, blue representing customers that did not churn (stayed), while red representing customers that churned. A bar chart was chosen for visualization as it provides a clear representation of the proportion of customers that left and stayed. It enables the two categories to be easily differentiable and compared, allowing the viewer to understand the objective of this diagram, which is to identify the number of customers that churned. At initial glance, it may be observed that the number of customers that churned amounted to 1,869 customers compared to the 5,174 customers that stayed. Based on this observation, it may be identified that the number of customers that churned from this Telco company is substantial, amounting to 26.54% of their overall customer base.

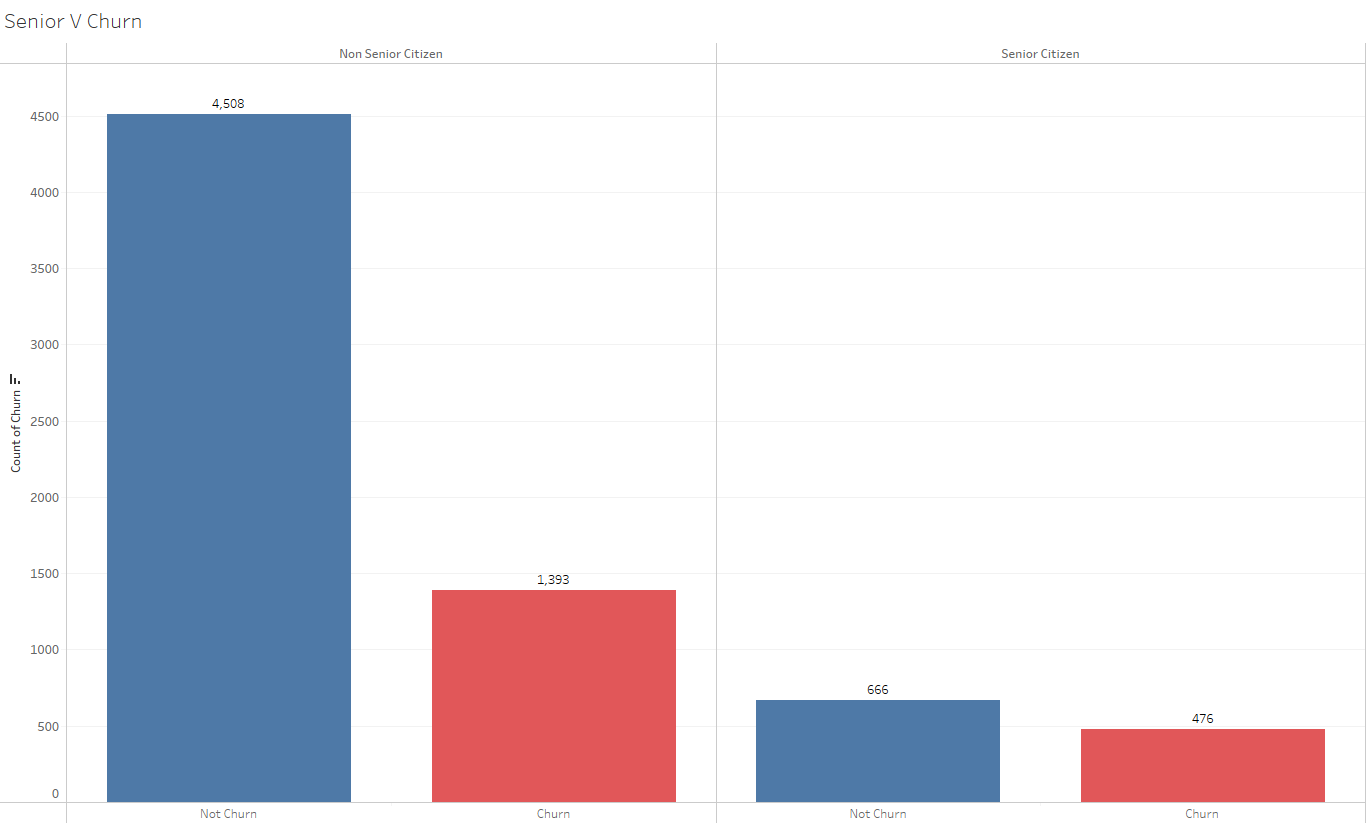
### **Bivariate Analysis**

The factors that affect the outcome of customers churning or not may be further analyzed with the use of bivariate analysis. Bivariate analysis is an effective method of determining how independent variables affect the dependent variable, in this case, the churn being the dependent variable. For the variables ‘Gender’, ‘Senior Citizen’, ‘Partner’, and ‘Dependents’, the bivariate analysis is visualized in the form of bar charts. Bar charts are most suitable for this kind of binary data due to its discrete nature (eg. “Yes” or “No”). Similarly to the univariate analysis bar chart, the churn labels are represented by the colors blue and red to indicate churn and not churn respectively. On the other hand, for the variables ‘Tenure’, ‘Monthly Charges’, ‘Total Charges’, and ‘Churn Score’, line charts are utilized. Line charts are most suitable for these variables that output continuous data, allowing the viewer to identify trends over a certain measurement, such as a given time period or price ranges. Similarly to the bar chart, the churn labels are represented by the colors blue and red, indicating churn and not churn respectively.



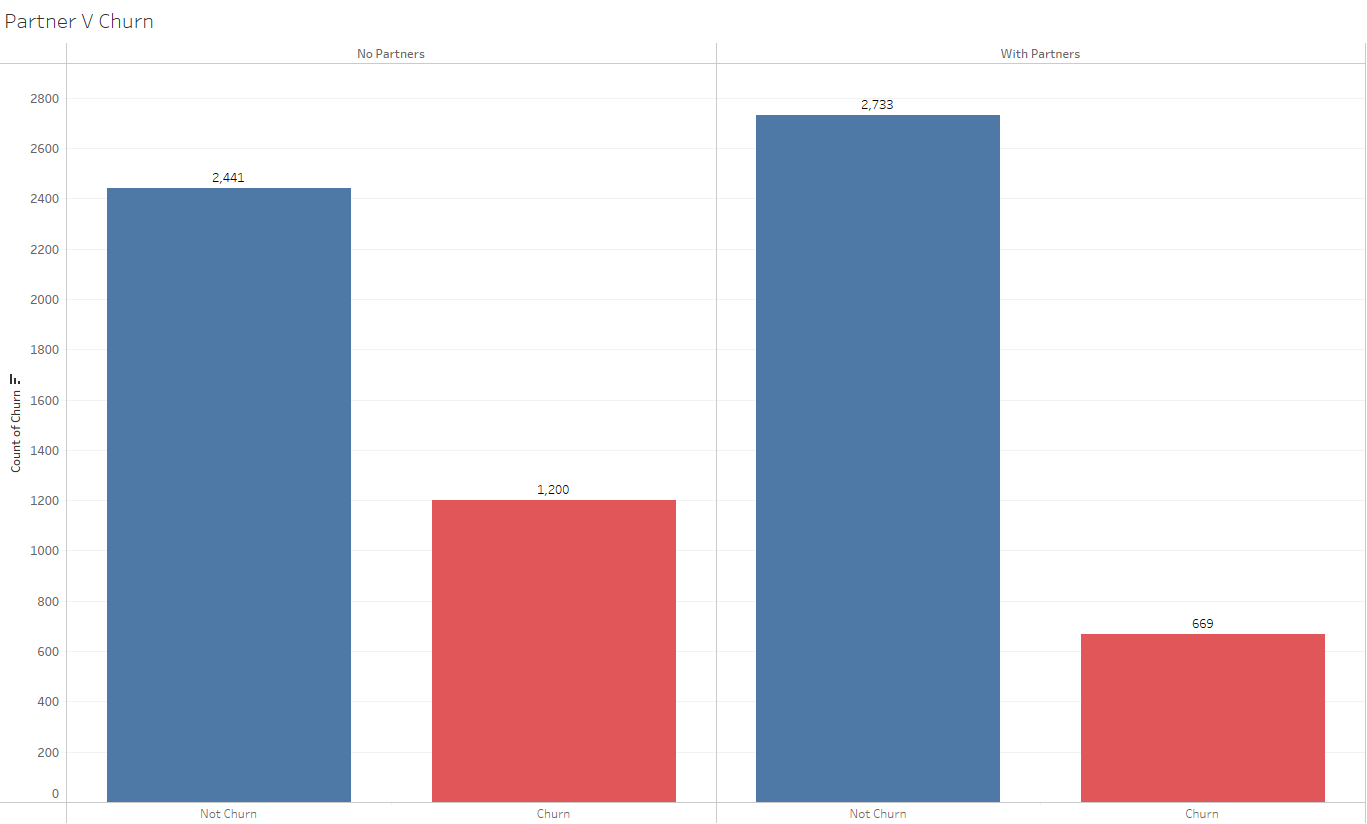
**Figure 2 – Gender vs Churn**

Figure 2 is a bar chart that indicates the distribution of customers that churn and did not churn separated by genders, Male and Female. From the visualization, it may be noted that there is a relatively equal number of male and female customers (3555 males and 3488 females). Furthermore, it may be observed that 2,625 males stayed and 930 males churned, while 2,549 females stayed and 939 females churned. Consequently, the customers that are male and churned make up 13.2% of the total customer base, while the customers that are female and churned make up 13.33%.



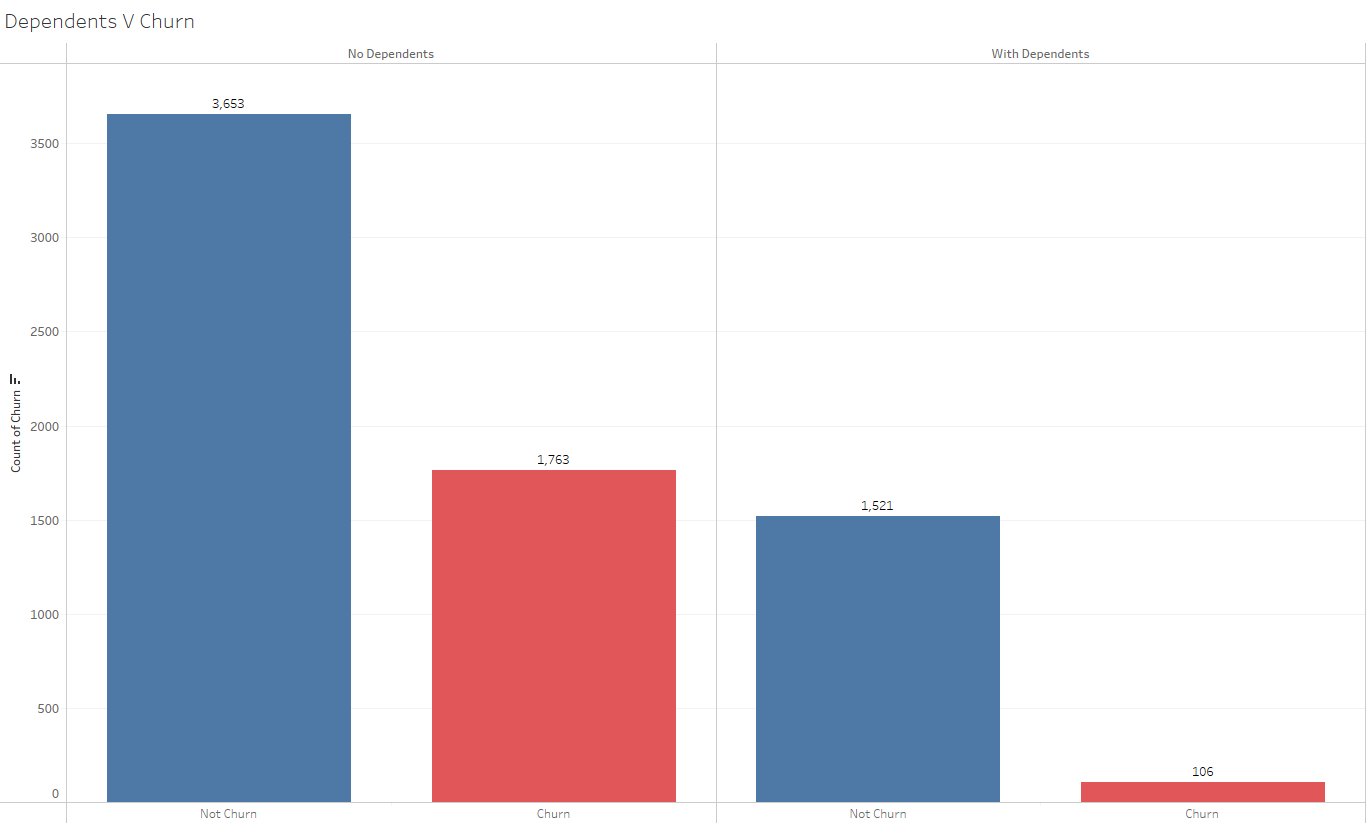
**Figure 3 – Senior vs Churn**

Figure 3 is a bar chart that indicates the distribution of customers that stayed and churned based on whether they are senior citizens or not. Upon first sight, it may be observed that the total number of customers that are senior citizens is significantly less than customers that are not senior citizens (1,142 customers are senior citizens, while 5,901 customers are not). Among the customers that are senior citizens, 666 of them stayed while 476 churned. Meanwhile, among the non-senior citizens, 4,508 of them stayed while 1,393 churned. Evidently, the customers that are not senior citizens and churned make up 19.78% of the total customer base, while customers that are senior citizens and churned only make up 6.76%. Overall, the difference between churned and not churned customers is far more significant among the non-senior citizens compared to the senior citizens.



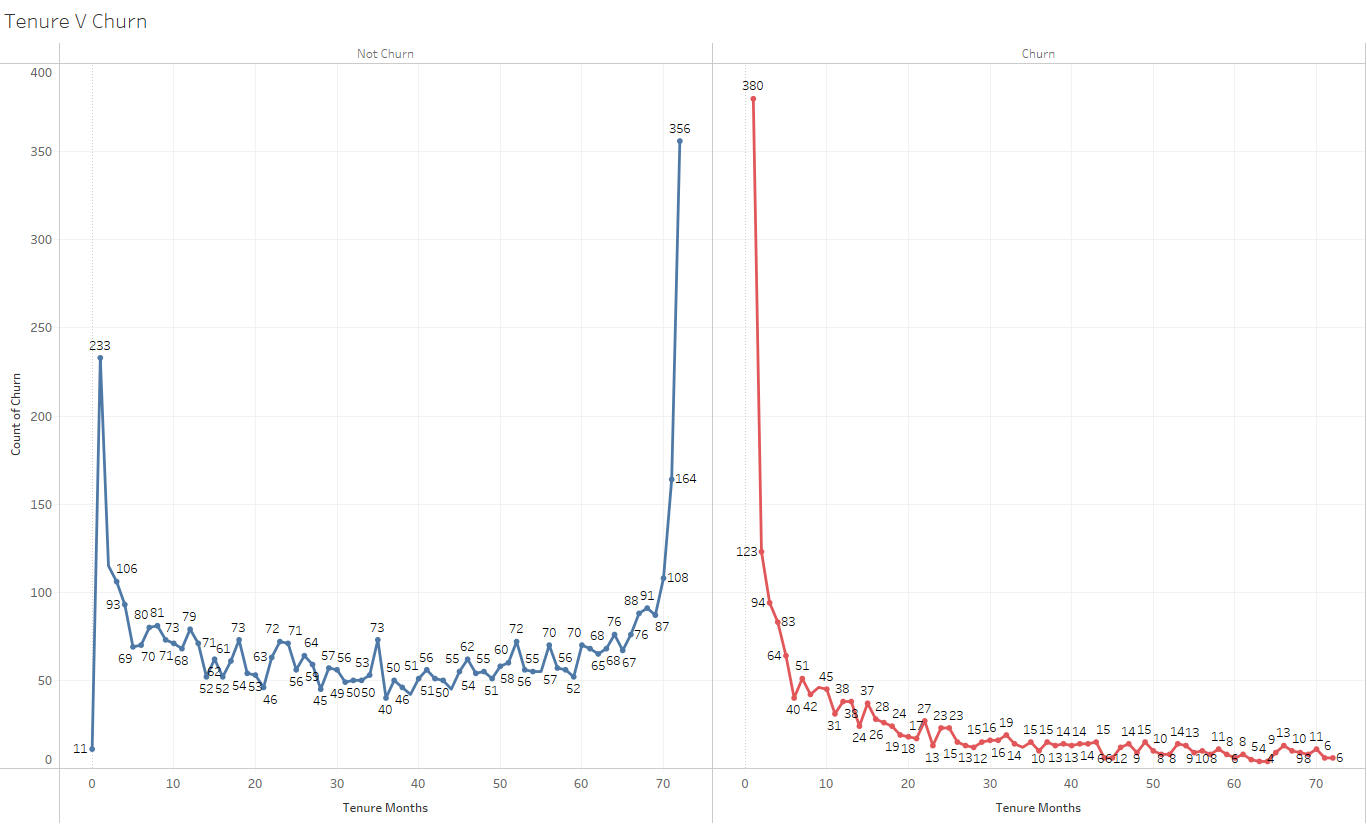
**Figure 4 – Partner vs Churn**

Figure 4 is a bar chart that represents the distribution of customers that stayed and churned based on if they have a partner or not. Based on the visual, it may be noted that there is a relatively equal number of customers that have partners and customers that do not have partners (3,555 customers have partners, while 3,488 customers do not). Among the customers who do not have a partner, 2,441 of them stayed, while 1,200 churned. On the other hand, among the customers with a partner, 2,733 of them stayed, while 669 churned. As a result, the customers that do not have a partner and churned make up 17.04% of the total customer base, while customers that do have a partner and churned make up 9.50%.



**Figure 5 – Dependents vs Churn**

Figure 5 is a bar chart that displays the distribution of customers that churned and stayed based on if they have dependents or not. From this visual, it is noticeable that the total number of customers with dependents is significantly larger than customers without dependents (5,416 with dependents and 1,627 without dependents). Among the customers with dependents, 3,653 of them stayed, while 1,763 of them churned. On the other hand, among the customers without dependents, 1,521 of them stayed, while 106 of them churned. As a result, the customers without dependents and churned make up 19.78% of the total customer base, while the customers with dependents and churned make up 6.76%.

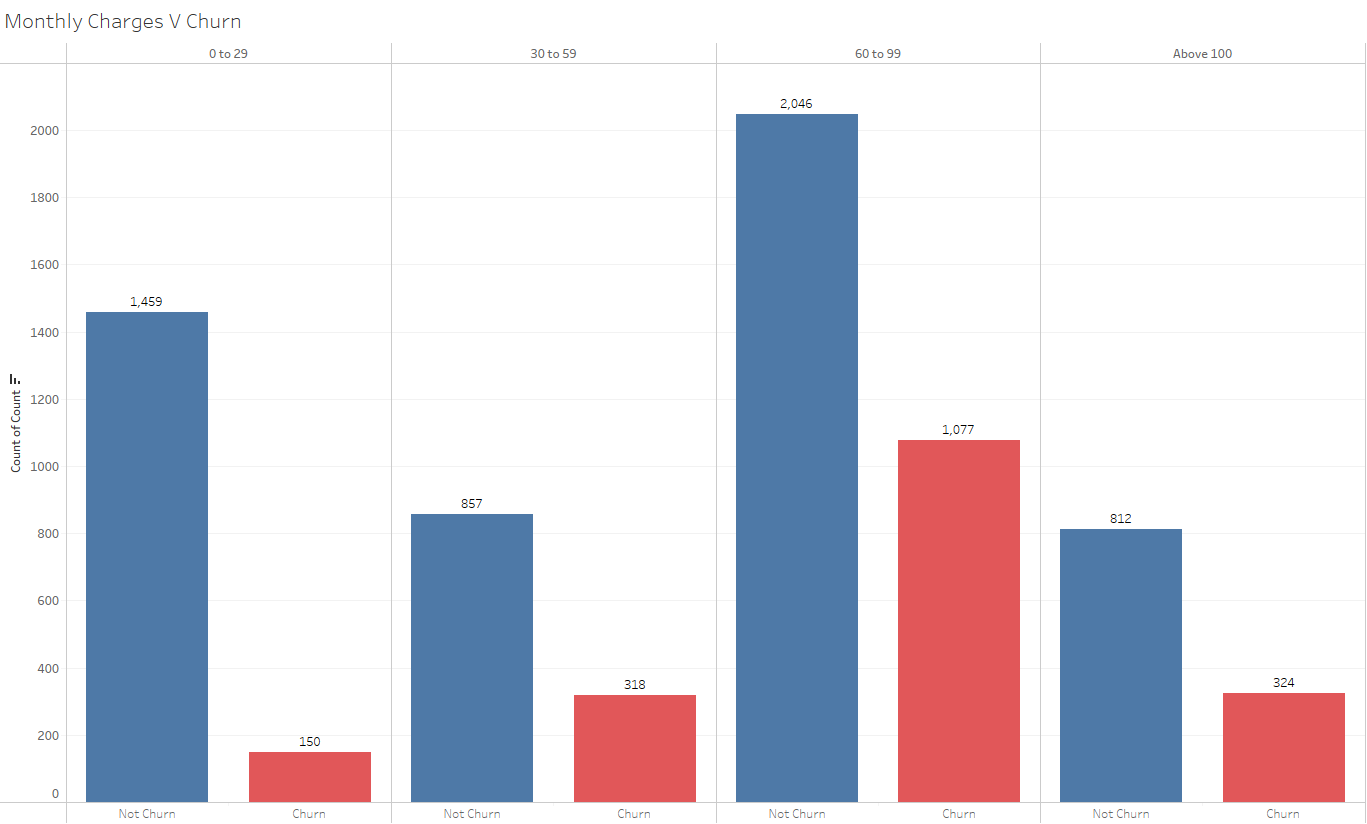


**Figure 6 – Tenure vs Churn**

Figure 6 is a line chart that counts the churn label of customers based on their tenure months. Observing the “Not Churn” line chart, there is a spike around the 1-month mark, where 233 customers are labelled that they stayed. This spike may indicate a large number of new customers and have not churned after a short period of time. After the peak, the churn label count stabilizes with some minor fluctuations. Around the 70-month mark, the number of staying customers start to increase significantly, achieving a peak at the 72-month mark (356 customers), indicating the long-term customers.

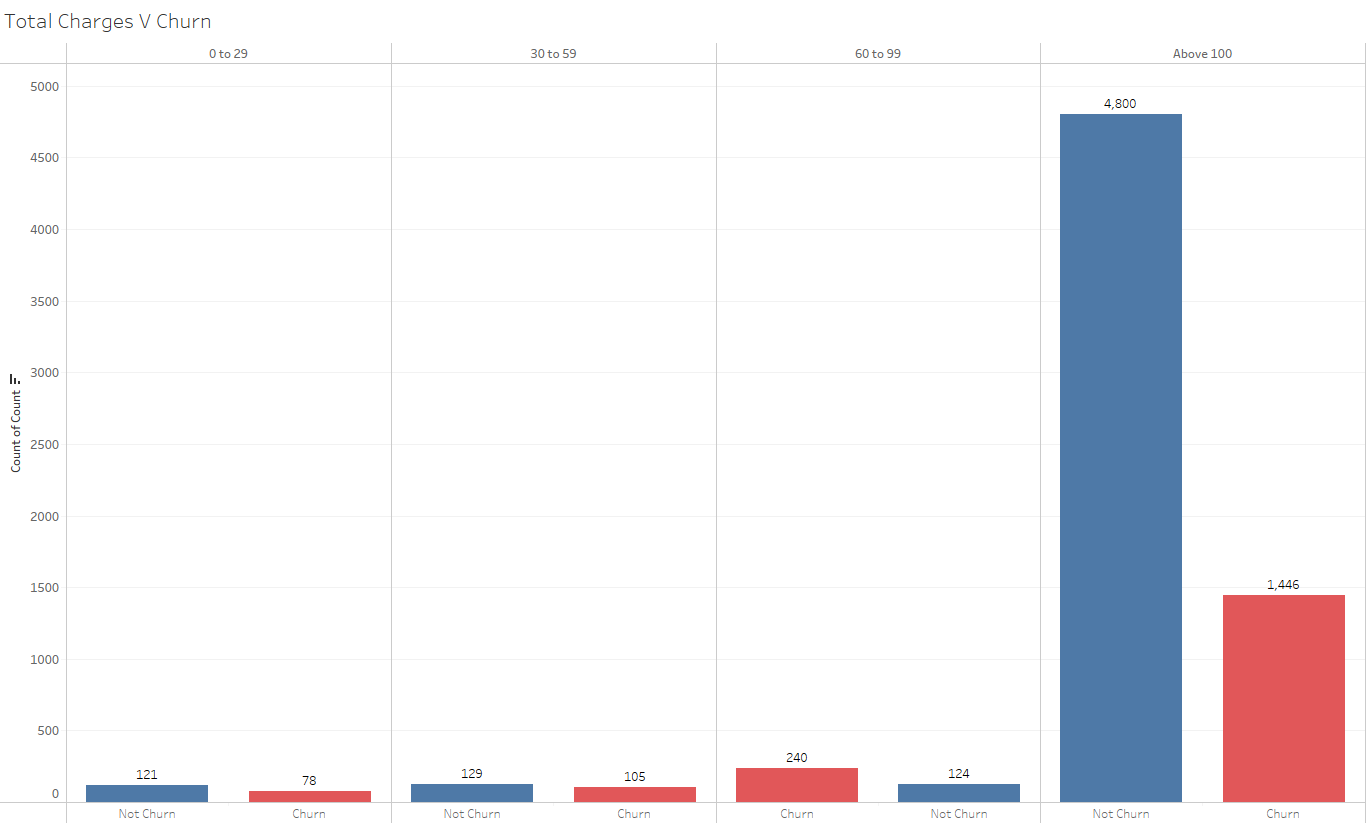
Observing the “Churn” line chart, there is an early peak at the 1-month mark, indicating 380 new customers that churned after a very short period of time. Following this peak, there is a significant drop in customers that churned, leading to a more stable number of customers that churned with some minor fluctuations.

Apart from the spike in the “Not Churn” line chart, it may be observed that the “Not Churn” and “Churn” line charts possess a relatively inverse relationship, meaning when the churn count is higher for one chart, it will likely be lower for the other. This is evident in the diagram above, for example at the 72-month mark, the “Not Churn” line chart indicates 356 customers that stayed, while the “Churn” line chart indicates 6 customers that churned. This inverse relationship may further be explored at the 1-month mark, where 233 customers stayed and 380 customers churned.



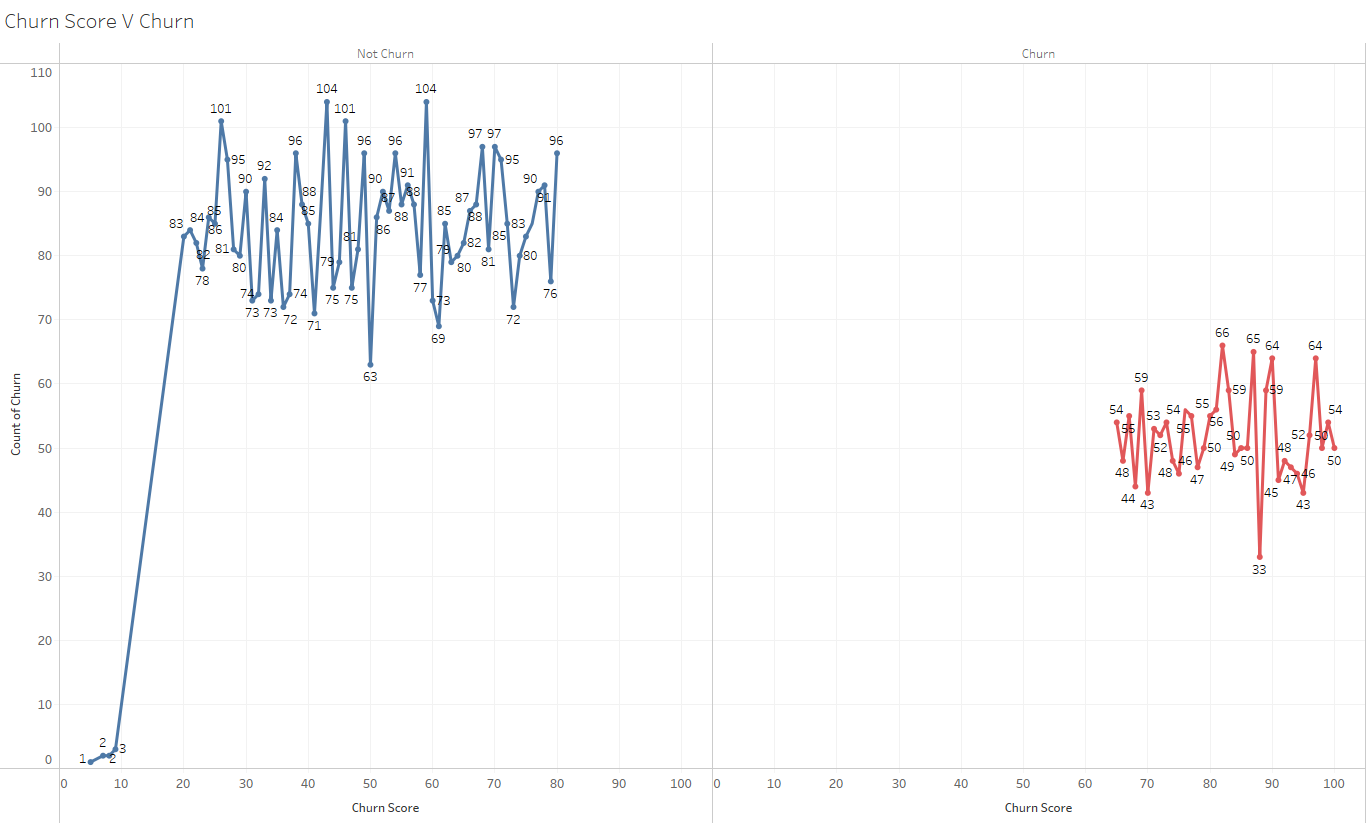
**Figure 7 – Monthly Charges vs Churn**

Figure 7 is a bar chart that counts the churn label on customers based on how much they are charged monthly. The monthly charges have been clustered into different price ranges, “0 to 29”, “30 to 59”, “60 to 99”, and “above 100”. The monthly charges have been clustered in this way, as it provides a better understanding of how monthly charges in a certain range may affect the churn. Furthermore, it is visually appealing and relatively easier to comprehend. By observing this clustered bar chart, it may be noticed that the majority of customers that both stayed and churned are within the monthly charge range of $60 to $99. Within this monthly charge range, 2,046 customers stayed, while 1,077 churned. Observing the remaining clusters, the number of overall customers is relatively distributed, with a noticeable peak of staying customers in the “0 to 29” cluster.



**Figure 8 – Total Charges vs Churn**

Figure 8 is a bar chart that counts the churn labels of customers based on how much they have been charged overall during their tenure. The total charges have been clustered into different ranges, “0 to 29”, “30 to 59”, “60 to 99”, and “above 100”. As observed in the bar chart, the majority of overall customers are within the “above 100” cluster, meaning their total charges during their tenure is $100 and above. Within the “above 100” cluster, 4,800 customers stayed, while 1,446 churned. Looking at the remaining clusters, it is noticeable that there is a relatively low number of customers overall compared to the “above 100” cluster (the number of overall customers does not exceed 364 in each cluster).



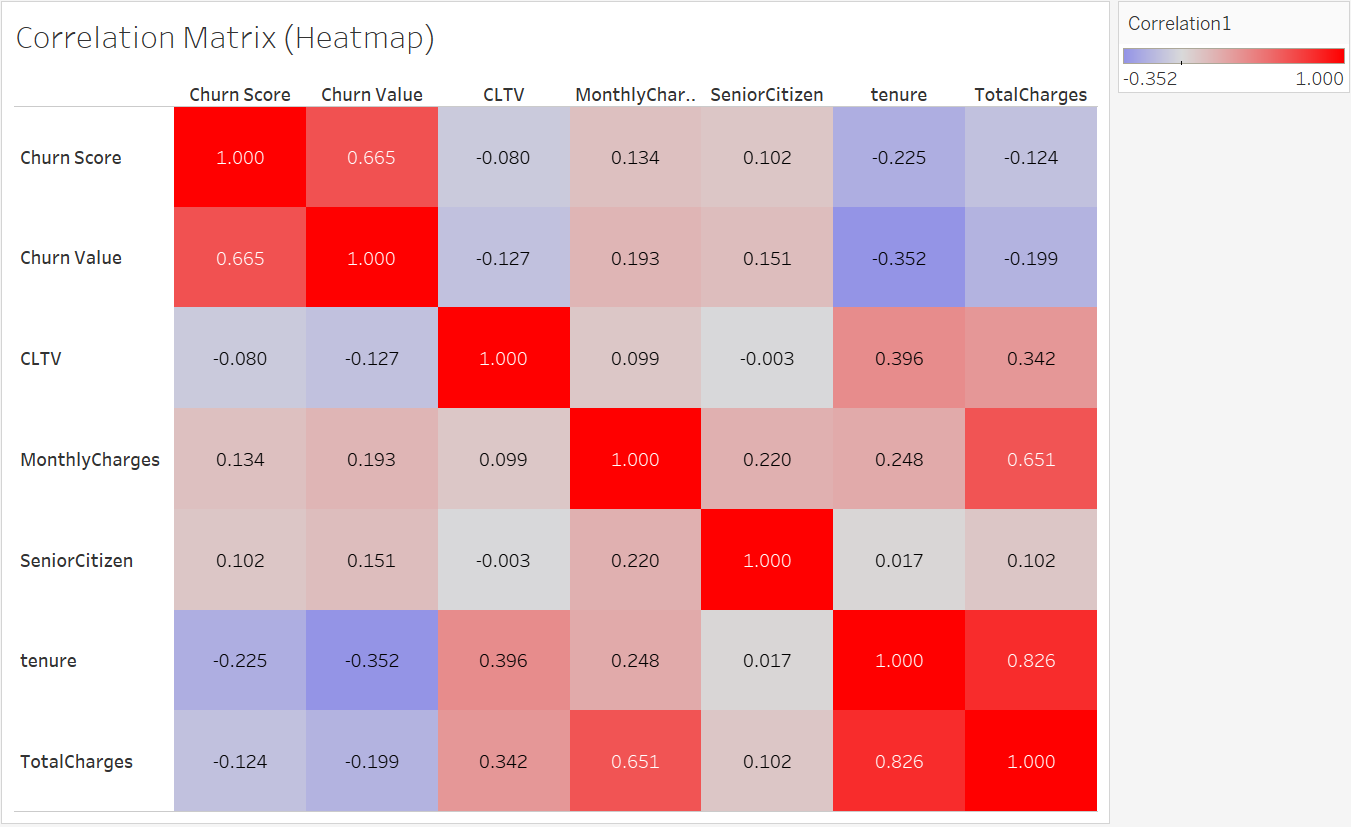
**Figure 9 – Churn Score vs Churn**

Figure 9 is a line chart that counts the churn label of customers based on churn score. Observing the “Not Churn” line chart, the line starts near the origin (0,0), indicating a low count of customers that stayed with a low churn score. The line sharply rises toward the churn score of 20, which precedes a significant amount of fluctuation in the count of customers that stayed as the churn score increases. On the other hand, observing the “Churn” line chart, it can be seen that the line does not start until the churn score is 65, which precedes a significant number of fluctuations in the count of customers that churned.

### **Multivariate Analysis**

Multivariate analysis provides a comprehensive understanding of complex data involving multiple variables as this approach is able to simultaneously observe and analyze more than 1 outcome variable. This analysis is the easiest yet most effective visuals to explain to the readers regarding the complex relationships and interactions between variables.

### **Correlation Matrix**

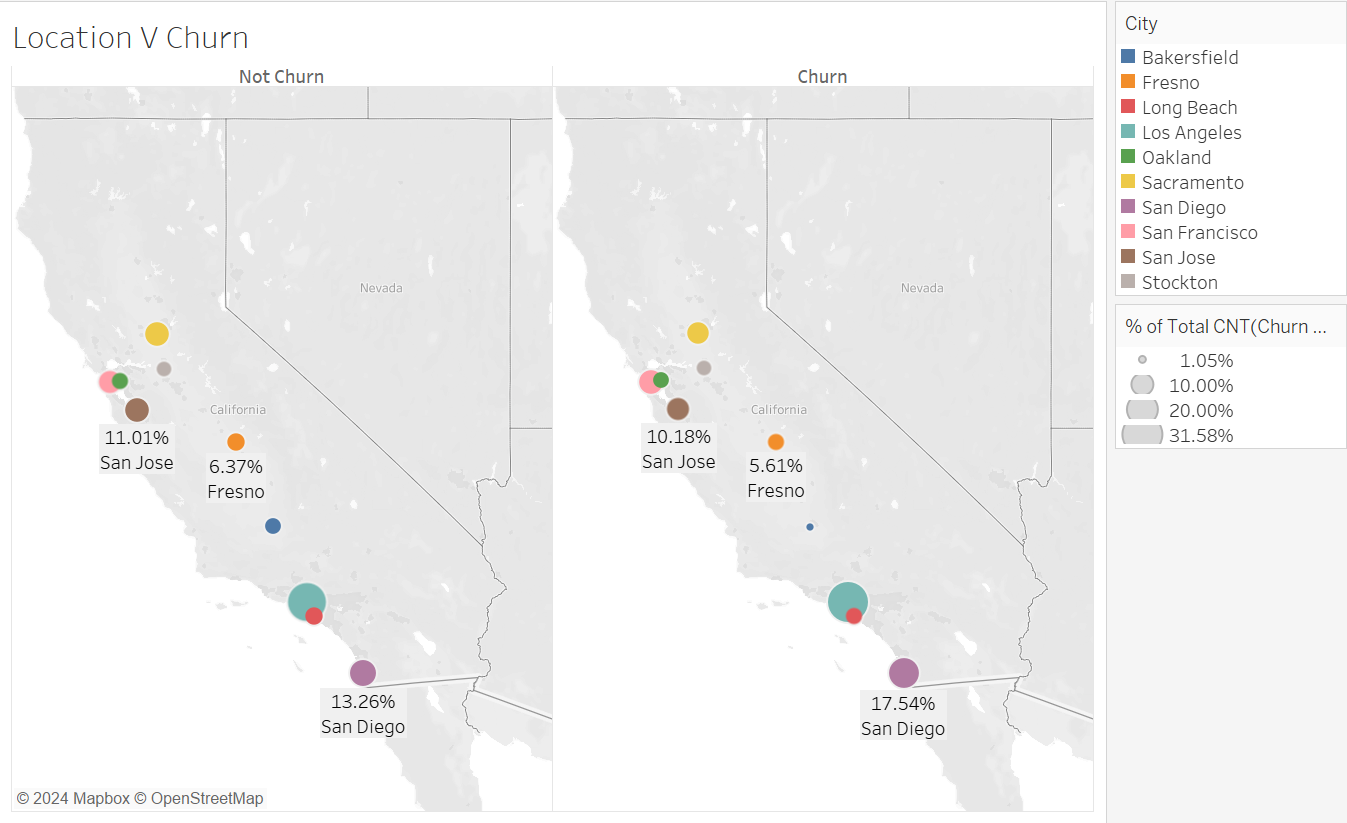


**Figure 10 – Correlation Matrix**

Figure 10 shows the correlation matrix between each variable in the Telco’s dataset. The heat map approach is chosen specifically for this scenario as this approach excels when we need to easily convey the message regarding the relationship between multiple variables. This approach effectively helps the researcher to simultaneously compare many variables as it is able to display a large dataset in a compact yet efficient format. The correlation value ranges between 1 to -1 and this value indicates the strength of the linear relationship between respective variables. For example, 1 explains that there is a positive linear relationship while -1 explains there is a negative linear relationship. Moreover, this diagram also uses a few varieties of colours which are blue to red, as this will indicate to any readers that red means a high correlation and reliable predictor to churning while blue means a weak correlation and reliable predictor to churning. With this, readers can purely understand through the visuals as it has been instilled that red means we must be more aware while blue can be taken less lightly.

From Figure 10, ‘Churn Value’ has a positive correlation with ‘Churn Score’ (0.665), ‘MonthlyCharges’ (0.193) and ‘SeniorCitizen’ (0.151) respectively. This means, as the variable number increases, the likelihood that churn rate increases and, in this case, the churn rate is increased the most with the ‘Churn Score’ (0.665), then by ‘MonthlyCharges’ (0.193) and followed by the ‘SeniorCitizen’ (0.151). On the other hand, the ‘Churn Value’ has a negative correlation with ‘tenure’ (-0.352), ‘TotalCharges’ (-0.199) and ‘CLTV’ (-0.127). While in this scenario, as the variable number decreases, the chances of churn will decrease so this shows that the churn rate decreases the most with ‘tenure’ (-0.352), then by ‘TotalCharges’ (-0.199) and lastly by ‘CLTV’ (-0.127). As a result, ‘Churn Score’ (0.665) is the strongest predictor for the Telco company churn rate, therefore they must be aware of this information. Besides that, they can conclude that the longer the customer is with them, then they are least likely to leave, therefore Telco has to create some strategies to ensure their customers stay with them for a long time. 

### **Location against Churn**

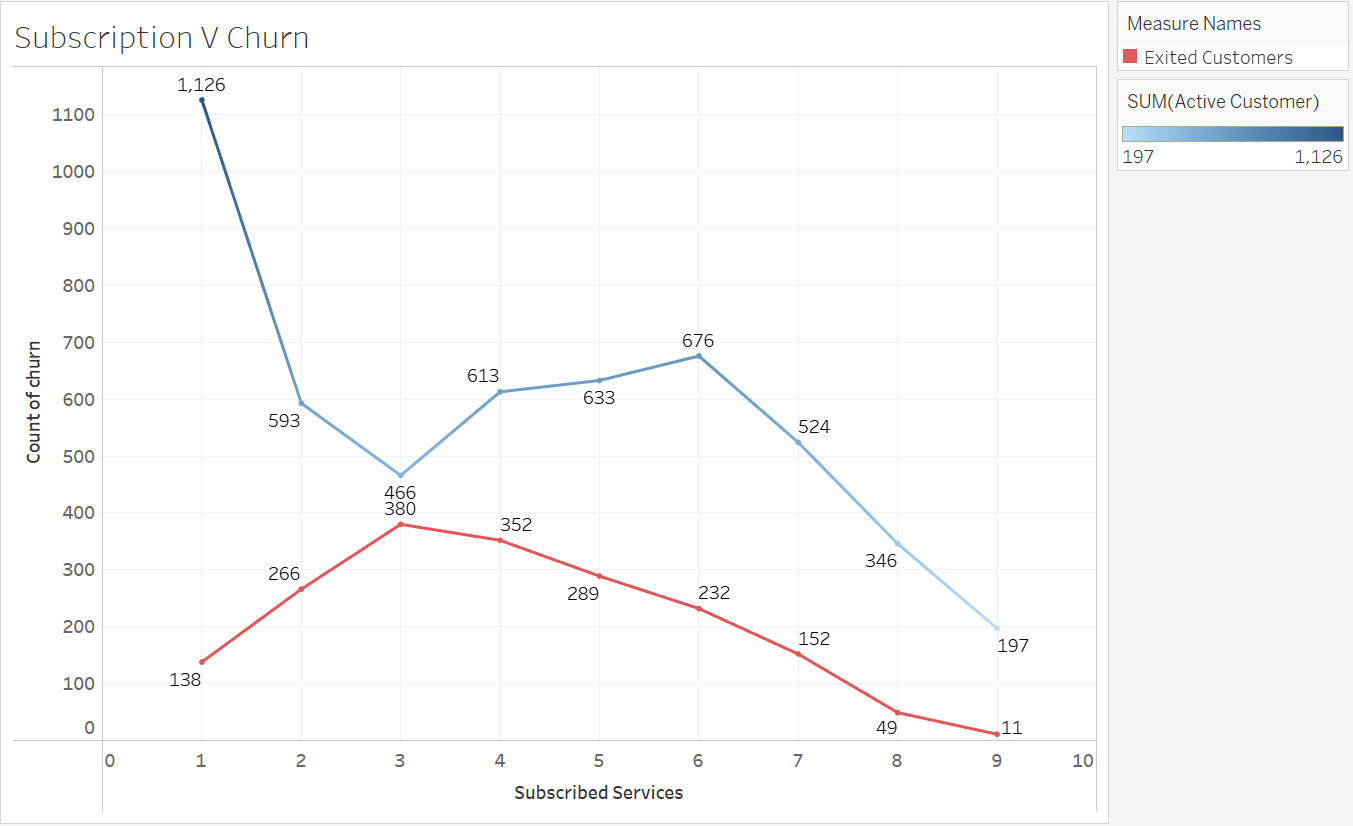


**Figure 11 – Location vs Churn**

Figure 11 shows the relationship between the location of the customers with the churn label/value. The suitable graph to represent the data is the symbol maps as this approach provides a clear location based visual representation of the points on a geographical map. This will help the reader to understand how the data varies across different locations. Moreover, each location is represented by their own respective colours to make it easier for analysis and prevent any confusions. If all the locations are in the same colour, then it would surely still be confusing to read even if there are labels. Lastly, the size of each location is based on the count of churn labels/rates. For example, the more the churn rate in the location, then the bigger the circle size. All of the additional assistance in visualization is included to ensure an easy understanding of the diagrams.

From Figure 11, we can see that the location with the most churn rate goes by Los Angeles (31.58%), San Diego (17.54%), San Francisco (10.88%), San Jose (10.18%), Sacramento (9.12%), Fresno (5.61%), Long Beach (5.26%), Oakland (4.56%), Stockton (4.21%) and Bakersfield (1.05%). Meanwhile the location with the most least churn follows by Los Angeles (28.51%), San Diego (13.26%), San Jose (11.01%), Sacramento (10.88%), San Francisco (9.68%), Fresno (6.37%), Long Beach (5.97%), Oakland (5.17%), Bakersfield (4.91%) and Stockton (4.24%). From this, we can see that Telco has to increase their promotion in the top least churn locations as these places tend to bring them more long-lasting customers and investigate why the churn rate at the respective locations are very high, then overcome it.

### **Services Subscribed against Churn**



**Figure 12 – Subscription vs Churn**

Figure 12 represents the relationship between the number of subscribed services with the churn label/value. This relationship is best approached by the line chart because it is effective for displaying continuous data thus, allowing the readers to easily grasp the flow and connection of the data points. Besides that, by plotting 2-line charts into 1 graph, it will make it possible to compare data in such a way that it is easily digestible by the reader. This chart is a very simple and straightforward comparison method for continuous data which enables the readers to interpret, thus making it an effective way to communicate the data to the required audience.

The variable of ‘Subscribed Services’ from the diagram represents the aggregate of all the individual services from the dataset, which are ‘Phone Service’, ‘Multiple Lines’, ‘Internet Service’, ‘Online Security’, ‘Online Backup’, ‘Device Protection’, ‘Tech Support’, ‘Streaming TV’ and ‘Streaming Movies’. The subscribed service against churn chart shows the pattern of increase until 3 subscribed services with a peak of 380 customer churn, then a continuous decrease till the end. Following that, the subscribed service against the not churn chart reveals an inconsistent pattern at the start but predictable end result. The pattern of the not churn chart is steep decrease until 3 subscribed services, then a slight increase till 6 subscribed services and eventually a steady decrease till the end. By comparing the charts, it can be concluded that the more the number of total subscribed services the less likely for the customer to churn, however the customers with 3 subscribed services are yet the most common to churn. Therefore, Telco companies have to suggest interesting ways to attract more customers to sign up for additional services to ensure longevity in their customer’s loyalty.

# **Section 3: Discussion**

Based on the univariate analysis, visualization of the overall count of churn labels, this Telco company suffers a 26.54% churn rate, meaning that 26.54% of their entire customer base churned. This implies that 26.54% of their customer base have stopped subscribing to this Telco company’s services. As a result, the company’s overall cash flow would be negatively affected in the long run, possibly causing issues in making profit or even covering costs. Ideally, this Telco company should aim to reduce this churn rate to sustain as a business in the long run.

For telco companies to improve their churn rate, they must identify the significant variables, which may most effectively be achieved by performing bivariate analysis. Based on the bivariate analyses, both bar chart and line chart variants, performed previously, the variables that this telco company should focus on are ‘Senior Citizen’, ‘Partner’, ‘Dependent’, ‘Tenure’, ‘Monthly Charges’.

This Telco company may greatly benefit from the bivariate analyses that utilize bar charts as it allows them to gauge a better perspective on which demographic they should focus on. First of all, from the findings in the ‘Senior Citizen vs Churn’ bar chart, non-senior citizens that churned made up 19.78% of the total customer base. As this percentage is higher than the percentage of customers that are senior citizens and churned (6.76%), it may be beneficial for this Telco company to focus on retaining non-senior citizens. A suggestion for this Telco company that may help them retain non-senior citizens is to capitalize on their social media, as non-senior citizens make up a large majority of social media’s demographic. This Telco company may achieve this by utilizing various social media marketing strategies, increasing their brand awareness toward this demographic. Furthermore, they could improve their social media presence, for example, by producing engaging content catered towards this demographic. If this demographic is more aware of this Telco company’s values and what they offer, it is highly likely that more non-senior citizens would continue using their services.

From the ‘Partner vs Churn’ bar chart, the customers without a partner and churned made up 17.04% of the total customer base. Since this percentage is higher than the percentage of customers that do have partners and churned (9.5%), this Telco company should also focus on retaining customers that do not have partners.

Lastly, looking at the ‘Dependents vs Churn’ bar chart, the customers that do not have dependents and churned make up 25.03% of the total customer base. This percentage is significantly higher than the percentage of customers that do have dependents and churned (1.51%). Due to this significant difference, this Telco company has an opportunity to capitalize on this demographic. As the variables ‘Partner’ and ‘Dependent’ are closely related, some suggestion for this telco company that may assist in the retention of these demographics is to implement perks for individual users and single user pricing models. For example, this Telco company could implement a pricing model that makes it more desirable for individual users to subscribe for a longer period of time. As a result, customers that do not have a partner and no dependents are likely to continue subscribing as the cost of the services are alleviated.

In terms of the analyses using line charts, the ‘Tenure vs Churn’ line chart illustrates that the majority of customers churn within the first month, amounting to 380 customers. This information is immensely useful for this telco company as they may want to consider implementing strategies to retain new customers for longer. This telco company may be able to achieve this by improving their onboarding process. For example, this telco company could introduce a special package catered to new customers that provide various perks and essential guides on how to use their services. This package could make the new customers feel more welcomed and may sway them to continue using their services in the future.

Observing the ‘Monthly Charges vs Churn’ bar chart, it may be observed that most customers churn within the $60 to $99, as illustrated by the “60 to 90” cluster. This high monthly charge rate may be caused by various reasons, the most prominent being the number of additional services subscribed. Generally speaking, if a customer subscribes to more additional services, their monthly charges would increase, indicating a direct relationship. As this telco company suffers high churn rates at this monthly charge range, this company should provide some form of incentive for customers at this monthly charge range to stay. With additional services in mind, this telco company could give a discount or various perks for customers that subscribe to a certain number of additional services. These provided discounts and perks aim to assist in justifying why these customers should continue using their services in the long term.

From the correlation matrix analysis, we can conclude that the ‘Churn Score’ (0.665) is the strongest positive correlation and is a reliable churn predictor. Telco companies have to leverage this insight to understand what exactly affects the churn score and how they can decrease it, because the higher the churn score, the higher the churn rate. A possible suggestion that can be implemented to reduce the churn score is by increasing customer engagement as this will allow Telco to hear their customer’s voice and work accordingly. When a company listens to their customers, surely the customers will be satisfied, thus reducing the chances of them churning.

Following that, the location against churn graphs have explained that locations such as Los Angeles, San Diego and San Francisco have the most churned customers while Los Angeles, San Diego and San Jose have the least churned customers. With this information, Telco have to analyze why those specific locations are doing well while the others are not. Once figuring out the reasons, they can start to work on it. As for now, Telco can maintain what they are doing in the top least churning locations while increasing their promotion plans and customer engagement in the locations that have more churns, as this will attract more thus, giving them more insights on what the customers are demanding. By gaining a more in-depth insight, Telco is able to work on their weakness and make it into their strengths.

Moreover, referring to the charts of subscribed services against churn, we gather that the higher the number of total subscribed services, the least likely for the customers to churn. However, the highest rate of churn is mainly when the customers have 3 subscribed services. Therefore, a recommendation for Telco’s company to decrease their churn rate is by instilling a penalty to the leaving customers as this will discourage any churning behaviours. Besides that, another suggestion is by providing customers exclusive rewards as they subscribe to more services because this reward system will definitely encourage more sales for the company.

By focusing on solving the specific research questions during the analysis, we are now able to identify the specific demographic segments that are influencing the churn rate which are age, location and marital status (dependents). Besides that, we gain a deeper understanding that the more the service subscribed by the customer, the more the customer retains. Finally, we also figured out that the shorter the length of service tenure, the quicker the customer churns. As the results from the analysis, we are able to gain sufficient helpful insights regarding the relationship between all variables with the churn label/rate, which will be able to reduce Telco’s churn rate thus increasing their profit margin.

# **Section 4: Limitations**

Conducting any data analysis, it is important to acknowledge and confront the limitations that exist in the datasets that are being used. These limitations would have a significant impact on the validity and reliability of the conclusion drawn.

Some of the limitations that occurred in this study are the data quality and completeness. One of the important concerns in any datasets are the presence of missing or erroneous data. In this study, the missing values are evident especially in critical columns such as ‘TotalCharges’. When there are missing values particularly in critical columns, it can lead to bias and affect the accuracy of the analysis if it is not handled correctly. However, some of the people will opt to manually key in the data which might lead to data entry errors and can cause inaccuracies. For example, ‘TotalCharges’ which should be a numeric field might contain non-numeric entries that necessitate careful cleaning. Addressing this issue is important to ensure the integrity of the datasets and validity of subsequent analyses.

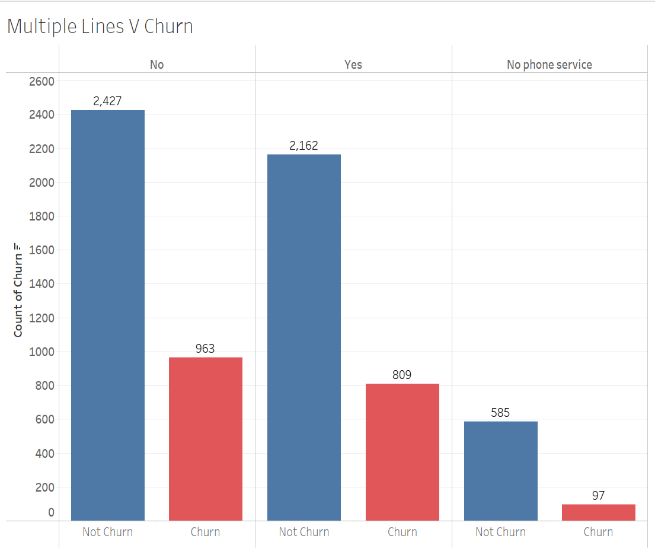
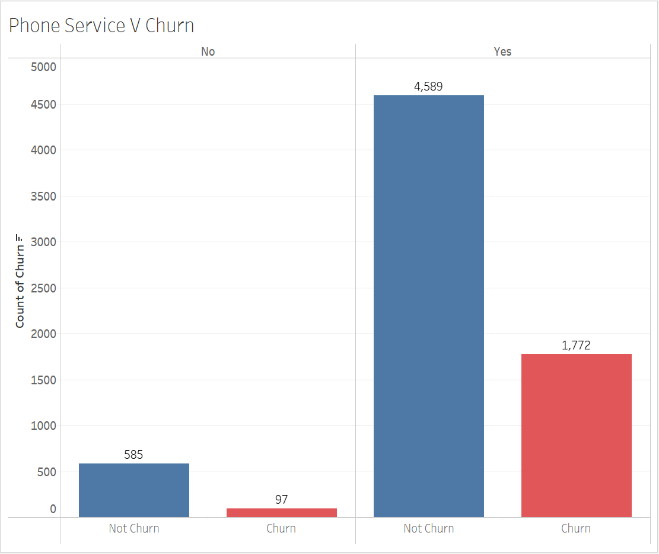
Another limitation that should be addressed is the representativeness of the sample. The datasets, which were obtained from Kaggle can be considered as the secondary data and it might be subjected to inaccuracies. The datasets may not accurately reflect the entire customer base, which can lead to skewed results. For instance, if the total number of customers stayed (not churn) are underrepresented, the analysis would not generalize well to the broader population. When conducting a data survey, it is important to have a well-balanced population as it can help in drawing conclusions that are applicable to the entire customer base. This can be done through stratified sampling or other techniques which can help to reduce bias of the sample representations.

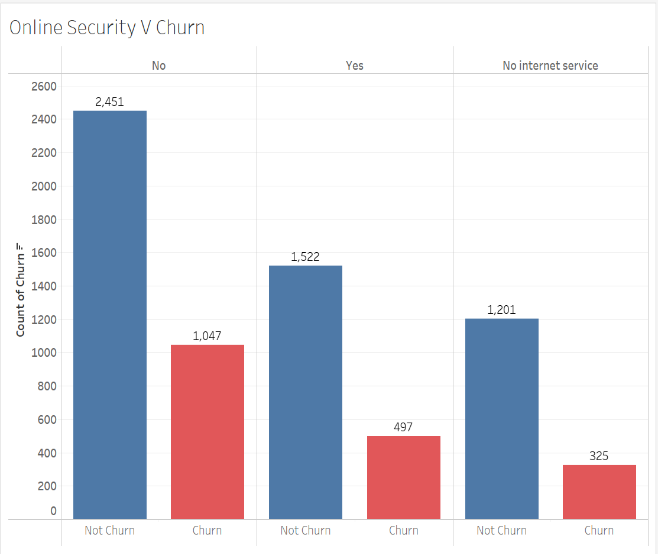
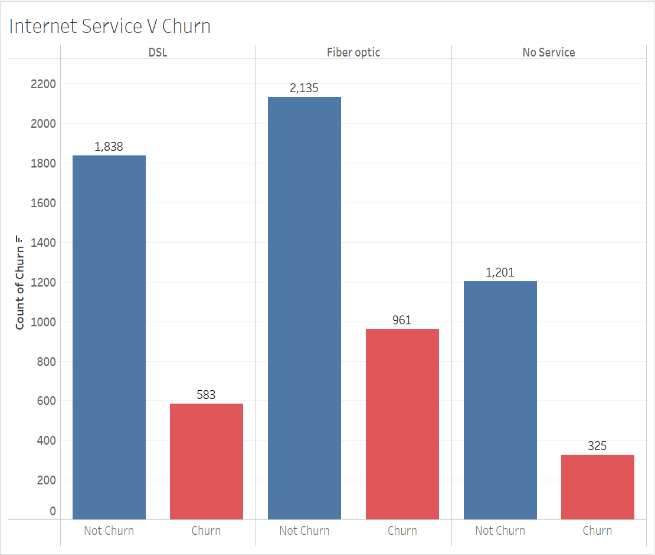
The coverage period of the datasets is another crucial consideration as it could lead to misinterpretation of the conclusions drawn. If the datasets only encompass a specific time frame, it will not capture typical customer behaviors. For example, special events, promotions or seasonal variations within the covered period may distort the analysis. This may lead to conclusions that may not be suitable in different contexts. In doing the data survey, it is important to consider the time period of the datasets and assess whether it is representative of regular business operations.

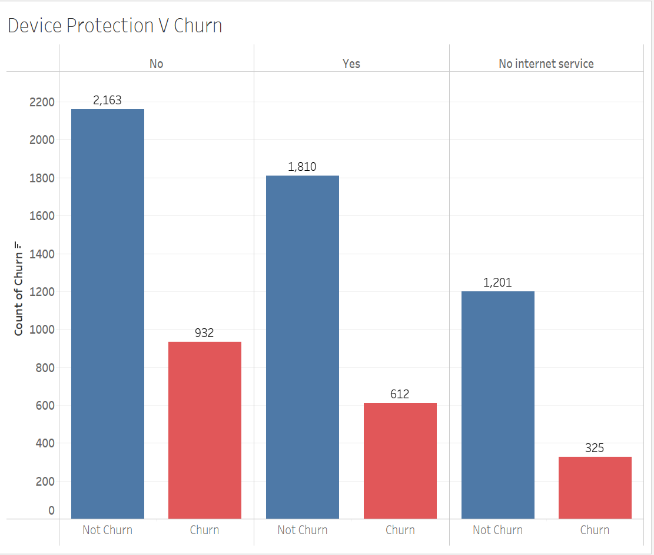
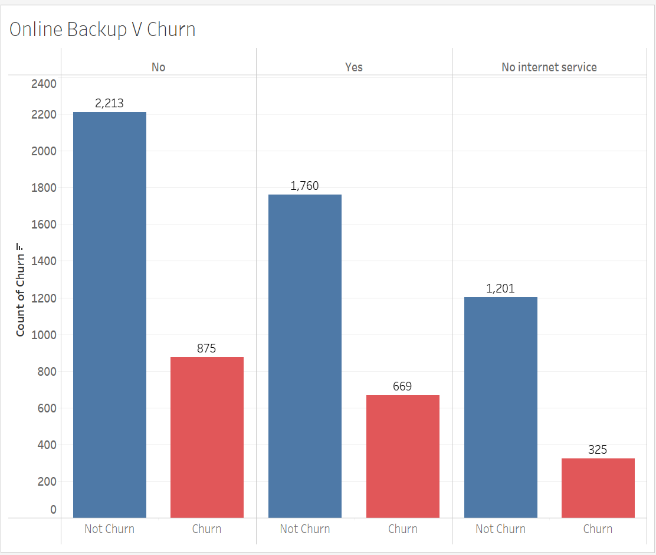
# **Section 5: Conclusion**

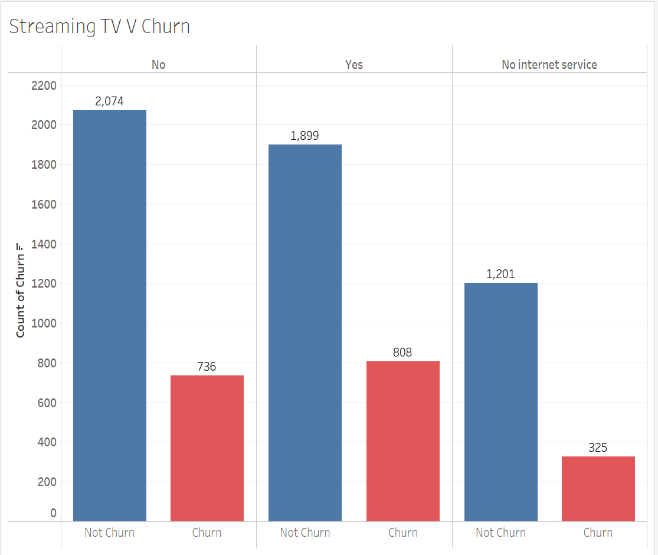
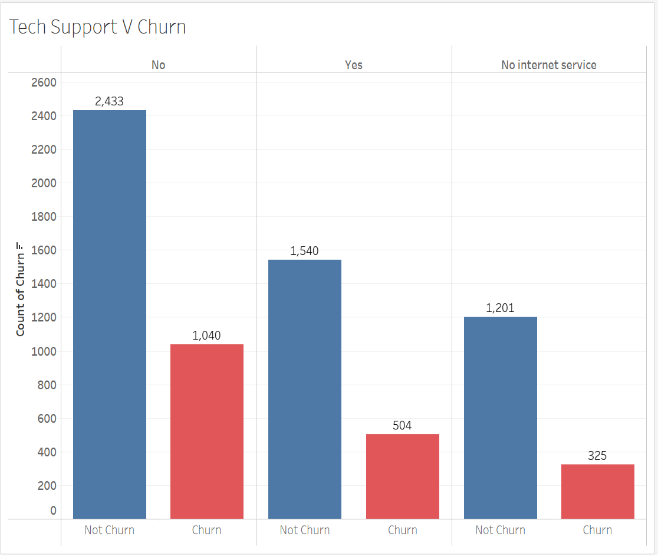
In this report an analysis of customer churn in a telecommunication company was presented. The analysis focused on studying the factors affecting the churn of the business. The telecommunication industry is a highly saturated market. Businesses have to constantly compete with other service providers to satisfy their customers and maintain their market share. The findings of this report may guide the business to focus on significant factors in order to reduce the churn rate. During the analysis, we have clustered the factors into three main categories involving demographic variables, service features variables, and tenure and charges variables. From this analysis, we have identified the key demographic segments that are influencing the churn rate which are age, location, and marital status (dependents). Furthermore, our analysis reveals that higher subscription levels lead to better customer retention, while shorter service tenure correlates with quicker customer churn. As a result, we as the consultant services are now able to provide the telecommunication company with the assistance that they require, which is to promote their products and plans to the range of customers which do not churn and create new plans and strategies to attract the group of customers which are churning. By following the given recommendations and understanding the analysis done, then surely without a doubt the telecommunication company will be able to hike up their profits and dominate the market.

# **Section 6: Additional Work**

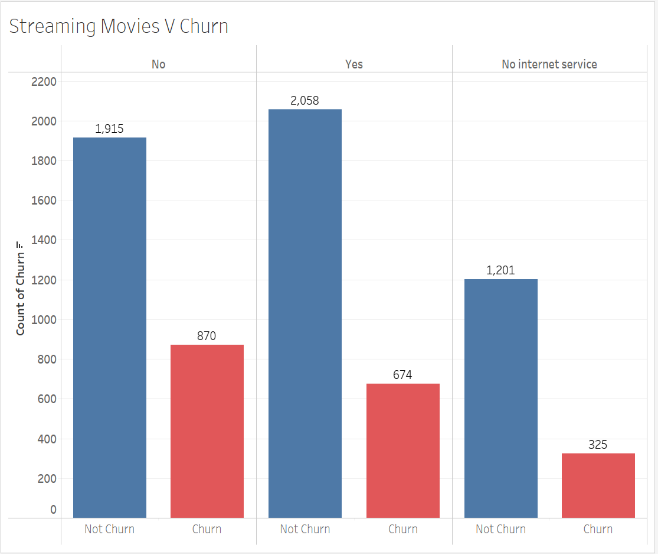


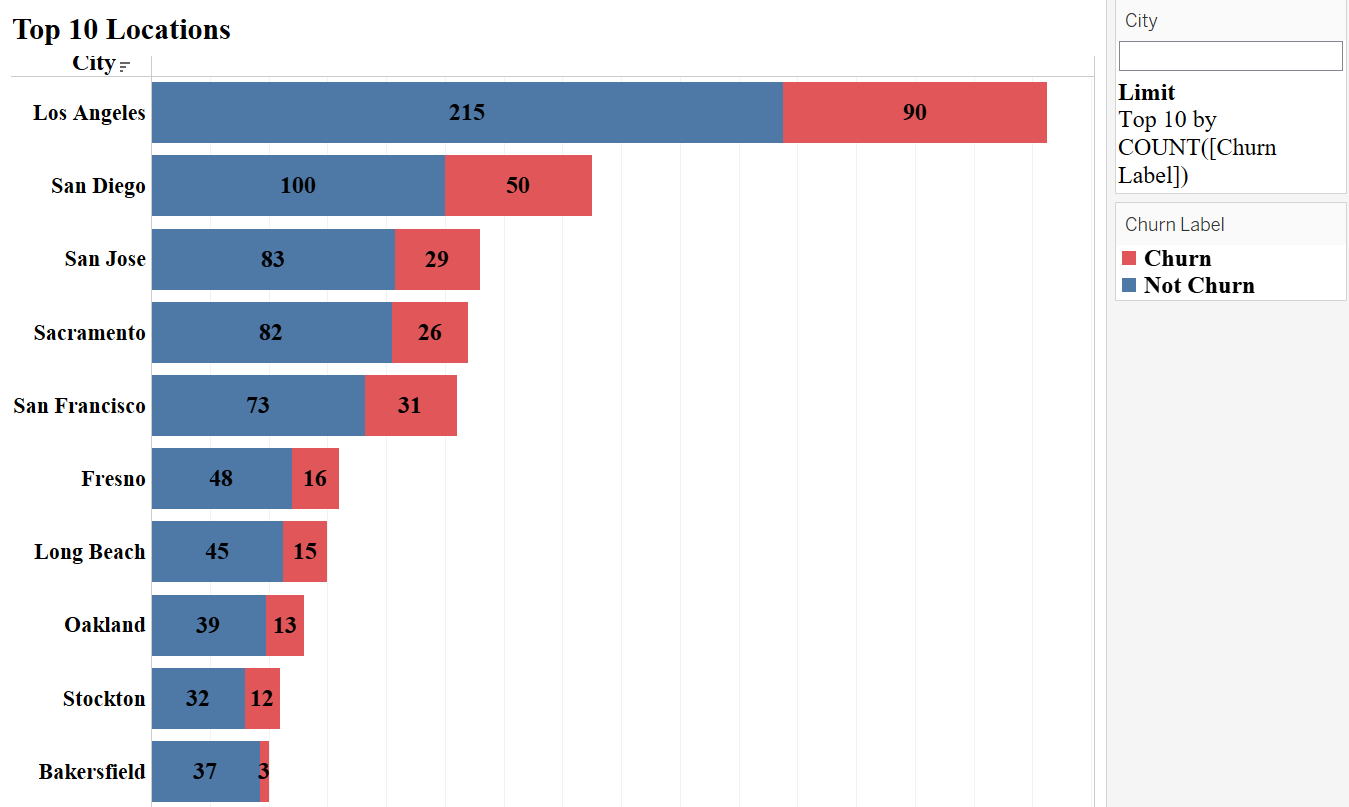


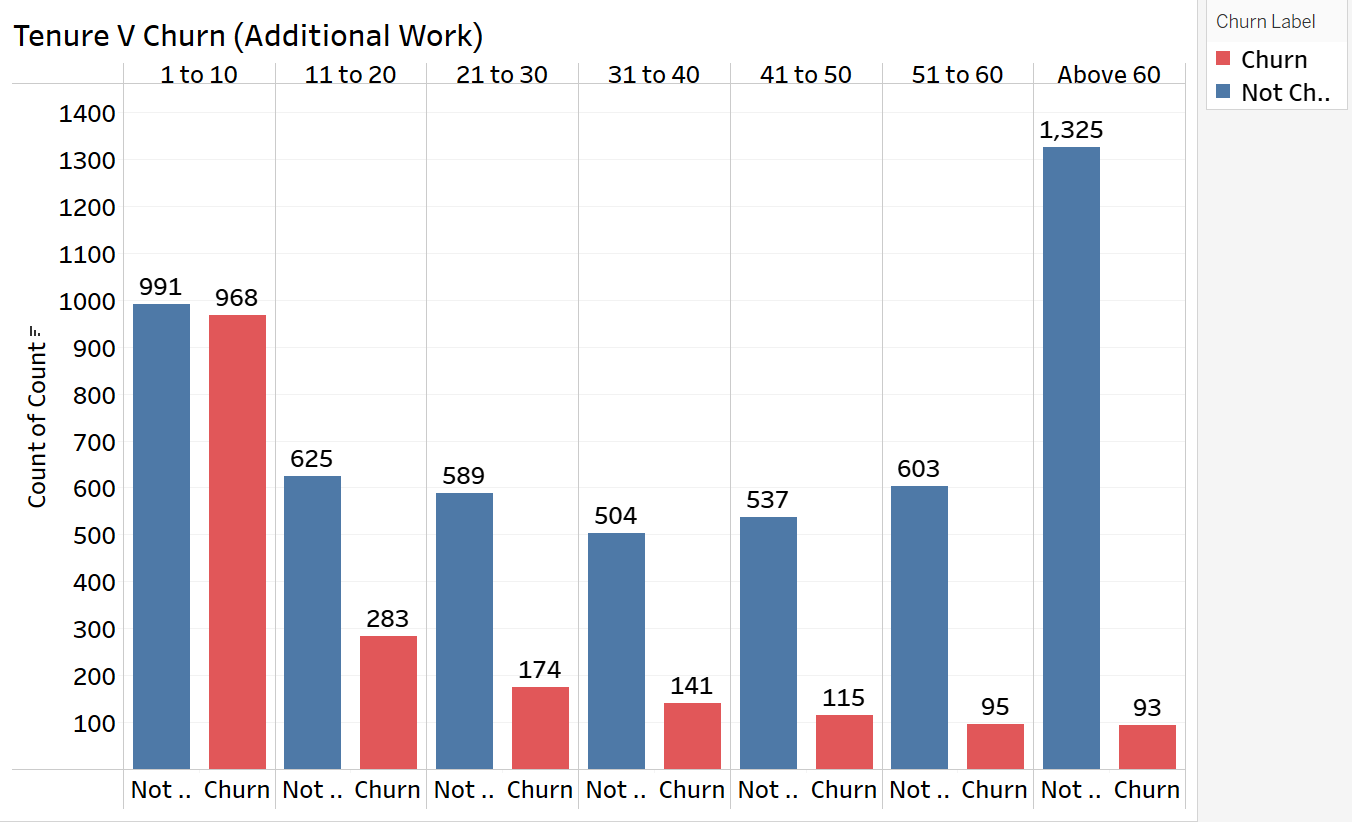




A screenshot of a graph

Description automatically generated





# **References**

Almana, A. M., Aksoy, S., & Alzahran, R. (2014). A Survey On Data Mining Techniques In Customer Churn Analysis For Telecom Industry. In *Journal of Engineering Research and Applications www.ijera.com* (Vol. 4, Issue 5). www.ijera.com

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414–1425. https://doi.org/10.1016/j.eswa.2011.08.024

Oghojafor, B., Mesike, G., Bakarea, R., Omoera, C., & Adeleke, I. (2012). Discriminant Analysis of Factors Affecting Telecoms Customer Churn. *International Journal of Business Administration*, *3*(2). <https://doi.org/10.5430/ijba.v3n2p59>

Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A., & Rehman, A. (2013). Telecommunication subscribers’ churn prediction model using machine learning. *8th International Conference on Digital Information Management, ICDIM 2013*, 131–136. <https://doi.org/10.1109/ICDIM.2013.6693977>

# **Tableau Project: Proposal**

1. **Narrative**

In the 21st century era, telecommunication has become the dominant medium of communication as many people have access to the mobile phone. Whether it is a brand-new model iPhone, or an inexpensive China branded smartphone, many people are using their mobile phone to communicate, surfing social media, playing games, streaming videos and many more. This has made the telecommunication companies realize the importance of giving the customers the possibility of having a good telco plan at a reasonable price. In the perfect scenario case, telecommunication companies should be able to maximize their profits in today’s market as more than half of population in the world are using smartphones (Almana et al., 2014)

However, according to Qureshi et al. (2013), the market for telecommunications services is saturated to the extent that each new consumer must be won over from competitors. The situation has been made worse by the government as the public policies and standardization of mobile communication has made the consumers to easily switch from one carrier to another, resulting in a fluid market. Thus, in the highly competitive industry, retaining customers is seen as an important approach for the telco company to sustain.

As the telecommunications industry has become more competitive, customer retention is seen as a crucial focus in sustaining growth and profitability. The issue of customer churn, where customers stop or discontinue their service subscription has posed a significant challenge. In this study, it will focus on how the consultant service could help the telco companies to understand their datasets and provide suggestions on how to minimize customer churn.

Other than that, this study will focus on analyzing customer churn within a telecommunications company using a given dataset. These datasets included records of customer demographics, service usage and churn status which would provide a strong foundation for in-depth analysis. By combining the datasets, the analysis will give a better insight that can help identify high-risk churn segments and develop targeted retention strategies.

The study seeks to provide the consultant service an actionable insight that the telco companies can use to mitigate churn and improve customer retention. By identifying high-risk segments and understanding the factors that contribute churn, telco companies can develop strategies that can increase customer satisfaction and loyalty. In return, this will contribute to a more stable customer based and sustained business growth in the highly competitive telecommunications industry.

1. **Datasets**

The datasets were collected by the IBM teams to identify the customer churn in a telco company. These datasets contained information about the customers demography and the type of mobile and internet services that were provided by the telco company.

The total number of 7043 customers data in California were collected in the datasets for the Q3. The datasets were then grouped into two tables which are: -

Dataset 1 (Customer ID, Gender, Senior Citizen, Partner, Dependents, Tenure, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tec Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges, Churn)

Dataset 2 (Customer ID, Count, Country, State, City, Zip Code, Lat Long, Latitude, Longitude, Gender, Senior Citizen, Partner, Dependents, Tenure in Months, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges, Churn Label, Churn Value, Churn Score, CLTV, Churn Reason)

1. **Research Questions**
2. How do demographic factors (age, gender, geographic location, marital status) influence customer churn rates?
3. How do additional service features (Online Security, Online Backup, Device Protection, Tech Support) affect customer retention?
4. What is the relationship between the length of service tenure and customer churn?
5. **Research Objectives**

The plan of this study is to identify demographic segments that are more likely to leave the telco company. Demographic factors such as gender, age, dependent and location will be tested to see the customers behaviour and satisfaction. Identifying the churn rate across the demographic segments will help to pinpoint the specific groups that are at higher risk to leave. This will be crucial information for tailoring retention effects, and it also will help in creating effective marketing campaigns that can resonate with the needs and preference of these segments (Oghojafor et al., 2012).

The second plan is to measure the impact of subscribing to additional services on the likelihood of churn. In this study, the telco company that will be analyzed offers a variety of services including internet, phone, and television packages. It is important to understand how the adoption of these services will affect churn and it will provide valuable information into customer satisfaction and utilization. For example, customers who subscribed to bundle services may have different churn behaviors compared to the customers who use the standalone services. Analyzing this pattern can identify the opportunities to enhance service offerings and promote service packages that can increase customer loyalty.

The third plan is to identify if longer-tenured customers are more or less likely to churn and identify the points in their tenure where churn risk increases or decreases. In churn analysis, customer tenure or the length of time a customer has been with the company is a critical factor (Huang et al., 2012). Customers who have stayed long in the subscriptions are less likely to churn as he has established loyalty and satisfaction. However, there might be a period in the customer’s lifecycle that the risk of churn spikes. Analyzing the churn rate over different tenure periods will help to develop strategies to address these critical points and increase customer retention.

1. **Hypotheses**

In the telecommunication industry, customer churn possesses a critical issue where retaining customers is often sought as more cost-effective than recruiting new ones. Understanding the customer churn will provide valuable insight for improving customer satisfaction and loyalty. The plan of this study is to explore various hypotheses related to factors influencing customer churn using the datasets given. These hypotheses address different variables of customer demographics, account information, services, payment methods, customer support and overall satisfaction.

**Hypothesis 1: Senior citizens are more likely to churn.**

One of the factors that play a significant role in customer behavior is demography. The hypotheses that can be made from this factor is that senior citizens are more likely to churn compared to younger customers. This assumption could be on the fact that senior citizens could be less tech-savvy or more price sensitive. This hypothesis will be tested by comparing the churn rates between senior citizens and younger customers using bar graphs. If the hypothesis is true, it indicates that targeted strategies are needed to retain senior citizens.

**Hypothesis 2: Customers with longer tenure are less likely to churn.**

A common conception is that customers who have been with a company for a longer period are less likely to leave. This hypothesis can be justified that longer tenured customers have higher loyalty and a greater investment in the service. Analyzing the relationship between tenure and customer churn could provide a better understanding of customer loyalty.

**Hypothesis 3: Monthly charges influence churn rates.**

One of the most crucial aspects of customer loyalty is the cost of services. Customers with higher monthly charges are more likely to leave as it can cause financial burden. If the hypothesis is true, it would mean that the customers are price sensitive. The telco company could consider pricing strategies or value-added services to reduce churn among high-paying customers.

**Hypothesis 4: Service subscription combination and churn.**

Customers who subscribe to multiple service subscriptions (e.g., phone, internet, TV) are less likely to churn due to bundled value and convenience. In other words, the combination of services that the customers subscribe to may influence their likelihood of churning. If the hypothesis is supported, the telco company may need to focus on promoting bundled services as it can be an effective strategy to enhance customer retention.