IE6600 Computation and Visualization for Analytics

Final Project Report

Group 9

TELCO CUSTOMER CHURN ANALYSIS

Team:

- **Naveen Kumar Pasala**
- **Apoorv Vats**
- **Shiva Naga Jyothi CH**
- **Vikrant Shelke**
- **Mrudula Tanniru**

Table of Content:

Introduction:

In today's highly competitive world, every company strives to be at the top of their competitors, the most important job for any company of any industry is to retain their existing customers and gain new customers to develop their market base.

Customer churn prediction is one of the most popular use cases of Data Science Marketing, and the major concern of a company. Companies incur a lot of costs when the current user churns since it would be very expensive to replace an existing customer. Due to this, most large to mid-ranged organizations will have some sort of churn prediction mechanisms.

Churn is the measure of customers that leave the service or the product after a certain period of time. Clients can stop doing business with an entity for a variety of reasons, such as affordability, dissatisfaction with the offering, and bad customer service and many other factors. The telecom industry faces an average annual churn rate of 15 to 25%.

More often than not, Clients that churn from one company may probably start doing business with their competitor. For instance, if a client is not happy with the current mobile service provider due to reasons like low signal strength and slow Internet speed at their location, they are very likely to switch to an alternative competitor network which has good services.

Churning does not happen suddenly, if customers experience low network bandwidth, they are likely to contact customer support, check their network speed, and express their dissatisfaction on social media. During that phase, if the customer concern is not addressed correctly, they would tolerate it for a month or two even then if the issue is not resolved then they are more likely to churn the network and shift for the other networks.

The data scientists at the current provider should collect all the data and analyze the factors and behaviors of the customers that are more likely to churn and alert the marketing team to retain the customers whose behavior looks similar to that of the customers who had churned the network in the past. The marketing team should reach out to the customer and attempt to cater the needs in the best way possible by providing special promotions, upgrading the current plan, and work towards creating a satisfactory user experience for preventing the client from leaving the network.

Problem Statement:

TELCO CUSTOMER CHURN ANALYSIS: The main objective of this project is to understand the Telco data, pre-process and create an interactive dashboard using different types of graphs to identify the attributes that are leading to customer churn.

Research Questions:

We would like to address below research question using our analysis on telco customer churn data,

1. What is the major factor affecting customer churn?
2. Does the age of the customer impact the Churn Rate?
3. What analysis can be done on the basis of customer payments?
4. Does customer service affect customer retention in any way?
5. What does the data tell us about customer loyalty?

Summary of Results:

We would analyze all the rows in the dataset for each customer and get the insights from data using attributes such as churn category and churn reason to address our first research question on factors leading to customer churn.

For the impact of age on customer churn, we would divide the age of the customers into different bins and try to extract the insights from data. This would help us understand which age group customers are most likely to churn.

In the Customer payments research question, we would address which are common payment methods used by customers and try to understand which payment type customers are more likely to churn. In this part we would also cover Monthly charges to customers, Involvement in group plans and their impact, and study different contract types.

Understanding the quality of service provided to the existing customers is a crucial factor, hence we need to know the impact of service calls on customers, is there a positive or negative impact. To address this question, we would use the Customer service calls attribute and Churn category for each service call to get the insights.

Customer loyalty is a crucial factor, we need to understand the behavior or the trend in customers over the years which would help companies to improve their process or operations.

Data description:

The dataset that we will be using 'Telco Customer Churn' which was obtained from Kaggle, has 7043 instances and 21 attributes. We used the Python programming language for data cleaning, exploratory data analysis (EDA) and for data visualization we used Tableau.

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn Label
- Services that each customer has signed up for – Device protection
- Customer account information – how long they've been a customer, contract, payment method, monthly charges, and total charges
- Demographic info about customers – gender and age range

Kaggle Dataset:

https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Data Dictionary:

https://drive.google.com/file/d/1bpMrqbTY50trEc81nmHO69o6cYtiN0Gt/view?usp=share_link

Methodology and Design Process:

The dataset was imported using python's pandas library as a dataframe. After importing, exploratory data analysis was performed. Head function is used to get us the first 5 rows. It is useful for quickly testing what type of data is in the dataframe. Shape function returns a tuple that gives the number of rows and columns of a given dataframe. Info function provides information about a dataframe like columns, datatype, non-null values and memory usage. Describe function provides descriptive statistics of the data that includes the central tendency, dispersion and data distribution.

Churn rate is the percentage of customers leaving the company versus the total number of customers. It was calculated for the dataset and it shows that 26.85% of the customers churned.

Columns that are unique and repetitive were removed from the dataframe. The data was then checked for any duplicate rows. After that the boxplot graph was used to check for any outliers in the data. Outliers may cause problems while analyzing data so the top 2% of the outliers is removed from the dataframe. The data is then checked for imbalance and then plotted below graph which shows the correlation between churn label and other columns in the data frame.

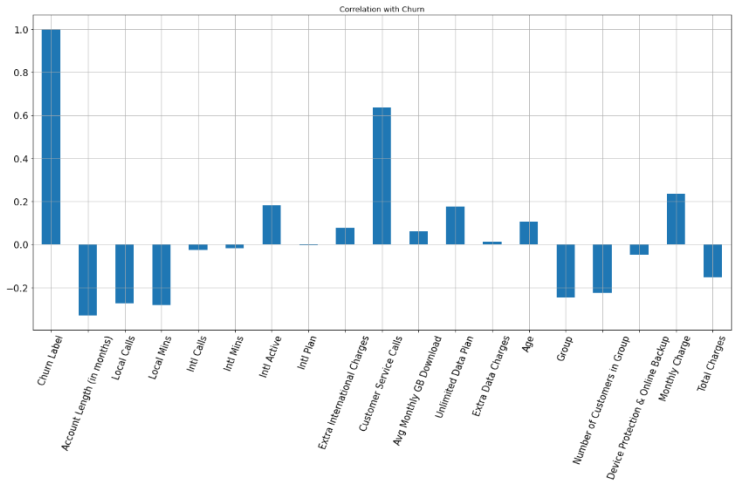


Fig 1. correlation between churn label and other columns

The correlation heatmap is plotted that contains correlation coefficients between variables as per below figure. In the end, the cleaned data is exported for Tableau.

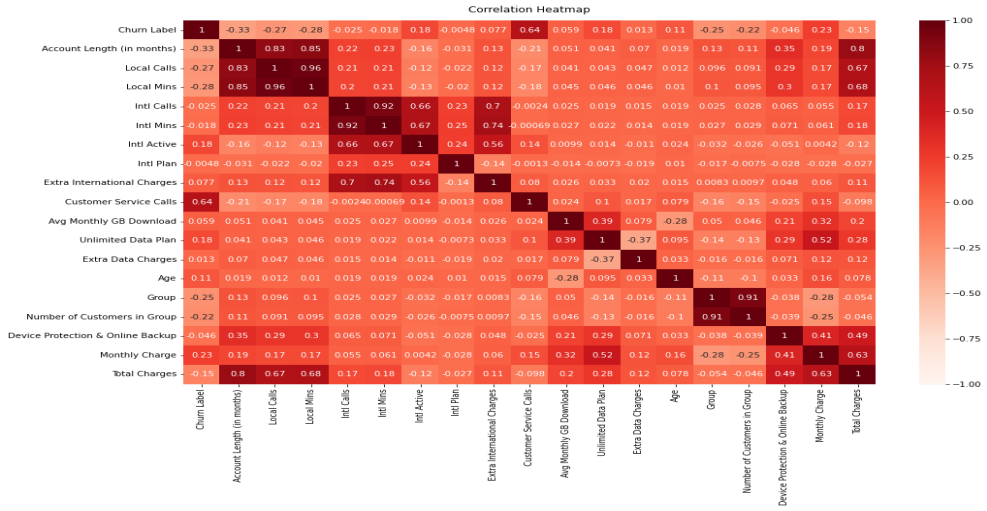


Fig 2. correlation heatmap

Python notebook:

https://colab.research.google.com/drive/1RuPYb3305w7AZAQQySTRK84joz-uxqM8?usp=share_link

Dashboard Visualization:

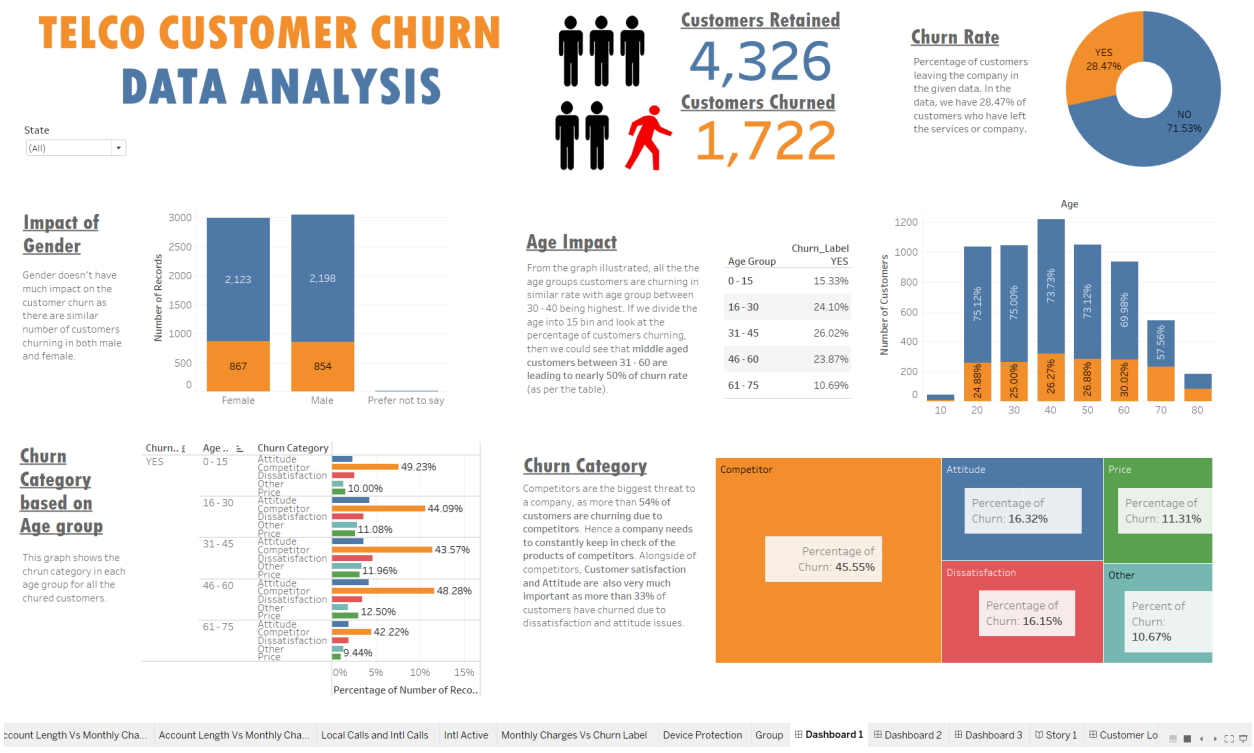
The cleaned telco data is imported into tableau as a data source and started to plot various graphs to study the relationship between each attribute and their effect on churn rate using workbooks. Also, we have performed a comparative study between the churned customers and non-churned customers which helped to extract the behavior of customers leaving the company.

Using the above created workbooks, we have built the below dashboards with our observations for each study.

Tableau Dashboard:

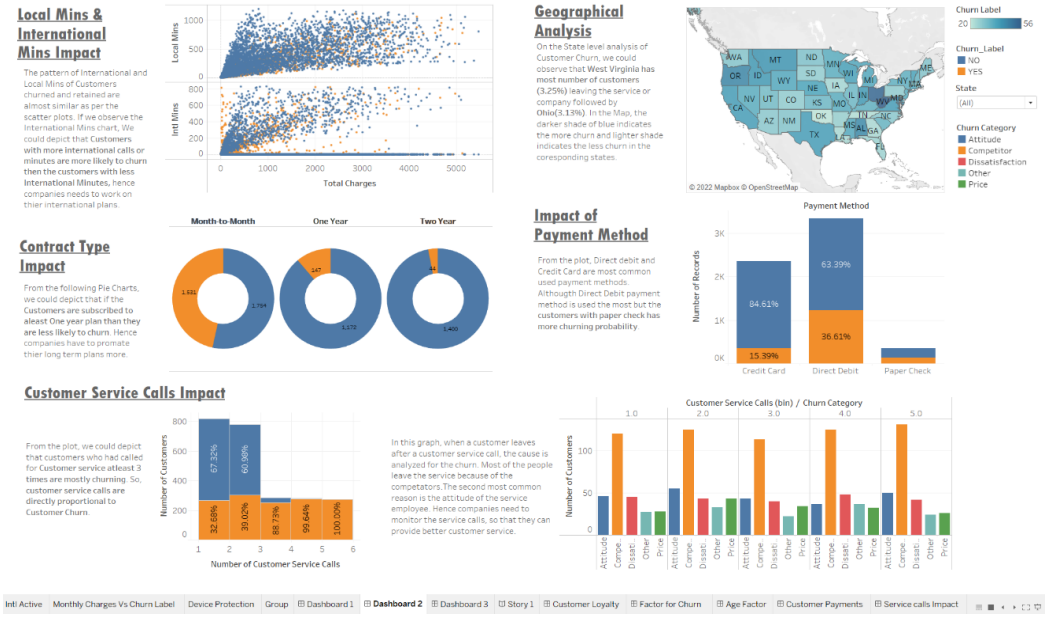
https://drive.google.com/file/d/1pR235_60WEtdbWYNPqYD6aHXgtWg9DV5/view?usp=share_link

Dashboard 1:



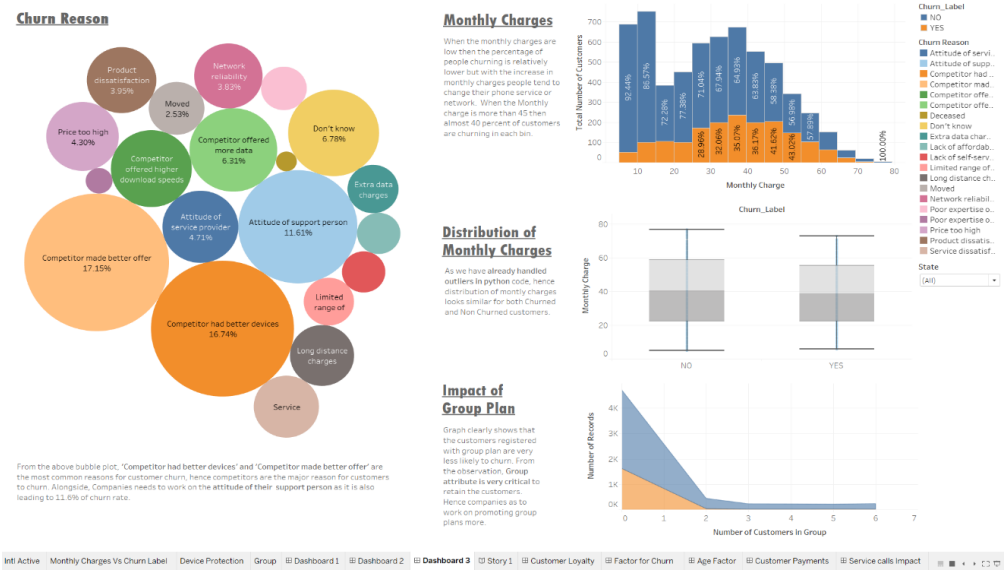
In this dashboard, we mainly focused on attributes such as Churn Rate, Gender, Age and Churn category.

Dashboard 2:



In this dashboard, we have performed a study on geographical analysis, Local and International Minutes, Total charges, Contract type, Payment Methods, Customer service calls.

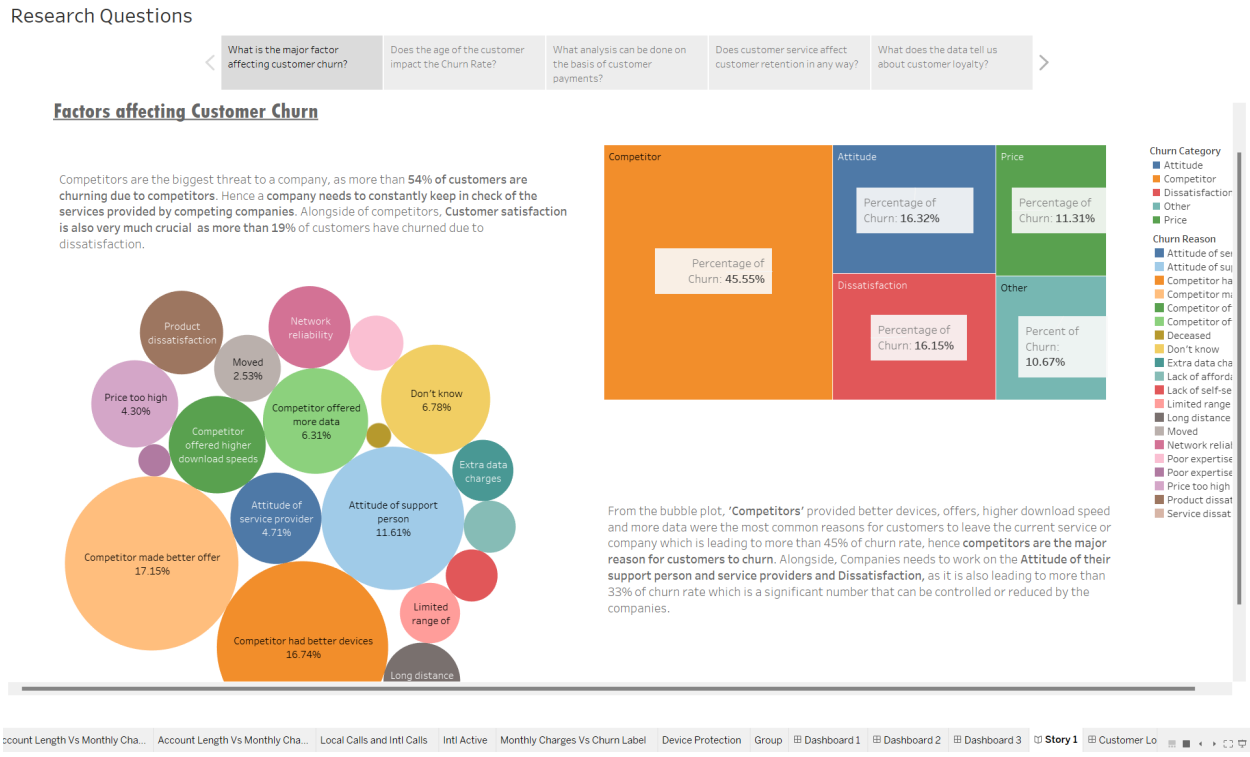
Dashboard 3:



In the dashboard 3, we have mainly focused on attributes such as Churn Reason, Monthly charges of customers and impact of group plan, which helped us conclude few observations.

Key Insights:

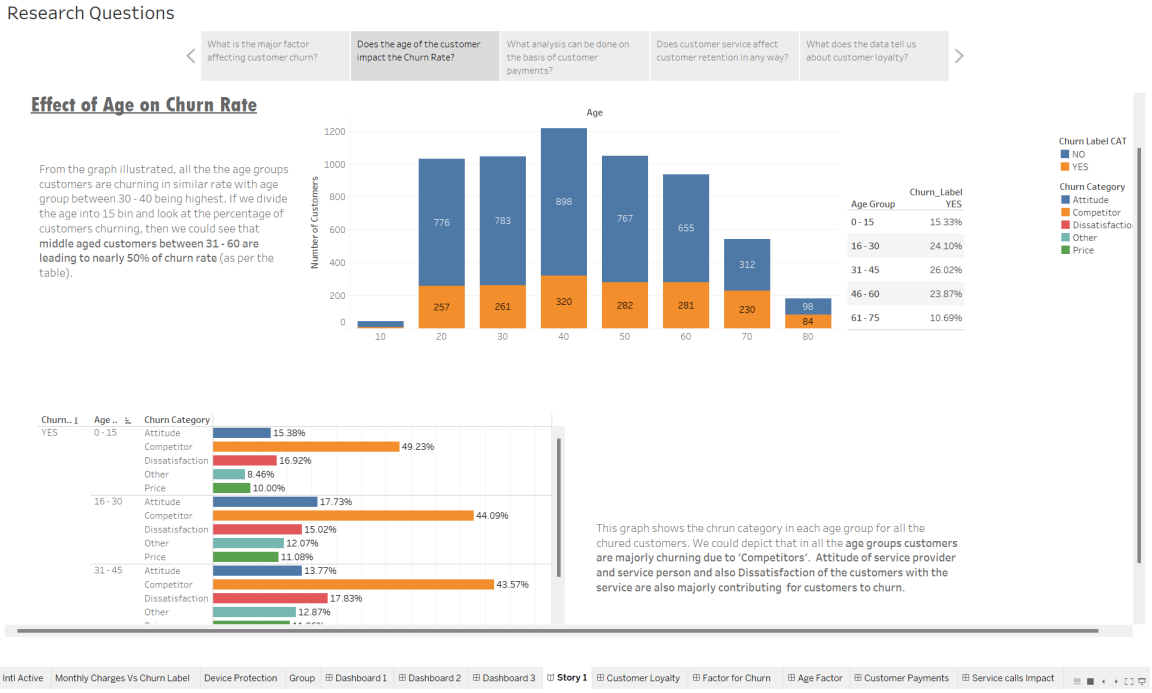
1. What is the major factor affecting customer churn?



Competitors are the biggest threat to a company, as more than 45% of customers are churning due to competitors. Hence a company needs to constantly keep in check of the services provided by competing companies. Alongside competitors, Customer satisfaction and attitude are also very much crucial as more than 33% of customers have churned due to dissatisfaction.

From the bubble plot, 'Competitors' provided better devices, offers, higher download speed and more data were the most common reasons for customers to leave the current service or company which is leading to more than 45% of churn rate, hence competitors are the major reason for customers to churn. Alongside, Companies need to work on the Attitude of their support person and service providers, as it is also leading to nearly 20% of churn rate which is a significant number.

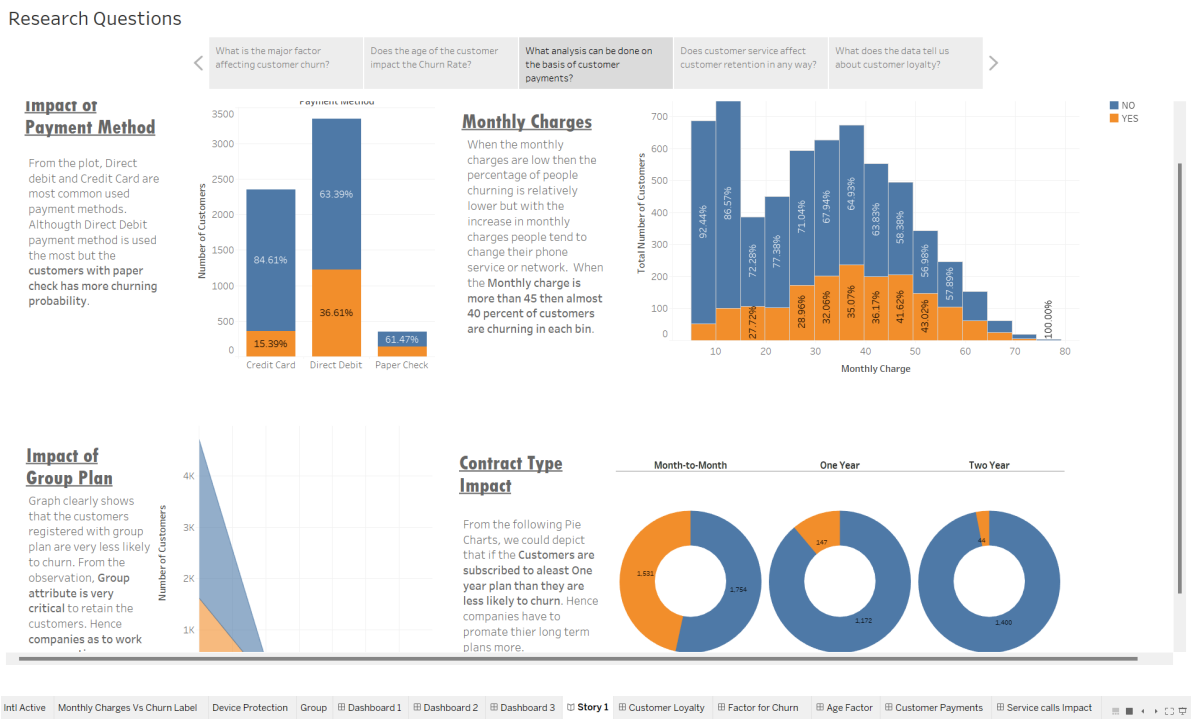
2. Does the age of the customer impact the Churn Rate?



From the graph illustrated, all the age groups customers are churning at a similar rate with the age group between 30 - 40 being highest. If we divide the age into 15 bins and look at the percentage of customers churning, then we could see that middle aged customers between 31 - 60 are leading to nearly 50% of the churn rate (as per the table).

This graph shows the churn category in each age group for all the churned customers. We could depict that in all the age groups customers are majorly churning due to 'Competitors'. Attitude of service provider and service person and also Dissatisfaction of the customers with the service are also majorly contributing for customers to churn.

3. What analysis can be done based on customer payments?



From the plot, Direct debit and Credit Card are most commonly used payment methods. Although Direct Debit payment method is used the most, the customers with paper checks have more churning probability.

When the monthly charges are low then the percentage of people churning is relatively lower but with the increase in monthly charges people tend to change their phone service or network. When the Monthly charge is more than 45 then almost 40 percent of customers are churning in each bin.

Graph clearly shows that the customers registered with group plans are very less likely to churn. From the observation, Group attribute is very critical to retain the customers. Hence companies have to work on promoting group plans more.

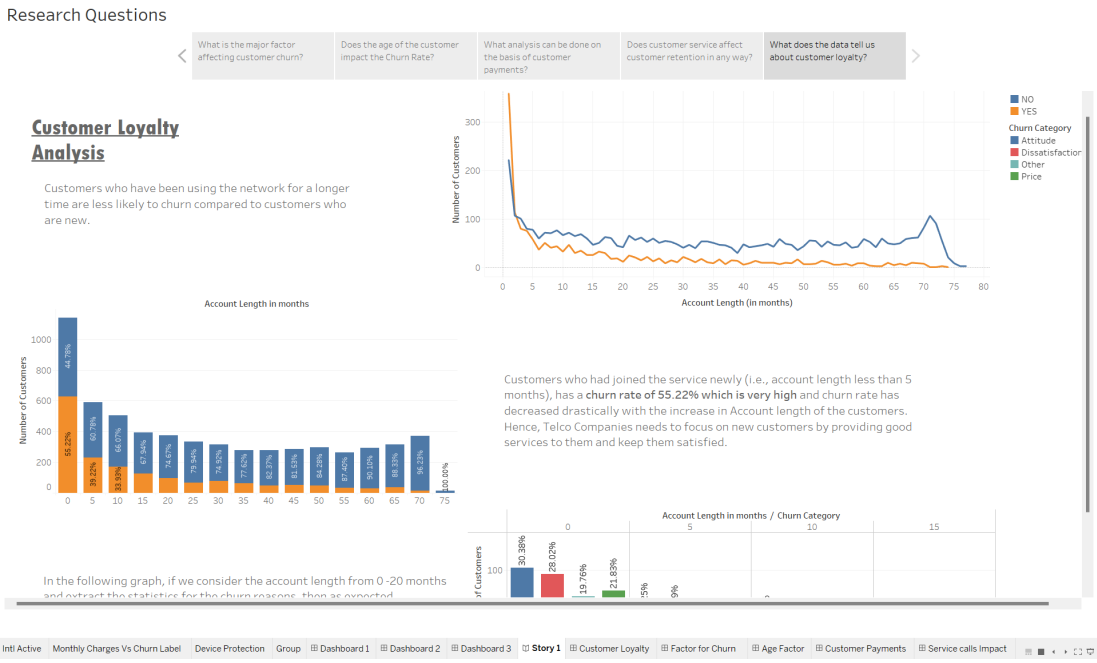
From the following Pie Charts, we could depict that if the Customers are subscribed to at least One year plan then they are less likely to churn. Hence companies have to promote their long term plans more.

4. Does customer service affect customer retention in any way?

From the plot, we could depict that customers who had called for Customer service at least 3 times are mostly churning. So, customer service calls are directly proportional to Customer Churn.

In this graph, when a customer leaves after a customer service call, the cause is analyzed for the churn. Most of the people leave the service because of the competitors.The second most common reason is the attitude of the service employee. Hence companies need to monitor the service calls, so that they can provide better customer service.

5. What does the data tell us about customer loyalty?



Customers who have been using the network for a longer time are less likely to churn compared to customers who are new. Customers who had joined the service newly (i.e., account length less than 5 months), had a churn rate of 55.22% which is very high and the churn rate has decreased drastically with the increase in Account length of the customers. Hence, Telco Companies need to focus on new customers by providing good services to them and keep them satisfied.

In the following graph, if we consider the account length from 0 -20 months and extract the statistics for the churn reasons, then as expected, Competitors are the most common reason but if we exclude the effect of 'competitors' and analyze the results then, 'Attitude of support person' and 'Attitude of Service provider' are the crucial factors for customers to leave the company during initial days.

Conclusion:

1. The biggest threat to a company is its Competitors, as more than 54% of customers are churning due to competitors as 'Competitors' are providing better devices, offers, higher download speed and more data were the most common reasons for customers.
2. Customer satisfaction is also very much crucial as nearly 20% of customers have churned due to dissatisfaction which is a significant impact.
3. Middle aged customers between the 30 to 60 age group are leading to nearly 50% of the churn rate. Hence companies have to provide better customer service experience.
4. If monthly charges are low then the percentage of people churning is relatively lower but with the increase in monthly charges people tend to change their phone service or network. When the Monthly charge is more than 45 then almost 40 percent of customers are churning in each bin.
5. Customers registered with a group plan are very less likely to churn. Hence companies have to work on promoting group plans more to retain their customers.
6. Customers who are subscribed to at least One year plan are less likely to churn. Hence companies have to promote their long term plans to retain their customer base.
7. Customers who have called for Customer service at least 3 times are most likely to churn. So, customer service calls are directly proportional to Customer Churn.
8. Customers who had joined the service newly (i.e., account length less than 5 months), had a churn rate of 55.22% which is very high and the churn rate has decreased drastically with the increase in Account length of the customers. Hence, Telco Companies need to focus on new customers by providing good services and keeping them satisfied.

Limitation and Future Scope:

The main goal of this project is to help the companies in the telecom market to make more profit. It is known that predicting churn is one of the most essential sources of income to telecom companies. Hence, this project is aimed to build an interactive dashboard that analyzes the churn rate of customers in telecom industries.

The use of telecom data imposed some limitations on the present study. For example, we could examine only a few factors that were recorded in the telecom company's data. If we could have got some data specific to a particular company, our study would have been more accurate. Also,

if we see in the current data, Gender attribute is mostly balanced but it would not be the case in the real world data.

Future research will further investigate the implementation of results and will also identify customer requirements using different methods and propose some techniques to prevent them from churning out. There is a scope to build a machine learning model using the attributes identified in predicting the chances of customer churning. Predicting customer's churning can be used to lower the initial costs and improve company's revenue, however this project does not cover such economical aspects.

References:

1. Kaggle Dataset:

   https://www.kaggle.com/datasets/blastchar/telco-customer-churn

2. Pandas documentation

   https://pandas.pydata.org/docs/

3. Matplotlib documentation:

   https://matplotlib.org/stable/index.html

4. Seaborn documentation:

   https://seaborn.pydata.org/

5. Tableau:

   https://help.tableau.com/current/pro/desktop/en-us/gettingstarted_overview.htm