**Customer Churn EDA**

**Business Understanding and Overview**

A telecommunication company is concerned about high rate of customer churning. The company provided us the data about its customers and requested to analyse the data about the characteristics of its customers to identify the factors for churning. The expectation is that the company should take some measures and change its policy according to insights from the exploratory data analysis. By reducing the churning rate, the company would be able to increase its revenue and provide better customer services. We are taking the following steps for EDA. We first study the data to understand it, then we clean the data and then we perform exploratory analysis using univariate and bivariate methods.

**Understanding the Data**

The raw data is in the form of CSV format, containing 7043 records and 21 columns. Our target variable is Churn column which contains only Yes and No values. According to the data, 26.5% of customers are churning, that is more than a quarter of total customers which is a significant loss to the company.

Below is a screenshot of the general info about the columns:

A screenshot of a computer

Description automatically generated

Below chart displays the percentage of customers who churn comparing to those who are not churning.

A green and red chart

Description automatically generated

Data Cleaning

In order to analyse the data, we must clean the data, that includes dealing with Null values, outliners, wrong data type, incorrect entries and removing unnecessary columns or renaming them.

We have noticed the data type for the column “TotalCharges” is displayed as “Object”, as this column is dollar amount, the data type must be numerical. The first step in cleaning data process is to change the type from Object to Numeric.

After changing the data type for TotalCharges to numeric, we have noticed that there are 11 records of null values for TotalCharges. The reason is because the blank values are not considered Null values when the data type is object, however by converting it into numeric data type, the blank values are displayed as null values.

Missing Values

We must figure out what percentage of data contains Null values and how we can address them.

The following graph dhows that only 0.15% of our record contains Null values.

A graph with blue dots and white text

Description automatically generated

Given the fact that the percentage of Null Values are very low, dropping the records containing Null values will have insignificant impact on overall result, we have decided to drop them. The alternative option was to use mean values of the column to replace the null values.

We must group the records into bins based on tenure column, because analysing the data on a monthly bases becomes cumbersome, as the tenure ranges from 1 to 72 months, we have decided to use increment of 12 months to group the records. Below is the result of counting in each group:

A close-up of numbers

Description automatically generated

After grouping the records, we have determined that two columns of “customerID” and “tenure” are not relevant to our analysis, so we exclude them from our data frame.

**Initial Intuition from Data**

* The only two columns that are numeric are “MonthlyCharges” and “TotalCharges”, the rest are all categorical data.
* The number of customers who churn are 1869, which is 26.5% of total customers in the dataset.
* The target variable is the column titled “Churn”, in order to analyse the data more efficiently, we have grouped the data using “tenure” column using increment of 12 months

**Univariate Analysis**

**Gender Based Analysis:** Considering only gender, we cannot get any insight about if gender is a contributing factor in churning. Both males and females have almost the same rate of churning. Please refer to below graph:

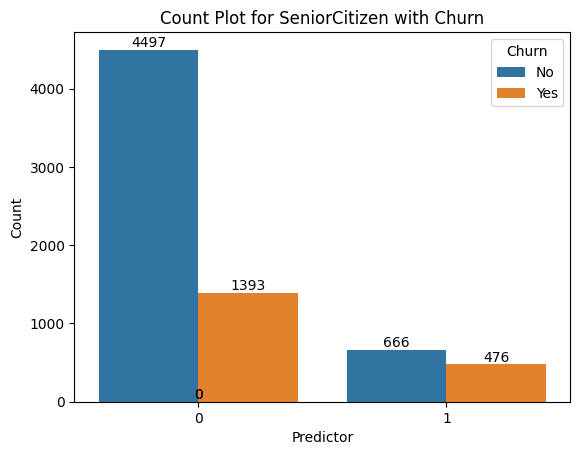
A graph with blue and orange bars

Description automatically generated

As it is shown in the graph, both male (26.2%) and female (26.9%) have almost equal rate of churning.

**Analysis based on Seniority:**

The rate of churning are much higher for senior citizens compared to young people (41.7 % vs 26.2%) as it is displayed in below graph:



**Analysis Based on Having Partner:**

People with partners are less likely to churn 19.7 vs 32.97

A graph with numbers and a number of columns

Description automatically generated with medium confidence

**Analysis based on having dependants:**

Customers without dependants are more likely to churn 31% vs 15.5%

A graph with blue and orange bars

Description automatically generated

**Analysis Based on Having Phone Service:**

Customers with phone service or without phone service does not give us an insight if having phone service is a contributing factor in churning as the percentages are very close 26.7% vs 25%

A graph with blue and orange bars

Description automatically generated

**Analysis Based on Having Internet Services:**

Customers with Fiber Optics Internet Service are more likely to churn comparing to those who have DSL internet or have no Internet. It is likely because Fiber Optic Internet is more expensive which is inline with those who have higher monthly bills.

A graph of a number of blue and orange bars

Description automatically generated

**Analysis Based on Having Online Security Service:**

Customers with no online security services are more likely to churn.

A graph of numbers and a number of blue and orange bars

Description automatically generated

**Analysis Based on Having Online Backup Services:**

Customers with no online Backup Services are more likely to churn.

A graph with numbers and a bar

Description automatically generated with medium confidence

**Analysis based on Having Device Protection Plan:**

Customers with no Device Protection Plan are more likely to churn.

A graph of a number of blue and orange bars

Description automatically generated

**Analysis Based on Having Tech Support:**

Customers who have received no tech support are more likely to churn, probably because they did not ask for support, rather they just churned.

A graph of blue and orange bars

Description automatically generated

**Analysis Based on Having Streaming TV And Streaming Movie Services:**

Customers with streaming TV service or without it does not give us a good insight about the impact of this service in churning. The difference between the churning rate are negligible.

A graph of a number of blue and orange bars

Description automatically generated A graph of a number of blue and orange bars

Description automatically generated

A graph of blue and orange bars

Description automatically generated

**Analysis Based on Contract Terms:**

Customers with longer contract are less likely to churn as the graph displays, the reason could be that customers with longer contract are bond to contract and it would cost them more to churn.

A graph of blue and orange bars

Description automatically generated

Month to month 12.71% One Year 11.28% Two years 2.84%

**Numerical Analysis - Monthly Charges and Total Charges**

Customers with higher monthly charges are more likely to churn as the graph indicates.

A graph of a graph of charge

Description automatically generated with medium confidence

Customers with lower total charges are more likely to churn, it can be explained that customers who are churning receives less number of monthly bills resulting to lowering the total charges even though the monthly charges could be higher.

**Note: This analysis is incomplete, the github will be updated soon**

README: Business Problem … Data Understanding … Insight Found… Next Step … (Creating a model to predict who is going to churn)

Categorical Analysis – Univariate Analysis – Bivariate Analysis – Numerical Analysis – Final Thoughts

A screenshot of a computer

Description automatically generated