TELECOM CHURN ANALYSIS

❖ DESCRIPTION :-

"**Churn**" or "**Customer Attrition**" or "**Customer Churn**" are few of the terms that are used to refer to the situation that occurs when customers stop doing business with the company. The companies are interested in identifying segments of these customers because the price for acquiring a new customer is usually higher than retaining the old one. In this project, we will be using Telecom Customer Churn Dataset to predict the factors that lead to customers changing their network provider.

So far we have performed below activities on the dataset :

❖ PRELIMINARY DATA ANALYSIS :-

To gain an insight of the data in the dataset, we have imported the dataset and have done some preliminary overview of dataset by fetching dimensions of the dataset, viewed the first few rows of the dataset, examined kind of data stored in each column.

❖ DATA CLEANING AND FORMATTING :-

- By doing the above preliminary data analysis, we analyzed that 'SeniorCitizen' and 'Tenure ' variables were stored as numerical which need to be converted to categorical variables and hence we have converted '**SeniorCitizen**' numerical variable into Categorical variable and '**Tenure**' value of 0-72 months into years with ranges of 12 months and represented this new column as '**Tenure_Range**'**.**
- Checked if there are any 'NA' or missing values in the dataset and found that there were 11 rows out of 7043 rows in 'TotalCharges' column which had null values and since these values are only 0.15% of the total hence we could afford to drop them.
- Noticed that there is an extra factor for few of the columns in the dataset. So replaced '**No Internet Service**' values in OnlineSecurity,OnlineBackup, DeviceProtection,TechSupport, StreamingTV and StreamingMovies columns with 'No' and have deleted the unused levels from the factor variables.

❖ DEPENDENT AND INDEPENDENT VARIABLES:-

- **Dependent Variable** :- Dependent variables are target or output variables. In this case 'Churn' is a dependent variable.

- **Independent Variables** :-An Independent variable is the variable that is changed or controlled to test the effects on dependent variable. Independent variable helps in predicting the dependent variables and hence they are also known as Predictors.

  Below is the list of Independent variables or predictors :-

  - Gender
  - SeniorCitizen
  - Partners
  - Dependents
  - Tenure
  - MultipleLines

- OnlineSecurity
- StreamingTV
- StreamingMovies
- Contract
- PaymentMethod
- MonthlyCharges
- TotalCharges

❖ EXPLORATORY DATA ANALYSIS:-

We have 2 types of features in the dataset: categorical (two or more values and without any order) and numerical.
Most of the feature names are self-explanatory, except for:

- Partners: Whether the customer has a partner or not (Yes, No)
- Dependents: Whether the customer has dependents or not (Yes, No)
- OnlineBackup: Whether the customer has an online backup or not (Yes, No)
- Tenure: Number of months the customer has stayed with the company
- MonthlyCharges: The amount charged to the customer monthly
- TotalCharges: The total amount charged to the customer
- Contract : The contract term of the customer
- MultipleLines: Whether the customer has multiple lines or not
- StreamingTV: Whether the customer has Streaming TV or not
- StreamingMovies: Whether the customer has Streaming Movies or not
- PaymentMethod: The customers payment methods.

As part of this step, we have done the below steps so far :

- Data Visualizations :

  - Bar Chart
  - Box Plot
  - Scatter Plot
  - Histogram
  - Correlation Matrix
  - Histogram