# INTRODUCTION

## Definition:

Churn Customer refers to the number of existing customers who may leave the service provider over a given period. These customers can be called as churners. The main aim of churn is to predict the churnable customers at the earliest, to identify the reason for churning. The primary goal of churn analysis is to identify and anticipate churnable consumers as soon as possible. This will help us to rectify the issues of the customer. This will be helpful to satisfy the customer needs and will continue to use that service. This will help to meet the needs of the customers, and they will continue to utilize the service.

### Objective of the Report:

The primary objective of this report is to predict customer churn in the telecom industry using machine learning techniques. Churn, in the context of telecommunications, refers to the phenomenon where customers discontinue their subscription or switch to a competitor. Accurately predicting churn allows telecom companies to identify at-risk customers and implement targeted retention strategies, ultimately reducing customer attrition and increasing profitability.

# Data Description

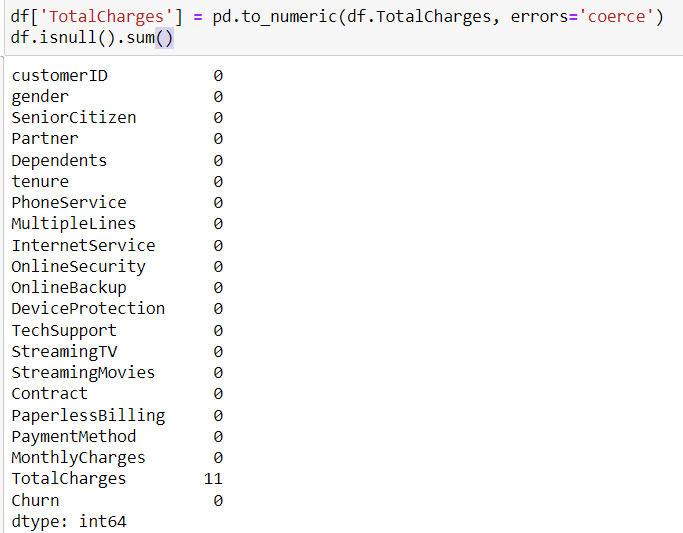
## Overview of the dataset used

* Source of the data: https://www.kaggle.com/datasets/blastchar/telco-
* Description of features/variables:

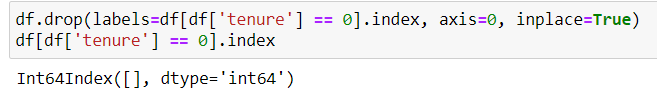
## Data pre-processing steps

* Handling missing values

The dataset used for the telecom churn prediction analysis consists of 7,043 customer records. Ensuring data quality is crucial for building reliable machine learning models, and part of this process involves handling missing values appropriately.Upon examining the dataset for missing values, it was found that there were 11 missing values in the “TotalCharges” column. These missing values were identified using the following code:



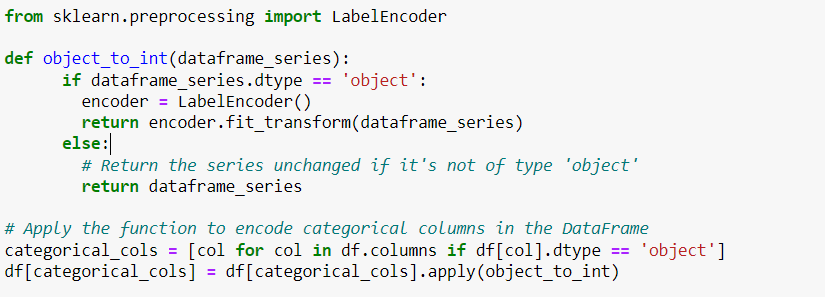
Additionally, it was observed that some records had a tenure value of 0. These records were also dropped from the dataset to avoid inconsistencies, as a tenure of 0 might indicate new customers who have not yet been billed, which could skew the analysis. The removal of these records was carried out using the following code:



* Encoding categorical variables

In the telecom churn prediction analysis, it was essential to encode the categorical variables into numerical values to make them suitable for machine learning models. Categorical variables often contain non-numeric data which needs to be converted into a numeric format before they can be utilized by most machine learning algorithms. To achieve this, the LabelEncoder from the sklearn.preprocessing module was used.

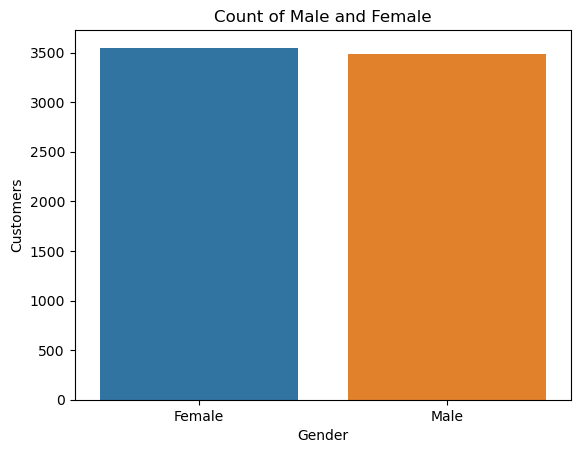
A custom function object\_to\_int was defined to transform the categorical columns in the DataFrame. This function checks if the data type of a series is 'object', which indicates a categorical variable, and then applies LabelEncoder to convert it into numerical values. If the data type is not 'object', the series is returned unchanged.



# Exploratory Data Analysis (EDA)

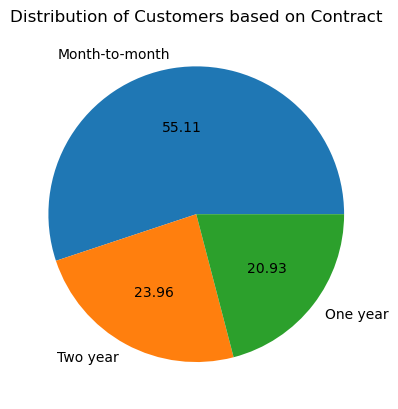
## Gender Distribution

Upon analyzing the dataset, it was found that the number of male and female customers is almost equal. This balanced distribution suggests that any gender-specific patterns observed in churn behavior can be considered representative of the overall customer base, without significant bias towards either gender.



#### Distribution of Customers Based on Contract

Analyzing the distribution of customers based on their contract type is crucial for understanding customer preferences and identifying potential factors that may influence churn. The provided pie chart illustrates the distribution of customers across three different contract types: Month-to-month, One year, and Two year

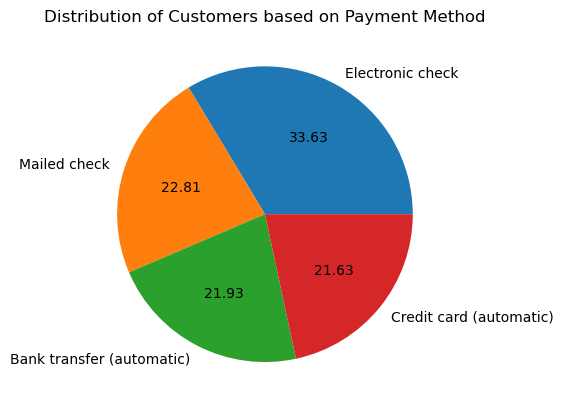
.

The pie chart shows the following distribution:

* **Month-to-month**: 55.11%
* **One year**: 20.93%
* **Two year**: 23.96%

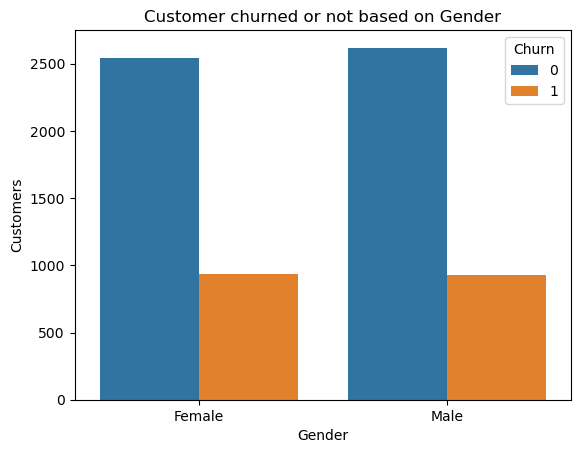
#### Distribution of Customers Based on Payment Method

Understanding the distribution of customers based on their payment methods provides valuable insights into customer preferences and behavior. This information can help telecom companies tailor their services and address potential churn more effectively. The pie chart below illustrates the distribution of customers across four different payment methods: Electronic check, Mailed check, Bank transfer (automatic), and Credit card (automatic).



## Churn Analysis Based on Gender

1. **Non-Churned Customers**:
   * Both female and male customers show a similar pattern in the number of non-churned customers.
   * Approximately 2,500 female and 2,500 male customers did not churn, indicating a nearly equal retention rate across genders.
2. **Churned Customers**:
   * The number of churned customers is also similar between genders.
   * There are about 1,000 churned customers for both females and males, suggesting that the likelihood of churn is not significantly different between genders.



## Churn Analysis Based on Internet Service

The graph shows the number of customers who churned based on their internet service. The y-axis represents the number of customers and the x-axis represents the customer’s internet service (DSL, Fiber Optic, or No Internet Service).

* Customers who have DSL internet service churn the least at around 450.
* Customers who have Fiber Optic internet service churn a little more than those with DSL is more than 1300.
* Customers who have no internet service churn is less than 200.

