Reducing Telecom Customer Churn: A Predictive Analytics Approach

# **Problem Statement:**

Customer churn is an enormous challenge for telecom firms as consumers switch to competitors or end their contracts early. The task is to create predictive models capable of identifying and comprehending the aspects contributing to client attrition. This project will address the following questions:

1. What are the primary factors influencing telecom customer churn?

2. Can we create an accurate predictive model to anticipate which customers would churn?

3. How may this information be used to build targeted retention tactics and lower client churn?

# **Background:**

Customer retention is crucial in the telecommunications sector since it is exceptionally competitive and consumer-centric. The rising cost of client acquisition and the possible revenue loss associated with churn have increased the importance of solving this issue. The origins of this problem may be traced back to the rising availability of telecom service providers, which provides consumers with more options, as well as customers' growing expectations for improved service quality, competitive pricing, and tailored experiences.

# 

# 

# 

# **Significance:**

Reduced telecom customer turnover is critical for various reasons, including:

1. Economic Impact: Customer turnover results in considerable revenue loss since gaining new customers is more expensive than maintaining existing ones.

2. Market Competition: The telecom sector is very competitive, with several service providers competing for clients. Reduced turnover can provide you with a competitive edge.

3. Customer happiness: Churn is a reflection of discontent with services, and resolving it leads to increased customer happiness and loyalty.

4. Data-Driven Insights: Using predictive analytics to understand consumer behavior better may help telecom firms customize their services and marketing efforts more successfully.

# **Contribution:**

This initiative has the potential to benefit the telecom industry significantly:

1. Predictive Modeling: By developing accurate predictive models, this project will provide telecom companies with valuable insight into which customers are at risk of churning, allowing for pre-emptive retention efforts.

2. Data-Driven Decision-Making: The initiative will enable telecom firms to make data-driven choices by identifying and prioritizing the most significant churn drivers.

3. Cost Savings: Lowering churn rates can result in significant cost savings since businesses can devote more energy to maintaining existing customers rather than constantly gaining new ones.

4. Improved Customer Experience: Addressing churn reasons improves the customer experience, which can lead to long-term loyalty and advocacy.

**Data Source:**

This study's dataset was taken from Kaggle, a library of publicly available datasets for different data science and machine learning projects. The dataset was chosen for its relevance to the topic at hand, forecasting client attrition. Here are the data source's specifics:

• **Source:** Kaggle / IBM Sample Data Sets

• **Dataset Name**: Telco Customer Churn

• **Description**: This dataset contains customer information about their churn behavior and numerous factors such as subscribed services, customer account details, and demographic information.

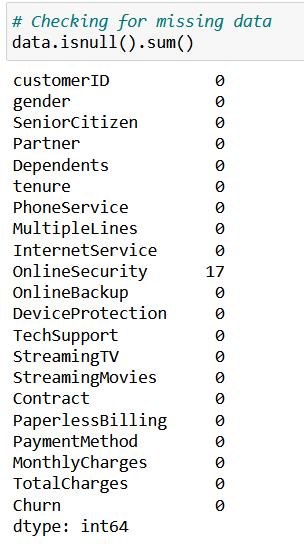
**Rationale for Dataset Selection**

The necessity for a comprehensive dataset that could be utilized to investigate client retention tactics drove the choice to use this dataset. Customers who departed within the last month are included in the dataset, as are the services they signed up for, account information, and demographic information. It has about 7000 rows and 21 columns, which provides a big enough sample size for significant data analysis and predictive modeling.

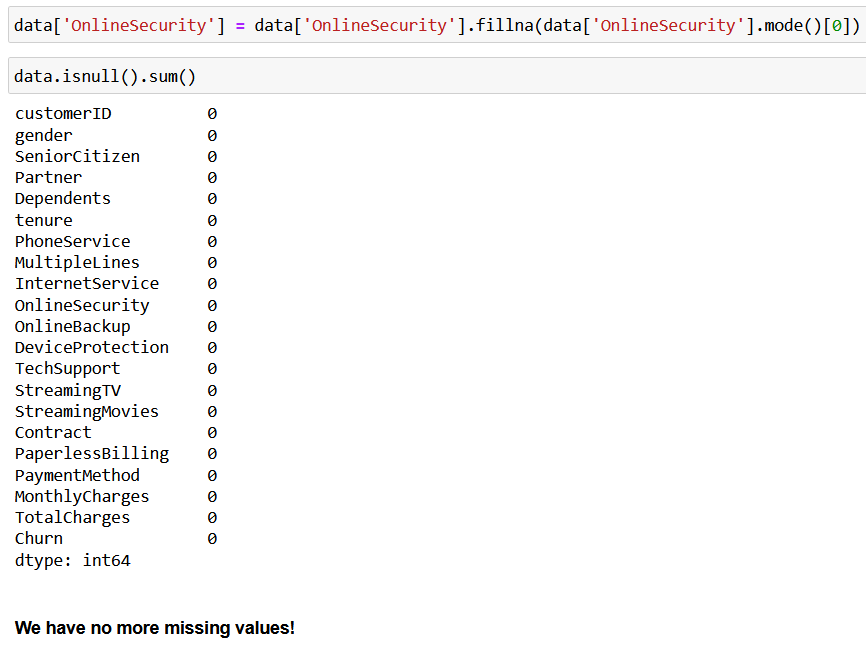
In summary, the dataset utilized in this study was obtained from IBM Sample Data Sets and was selected due to its relevance to customer churn analysis. It satisfied the criterion of having sufficient records (over 2000 rows) and the relevant columns to satisfy the Phase 1 objectives.

# **Data Cleaning**

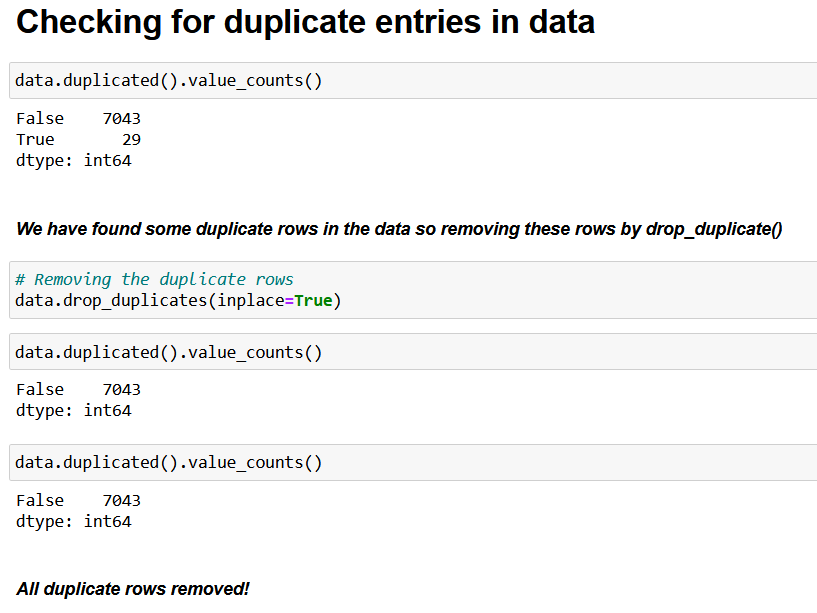
* **Dealing with missing data**
* We have missing data in the column 'OnlineSecurity.'



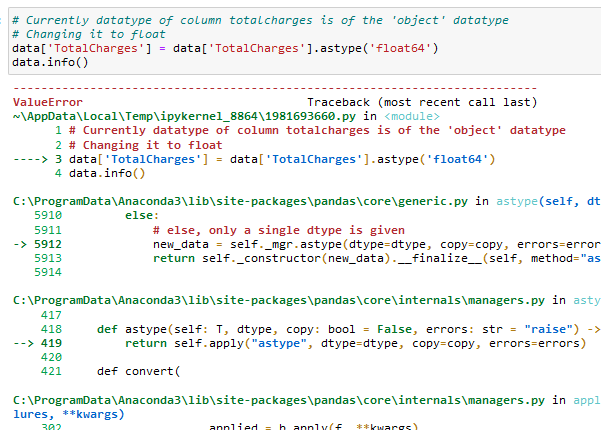
* Column OnlineSecurity is categorical data and 50% of the people have No as the categorical value so replacing missing values with No. Using the mode of the column for data imputation.

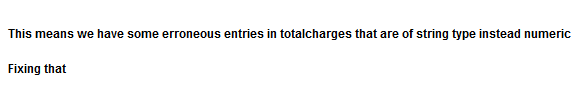


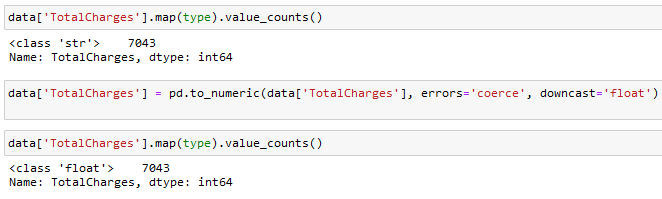
* **Removing Duplicate Entries**



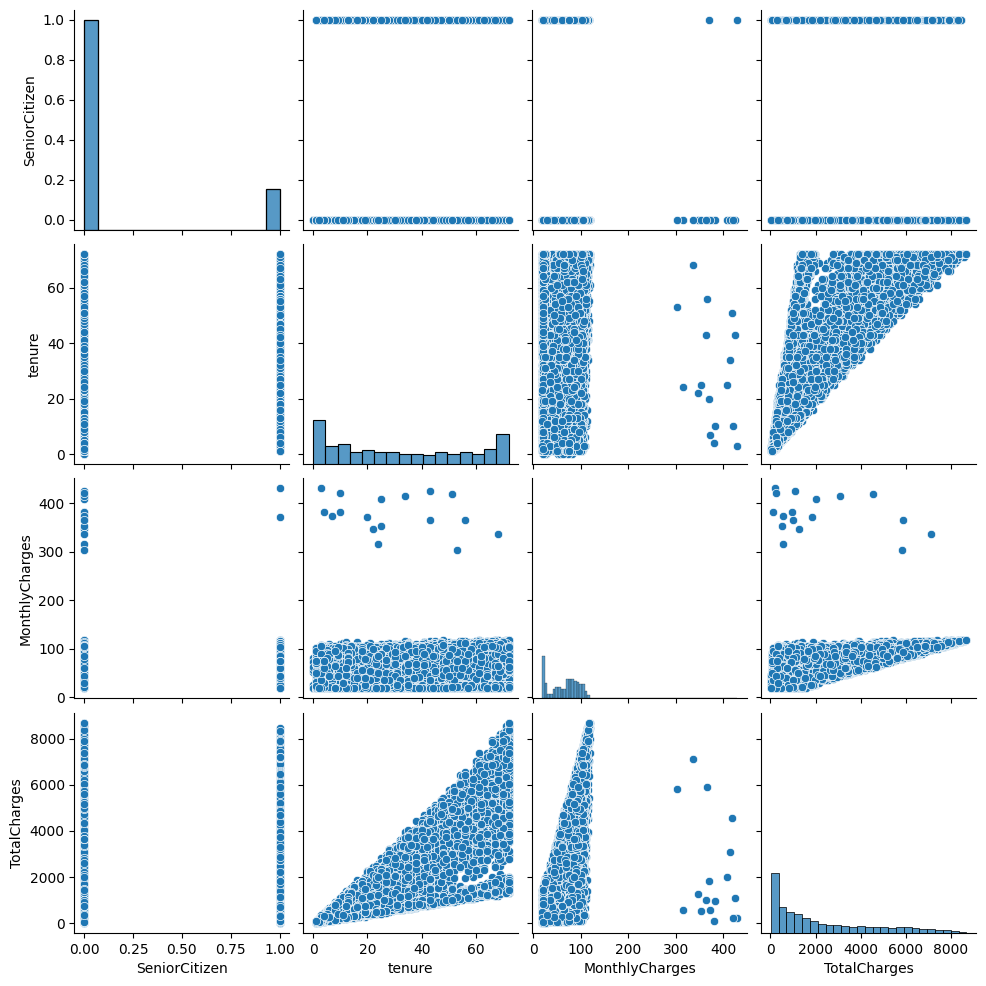
* **Fixing Inconsistencies in column**

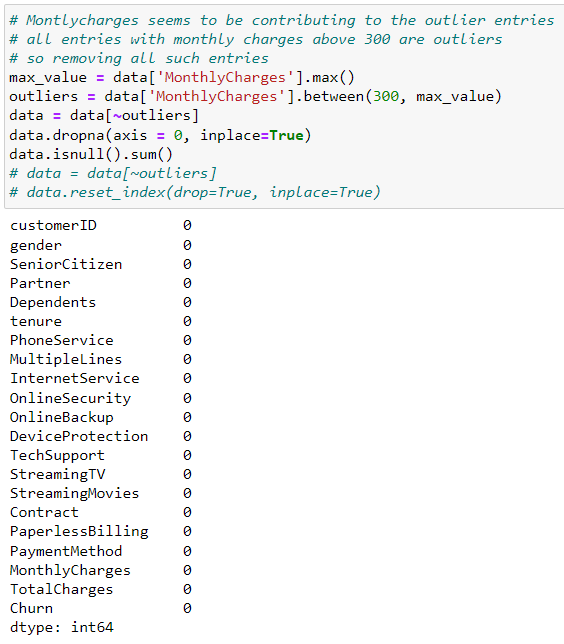




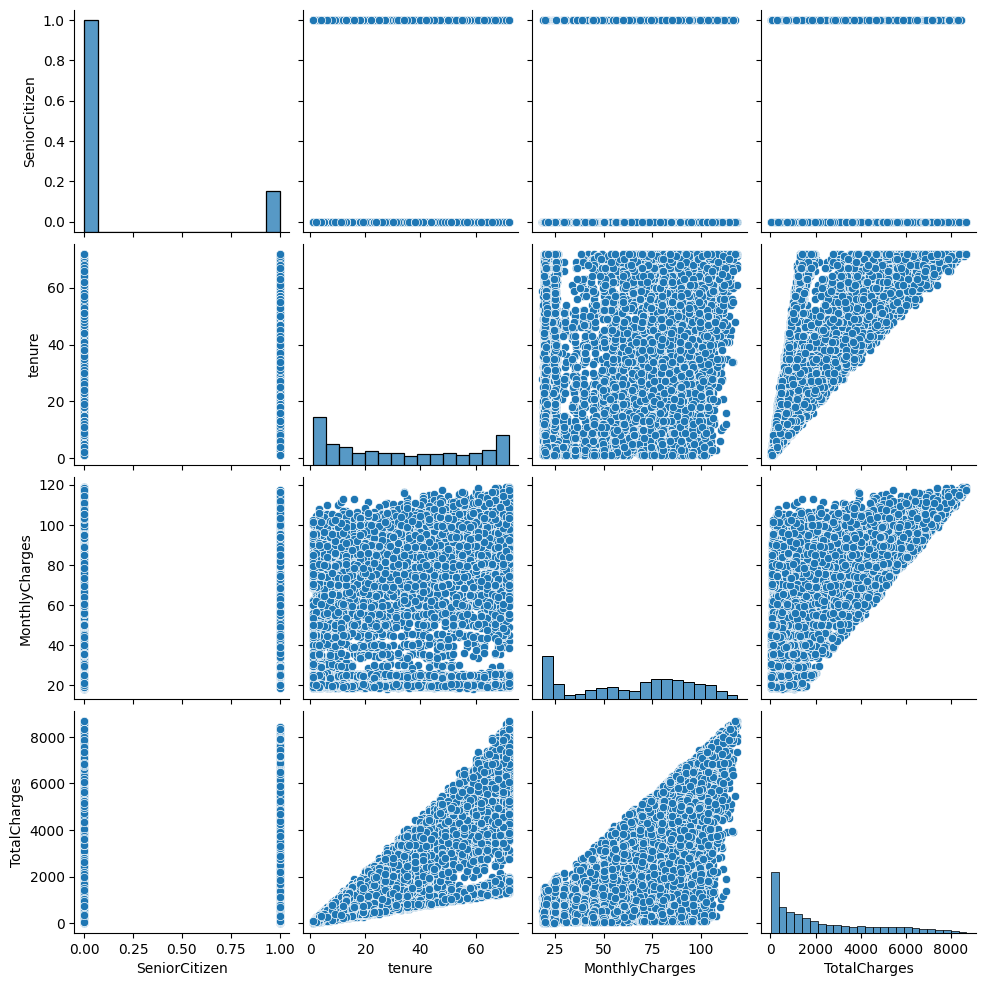


* **Removing Outliers**
* We have outliers in our data, which are arising due to high-paying customers.
* This can negatively affect our predictive analysis during the modeling phase, thus removing these outlier customers.

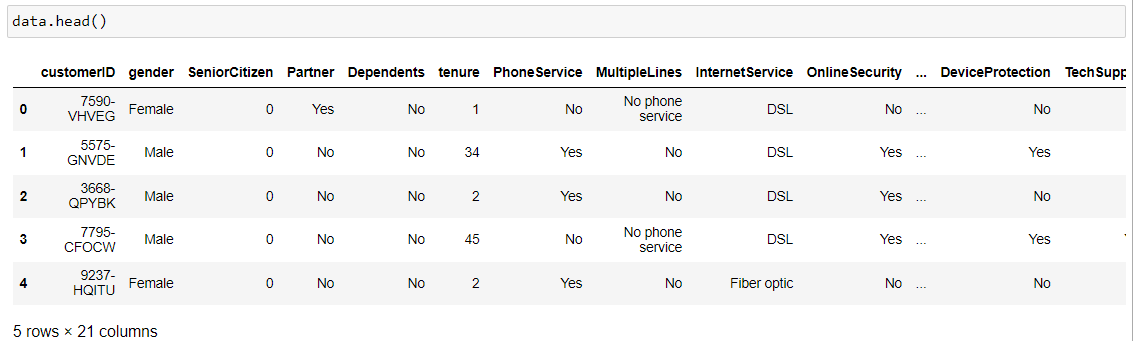


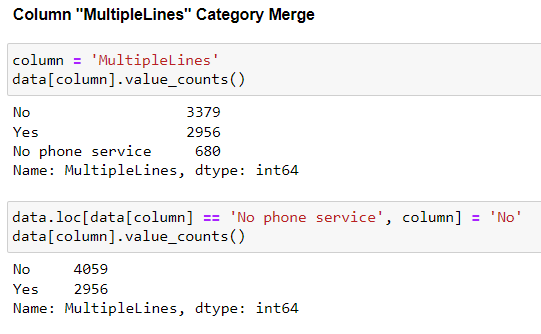


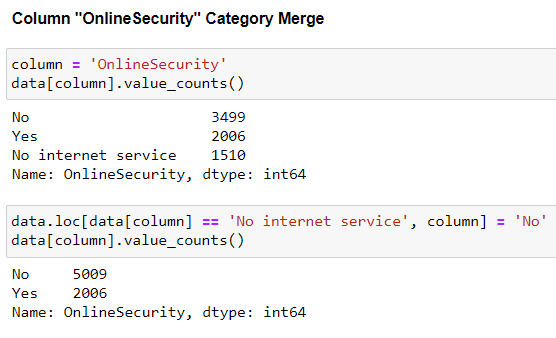
* After removing outliers, our graph looks like this:

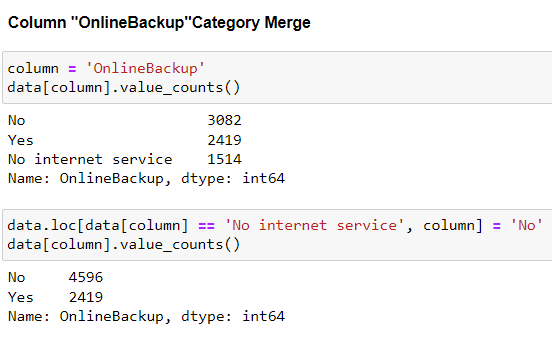


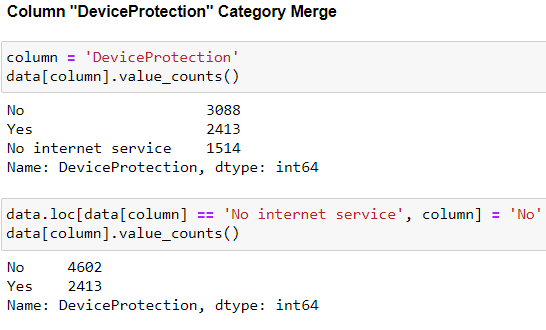
* **Addressing Inconsistencies in Categorical Data**
* Multiple columns have redundant category names that can be merged into one. This will reduce the complexity of the data.
* Before:



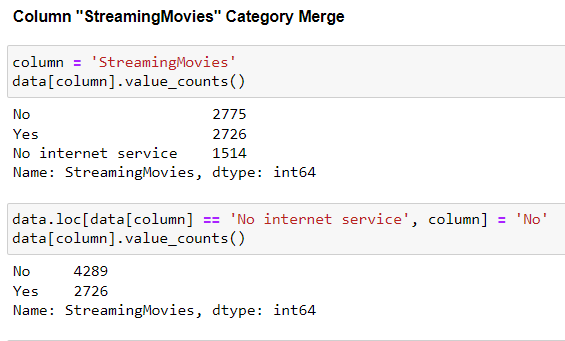




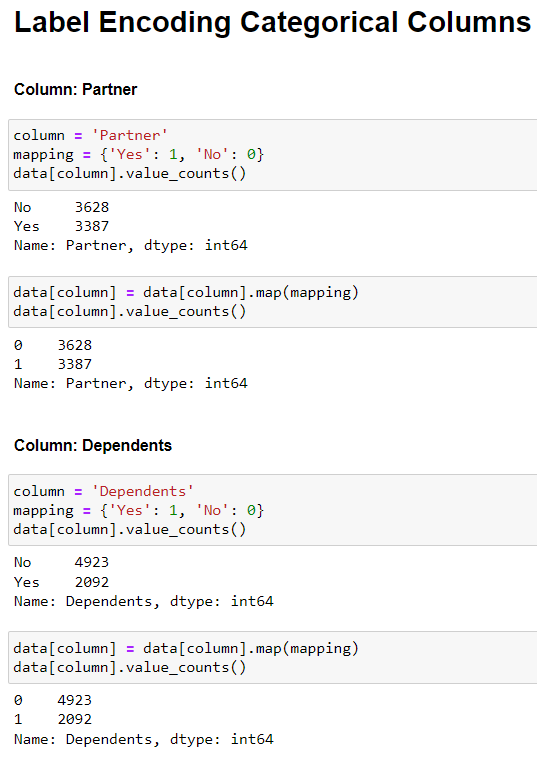


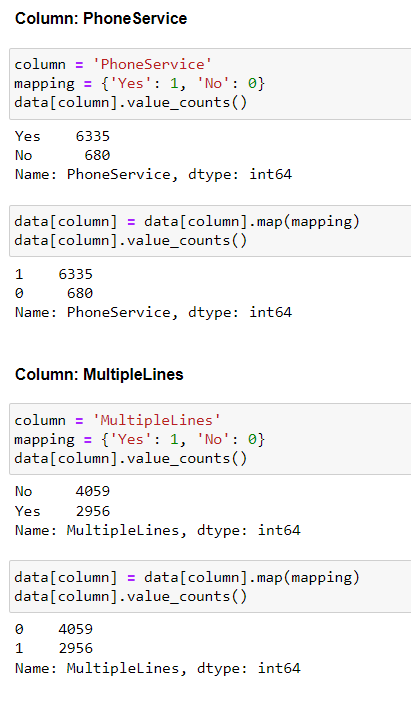


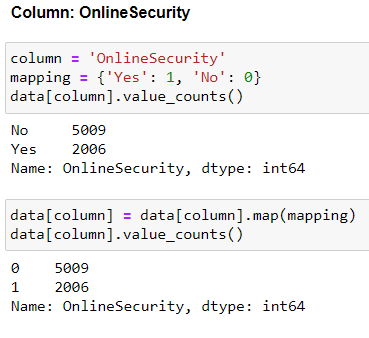


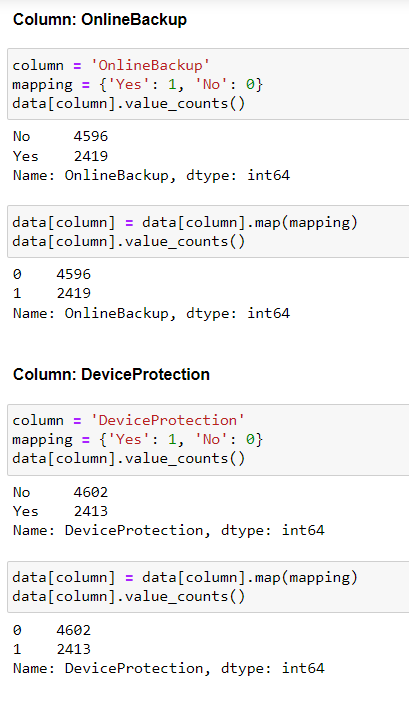


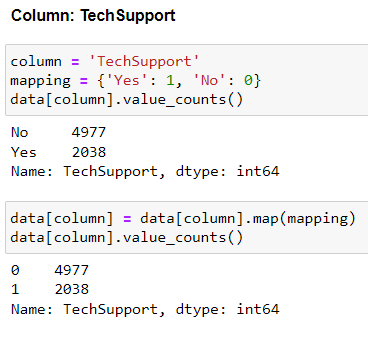
* **Label Encoding Categorical Columns**

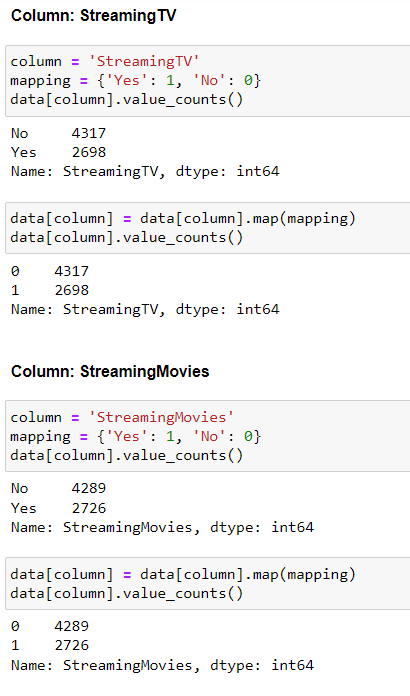


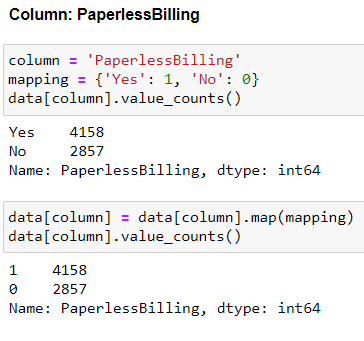


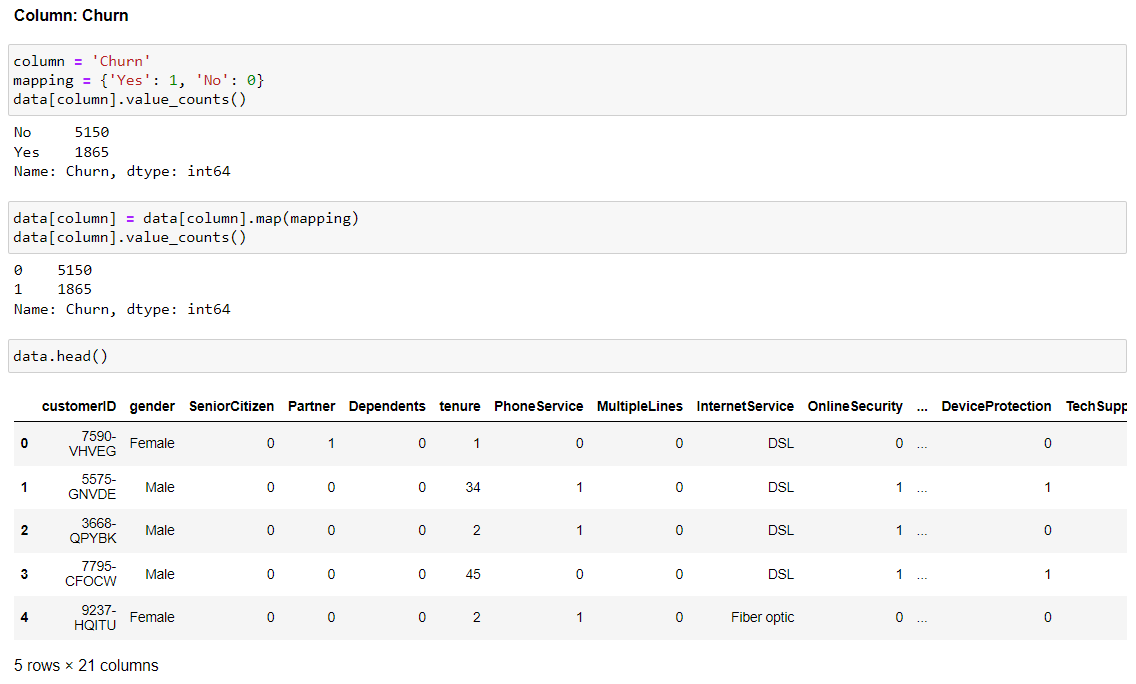




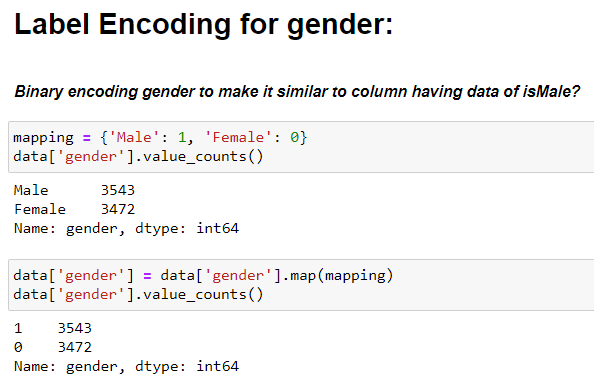






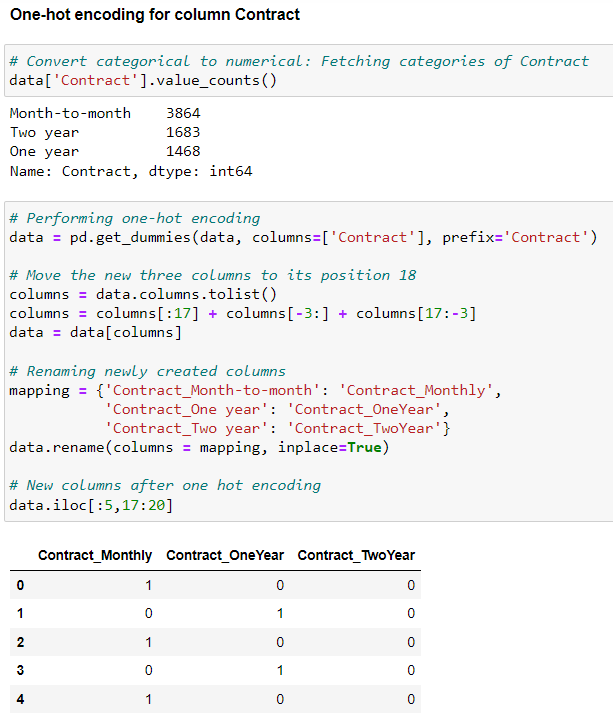


* **Binary Encoding for gender**



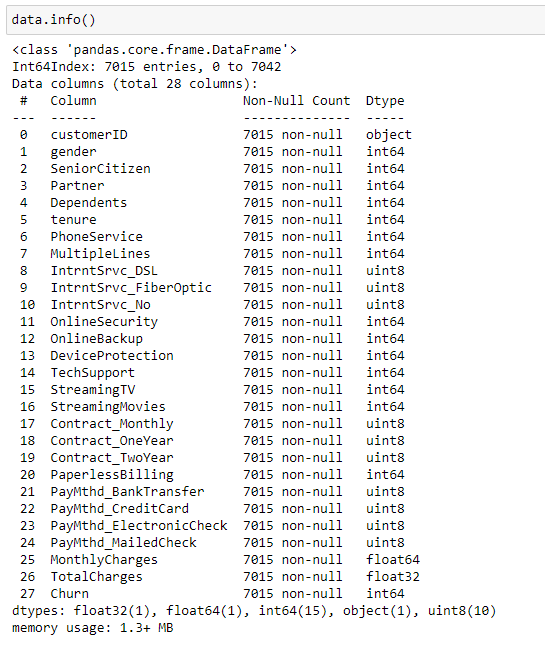
* **One - Hot Encoding**
* One-hot encoding is essential for classification jobs because it converts categorical data into a format machine-learning algorithms can understand. To make predictions in classification tasks, algorithms require numerical input characteristics, which one-hot encoding does by expressing category variables as binary columns. Each category is transformed into a distinct binary column, allowing the algorithm to consider each individually without indicating any ordinal connection. This encoding prevents the algorithm from assigning unwanted meaning or hierarchy to the categories and increases the model's ability to catch patterns and generate accurate predictions on categorical data.



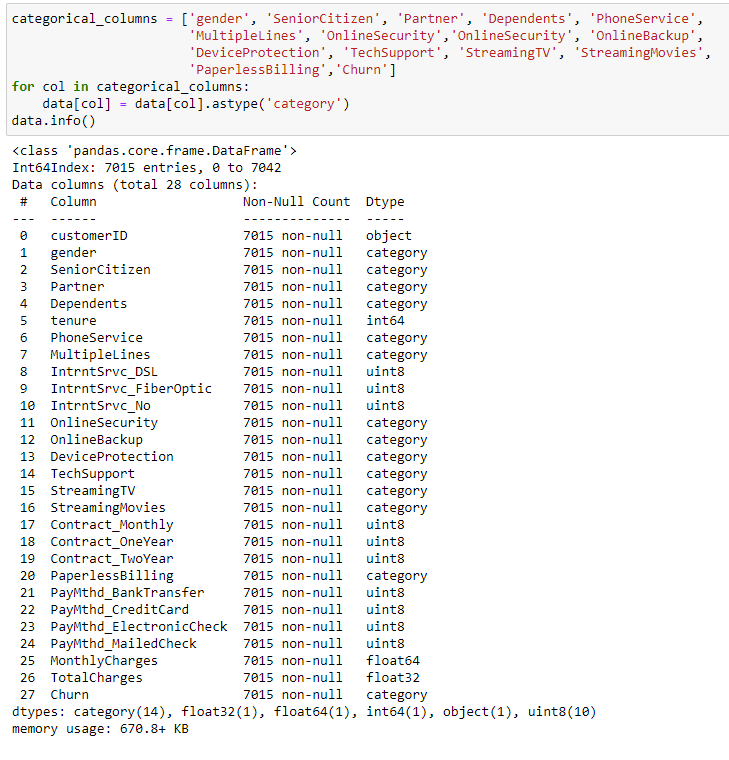




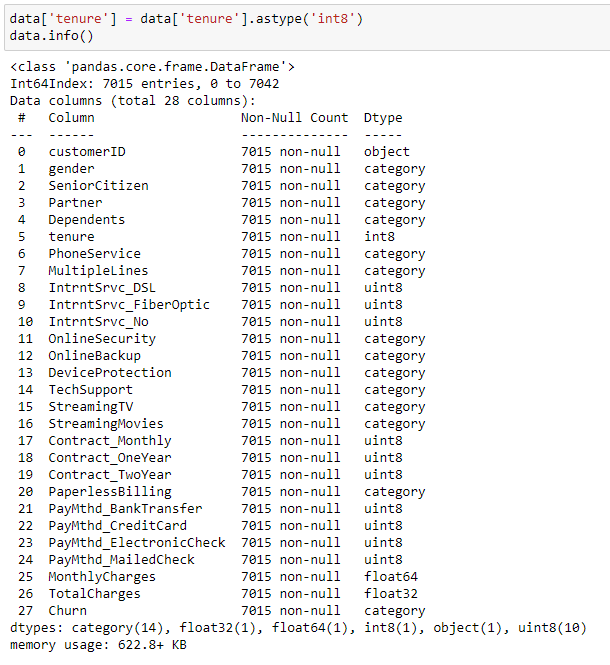
* **Typecasting Categorical columns to data type 'category'**
* Since many of our columns are of the categorical type, storing them as 'object' type is poor practice, thus converting them to the dtype 'category.'
* Before



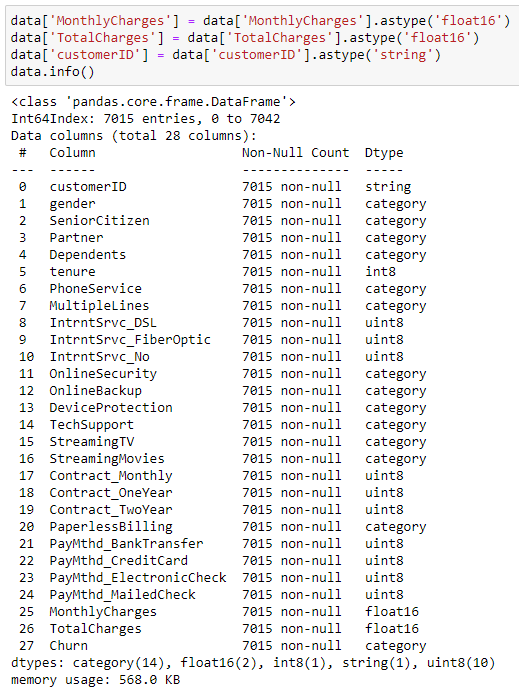
* After



* **Datatype downcasting**
* Storing data such as int64,float64, etc., not only requires more space but also increases processing times.
* In such scenarios, downcasting will stop the wastage of space and improve data processing times during the training phases of our predictive models.
* Before:



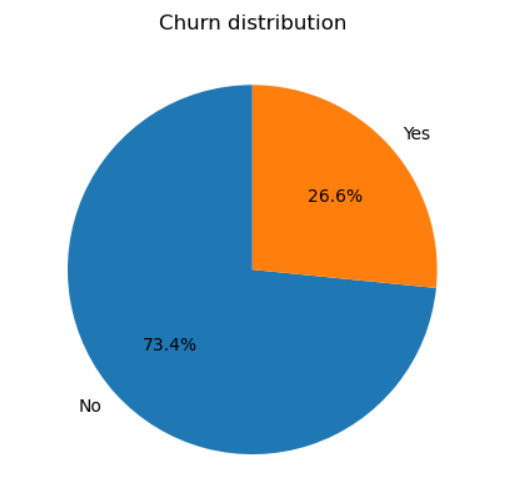
* After:



# **EDA**

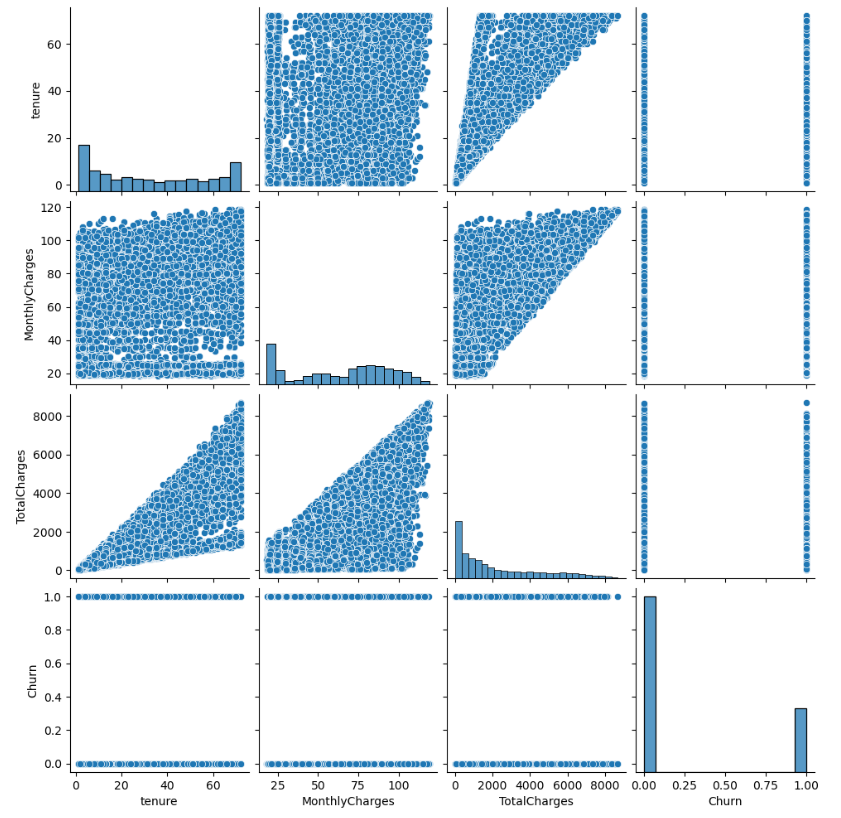
Exploratory Data Analysis (EDA) for the Telecom Customer churn dataset comprises initial data exploration, pattern identification, and understanding of significant factors affecting customer retention, i.e., customer churn. EDA provides insights such as the distribution of churn, correlations between features, and possible predictors for the churn utilizing graphical representations, analysis of data, and various plots, charts, and graphs.

# **Churn Distribution**



We can see from the churn distribution pie chart that 26.6% of customers churned while the rest 73.4% of the customers stayed with the company. It can be induced from this pie chart that more than a quarter of the customers have left in the past month.

# **Pair plots of some crucial features**



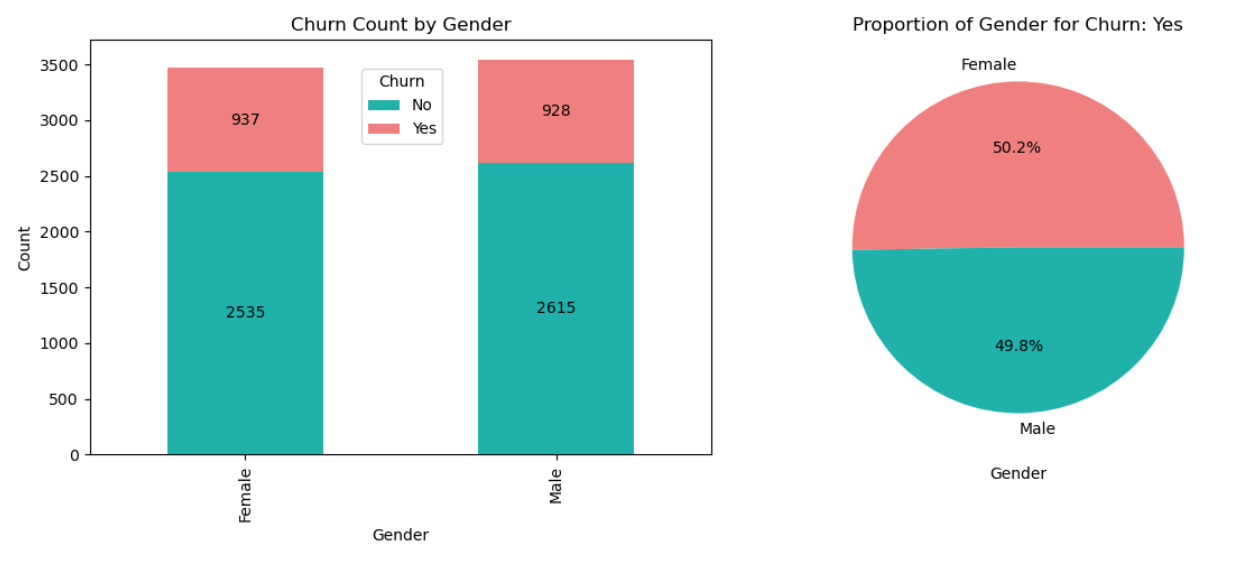
# **Customer's analysis**

The data has various features of the Customer. These features include their:

* **Gender**: Customer is male or female
* **Dependent**: Does the Customer have dependents or not
* **Partner**: Does the Customer have a partner or not
* **Senior Citizen**: Is the Customer is a senior citizen or not

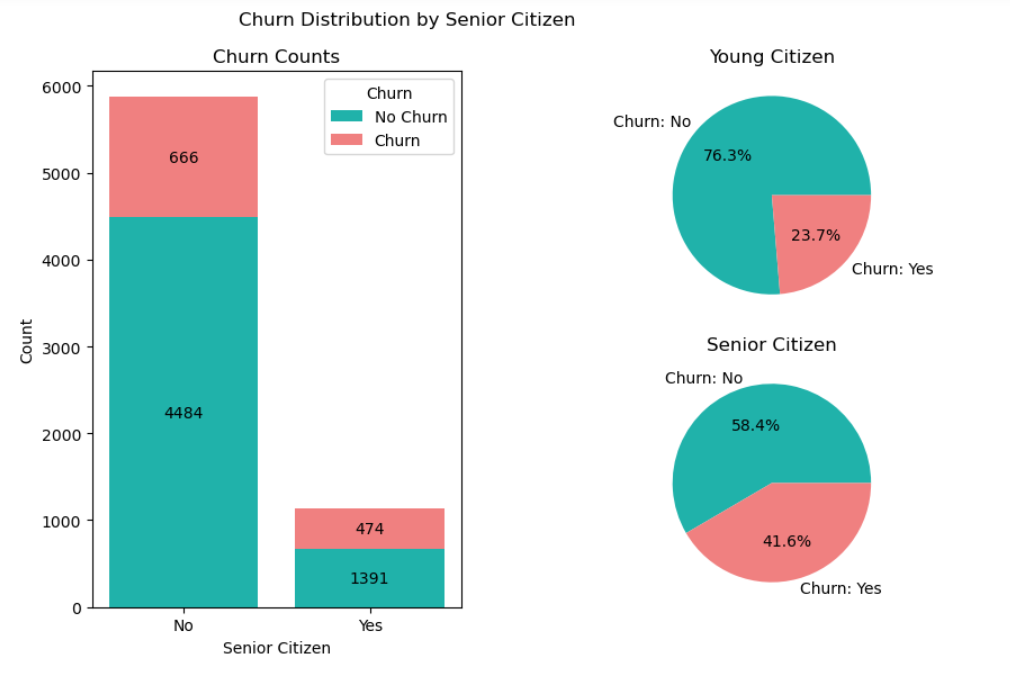
Analyzing each feature separately:

* **Churn distribution by Gender**



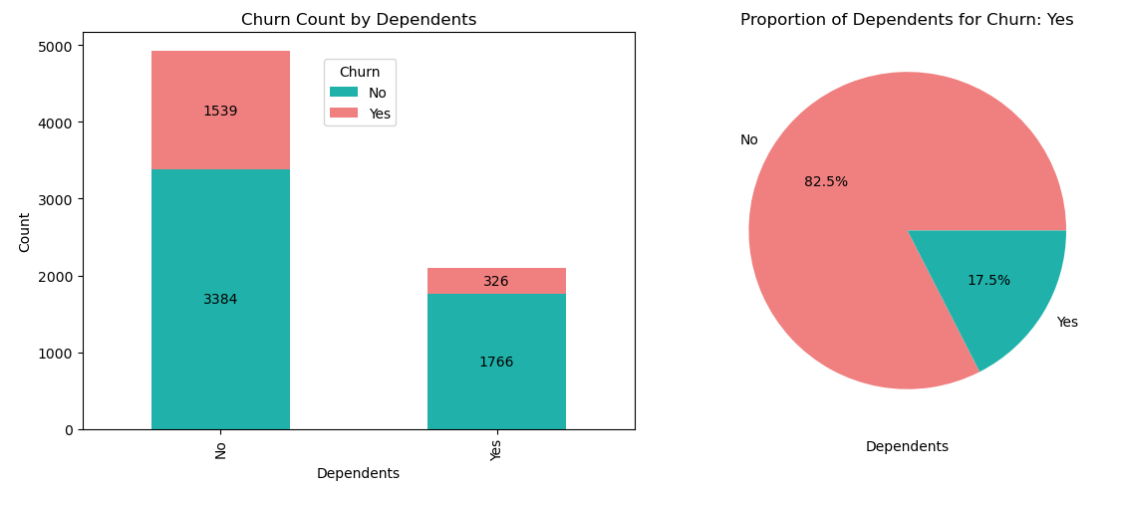
#### No particular trend was observed based on Gender for churn prediction. Both the genders are equally likely to churn.

* **Churn distribution by Senior Citizen**



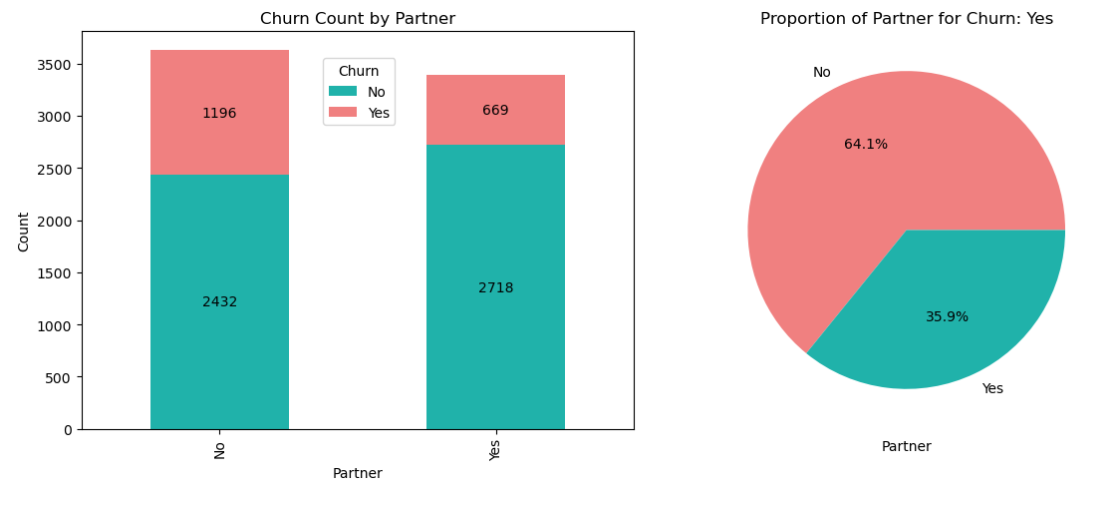
The bar plot unmistakably illustrates that young people, as opposed to senior citizens, have a higher contribution to churn in terms of count. However, the pie charts reveal a distinct pattern when we examine each category individually. Approximately 42% of senior citizens left, nearly double that of young citizens who left (Senior Citizens = No).

* **Churn distribution by Dependent**



#### The majority of the customers do not have dependents and customers who do not appear more likely to churn than those who have dependents.

* **Churn distribution by Partner**



Customers who do not have partners (single) appear to contribute more towards the churn.

# **Insights from Customer's Analysis**

It was discovered that gender and relationship status are pretty evenly distributed within the client base, with approximate percentage values based on the study. While females have a somewhat greater turnover rate, this difference is modest and may not be statistically significant.

When diving further into the details, though, a noteworthy tendency emerges. Younger consumers, consumers without partners, and consumers without dependents have a greater turnover rate. Based on the data study, these specific categories of the consumer population stand out as being more prone to churn.

Our findings, in particular, highlight the importance of non-senior citizens without partners or dependents as a separate client niche worthy of consideration when developing customer retention tactics.

# **Customer’s Subscription Service Analysis**

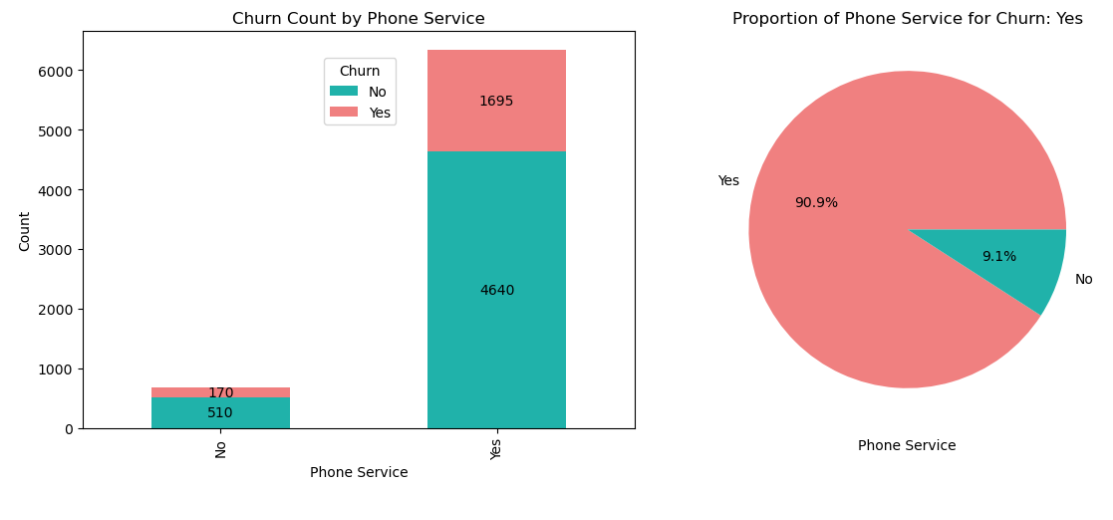
The dataset has various details of the various services subscribed by the Customer.

These subscription services include various columns as follows:

* **Phone Service**: Does the Customer have phone service or not
* **Multiple Lines**: Does the Customer have multiple lines or not if he has phone service
* **Online Security**: Does the Customer has online security or not if he has internet service
* **Online Backup**: Does the Customer have online backup or not if he has internet service
* **Device Protection**: Does the Customer have device protection or not if he has internet service
* **Tech Support**: Does the Customer have tech support or not if he has internet service
* **Streaming TV**: Does the Customer have streaming TV or not if he has internet service
* **Streaming Movies**: Does the Customer have streaming movies or not if he has internet service

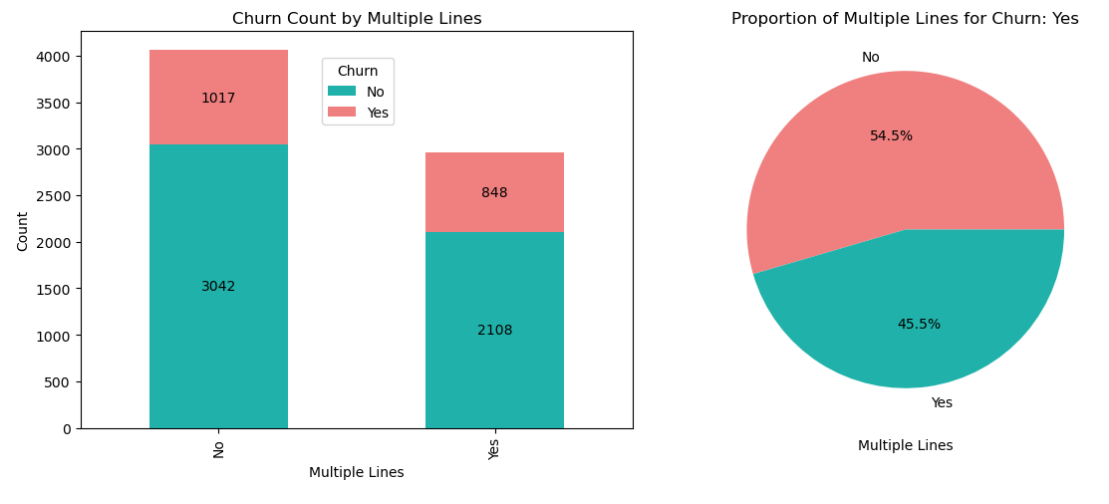
Analyzing each feature separately:

* **Churn distribution by Phone Service**



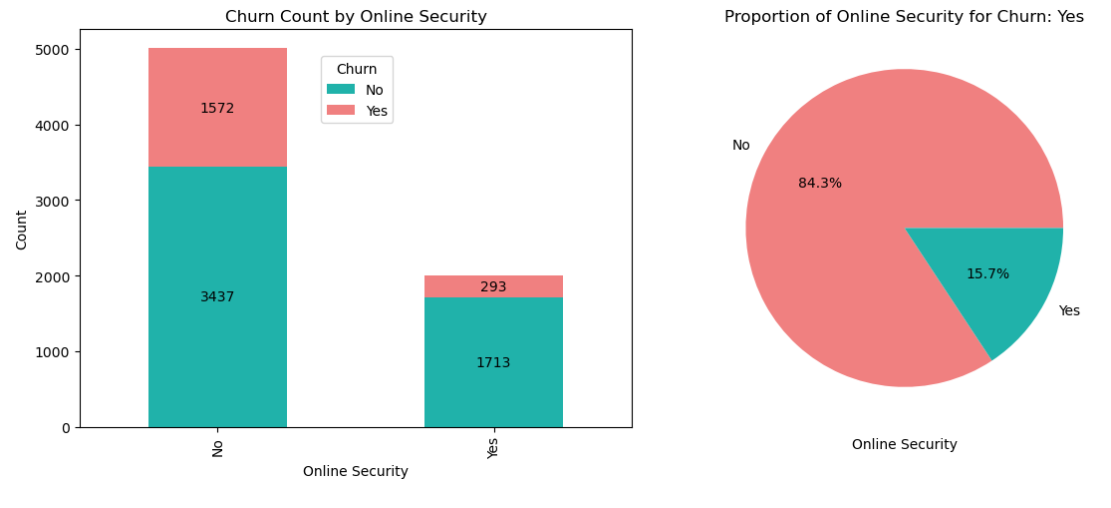
#### More than 90% of the customers have phone service.

* **Churn distribution by Multiple Lines**

****

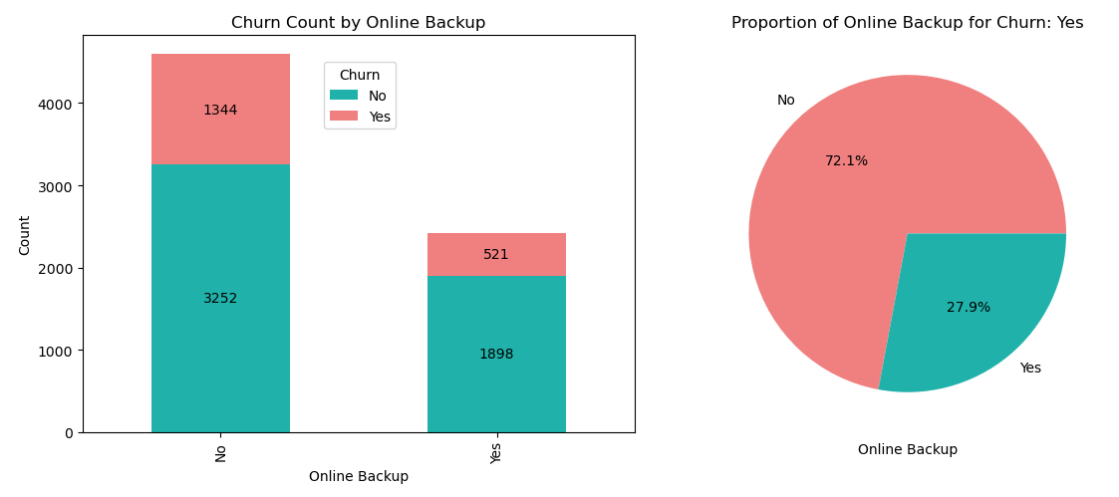
#### Customers who do not have multiple lines are more likely to be retained than those with multiple lines.

* **Churn distribution by Online Security**

****

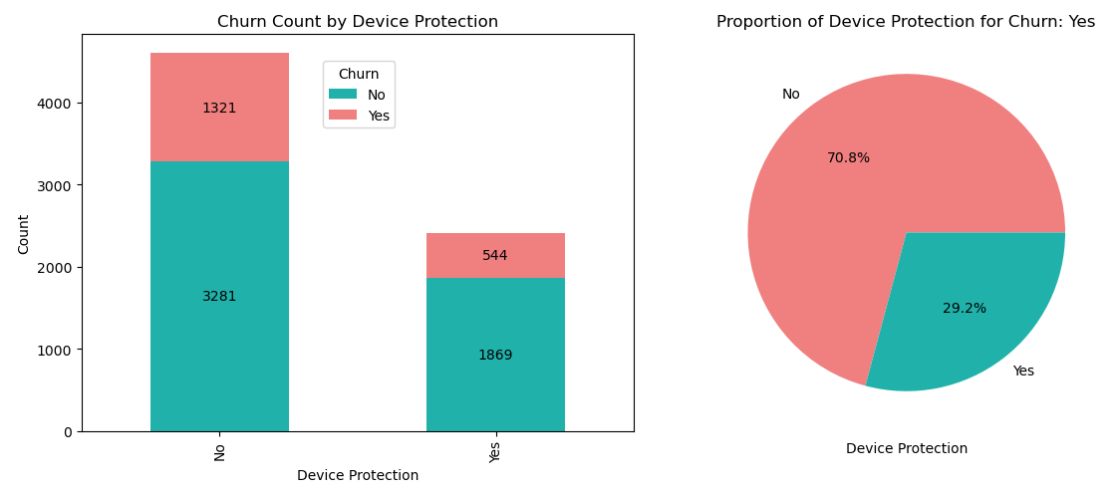
#### Roughly 5 out of 7 customers need online security, and these customers have higher chances of churning than those with online security.

* **Churn distribution by Online Backup**

****

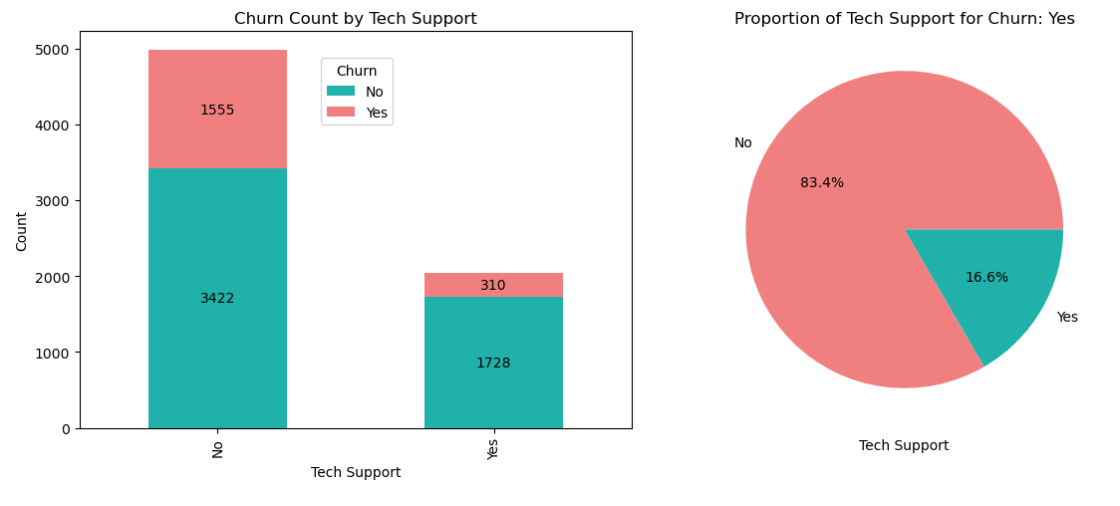
#### Just like Online security, customers who do not have online backup are churning more.

* **Churn distribution by Device Protection**

****

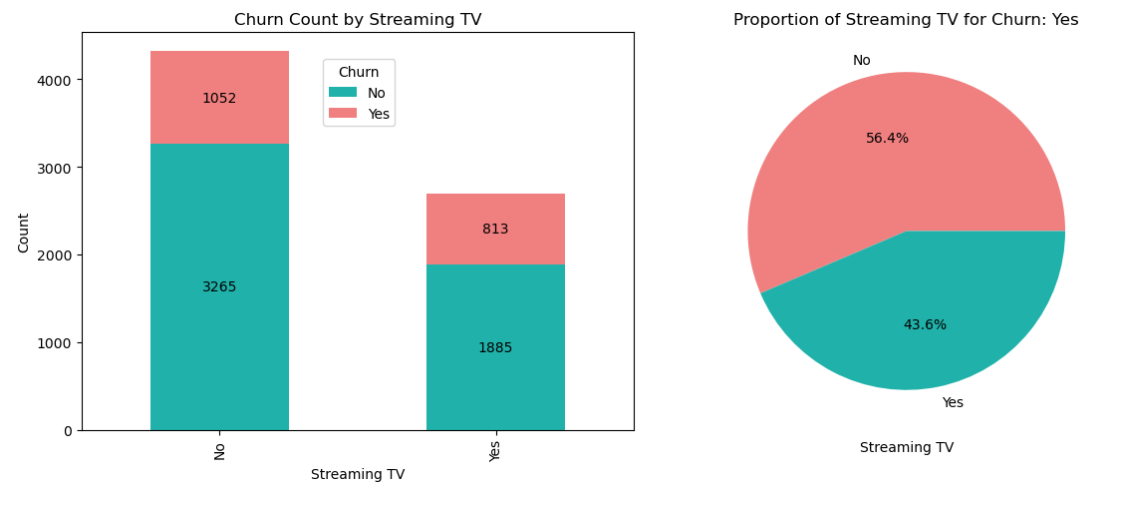
#### More than half of the customers do not have device protection, and such customers are more likely to churn.

* **Churn distribution by Tech Support**

****

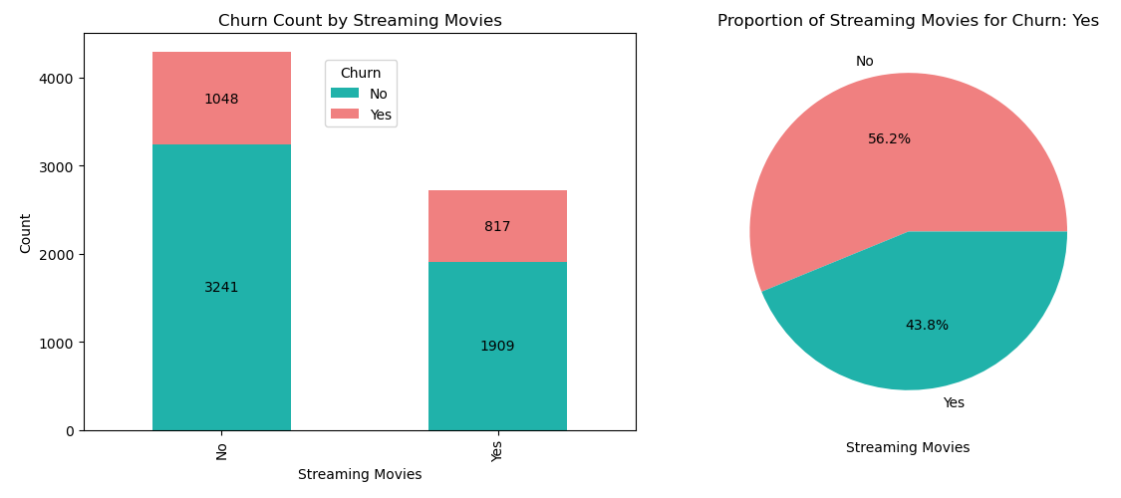
#### Approximately 5 out of 7 customers opt for something other than tech support. These customers are more likely to churn.

* **Churn distribution by Streaming TV**

****

#### More than half of the customers (exactly 61.5%) do not have Streaming TV service, and such customers are slightly more likely to churn than the rest.

* **Churn distribution by Streaming Movies**

****

#### For Streaming Movies, the same trend can be observed as streaming TV, where more than half (precisely 61.1%) of customers do not have Streaming Movies and are slightly more likely to churn.

# **Insights from Customer Subscription Service**

Our examination of customer service subscriptions found substantial differences among different service offerings. Notably, the following tendencies may be identified:

* **Dependency on Phone Service**: It should be noted that clients need phone service to have several lines. Phone services are used by about 90.3% of our consumers, and they have a higher turnover rate. This discovery may point to the necessity for more investigation into the causes of this unanticipated trend.
* **Fibre Optic Internet and Churn:** Customers who have chosen fiber optic as their internet service provider are more likely to churn. This can be attributable to various variables, including prospective price increases, greater competition, customer service quality, and other underlying causes. Notably, fiber optic connection is substantially more expensive than DSL, which may contribute to customer turnover.
* **Reduced Turnover Services:** Customers who have subscribed to extra services such as OnlineSecurity, OnlineBackup, DeviceProtection, and TechSupport, however, are less likely to churn. These services are essential to client retention, stressing their importance in customer retention tactics.
* **Neutrality of Streaming Services**: Surprisingly, the availability of streaming service subscribers does not predict attrition. This service is evenly distributed among consumers who select both "yes" and "no" choices, indicating that it is not a significant churn factor.

# **Customer's Contract and Payment Analysis**

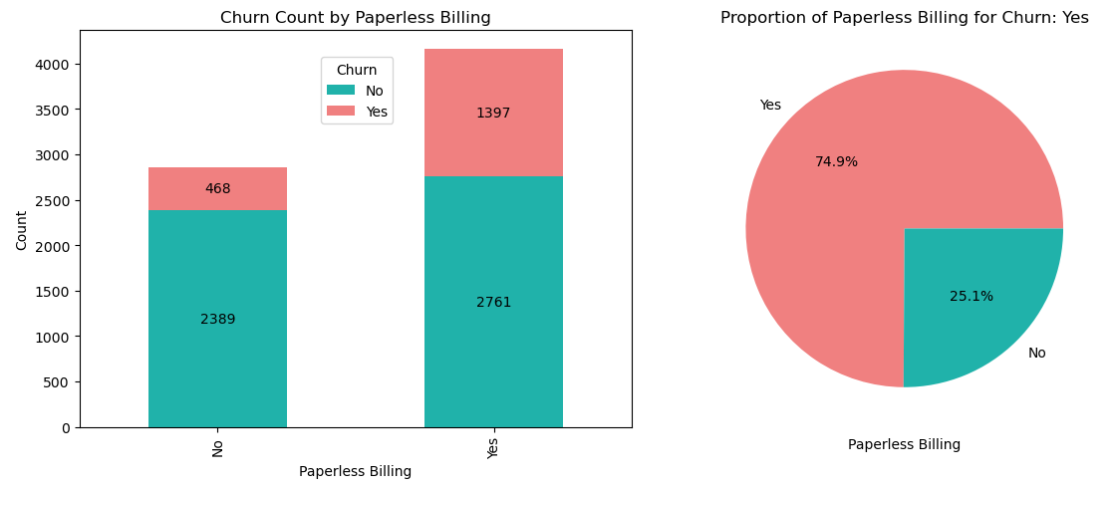
The data includes the Customer's contract duration and payment details.

These features include below columns:

* **Paperless Billing**: Does the Customer have paperless billing or not
* **Internet Services**: Customer's internet services provider (DSL et al., No ISP)
* **Contract**: Customer’s contract term (Month-to-Month, One Year, Two Year)
* **Payment Method**: Customer's Payment method (Electronic et al. (automatic), credit card (automatic))

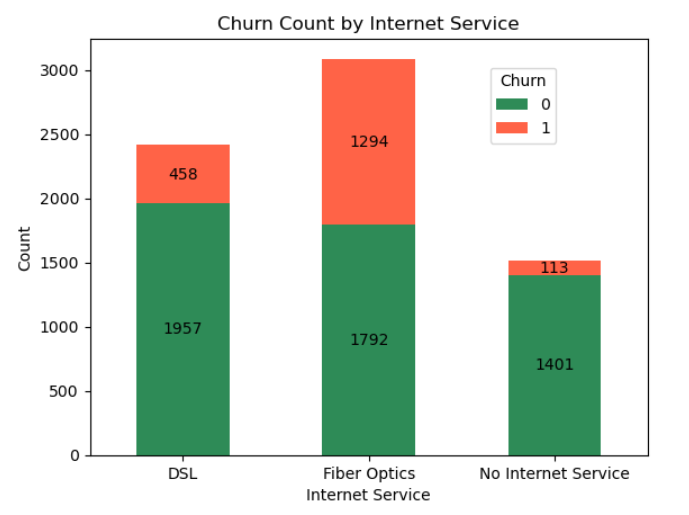
Analyzing each feature separately:

* **Churn distribution by Paperless Billing**

****

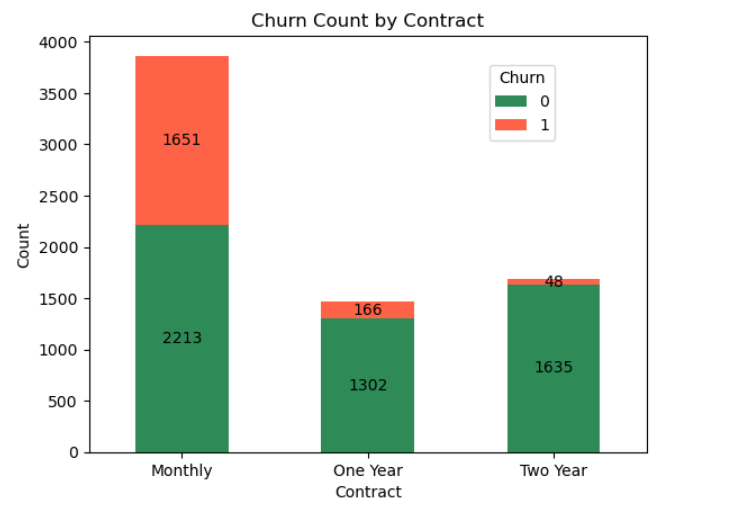
#### Nearly 6 out of 10 customers have gone for paperless billing. These customers are more likely to churn.

* **Churn distribution by Internet Services**

****

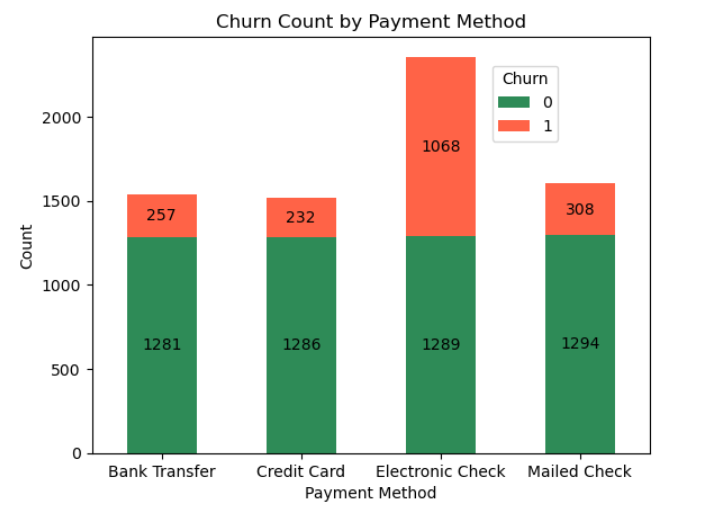
#### Customers who have Fiber Optics churned the most compared to people with DSL. However, people with no internet service are much less likely to churn.

* **Churn distribution by Contract**

****

#### Customers prefer short-term contracts (monthly contracts) to longer-term ones (one-year and two-year contracts). These short-term customers are majorly contributing to the churn. Customers with more extended Contract with the company are significantly less likely to churn.

* **Churn distribution by Payment Method**

****

#### Customers who pay through the Electronic check are more likely to churn than the rest of the payment methods.

#### 

# **Insights from Customer Contract and Payment**

Our examination of Customer's payment and contracts with the company revealed below trends:

* **Contract Length and Churn:** One intriguing finding is the negative association between contract length and turnover rate. Customers who have shorter contract terms are more likely to leave. On the other hand, those with longer-term obligations face extra obstacles when seeking to cancel early. This research emphasizes the need to develop long-term client connections to lower churn rates since such ties appear more robust.
* **The Impact of Paperless Billing:** It is worth noting that clients who choose paperless billing have a greater turnover rate. Paperless billing has been implemented by about 59.2% of our clients. The reasons for this correlation need further examination since it gives insight into consumer billing preferences and habits.
* **Electronic Checks and Churn**: According to one fascinating result, customers who pay with electronic checks are more likely to churn. This payment type is popular among our clients. Understanding the causes behind this correlation might be critical in devising ways to decrease attrition among electronic check consumers.

# **Customer’s account information analysis**

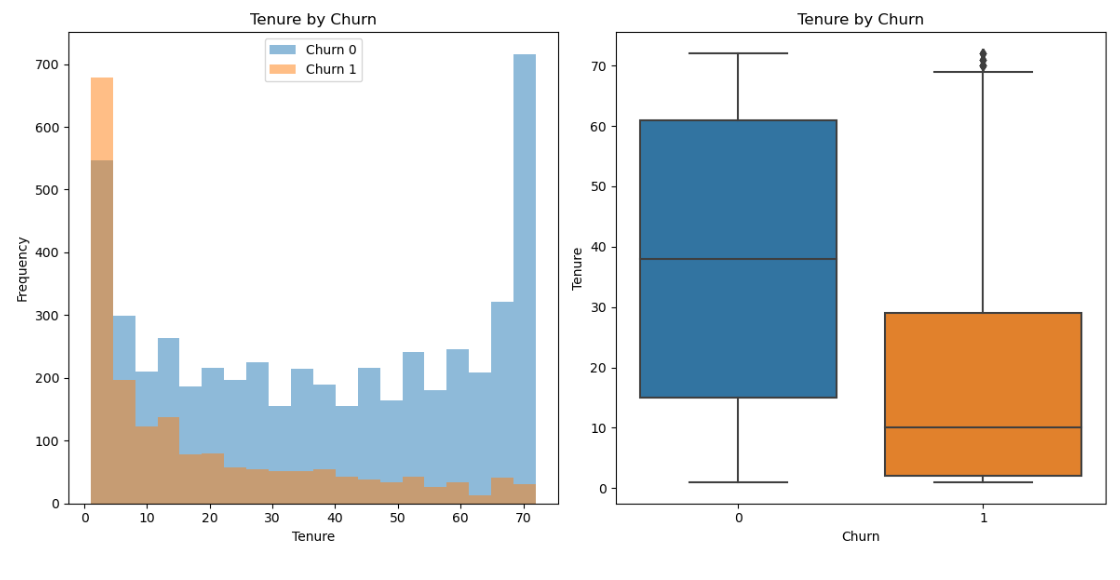
The data has features related to the account information of the Customer.

These features include their:

* **Tenure**: Number of months the Customer has stayed with the company
* **Monthly Charge**: The amount charged to the Customer monthly
* **Total Charge**: The total amount charged to the Customer

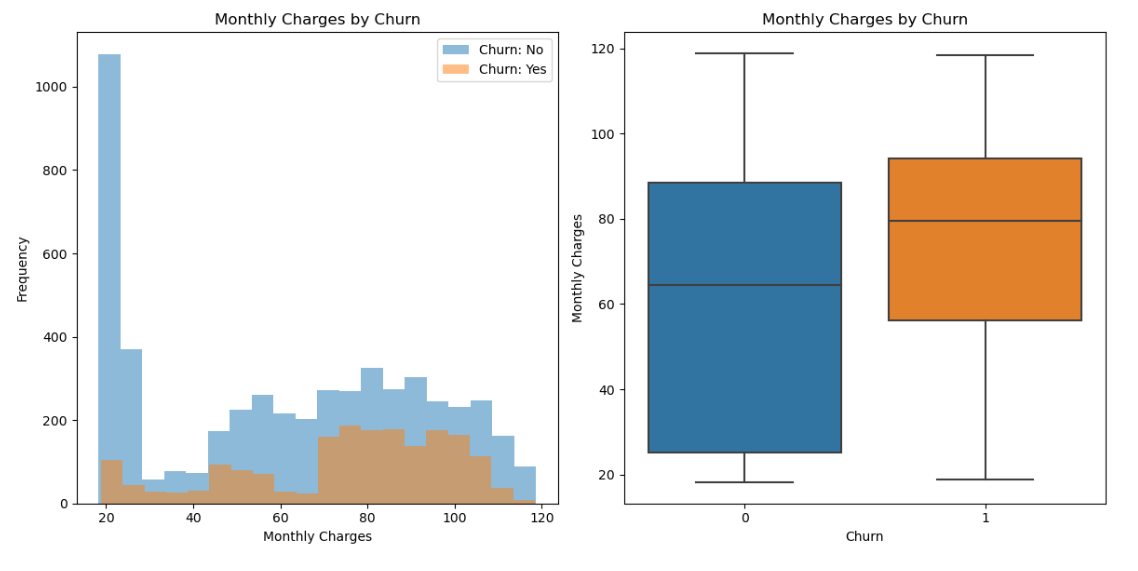
Analyzing each feature separately:

* **Churn distribution by Tenure**

****

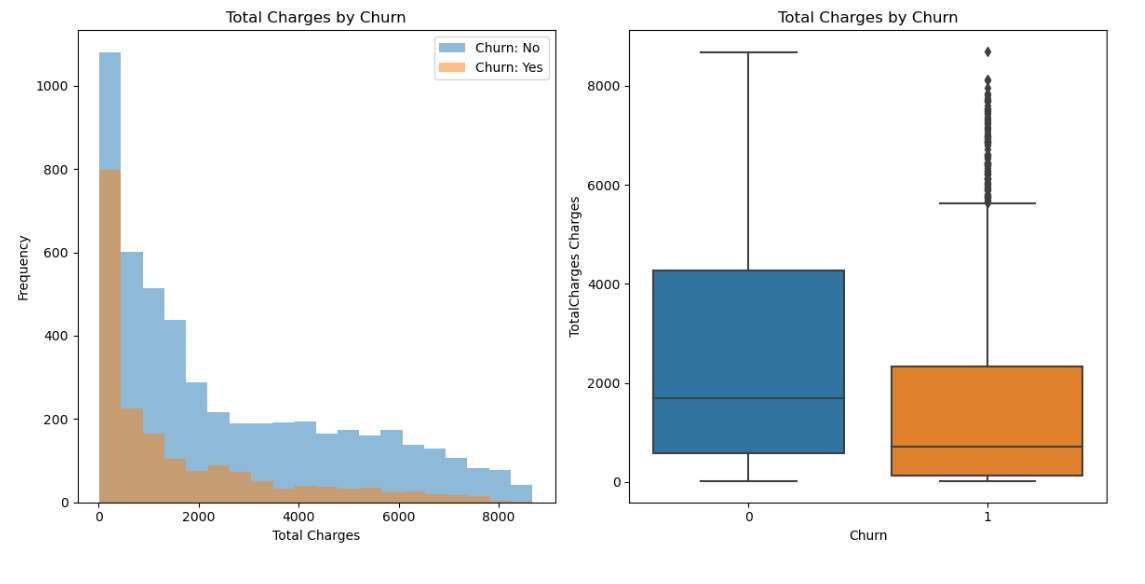
#### It is evident from the graphs that once people stay more than 20 years, then they are less likely to churn based on the Tenure of the customers compared to their churn rate.

* **Churn distribution by Monthly Charge**

****

#### Monthly charges are directly proportional to the churn rate, meaning the lower the monthly charges the Customer pays, the less likely the Customer is to churn.

* **Churn distribution by Total Charge**

****

#### The churn rate is inversely proportional to the total charges the Customer pays. That is, the higher the total charges paid by the Customer lower the chances of their churning.

# **Insights from Customer account information**

Our examination of the Customer's account information has shown the below trends:

* **Tenure Distribution:** The customer tenure histogram shows a right-skewed distribution, indicating that most consumers have only been with the telecom business for the first few months (0-9 months). This realization emphasizes the significance of efficiently maintaining consumers throughout their first few months of involvement.
* **Churn Timing**: Surprisingly, the most significant percentage of churn happens within the first few months (0-9 months). This discovery underscores the critical period during which customer retention efforts should be concentrated to minimize churn rates effectively.
* **Early Churn Concentration**: One significant conclusion is that around 75% of consumers who eventually quit the Telco firm do so during their first 30 months of employment. This statistic emphasizes the importance of early client interaction and satisfaction in developing long-term connections.
* **Monthly Fees and Churn:** Our examination of the monthly charge histogram indicates an interesting pattern. Customers who pay more significant monthly fees are more likely to leave. This implies that discounts, promotions, or competitive pricing effectively motivate customers to stay loyal. Pricing methods that consider these findings may be beneficial in keeping consumers.