Customer Churn Analysis Case Study for a Telecom Company

Sajan Bopaiah

Business Understanding

The dataset includes statistics on telecom provider customers, including whether or not they have churned (i.e., discontinued their subscription). This analysis's goal is to find trends and key information that will help us understand why customers leave. We will perform Exploratory Data Analysis to check the trends and insights.

Data Collection and Description

The dataset contains 21 variables and 7,043 observations. There are both numerical and categorical variables. The variables include demographic information such as age, gender, and income, as well as behavioral information such as length of tenure, internet service type, and contract type. The target variable is "Churn", which indicates whether the customer has churned or not.
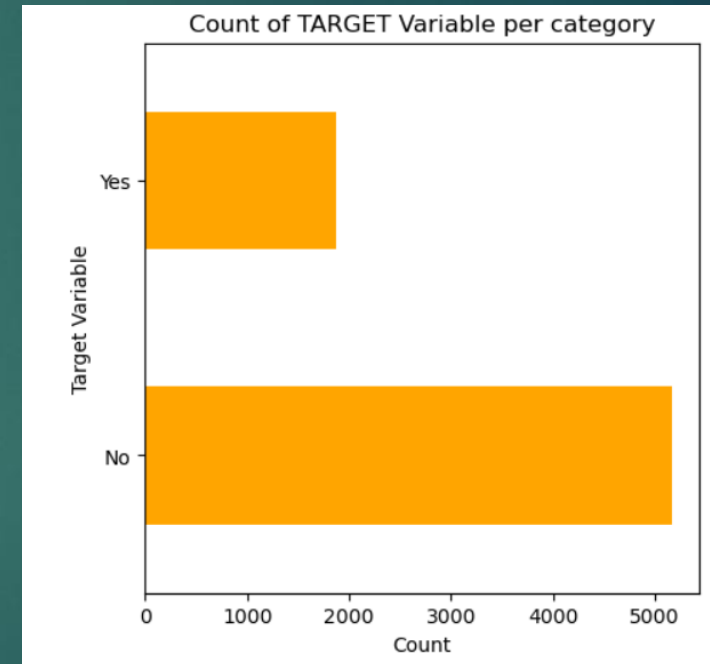
Understanding the Data

Target Variable

Findings

- Data is highly imbalanced, ratio = 73:27
- We will analyze the data with other features taking the Target values separately to get some insights.

```
No      73.463013
Yes     26.536987
Name: Churn, dtype: float64
```


Count of TARGET Variable per category

Missing Data - Initial Intuition

● Here we don't have any missing value.

● If there are fewer missing values, we use regression to predict the missing values or fill the missing values with mean of the values present.

● If there are more missing values, it is better to drop those columns as they give less insight to the analysis.

● Also, we can see that total charges is under object datatype. Next steps would be to convert the datatype to integer or float.
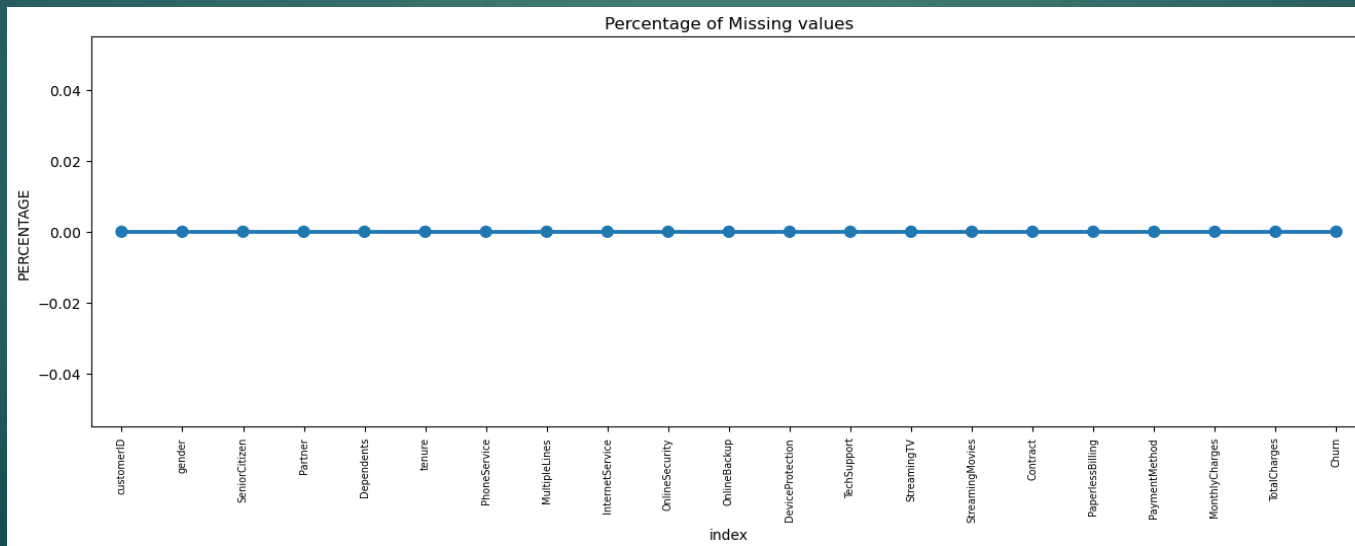
Data Cleaning

● Here we are converting the total charges from categorical to numerical data type using the following:

```python
new_df.TotalCharges = pd.to_numeric(new_df.TotalCharges, errors='coerce')
```

```
 #    Column            Non-Null Count   Dtype
---   ------            --------------   -----
 0    customerID        7043 non-null    object
 1    gender            7043 non-null    object
 2    SeniorCitizen     7043 non-null    int64
 3    Partner           7043 non-null    object
 4    Dependents        7043 non-null    object
 5    tenure            7043 non-null    int64
 6    PhoneService      7043 non-null    object
 7    MultipleLines     7043 non-null    object
 8    InternetService   7043 non-null    object
 9    OnlineSecurity    7043 non-null    object
 10   OnlineBackup      7043 non-null    object
 11   DeviceProtection  7043 non-null    object
 12   TechSupport       7043 non-null    object
 13   StreamingTV       7043 non-null    object
 14   StreamingMovies   7043 non-null    object
 15   Contract          7043 non-null    object
 16   PaperlessBilling  7043 non-null    object
 17   PaymentMethod     7043 non-null    object
 18   MonthlyCharges    7043 non-null    float64
 19   TotalCharges      7032 non-null    float64
 20   Churn             7043 non-null    object
dtypes: float64(2), int64(2), object(17)
memory usage: 1.1+ MB
```
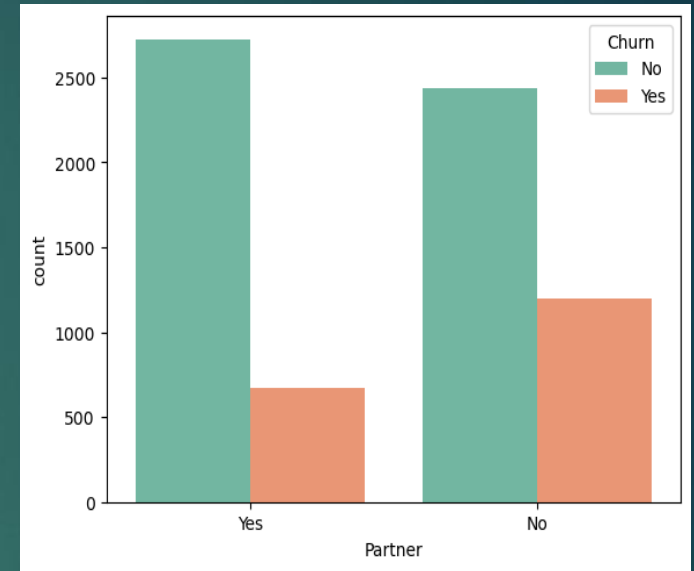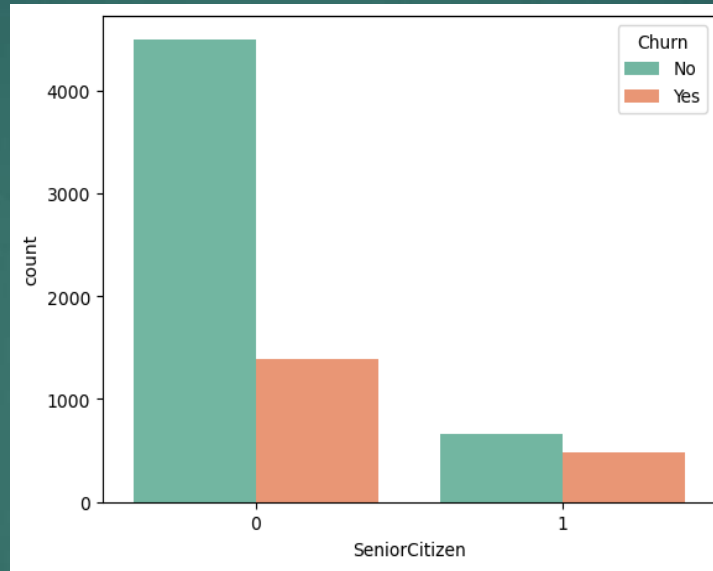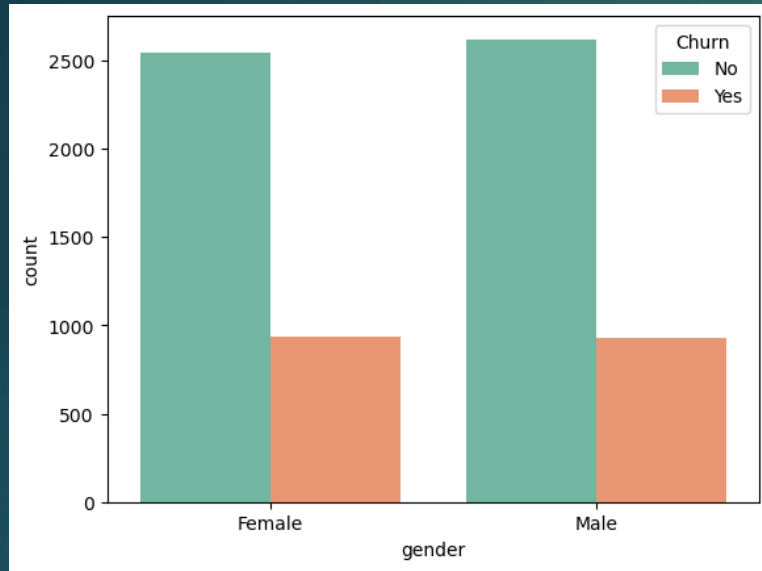
Percentage of Missing Value

● We see that we have 11 missing values present in Total
Charges column.
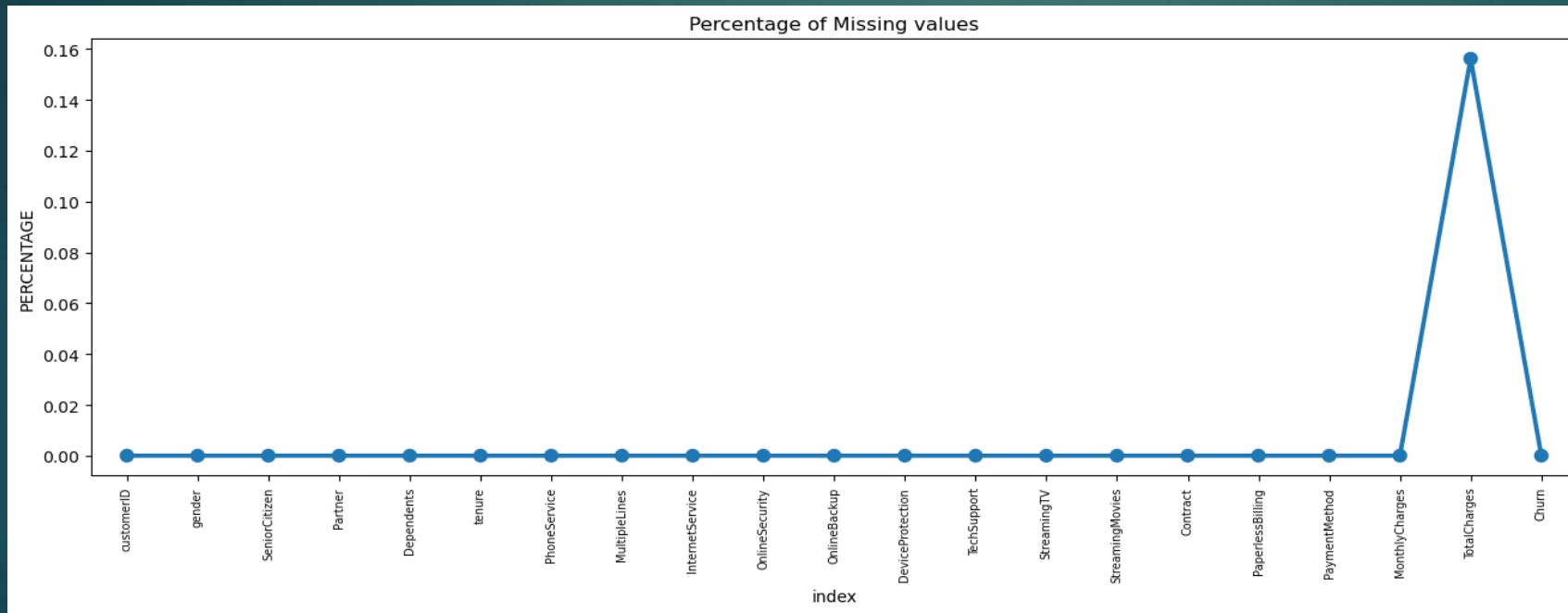
● Since the missing value percentage is ~0.15% it is safe
to ignore them by dropping the missing values..
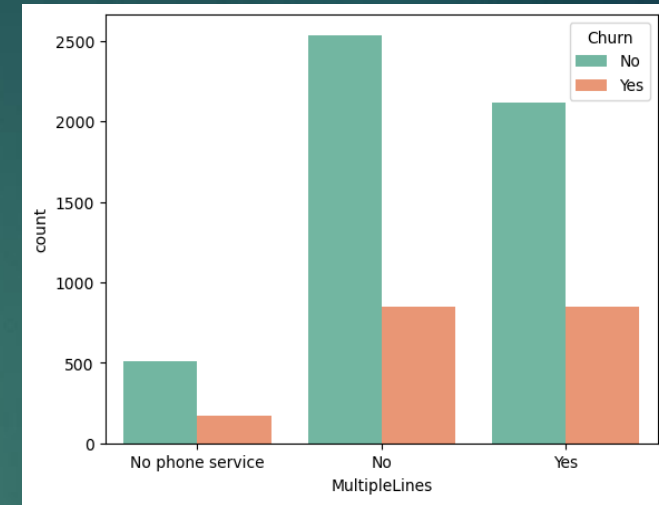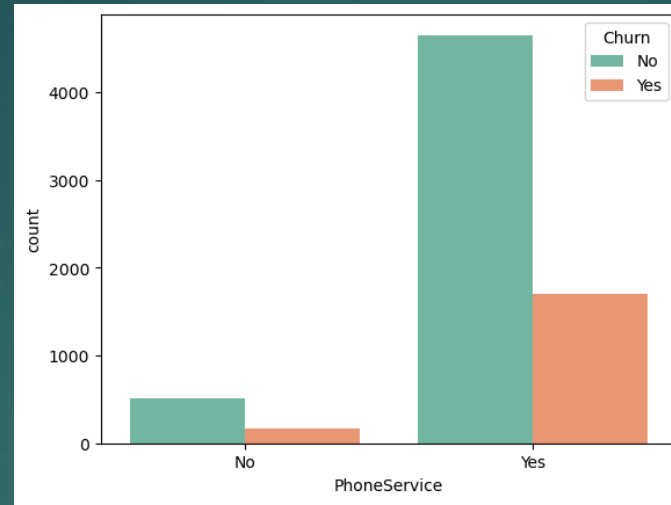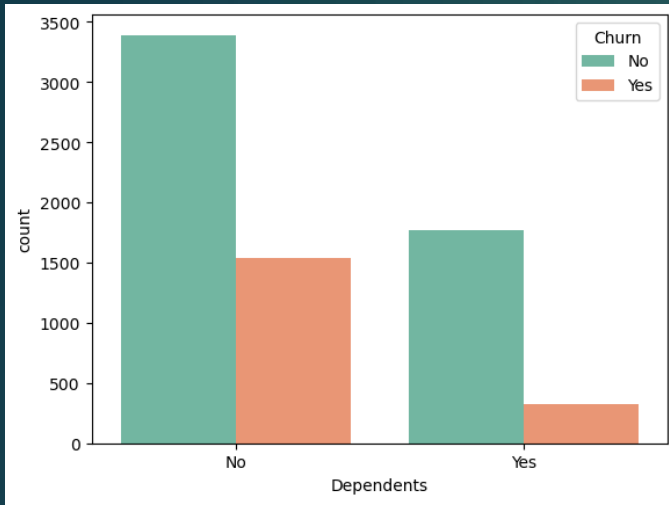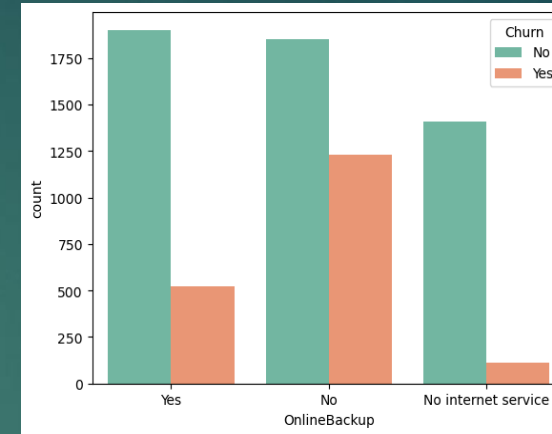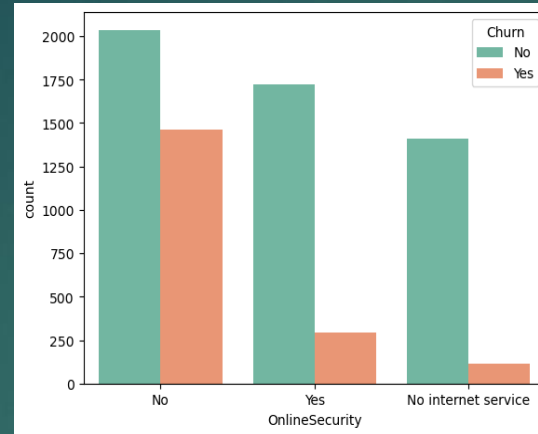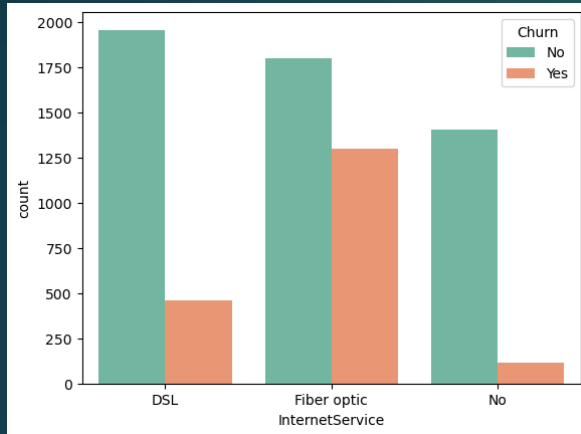
Data Exploration

Univariate Analysis

Insights:

● Since the distribution of churners is more or less equal, gender has no impact on churning.
● Senior Citizens are more likely to churn, they have a churn rate of ~29%

Insights:

- Phone Service has no impact on churning
- People with multiple lines are also less likely to churn

Insights:

● Customers with fiber optic internet are more likely to churn with a churn rate of ~29% whereas customers without internet service are very less likely to churn

● Customers with no online security are likely to churn at a churn rate of ~28% and customers with internet security have a churn rate of 12%

● People with no online backup have a churn rate of ~28% compared to people with online backup wherein the churn rate is ~18%.
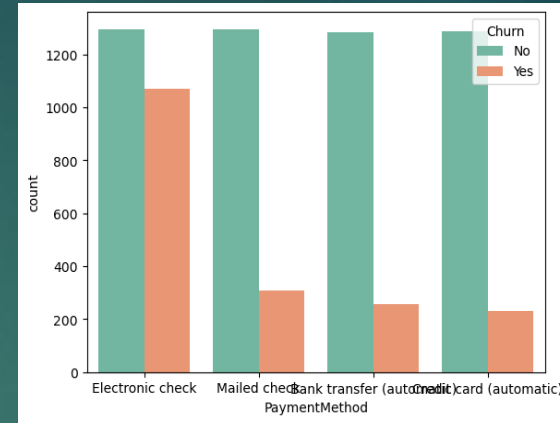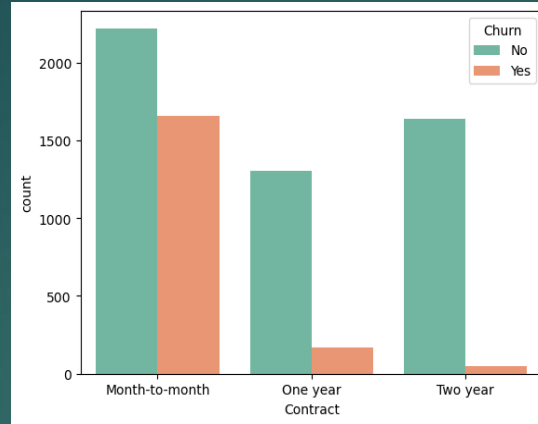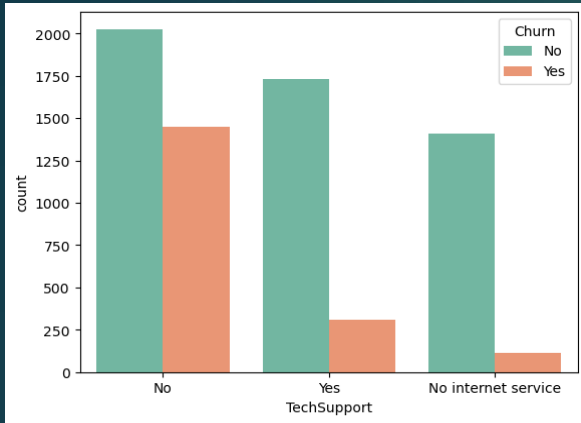
Insights:

● Customers with internet service for tech support are high churners compared to customers without internet service
● Contract Type - Monthly customers are more likely to churn because of no contract terms, as they are free to go customers.
● Almost 44% of the customers paying via Electronic check are churners.

Relationship between Monthly Charges and Total Charges
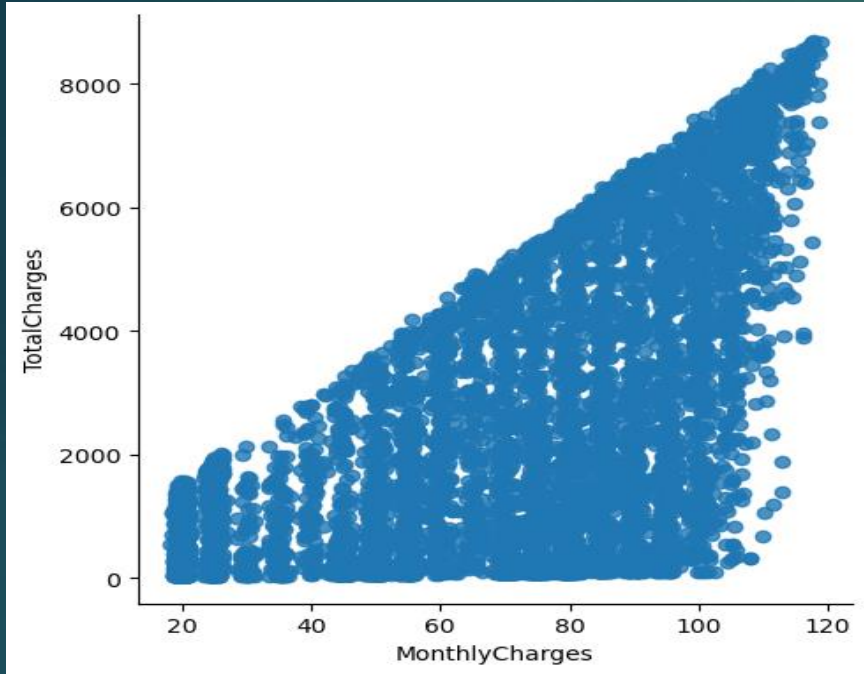

```
1 new_df_dummies['MonthlyCharges'].corr(new_df_dummies['TotalCharges'])
```

0.6510648032262024

Insights:

● We see that as the correlation is positive between Monthly charges and Total charges.

● Total Charges increases as Monthly Charges increases
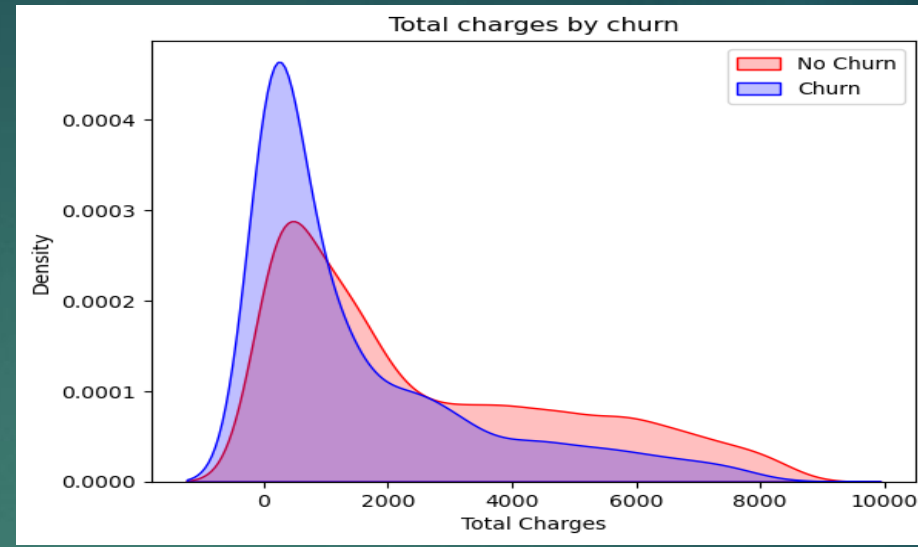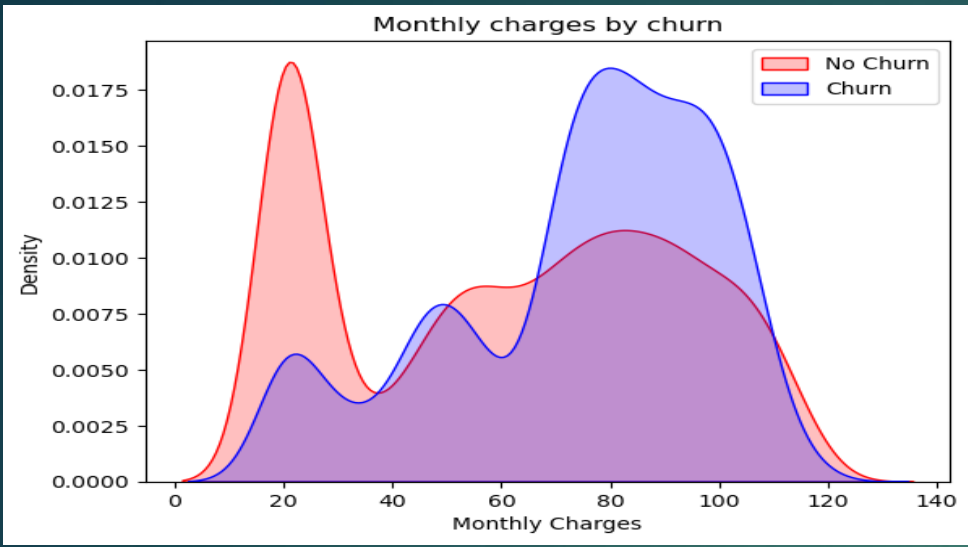
Insights:

● Higher the monthly charge, the churn rate is greater.

● Lower the total charge, the churn rate is higher

● Also lower the tenure , the churn rate is higher.

```
1 - 12       2175
61 - 72      1407
13 - 24      1024
25 - 36       832
49 - 60       832
37 - 48       762
Name: tenure_group, dtype: int64
```
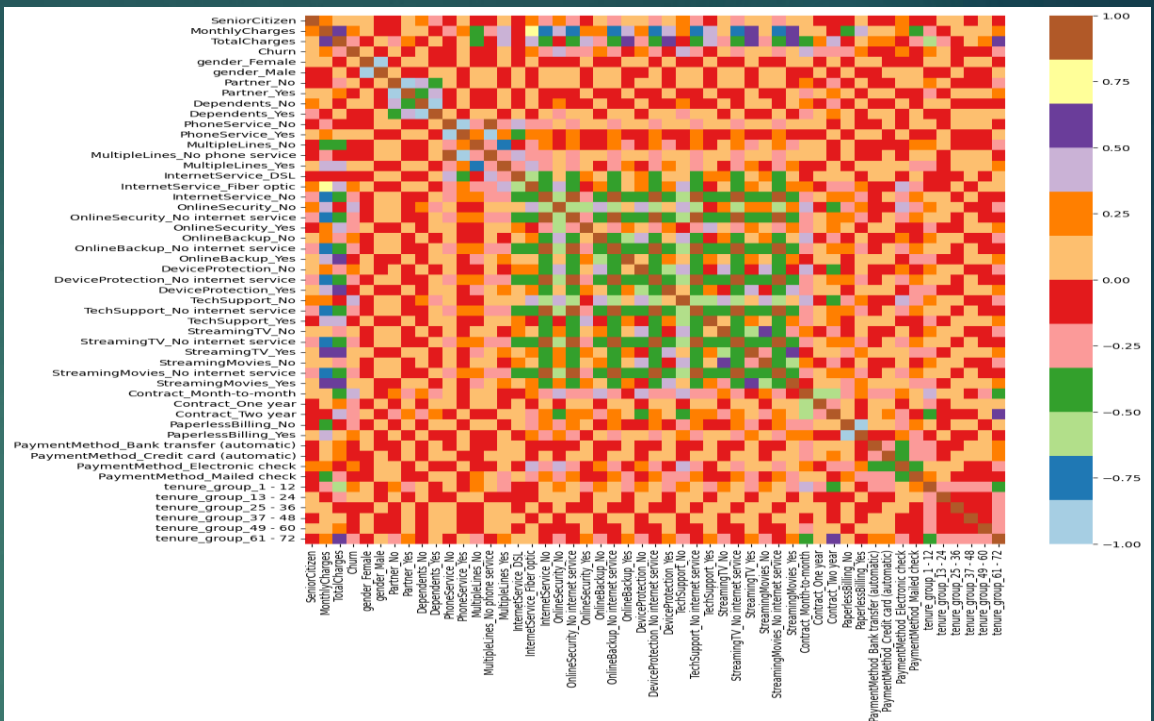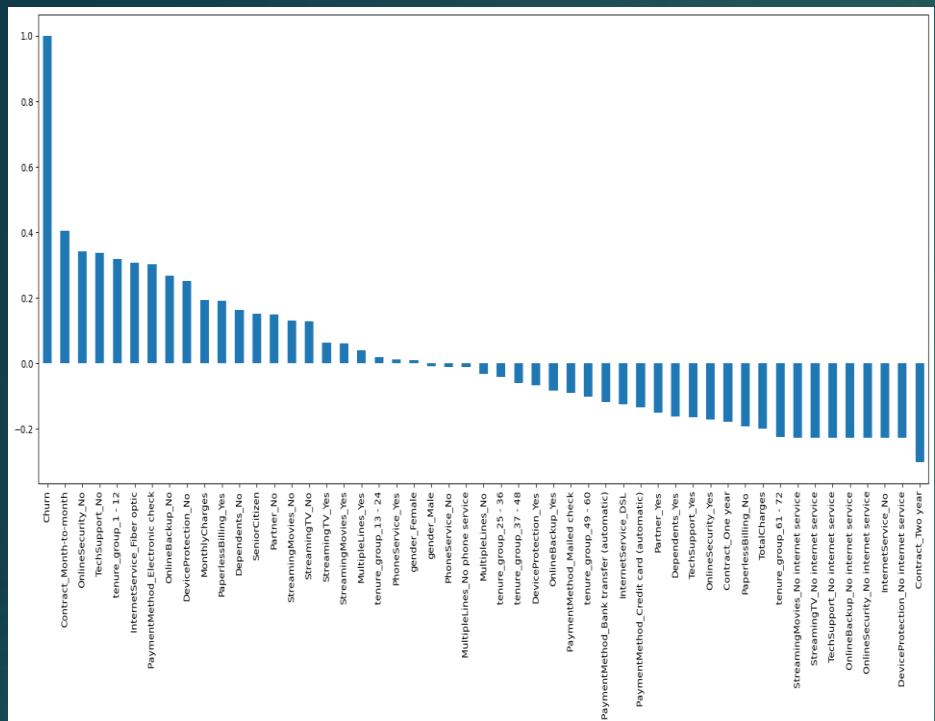
Insights:

● High churn rate is seen in Month-to-Month Contracts, No Online Security, No Tech Support, First year of subscription and Fiber Optics Internet

● Low churn rate is seen in case of Long-Term Contracts, Subscriptions without Internet Service.

● Gender, Availability of Phone Service and Multiple Lines have almost no churn rate.

● The Heat Map also depicts the same.

Bivariate Analysis
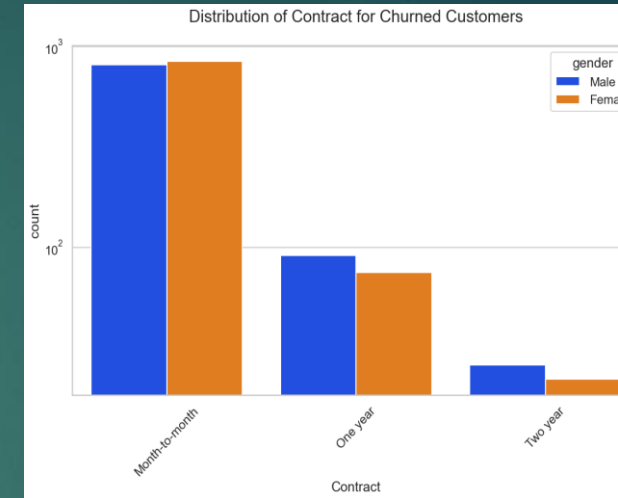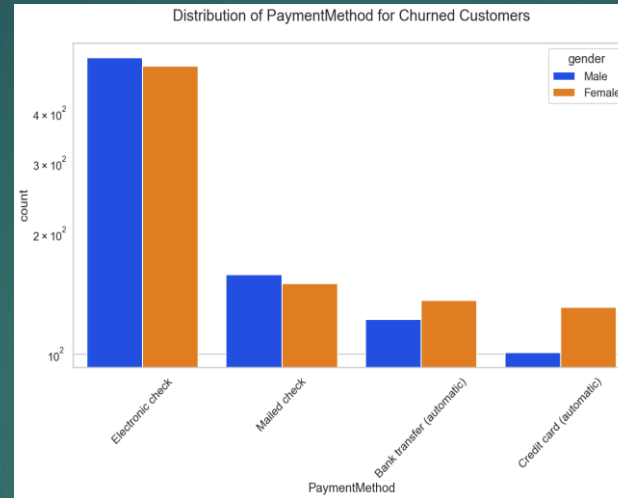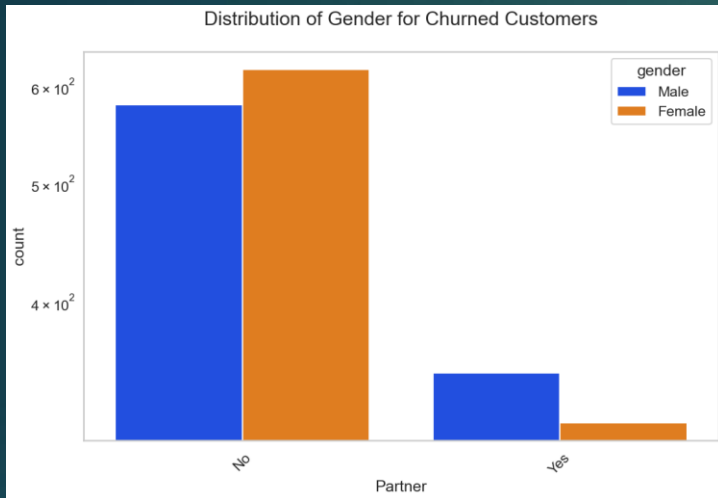
Insights:

● Among gender distribution for churned customers, males are most likely to churn who have partners and females are most likely to churn amongst the ones who does not have partners.

● For Electronic check distribution, the churn rate is nearly same for male and female. For bank transfer and credit card payments females tend to churn more.

● Month-to-Month distribution again male and female have nearly the same churn rate. For One year contract, males are likely to churn and for two-year contract, males are likely to churn.

Conclusion

1. Gender has no impact on churning
2. Senior Citizens are more likely to churn, they have a churn rate of ~29%
3. Phone Service has no impact on churning
4. People with multiple lines are also less likely to churn
5. Customers with fiber optic internet are more likely to churn with a churn rate of ~29% whereas customers without internet service are very less likely to churn
6. Customers with no online security are likely to churn at a churn rate of ~28% and customers with internet security have a churn rate of 12%
7. People with no online backup have a churn rate of ~28% compared to people with online backup wherein the churn rate is ~18%
8. Customers with internet service for tech support are high churners compared to customers without internet service
9. Contract Type - Monthly customers are more likely to churn because of no contract terms, as they are free to go customers.
10. Almost 44% of the customers paying via Electronic check are churners
11. People with less tenure i.e., 1-12 are very high churners