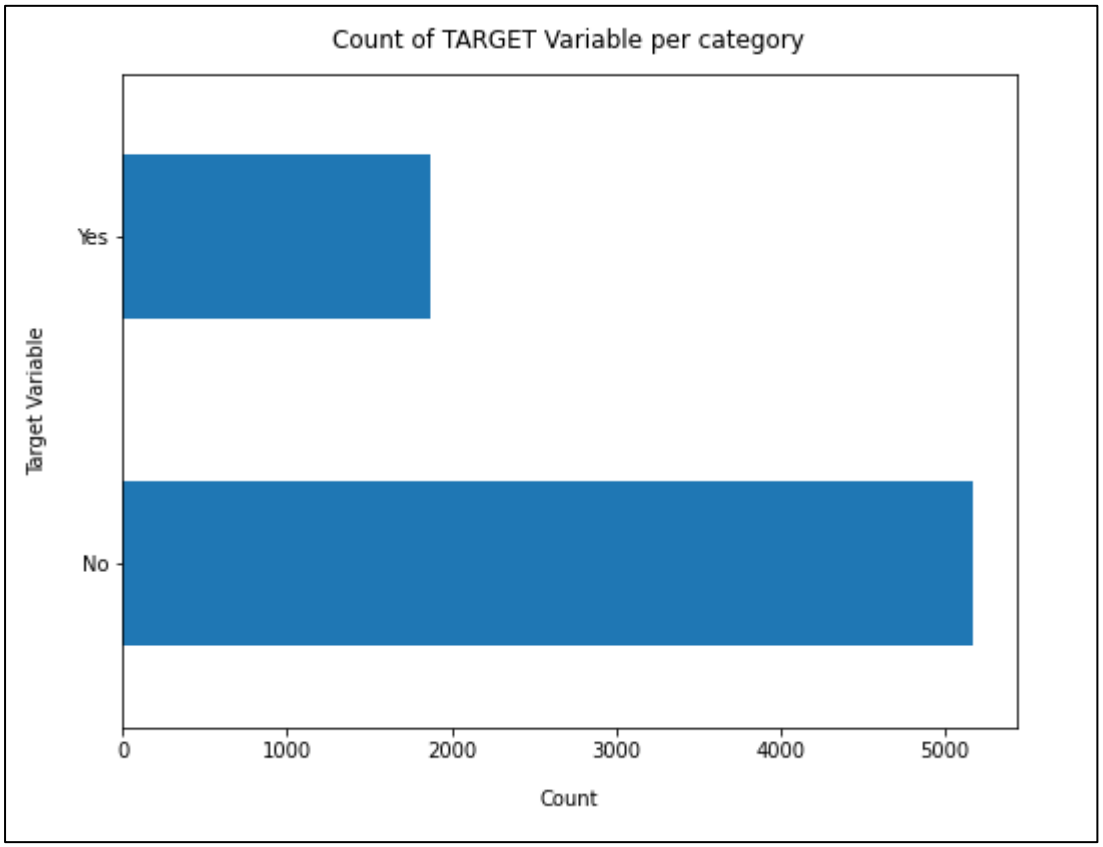
Telecom Churn Analysis

Business Understanding

Customer Churn in the telecom industry is the loss of customers to competing telecom service providers. This is a cause of major concern, as it leads to dwindling profits, overall loss of revenue and market capitalization. A thorough analysis of the data has been conducted, drawing out various insights, allowing us to delve deep into the factors leading to customer churn. These factors could play a big role in potentially shaping the business in future.

Understanding Data

The dataset in question is an open-sourced dataset, taken from Kaggle. Upon inspection, it is observed that there is a significant imbalance in the churned vs non-churned customers.




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

In other words, a little over 73% of the people in the dataset have not churned, while a smidgen over 26% of the people from the dataset have churned.

The dataset has 7043 records and 21 features.

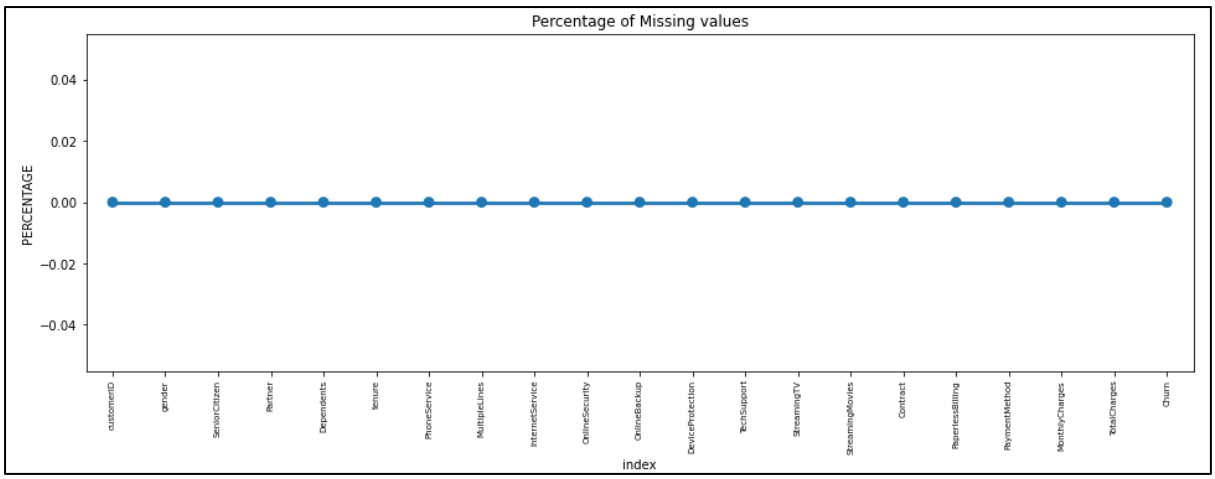
```
telco_base_data.shape
```

```
(7043, 21)
```

On first inspection, it appears that all features, apart from 'SeniorCitizen', 'MonthlyCharges', and 'tenure' are non-object data types, and there are no missing values.

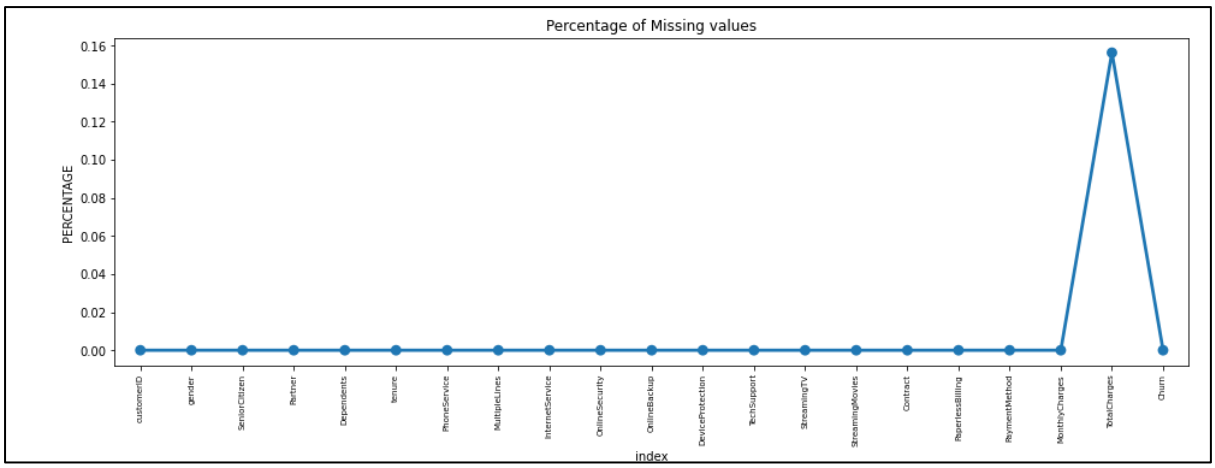
```
customerID           object
gender               object
SeniorCitizen         int64
Partner              object
Dependents           object
tenure                int64
PhoneService         object
MultipleLines        object
InternetService      object
OnlineSecurity       object
OnlineBackup         object
DeviceProtection     object
TechSupport          object
StreamingTV          object
StreamingMovies      object
Contract             object
PaperlessBilling     object
PaymentMethod        object
MonthlyCharges      float64
TotalCharges         object
Churn                object
dtype: object
```

```
 #   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      7043 non-null   object
 20  Churn             7043 non-null   object
```



However, upon further introspection, it is observed that there are 11 missing values in the 'TotalCharges' column.

```
customerID           0
gender               0
SeniorCitizen        0
Partner              0
Dependents           0
tenure               0
PhoneService         0
MultipleLines        0
InternetService      0
OnlineSecurity       0
OnlineBackup         0
DeviceProtection     0
TechSupport          0
StreamingTV          0
StreamingMovies      0
Contract             0
PaperlessBilling     0
PaymentMethod        0
MonthlyCharges       0
TotalCharges        11
Churn                0
dtype: int64
```



From the plot above, it is observed that there are missing values present on the 'TotalCharges' column.

As the ratio of missing values is extremely small, approximately 0.15%, we can simply ignore them by dropping missing values.

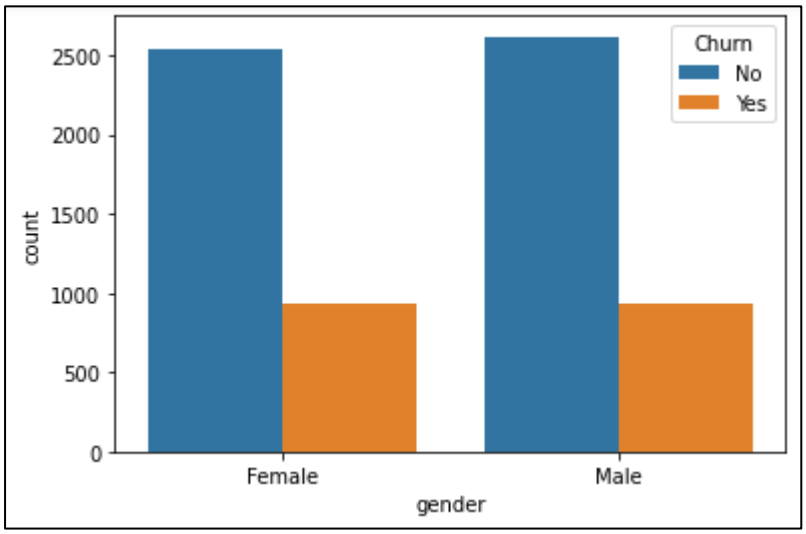
```
missing_data_ratio=(telco_data_missing.shape[0]/telco_data.shape[0])
missing_data_ratio*100
# avg=telco_data_missing['MonthlyCharges'].mean()
# avg
```

```
0.1561834445548772
```

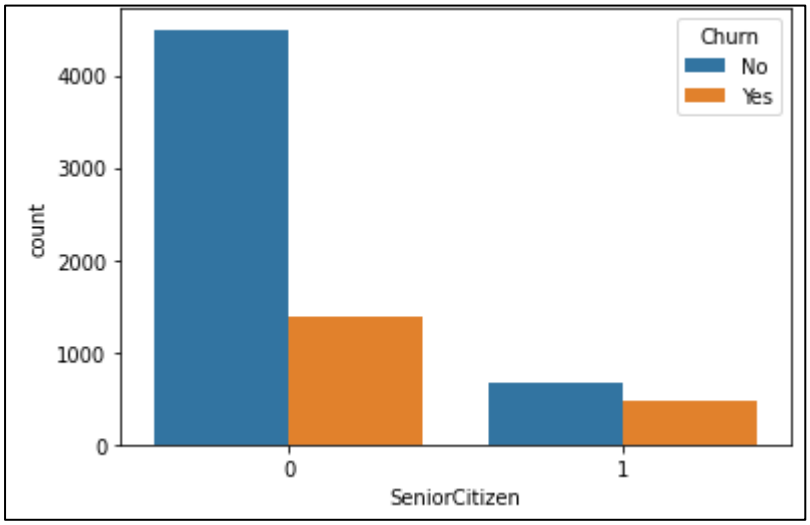
Unearthing Insights

1. Univariate Analysis:

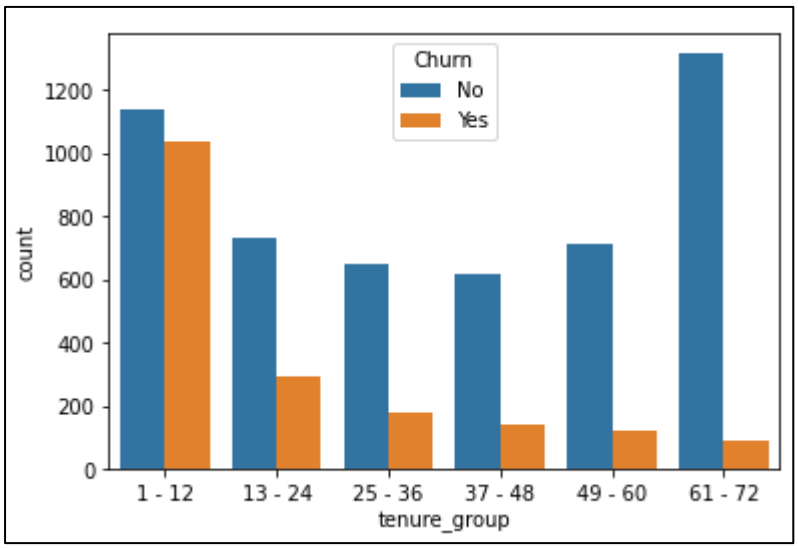
Insight: A significant, first insight obtained is there is an equal distribution in churners between males and females. Hence, gender on its own is not a good predictor for churn.



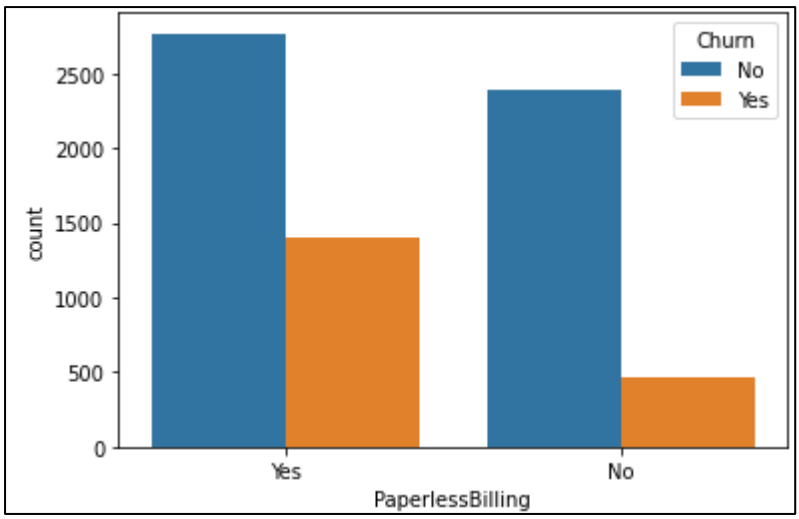
Insight: Non-Senior citizens seem to churn more, compared to senior citizens.



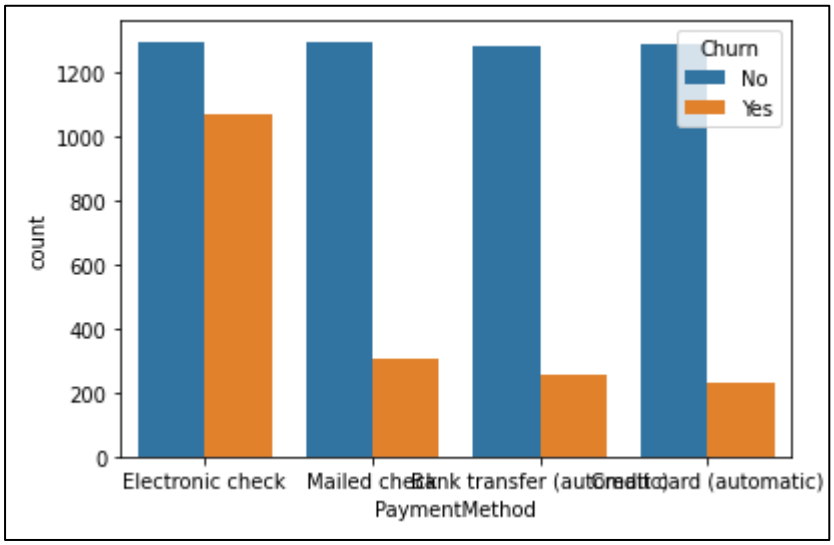
Insight: Customers who have stayed on for a significant amount of time are less likely to churn, and vice versa.



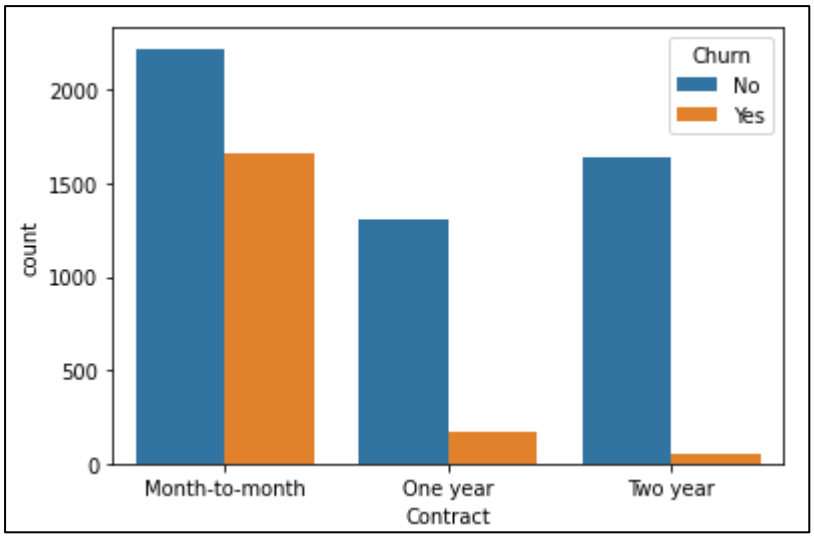
Insight: Customers who have opted for paperless billing are more likely to churn.



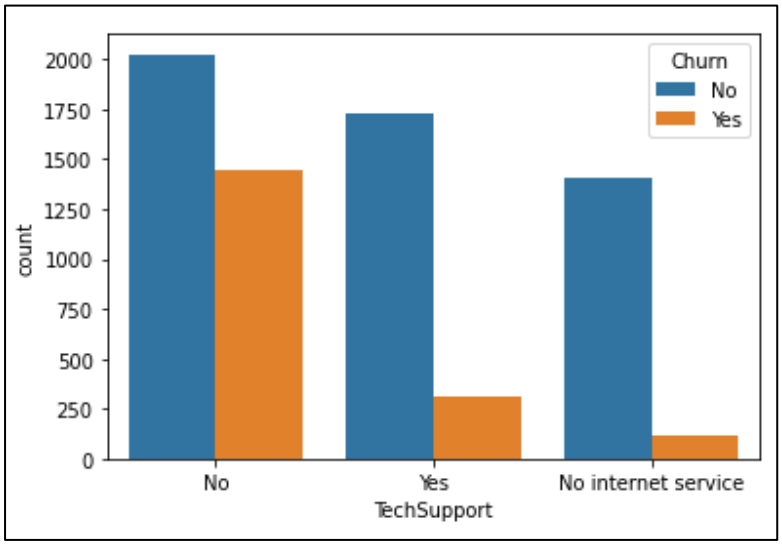
Insight: Customers having opted for electronic checks as payment methods are more likely to churn



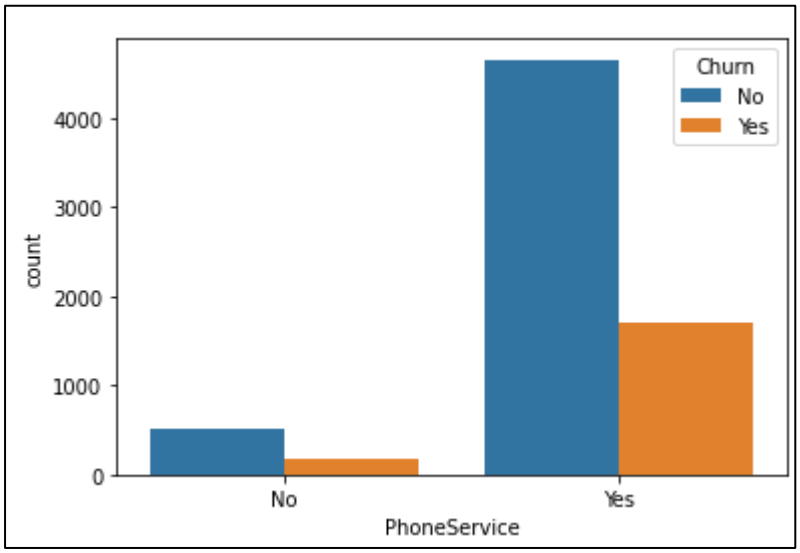
Insight: Customers with month-to-month plans are more likely to churn compared to contracted customers.



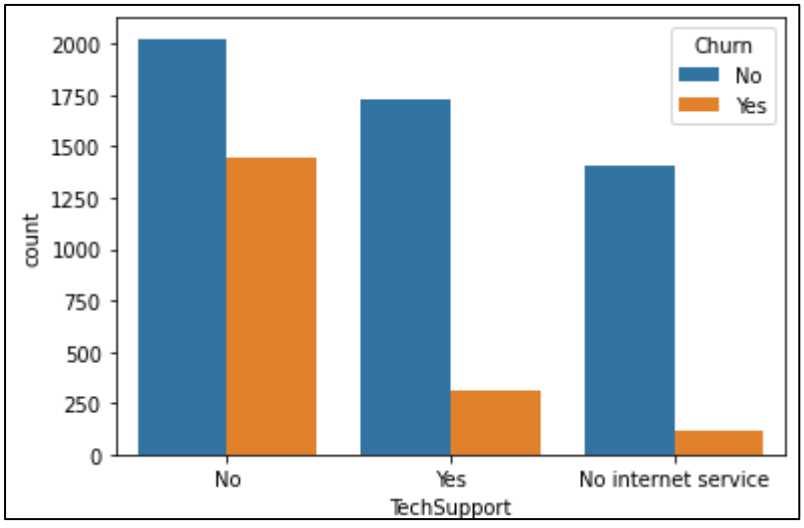
Insight: Customers without tech support are more likely to churn.

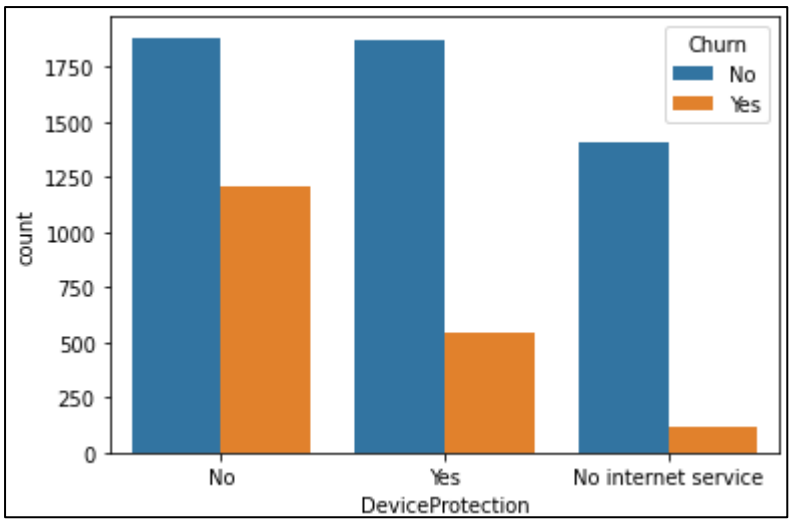


Insights: Customers with phone service are less likely to churn

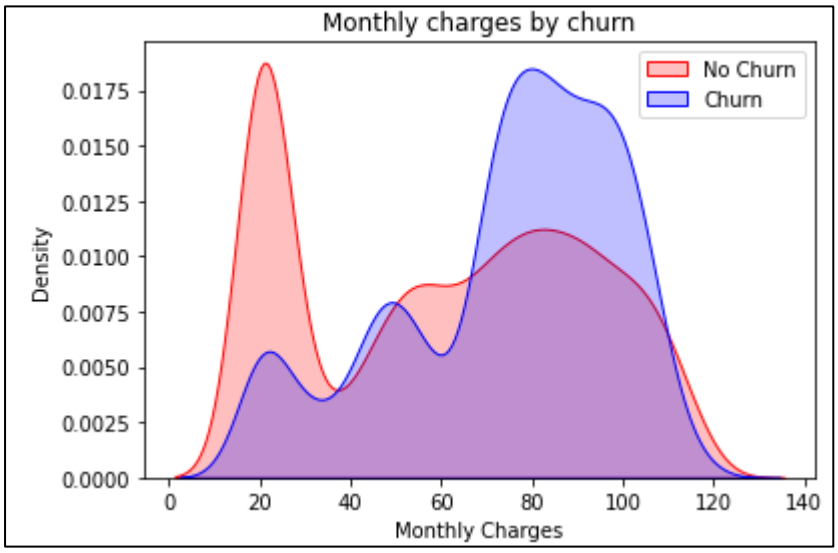


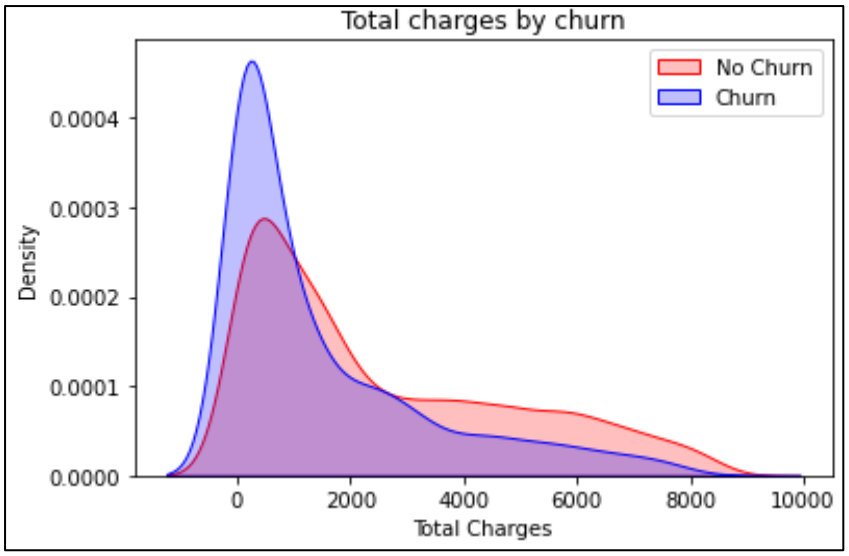
Insight: Lack of tech and device support seems to be another reason for customers churning.





Insights: As an expected trend, it is observed that the higher the monthly charges, greater is the churn. However, to everyone's surprise, the lower the total charges, the more the churn.

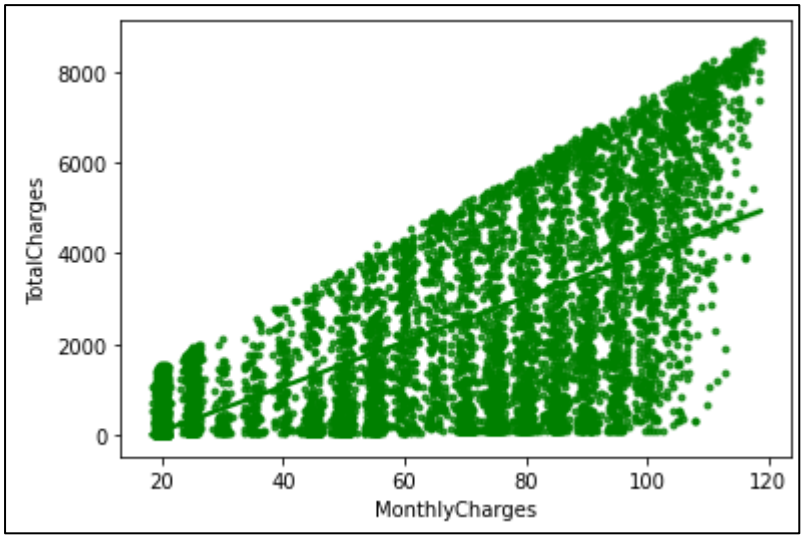




2. Bi-Variate Analysis

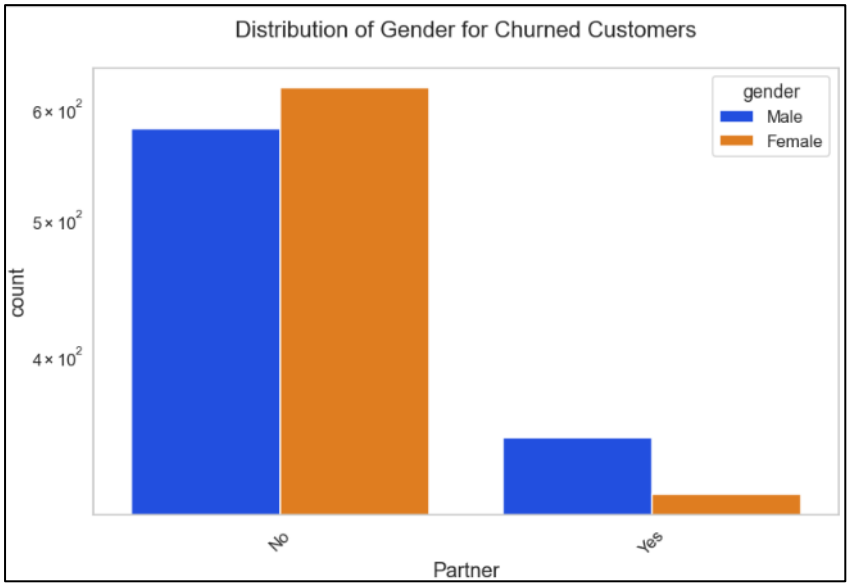
Now that a basic overview has been obtained as to which factors seem to contribute to customer churn, it is time to deep dive into some of the previously obtained factors and see how they correlate with other features.

Insight: As an expected trend, we see a fairly strong 'positive' correlation between 'monthly charges' and 'total charges.

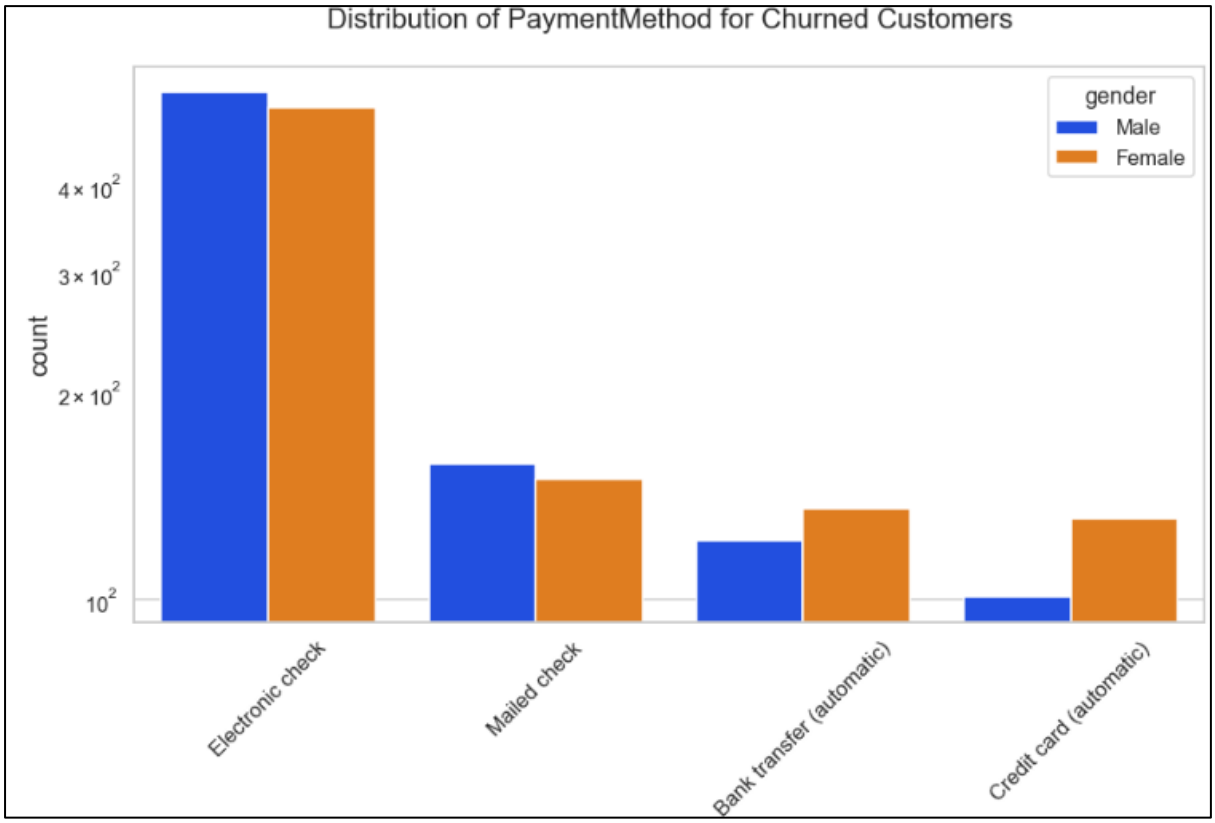


	MonthlyCharges	TotalCharges
MonthlyCharges	1.000000	0.651065
TotalCharges	0.651065	1.000000

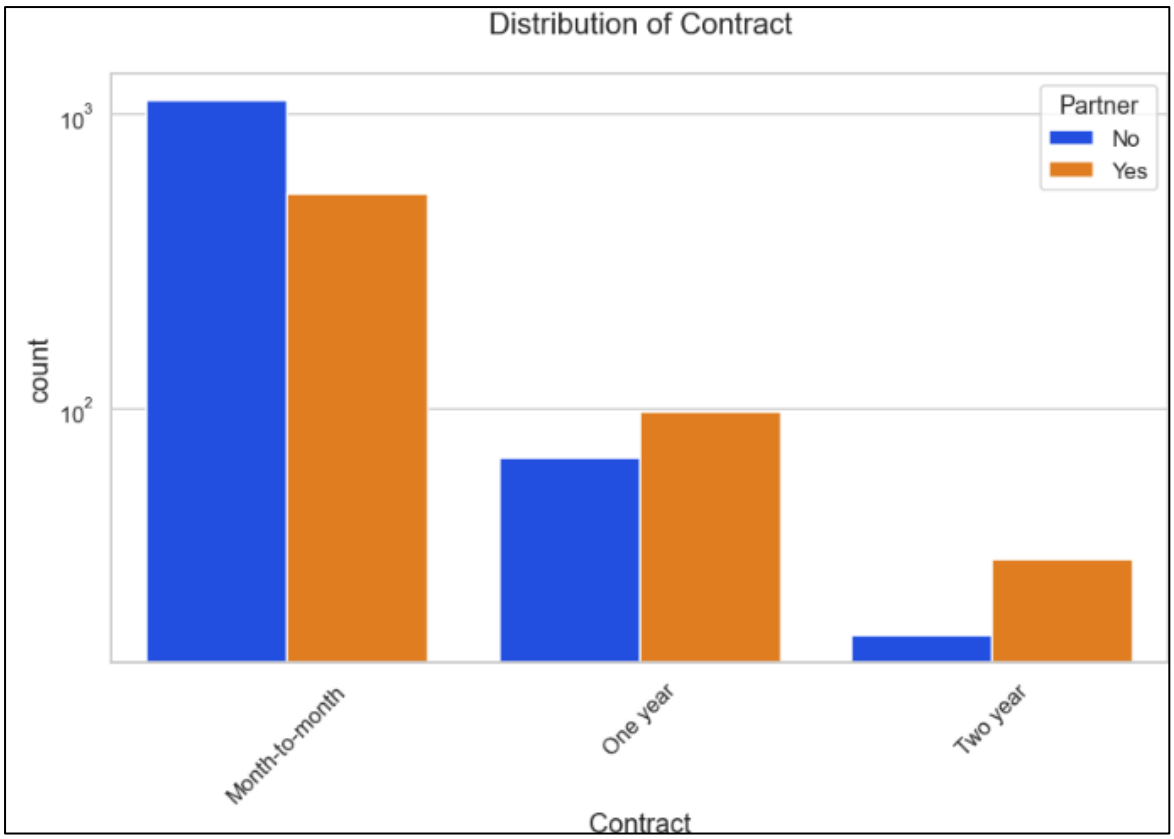
Insight: Introspecting into the gender distribution for churned customers who have partners, it is observed that among those who have partners, males are most likely to churn. Conversely, among those who don't have partners, females are most likely to churn.



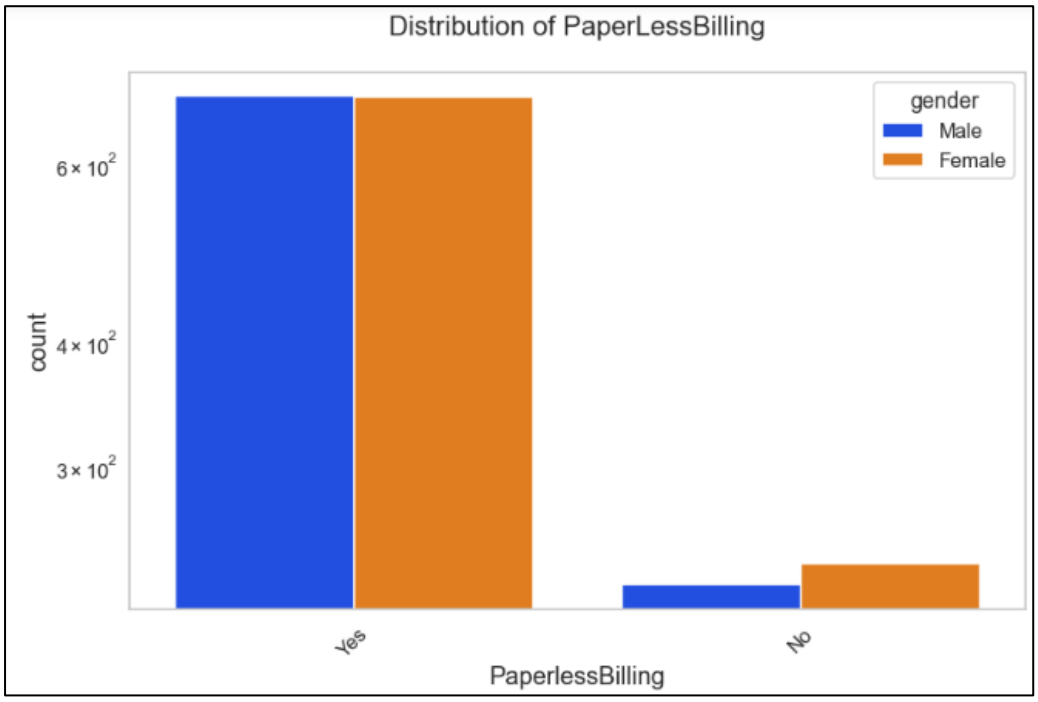
Insight: Among the customers who have churned, there is a near equal distribution between male and female for those using electronic check and mailed check as payment methods. However, there is a significant difference in distribution among those using bank transfer and credit cards as payment methods with females tending to churn more.



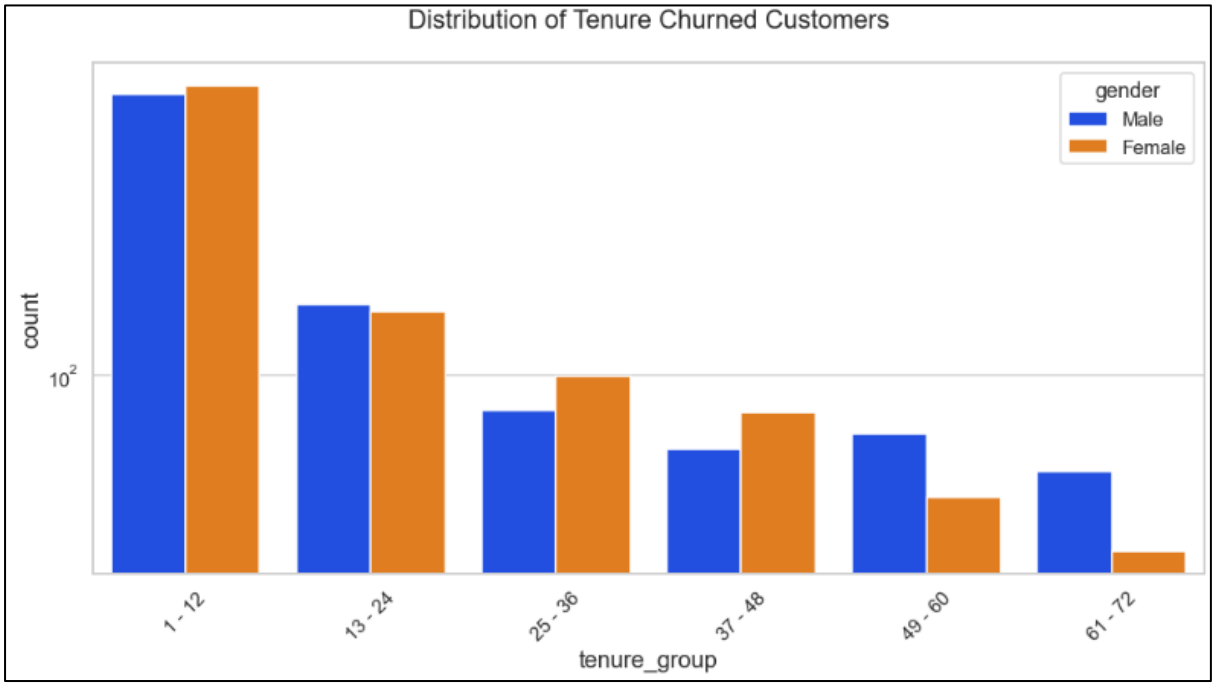
Insight: Among the customers who churn, people currently on a month-to-month billing cycle and who don't have partners are more likely to churn. Whereas those contacted and have partners are more likely to churn.



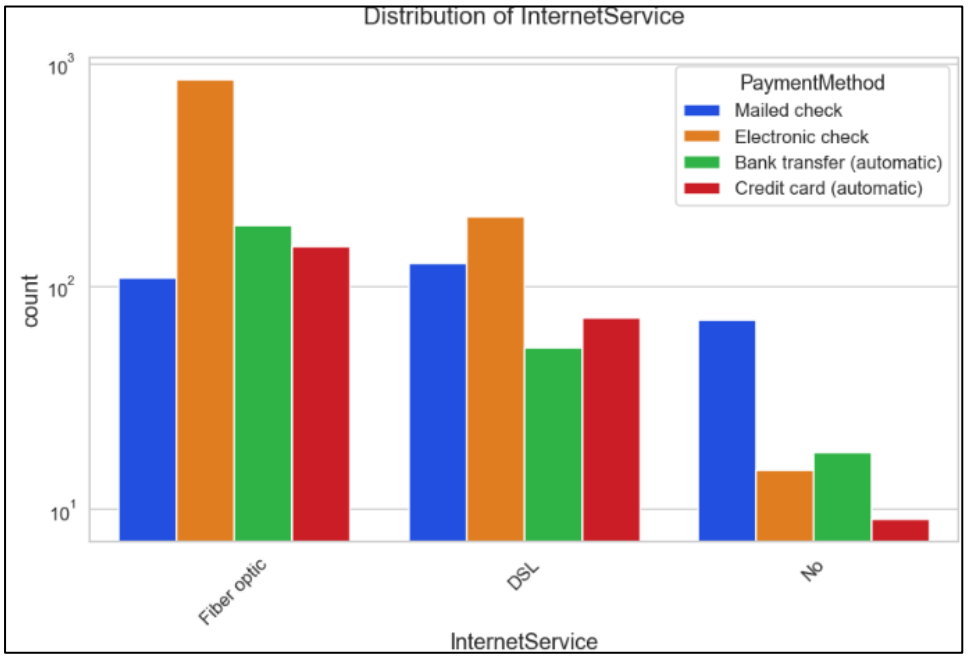
Insight: Among the churners who have not opted for paperless billing, females are mor likely to churn as compared to males. For those who have opted for paperless billing, there is an equal distribution between males and females.



Insight: Looking closely at the distribution of customers based on their tenure with the company, it is seen that there is a near equal distribution between male and female customers, who have a tenure between 1-12 months and 13-24 months. Among those customers who have been with the company for more than 25 months but less than 49 months, females are the majority churners. Finally, for those who have been with the company for more than 49 months, males are the majority churners.



Insight: Among those who have churned and use fiber optic and DSL services, most of them pay through electronic medium. However, for those who don't have internet service, they majorly pay through mailed checks.



Final Thoughts

The dataset and the EDA process has unearthed several insights. Some of them are as follows:

- Equal distribution between male and female churners.
- Non-Senior citizens seem to churn more than senior citizens. This could be a result of the younger audience being more aware of plans provided by competitors, or simply senior citizens more likely to pass away over time.
- Customers who have opted for paperless billing and electronic checks are more likely to churn.
- Non-Contracted customers, i.e., those who follow the month-to-month scheme are more likely to churn, as they are free to explore and aren't bound by contracts.
- Lack of tech and device support could be seen as further reasons for customer churn.
- Customers who have stayed with the provider for a longer duration are less likely to churn. This could be attributed to them being happy with the service or as a show of loyalty to the company.
- Customers with higher monthly charges are more likely to churn. However, as a surprise to all, customers with lower total charges are more likely to churn.
- Among churned customers who have partners, males are more likely to churn. Conversely, among those who don't have partners, females are most likely to churn.
- Dissecting into the payment modes among the churners, females using credit card and bank transfer as payment modes are more likely to churn as compared to males.
- Churned customers using fiber optics and DSL services, majorly pay through electronic medium. However, those who don't have internet service pay through mailed checks.