

1) Project Overview :-

Customer churn analysis involves examining the reasons and patterns behind customers leaving a business or discontinuing services. It's a critical process for businesses aiming to retain customers, optimize revenues, and improve their products or services.

In this project I have used the Python tool for my Exploratory Data Analysis (EDA) where I used the various features and libraries like Numpy, Pandas, Matplotlib, Seaborn and Scikit-learn for my data sourcing, data cleaning, data analysis, data visualization.

This helps me to get a good insight from the data and also helps to take actionable decisions accordingly. Similarly we can apply the Customer Churn analysis in the field of Subscriptions Based Services, E-Commerce, Banking and Financial Services, Healthcare, Gaming and so on by applying the EDA using Python tool.

2) Problem Statement :-

In the highly competitive telecommunications industry, retaining customers is crucial for sustaining revenue and market share. Despite efforts to deliver high-quality services and competitive pricing, many telecom companies face significant customer churn.

The objective of this analysis is to identify the factors driving customer churn and develop actionable strategies to reduce churn rates. By analyzing customer data, we aim to uncover patterns and behaviors associated with churn, such as contract types, payment methods, tenure, and service usage. This analysis will help the telecom company prioritize retention efforts, optimize marketing strategies, and ultimately reduce customer attrition, leading to sustained growth and profitability.

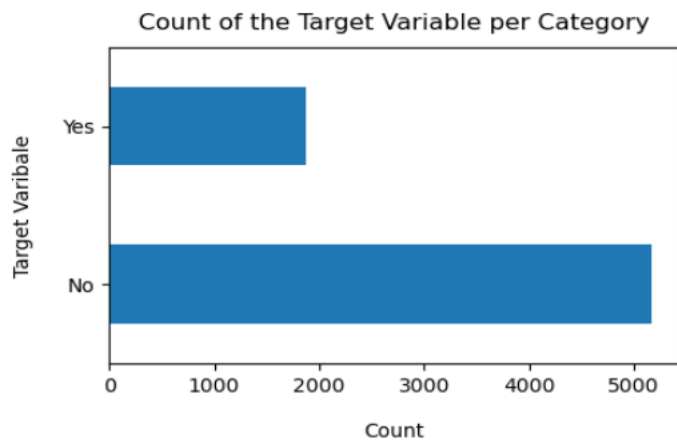
3) Data Understanding :-

3.1) Data Overview : -

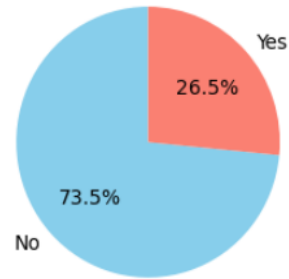
The dataset contains 7044 records and 21 columns, including customer demographics, various services like phone and internet, tenure, Payment methods, contracts, monthly and total charges and the churn status.

3.2) Churned Rate :

By plotting a pie-chart we can easily understand the churned rate which is shown below . Overall 26.5% of the people got churned.



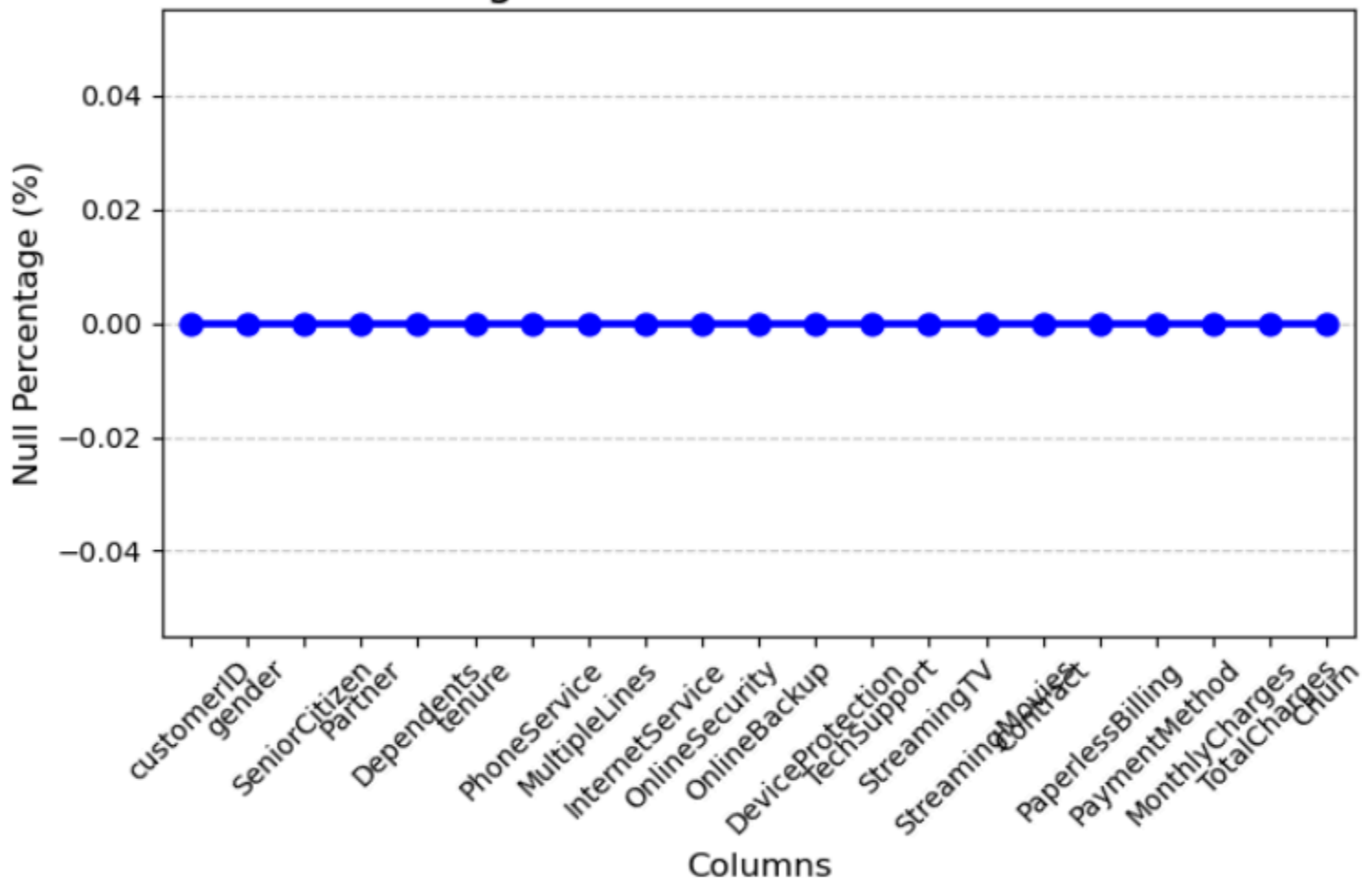
Percentage of Churned and Non-Churned Customers



3.3) Null Values :

By plotting the graph we can check if there are any null values in the columns or not.

Percentage of Null Values in Each Column



From this graph we can easily see that there are not any null values in our data set. So we can further proceed to our data cleaning part.

4) Data Processing :-

4.1) Finding Errors :-

By analysing the data I found an error in our data set, that the data type of the column name **Total Charges** is in string type but it should be in Integer or Float type. So firstly we have to convert the data type from string into an integer.

df.dtypes	
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

Data columns (total 21 columns):			
#	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)			

4.2) Dealing with Missing Values :-

By converting the data type of the total charges I found that there are 11 null values in our dataset as we can see above. It contains only 7032 numbers of data and all the other columns contain 7043 numbers of data.

Hence, the % of the missing values compared to the dataset is very less as 0.15%. It is safe to ignore them for further processing, else we can take the mean or median of all the total charges and replace them with the null values.

4.3) Feature Encoding :-

Tenure is given in the number of months the customer uses the services, it will be difficult to plot the graph for each user and also it gets tough to analyse the data. So we can divide customers into bins based on tenure e.g. for tenure <12 months - assign it to group 1-12, for tenure b/w 1-2 yrs assign it to group 13-24 and so on.

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

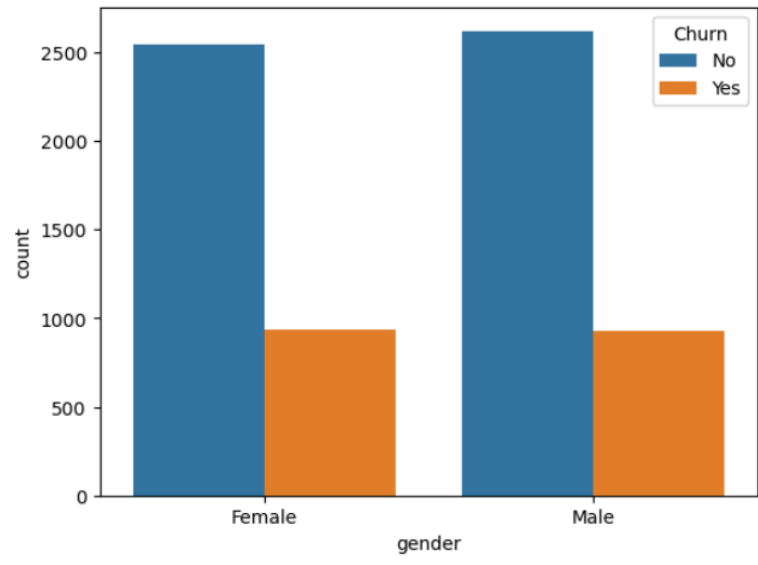
5) Exploratory Data Analysis (EDA) :-

5.1) Univariate Analysis :-

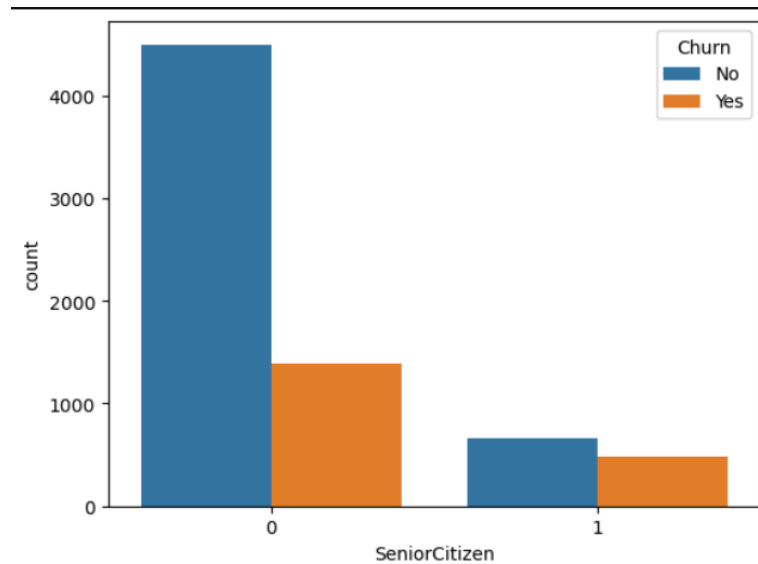
Plotting distribution of individual predictors for churn.

5.1.1) Demographics

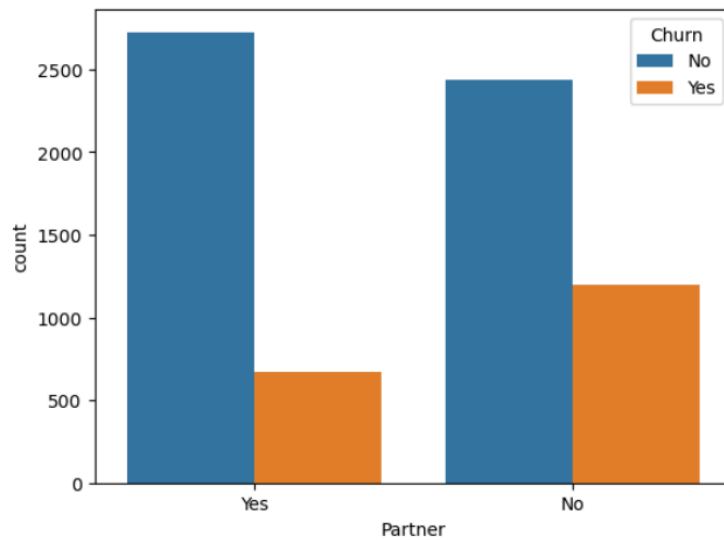
5.1.1.i) Gender - Insight - In genders graph the churn rate is almost same for both male and female, hence we don't get any insights from this particular graph.



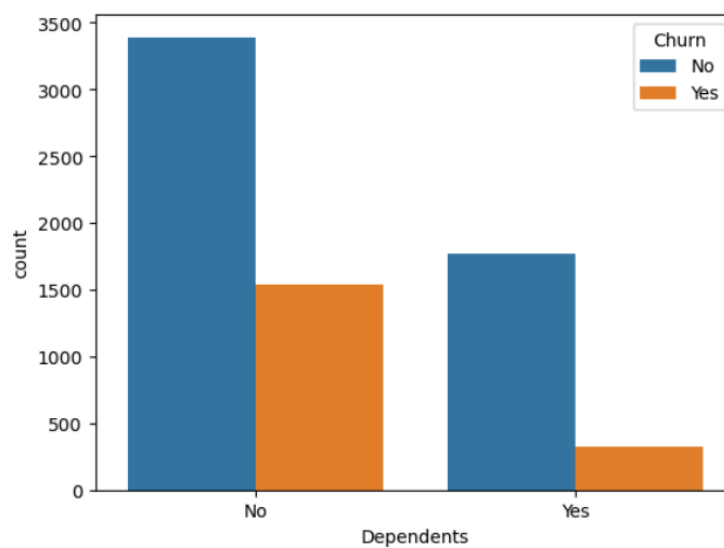
5.1.1.ii) Senior Citizen - Insight - The Churn rate is higher for the senior citizen people by almost 45%.



5.1.1.iii) Partners - Insight - As we can see in the graph, people with no partners having the higher churn rate by almost 34.28%.

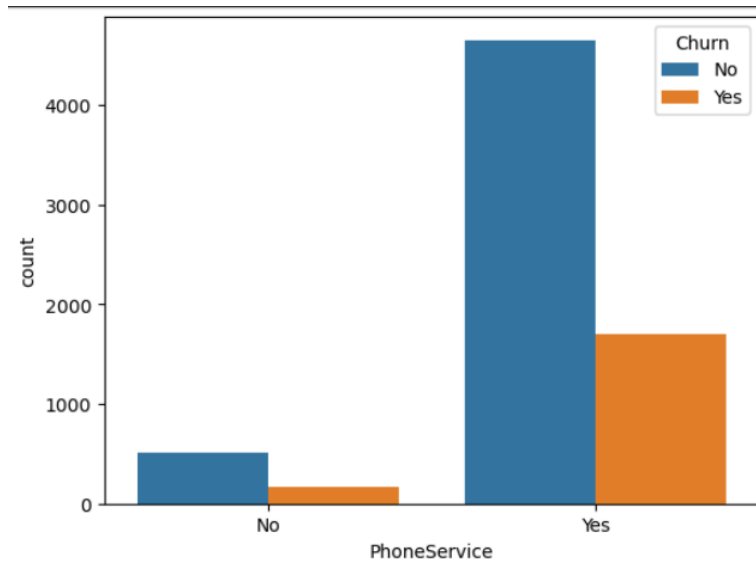


5.1.1.iv) Dependents - Insight - A person with no dependency has a higher churn rate almost 30%.

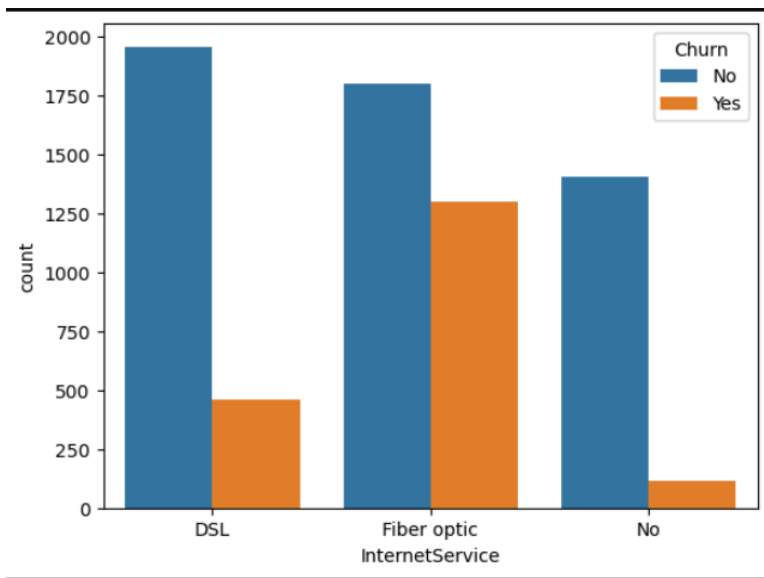


5.1.2) Services :

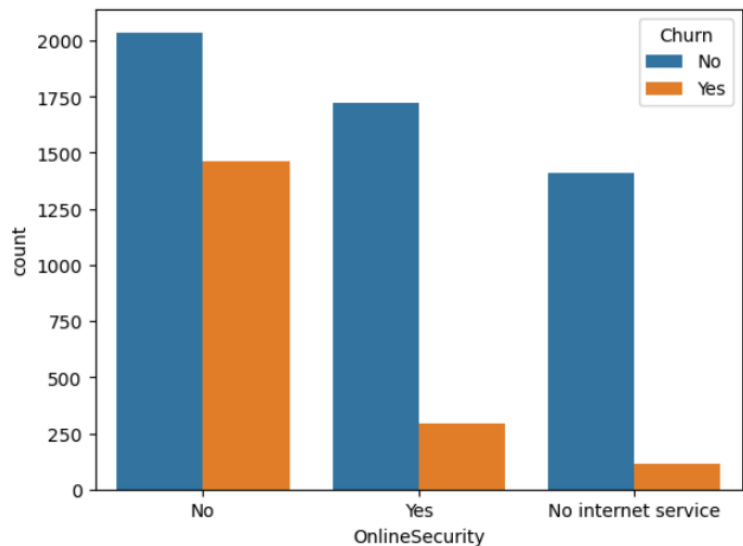
5.1.2.i) Phone Services-Insight - People with phone services have a higher churn rate of almost 29%.



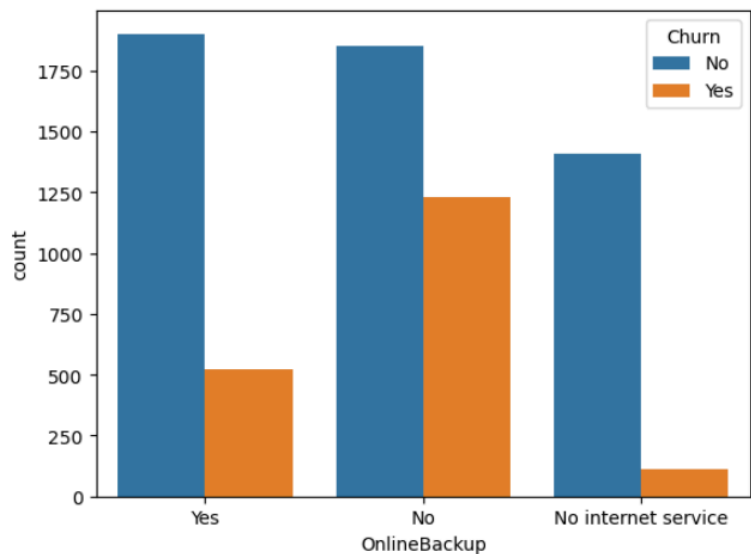
5.1.2.ii) Internet Services - Insight - People using internet service (fiber optic) have a higher churn rate by almost 41.66%.



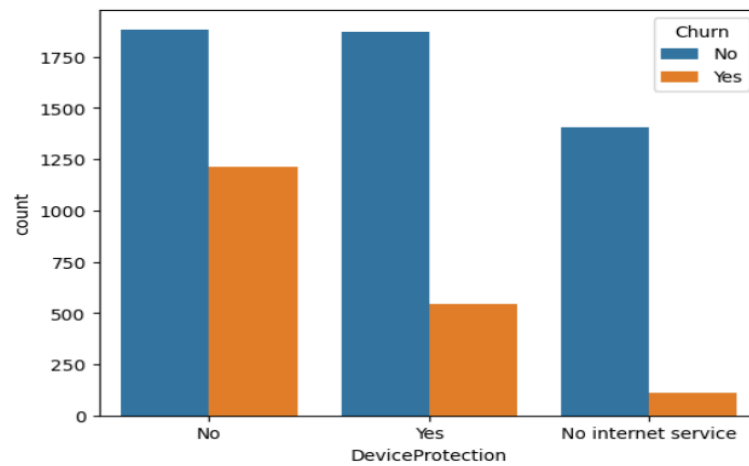
5.1.2.iii) Online Security - Insight - People with no online security have higher churn rate almost 41%.



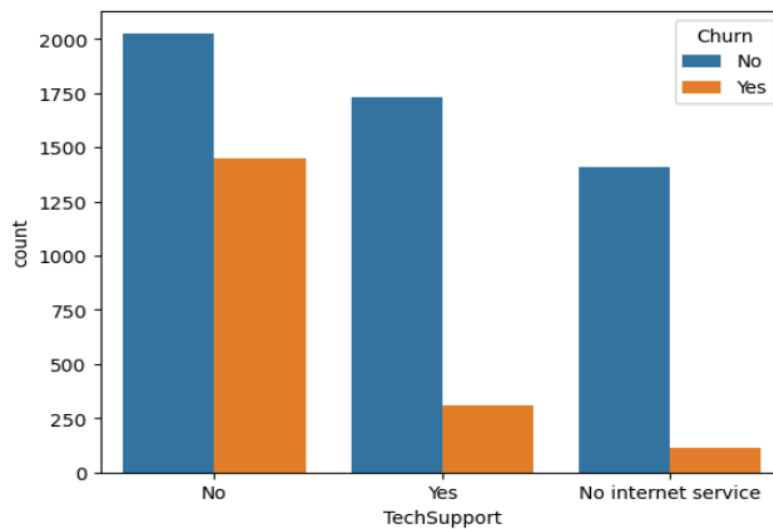
5.1.2.iv) Online backup - Insight - With no online backup has higher churn rate 41%.



5.1.2.v) Device Protection - Insight - No device protection has a higher churn rate around 40%.

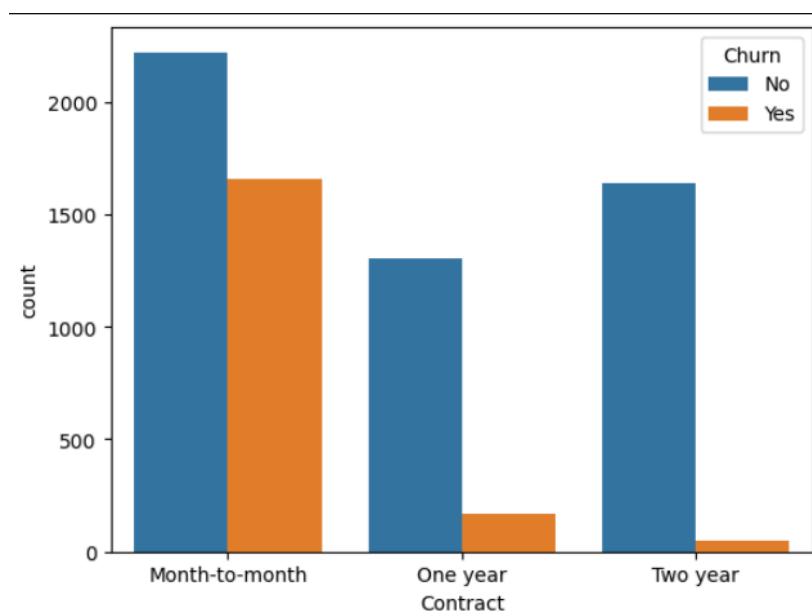


5.1.2.vi) Tech-Support - Insight - No tech support has a higher churn rate around 41%.

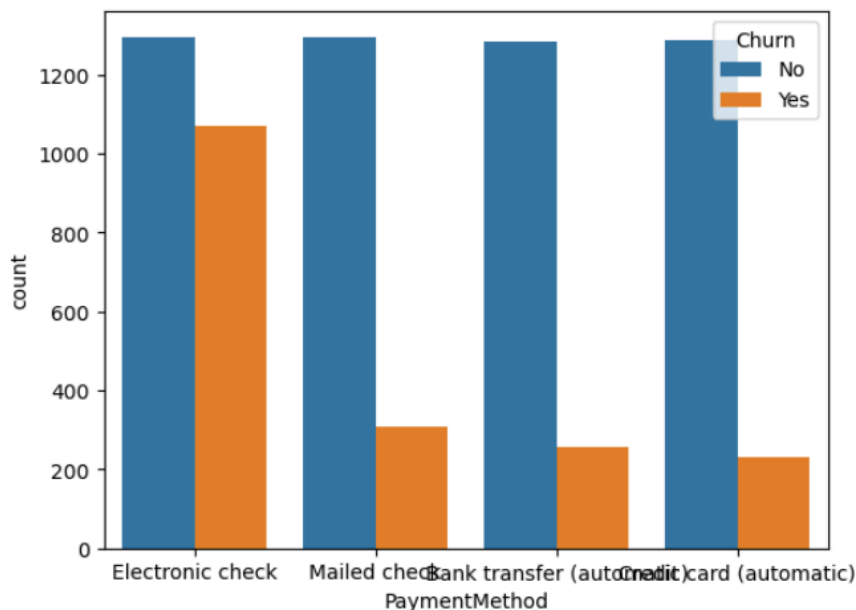


5.1.3) Others :

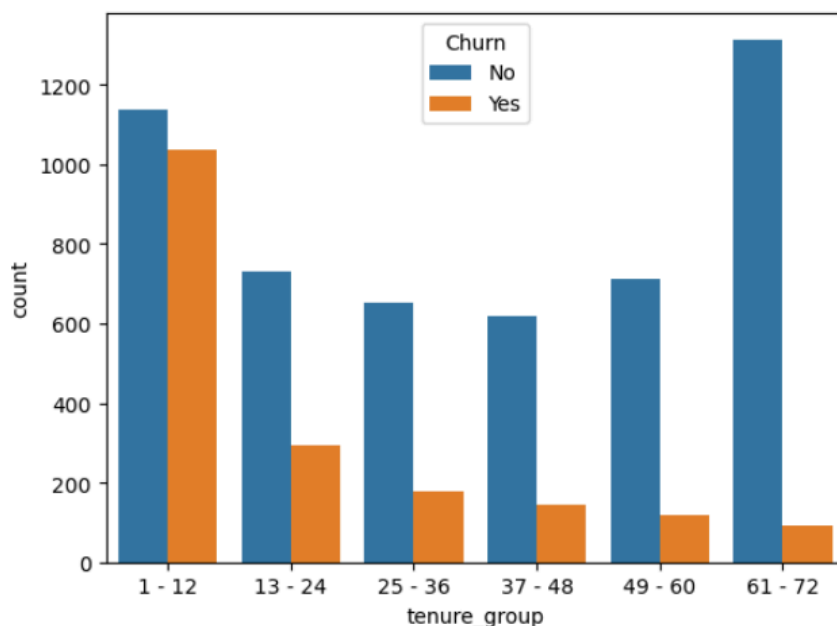
5.1.3.i) Contracts - Insight - People who have contracts or plans on a monthly basis their churn rate is much higher around 41% but as the contract periods extended from month on month to one or two years their churn rate decreases.



5.1.3.ii) Payment Method - Insight - The churn rate is higher via electronic check payment method around 45.83%



5.1.3.iii) Tenure - Insight - The churn rate is higher when is tenure is between 1-12 months around 45% and decreases as the tenure increases.



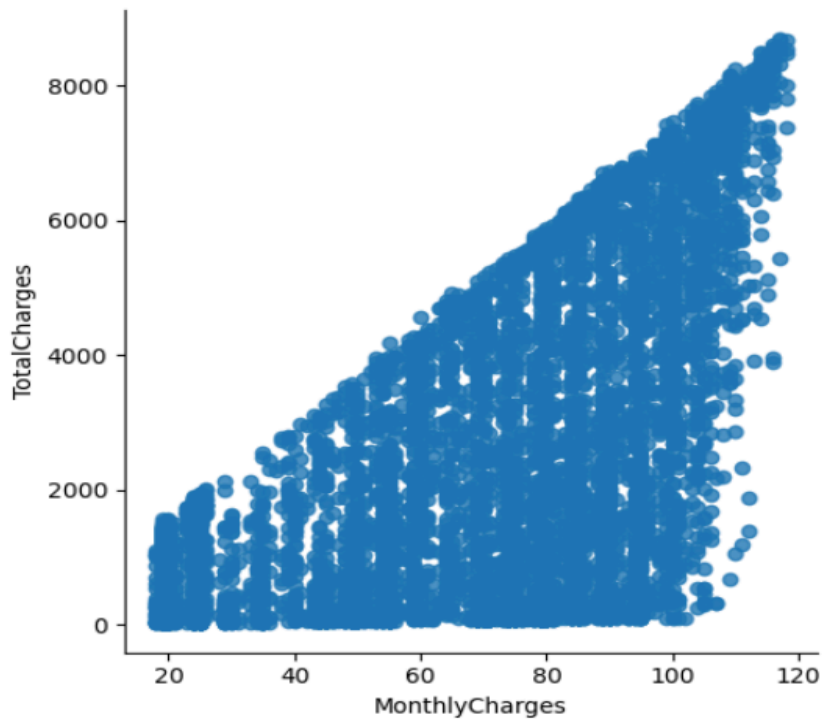
5.2 Numerical Analysis :

Converting the variable churn into binary value like yes = 1 and no = 0 because we are going to perform some correlation analysis and it is not possible in categorical values like yes and no.

5.2.1) Relationship between Monthly Charges and Total Charges :

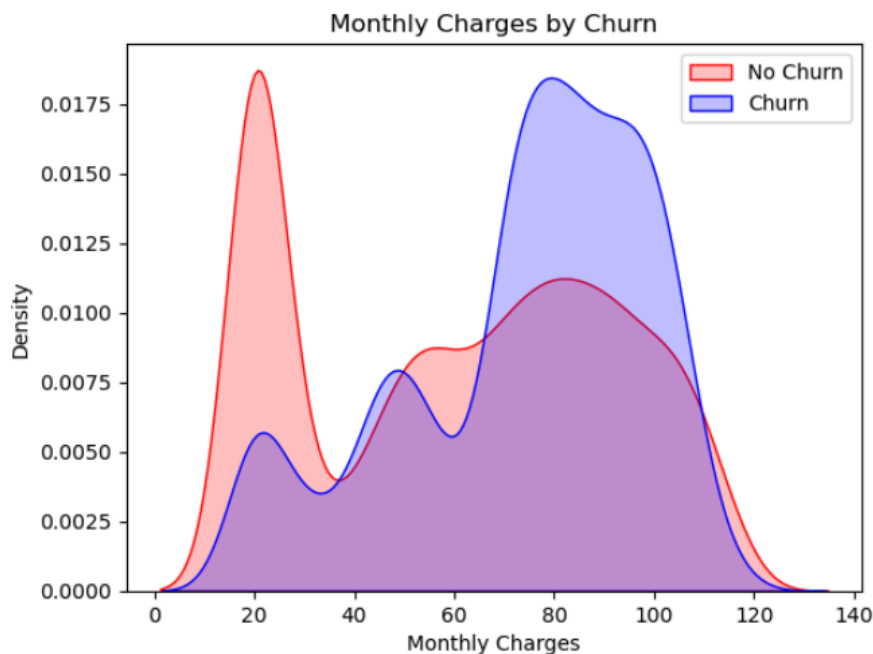
In this case the correlation is positively Correlated (Correlation is defined between -1 to +1, if the value is near -1 it means it is negatively correlated and when it is closer to +1 then it is positively correlated).

Total Charges increases as the Monthly Charges increases - as expected



5.2.2) Churn Rate by Monthly Charges :

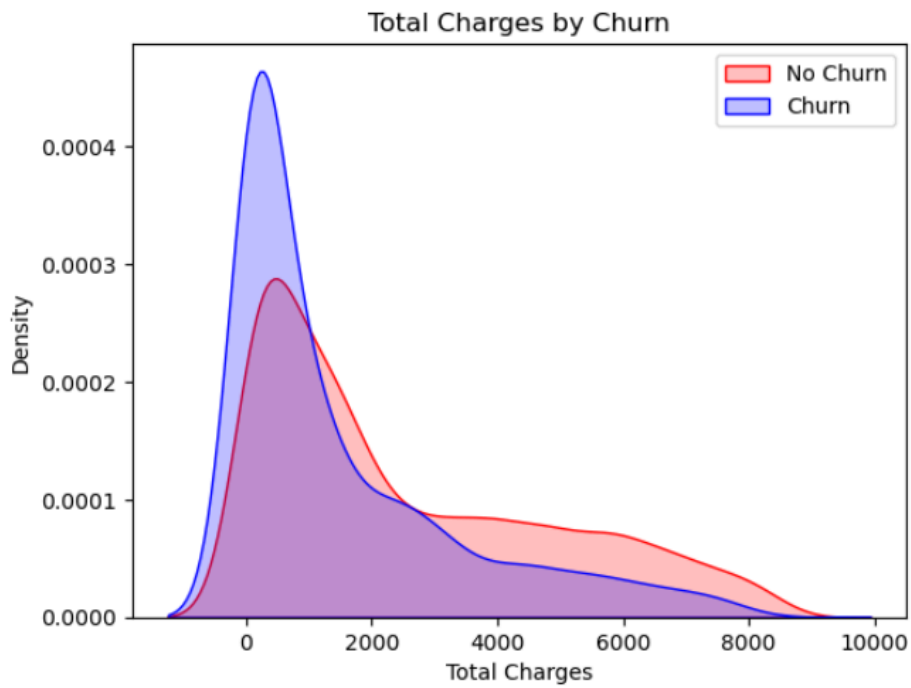
Insight - Churn is high when the monthly charges are high.



5.2.3) Churn rate by Total Charges :

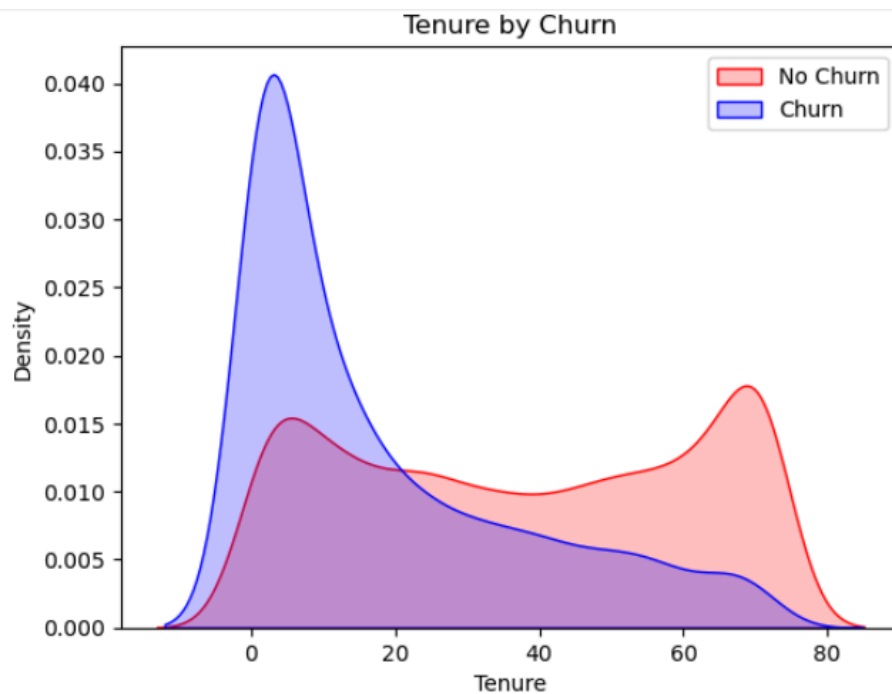
Insight - Surprising Insight : As higher Churn at lower Charges.

However if we combine the insights of 3 parameter i.e. Tenure, Monthly Charges & Total Charges then the picture is a bit clear :- Higher the Monthly Charges at lower tenure results lower Total Charges. Hence all these 3 factors viz Higher Monthly Charge, Lower Tenure and Lower Total Charge are linked to High Churn.

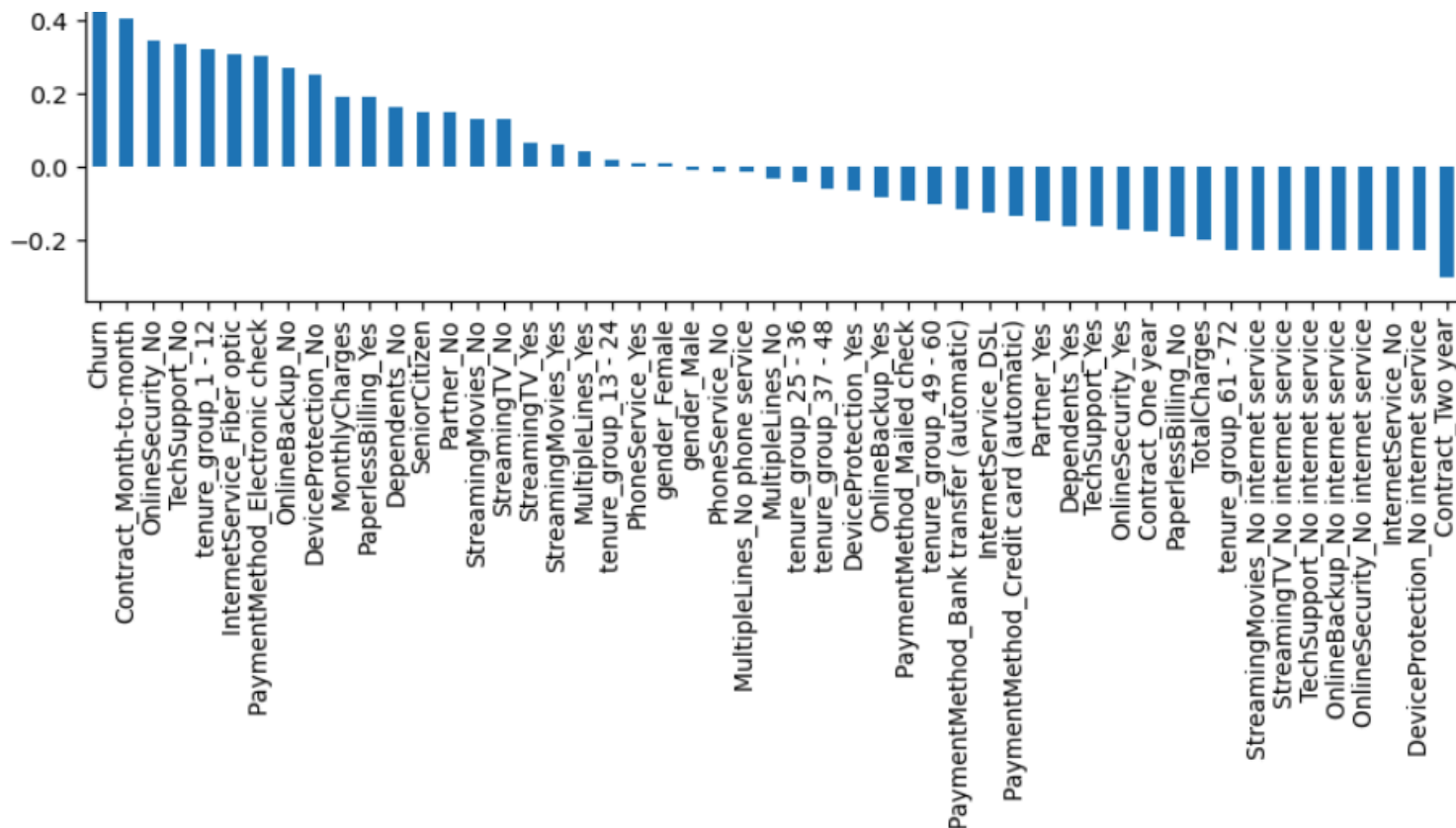


5.2.4) Churn Rate by Tenure :

Insight - As we can see in the graph the churn rate is higher in the lower tenure and it decreases as the tenure increases. In the beginning of the tenure from month 0-12 the churn rate is much more but as the tenure increases from month 60-80 the churn rate is lower.



5.2.5) Build correlation for all predictors :



Derived Insight :

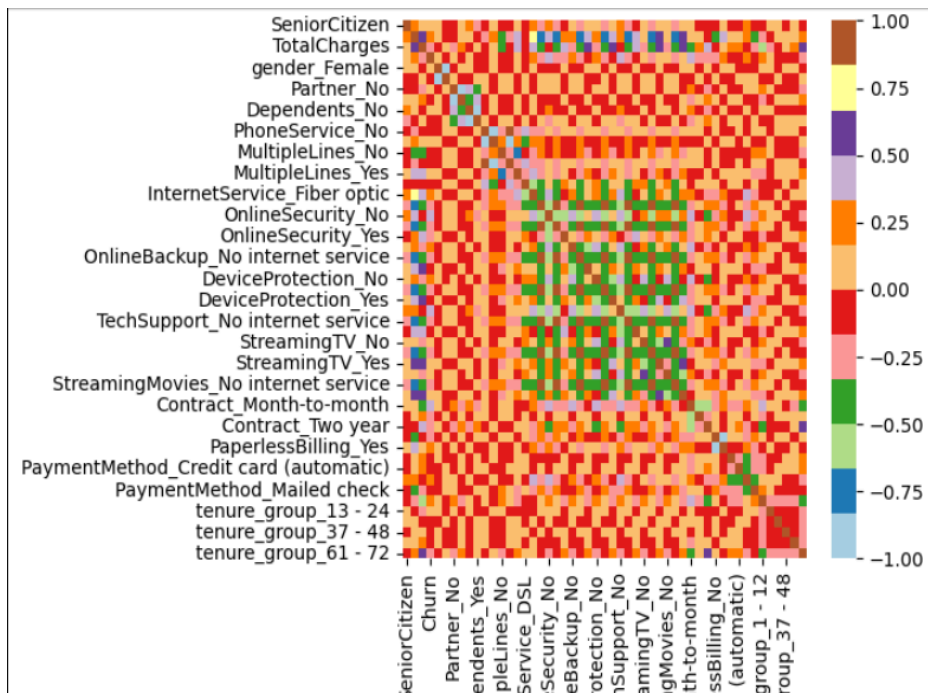
High Churn is seen in case of Month to month contracts, No Online security, No tecg Support, First Year of Subscription and Fibre Optics Internet.

Low Churn is seen in case of Long term contracts, Subscription without Internet Services and the customers engaged for 5+ years.

Factors like Gender, Availability of PhoneService and multiple lines have almost no impacts on churn.

This is also evident from Heatmap below.

We can also plot heatmap for the similar correlation :

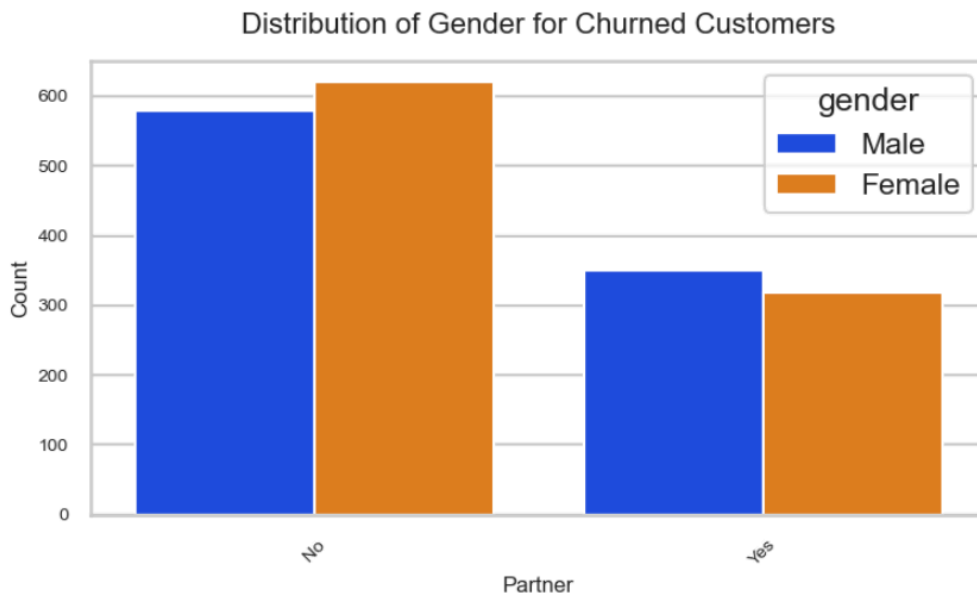


5.3) Bivariate Analysis :

In Bivariate analysis we take two variables together to get some more deep insights about the data set, Here we are taking the Gender and Partner column together and we found some more insights.

5.3.1) Partner and Gender comparison :

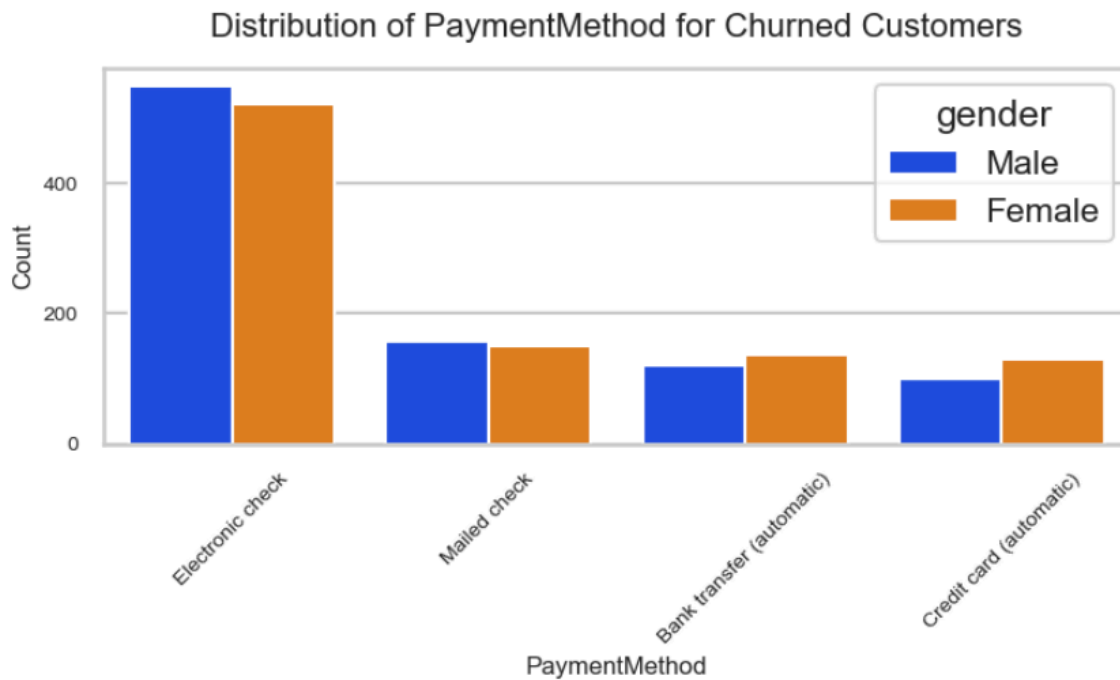
Insight - From this graph we get to know that females with no partner are churning more in comparison to the male one.



5.3.2) Payment Method and Gender comparison :

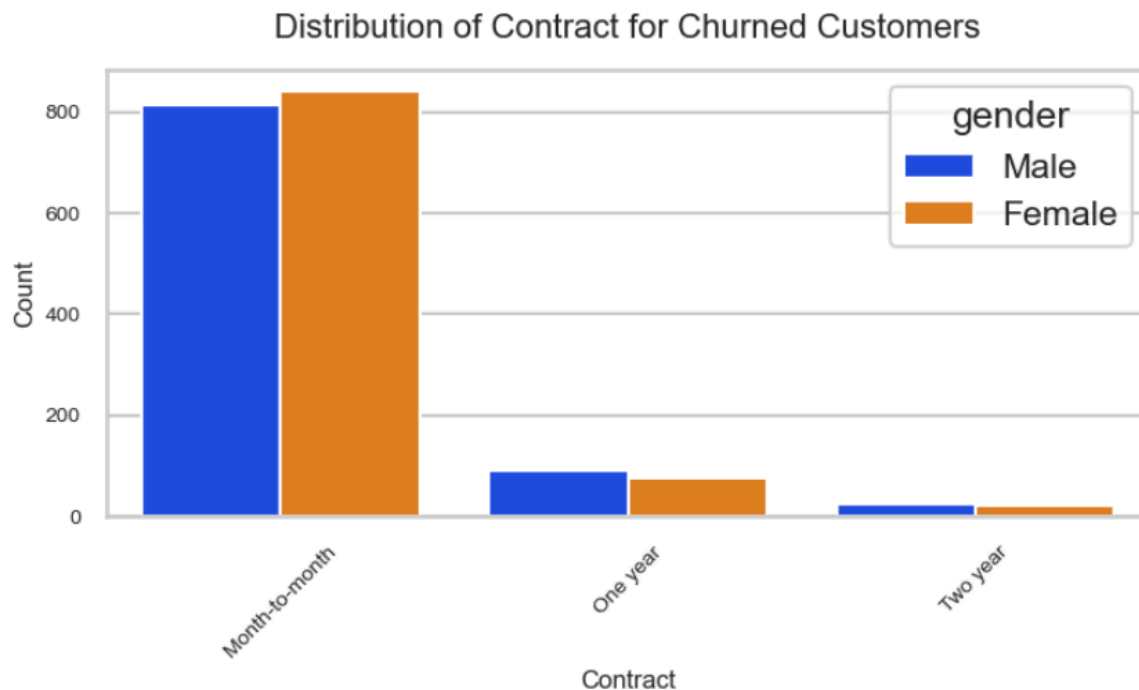
Insight - i) Male churners are slightly higher than the female churners via payment through electronic cheque.

ii) Female churners are higher via payment through credit card.



5.3.3) Contract and Gender comparison :

Insight - Female month on month churners are slightly higher than the male churners.



6) Key Insights :

- 1) The Churn rate is higher for the **senior citizens** by almost 45%.
- 2) People using **internet service (fiber optic)** have a higher churn rate by almost 41.66%.
- 3) People with no **online security,online backup, device protection and tech support** have a higher churn rate, almost 41%.
- 4) People who have **contracts** or plans on a monthly basis their churn rate is much higher around 41% but as the contract periods extend from month on month to one or two years their churn rate decreases.

- 5) As we can see in the graph the churn rate is higher in the lower **tenure** and it decreases as the tenure increases. In the beginning of the tenure from month 0-12 the churn rate is much more but as the tenure increases from month 60-80 the churn rate is lower.
- 6) Churn is high when the **monthly charges** are high.
- 7) Surprising Insight : As higher Churn at lower Charges.
- 8) However if we combine the insights of 3 parameter i.e. Tenure, Monthly Charges & Total Charges then the picture is a bit clear :- Higher the Monthly Charges at lower tenure results lower Total Charges. Hence all these 3 factors viz Higher Monthly Charge, Lower Tenure and Lower Total Charge are linked to High Churn.
- 9) We get to know that **females** with no **partner** are churning more in comparison to the male one.
- 10) i) Male churners are slightly higher than the female churners via **payment** through electronic cheque.
ii) Female churners are higher via payment through credit card.
- 11) Female **month on month** churners are slightly higher than the male churners.

Similarly by plotting more graphs we can make more deep insights about that particular data by doing multi-variate analysis.

7) Recommendations :

- 1) Offer incentives such as discounts or loyalty rewards to encourage long-term contracts.
- 2) Create affordable packages or adjust pricing to retain cost-sensitive customers.
- 3) Conduct customer satisfaction surveys to identify pain points.
- 4) Invest in infrastructure and technology upgrades to improve service reliability and speed.
- 5) Train support staff to handle complaints efficiently & Implement proactive support channels, such as chatbots or self-service options, to resolve common problems quickly.
- 6) Implement targeted marketing campaigns for at-risk customers, highlighting services they frequently use.
- 7) Introduce a points-based loyalty program that rewards tenure, service usage, or referrals.
- 8) Conduct periodic churn analysis to stay updated on emerging trends.
- 9) Adjust retention strategies dynamically based on real-time data.
- 10) Highlight unique value propositions in marketing (e.g., better network coverage or faster speeds).
- 11) Monitor competitor offerings and adjust strategies to remain competitive.

8) Limitations :

The dataset lacks behavioral data like customer feedback, which could provide deeper insights.

9) Conclusion :

This analysis highlights critical churn factors, enabling targeted retention strategies to enhance customer loyalty.

