Telecom Customer Churn Prediction

Introduction :-

Customer churn, also known as customer attrition, customer turnover, or customer defection, is the loss of clients or customers. Telephone service companies, e-commerce companies, internet service providers, pay TV companies, insurance firms, etc., often use customer churn analysis and customer churn rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Churn rate is the amount of customers or subscribers who cut ties with the service or company during a given time period. These customers have "churned".

Problem at hand :-

Predicting the behaviour of the customers leading to churn.

Value to client :-

A telecom company has been affected by the increasing number of customers subscribing to the services of a competitor. It is much more expensive to attract new customers than retaining old customers. At the same time, spending too much on or spending on the wrong factor for retaining a customer who has no intention to leave (or who was not leaving for that factor which was addressed) could be a waste of money. Therefore it is important to identify the customer who has high probability of leaving. An analysis of the past records of the customers can give great insights on who might leave and what is the cause.

We can predict behaviour to retain customers. We can analyze all relevant customer data and develop focused customer retention programs.

Data Source :-

www.kaggle.com/blastchar/telco-customer-churn

Methodology :-

- Data Wrangling
- Exploratory Data Analysis (EDA) and Visualization
- Data Storytelling
- Training and Testing Machine Learning models
- Recommendations to retain the customers
- Scope for future work

<u>**Data Wrangling**</u>

The data was downloaded from Kaggle. Each row represents a customer and each column contains customer's attributes described on the column Metadata. The "**Churn**" column is our target.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

The data was then imported into a Pandas DataFrame for ease of data manipulation. The DataFrame contains 7043 rows (customers) and 21 columns (features).

Initially on checking the information of the DataFrame, no nulls were shown. On further inspection I saw the column 'TotalCharges' had a few rows with spaces, 11 rows to be exact. Tenure for all these 11 exact rows is also 0, churn is "No" as well. One can interpret this as belonging to all new customers, if indeed these are the only rows with tenure = 0 too. Hence we can set 'TotalCharges' to 0, whenever tenure is 0. So I used the replace function to convert the spaces to 0. I also checked for the unique values and items of each column. The 'customerID' column is of no use to us, so I created a new DataFrame without the 'customerID' column using the iloc method. The new shape of the DataFrame is 7043 rows by 20 columns.

On the whole this data was very clean and required minimal data cleaning steps. As of now the DataFrame looks good to be used for further analysis. If any data manipulation is required in the later part of the project, it'll be done accordingly.

Exploratory Data Analysis

Data Story

Here we are going to ask and answer a few questions about the data with respect to our dependent variable and also visualize & get insights from it.

Target (Dependent) Variable: Churn

Feature (Independent) Variables: 19 variables out of 20.

We can divide the independent variables into:

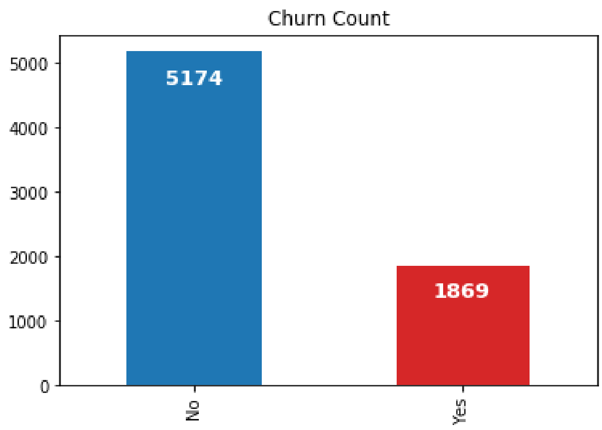
- **Person specific :** gender, SeniorCitizen, Partner, Dependents, tenure
- **Service specific :**
  - Phone : PhoneService, MultipleLines
  - Internet : InternetService, OnlineSecurity, OnlineBackup, StreamingTV, Streaming Movies, TechSupport, DeviceProtection
- **Money specific :** MonthlyCharges, TotalCharges, Contract, PaperlessBilling, PaymentMethod

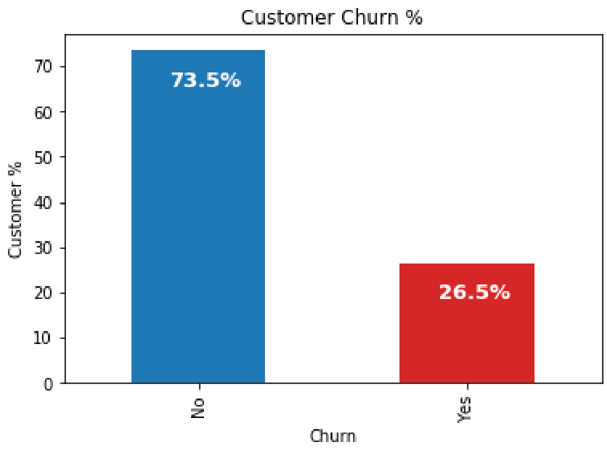
The questions to which we seek answers:

- Is there a gender specific to churn?
- Are there any person specific trends in churn?
- Is there a correlation between tenure and churn?
- Is there a correlation between certain types of services and churn?
- Is there a correlation between different types of contract and churn?
- Is there a correlation between paperless billing and churn?
- Is there a correlation between different types of payment methods and churn?
- Is there a correlation between monthly charges and churn?
- Is there a correlation between total charges and churn?
- Is there a correlation between monthly charges and total charge with respect to churn?
- Is there a correlation between monthly charges and tenure with respect to churn?
- Is there a correlation between total charges and tenure with respect to churn?

Customer Churn in the data :

First let us check out the number of customers who have churned or not churned.



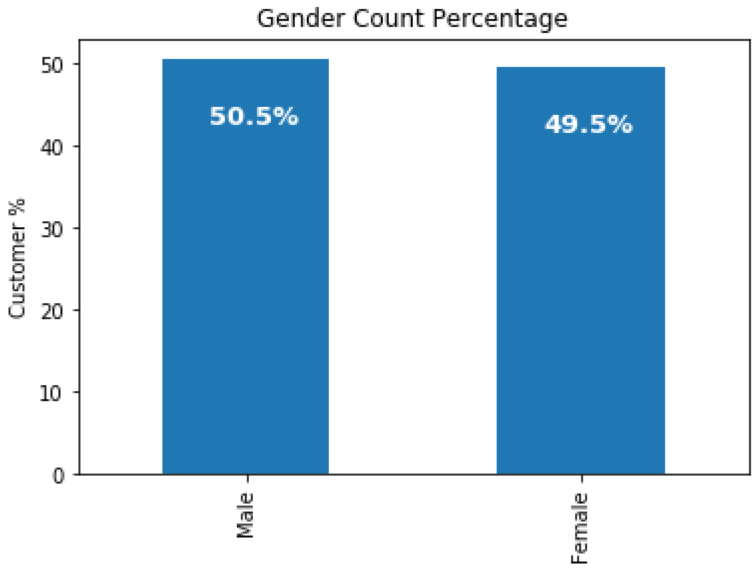


➢ **So, we can see from above:**

- Churn No - 5174 or 73.5%
- Churn Yes - 1869 or 26.5%

Gender distribution :

Let's see the gender distribution in the dataset.

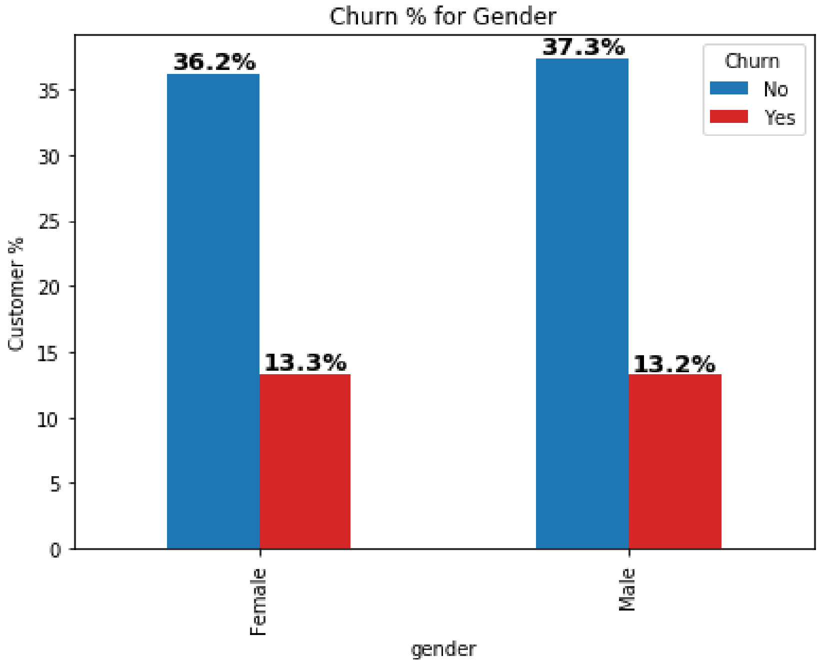


➢ We can see that the gender distribution looks balanced.

Visualizing the customer's attributes with respect to Churn :

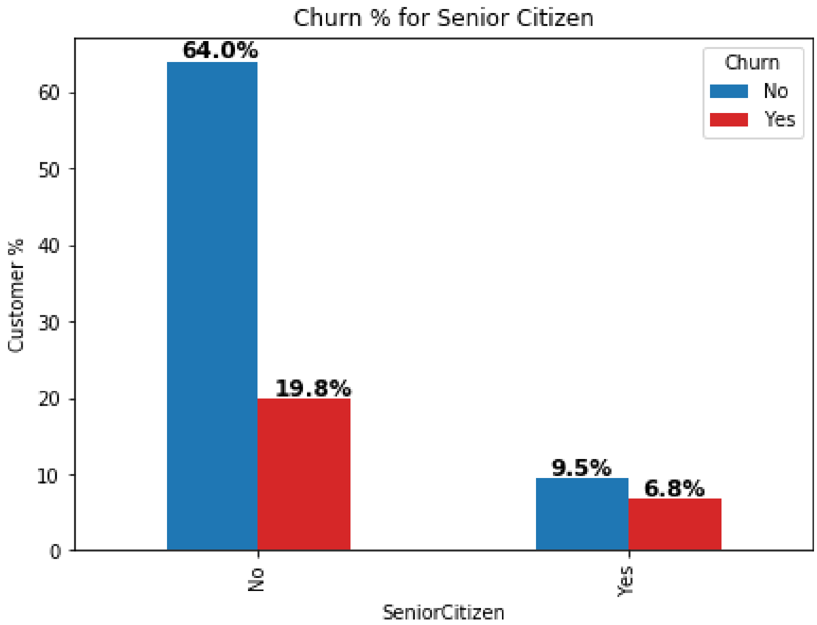
1. **Person specific attributes -**

Gender



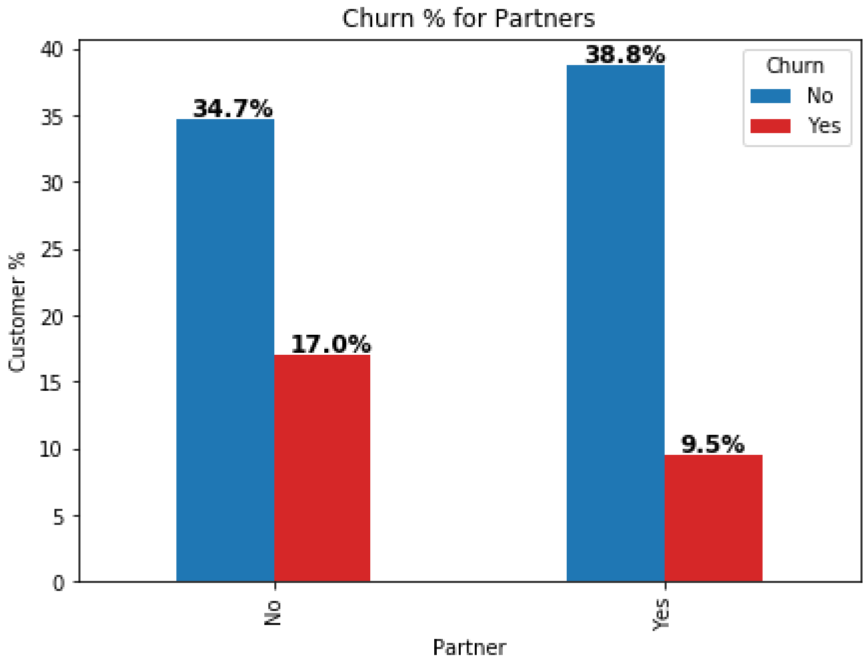
➢ The churn rate is not affected by the gender.

Senior Citizen



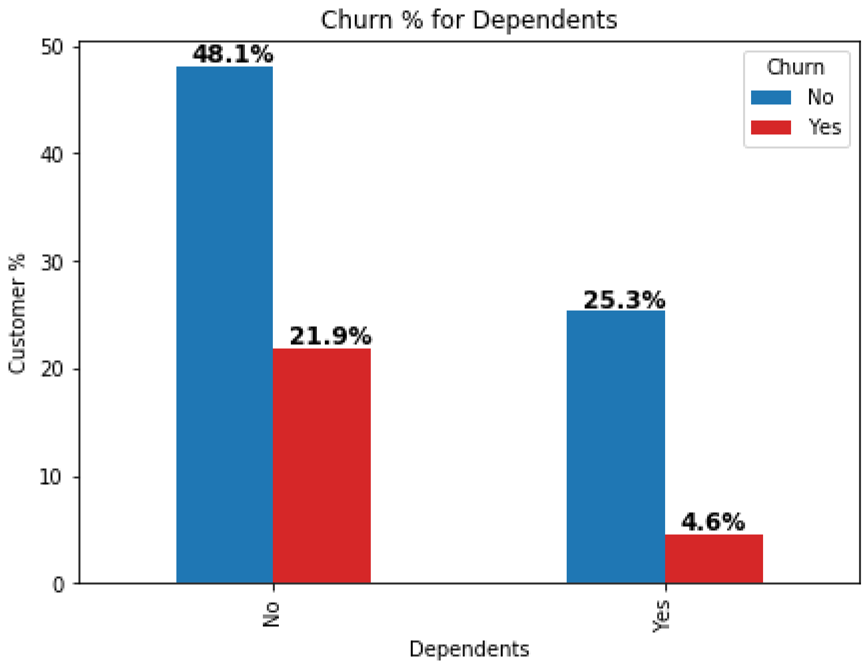
➢ Senior Citizens tend to churn less compared to non-senior citizens.

Partners



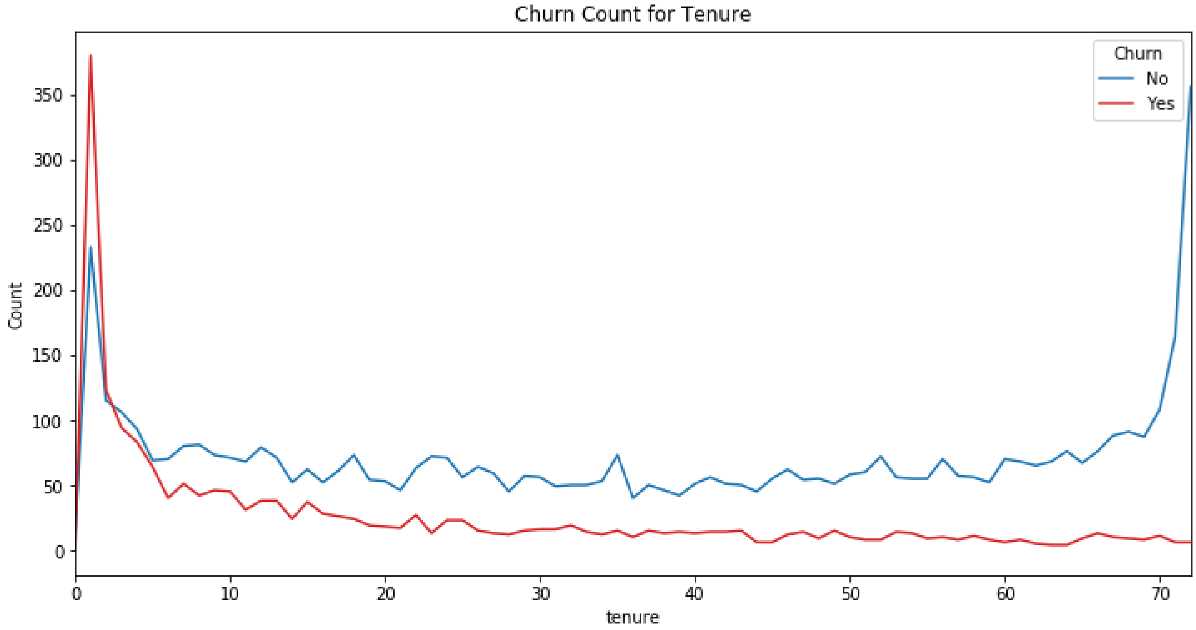
➢ Customers who don't have partners have higher churn rate.

Dependents



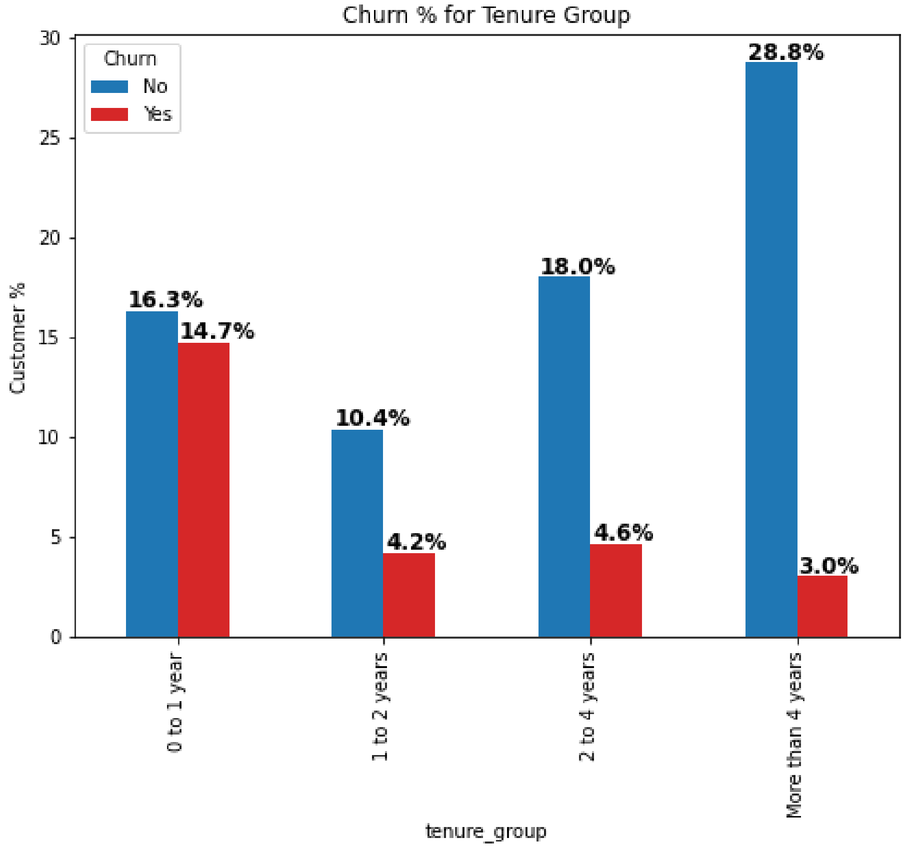
➢ Customers without dependents have higher churn rate.

Tenure



- Churn count decreases as the tenure increases.
- Customers tend to churn within the first few months or within a year.

Let's create 4 tenure groups to check the churn rate more clearly.

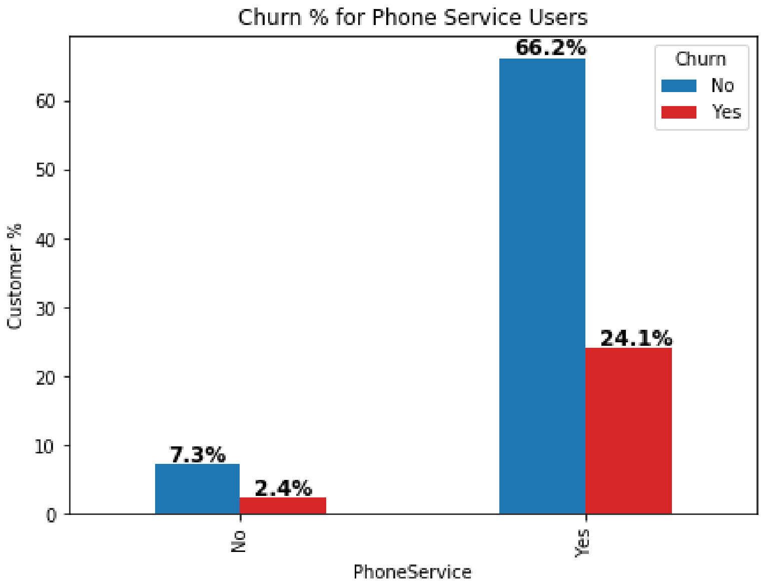


- Now we can clearly see that the churn rate is high within the 1st year.

2. Service specific attributes -

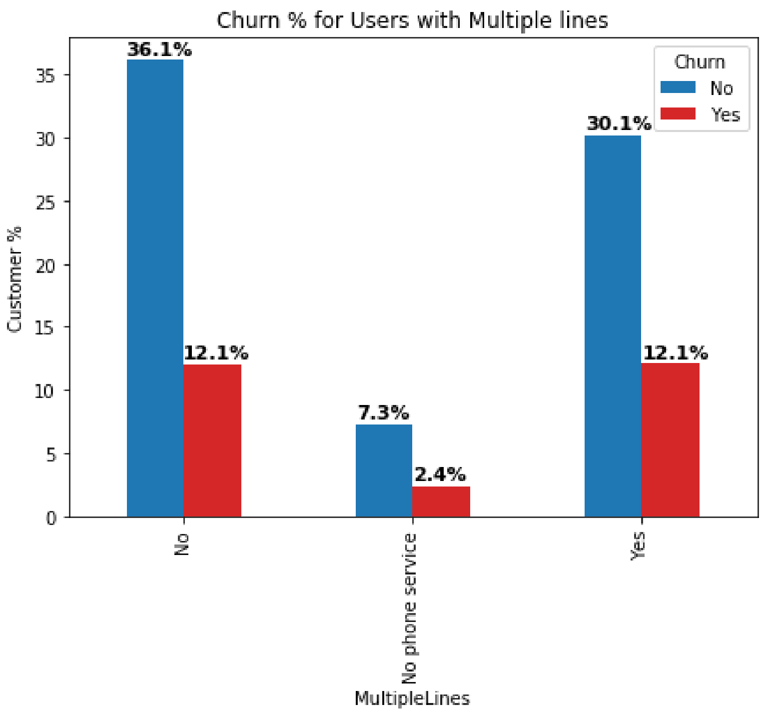
a. Phone:

Phone Service



- ➢ Customers having phone service have higher churn rate.

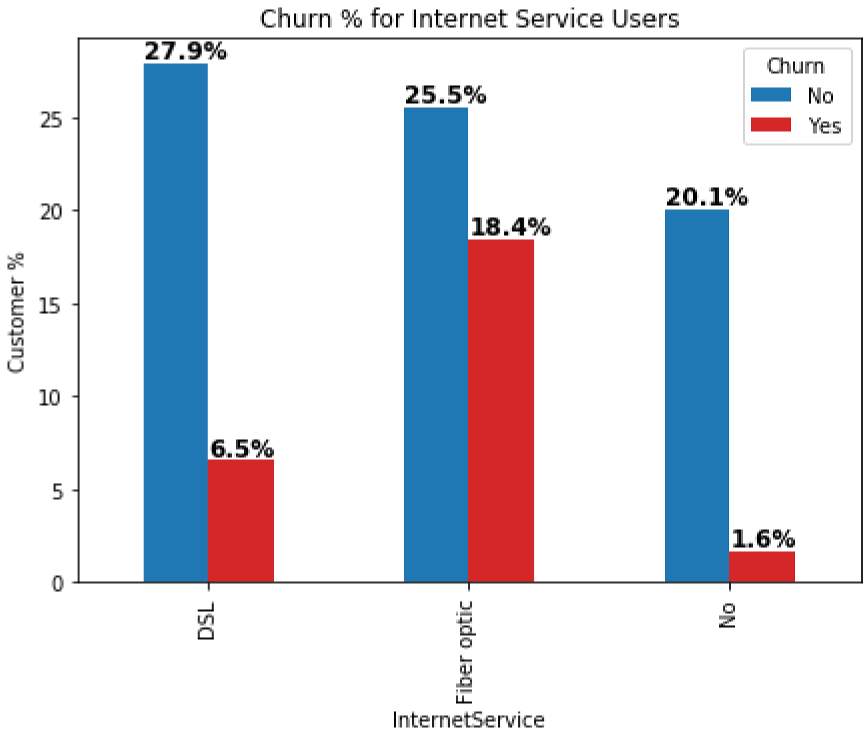
Multiple lines



- ➢ Customers having multiple lines or not does not affect the churn rate.
- ➢ Customers without phone service tend to churn less.

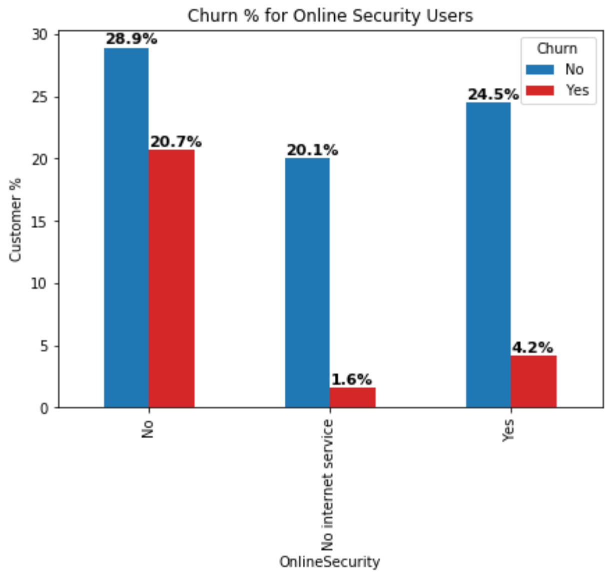
b. Internet:

Internet Service

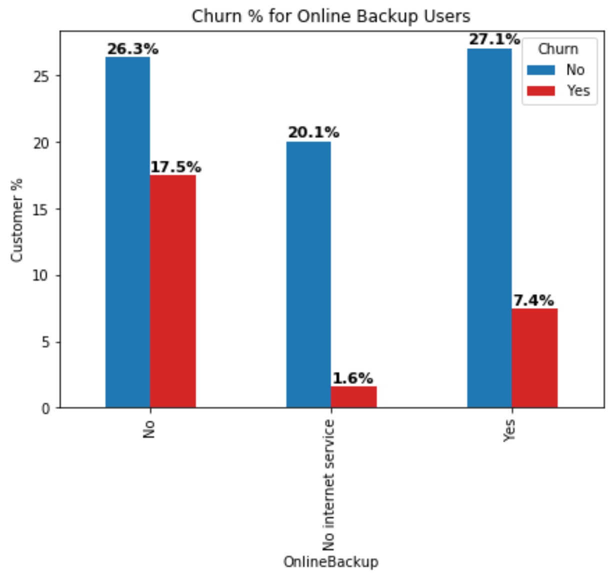


- Customers with fiber optic connection have higher churn rate.

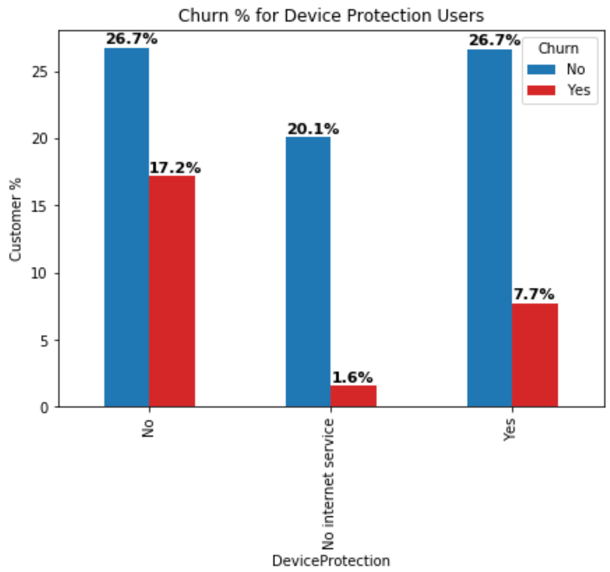
Online Security



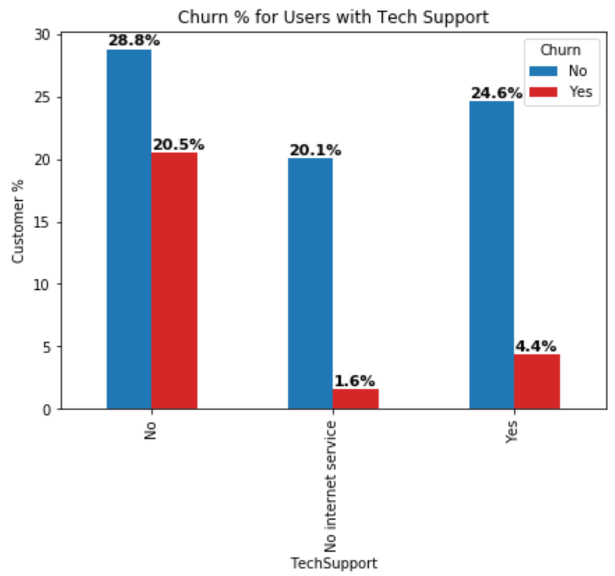
Online Backup



Device Protection

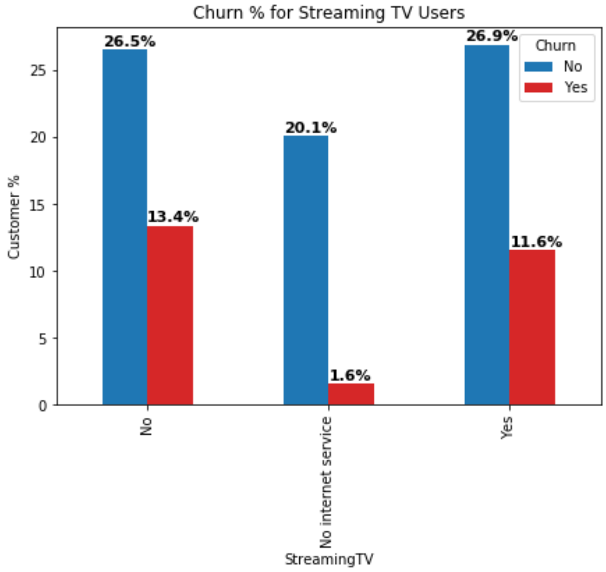


Tech Support

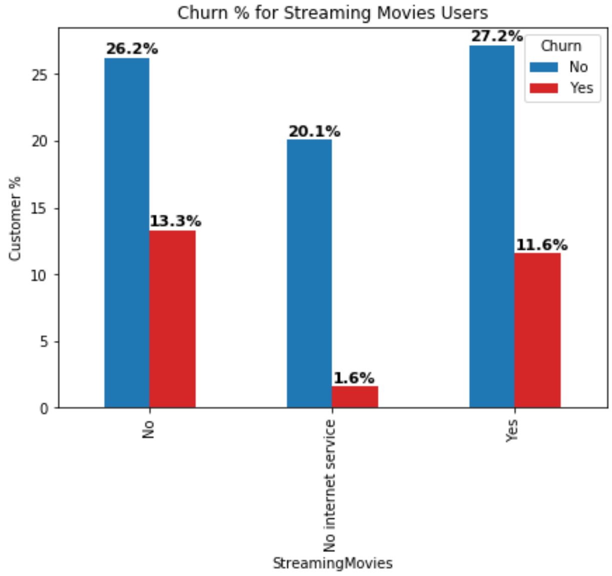


- Customers who do not have Online Security, Online Backup, Device Protection and Tech Support have higher churn rate.

Streaming TV



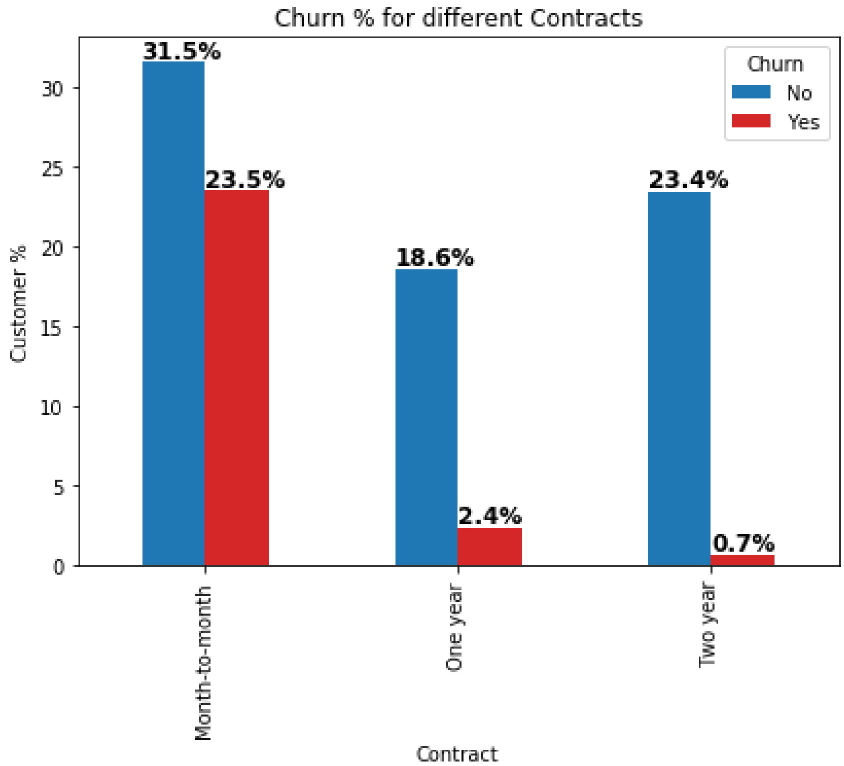
Streaming Movies



➢ The churn rate do not have a big difference between the customers having the service of Streaming TV & Streaming Movies or not.

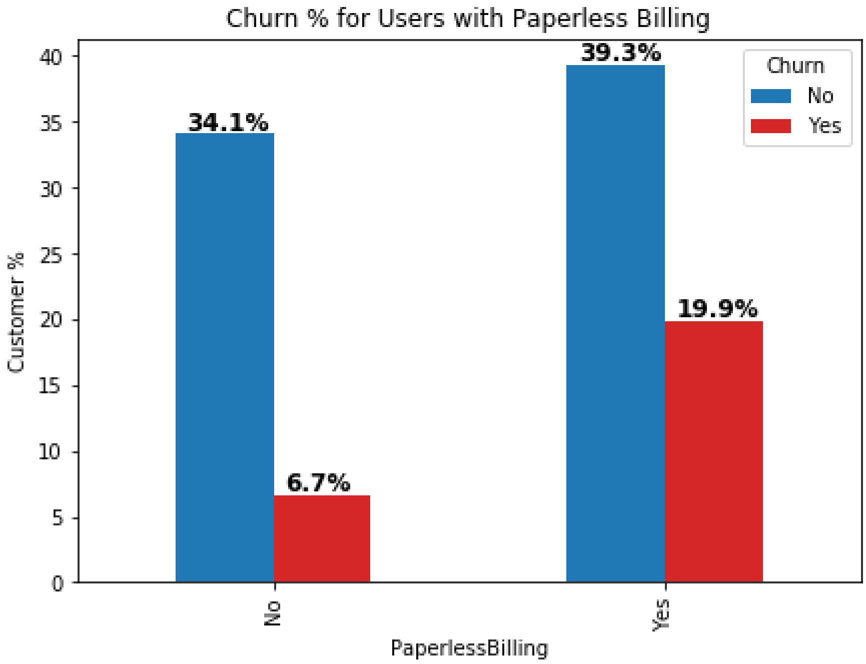
3. Money specific attributes -

Contract



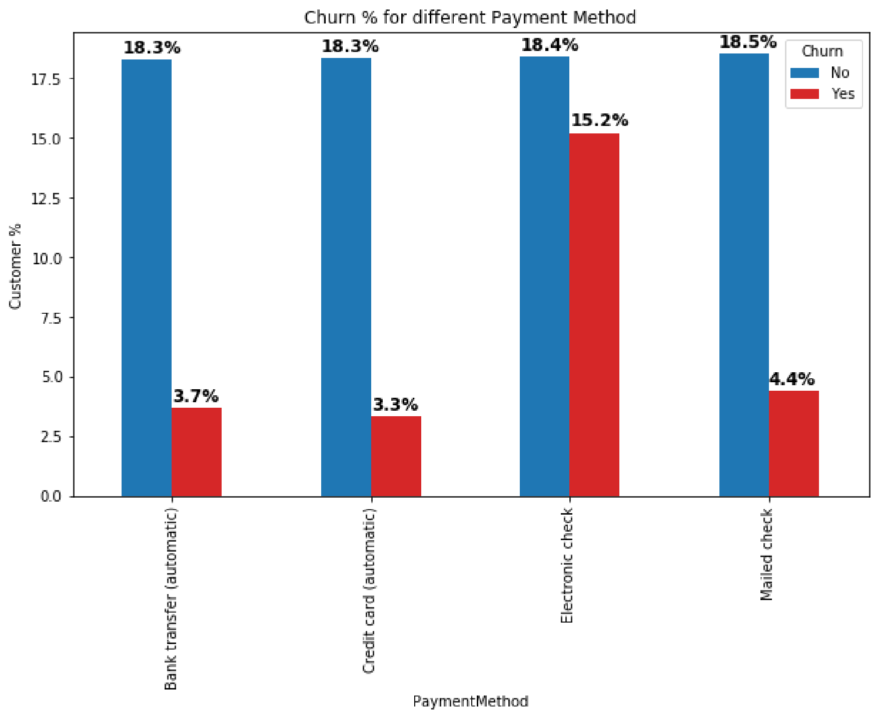
➢ Customers having Month-to-Month contracts have a high churn rate.

Paperless Billing



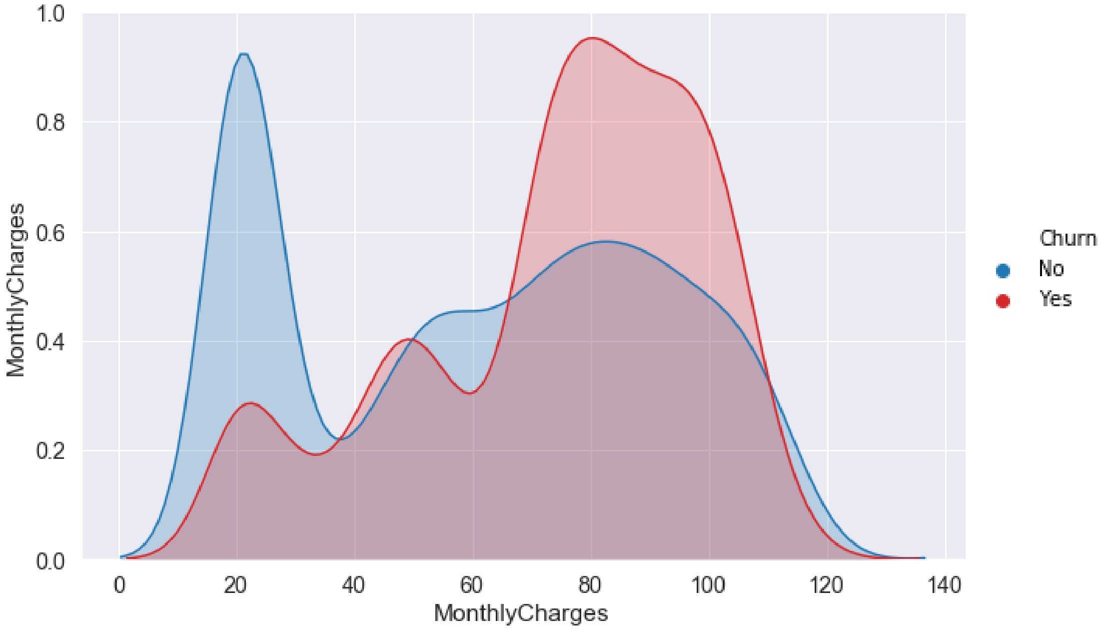
- Customers with paperless billing tend to churn out more.

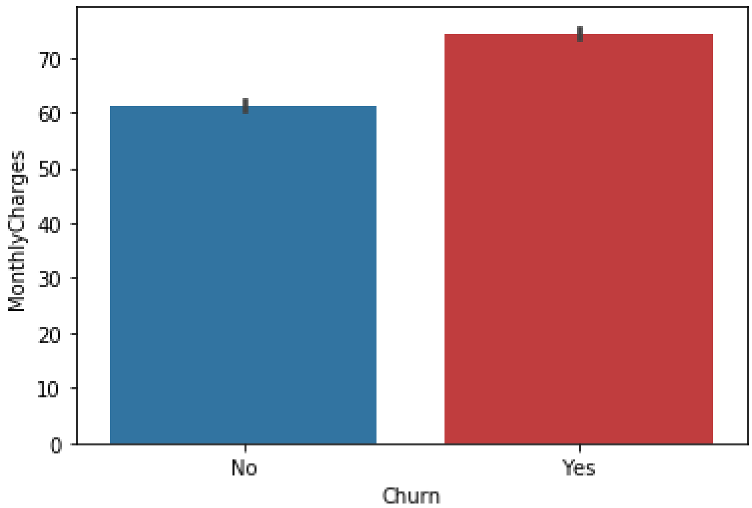
Payment Method



- Customers who pay electronic check have a high churn rate.

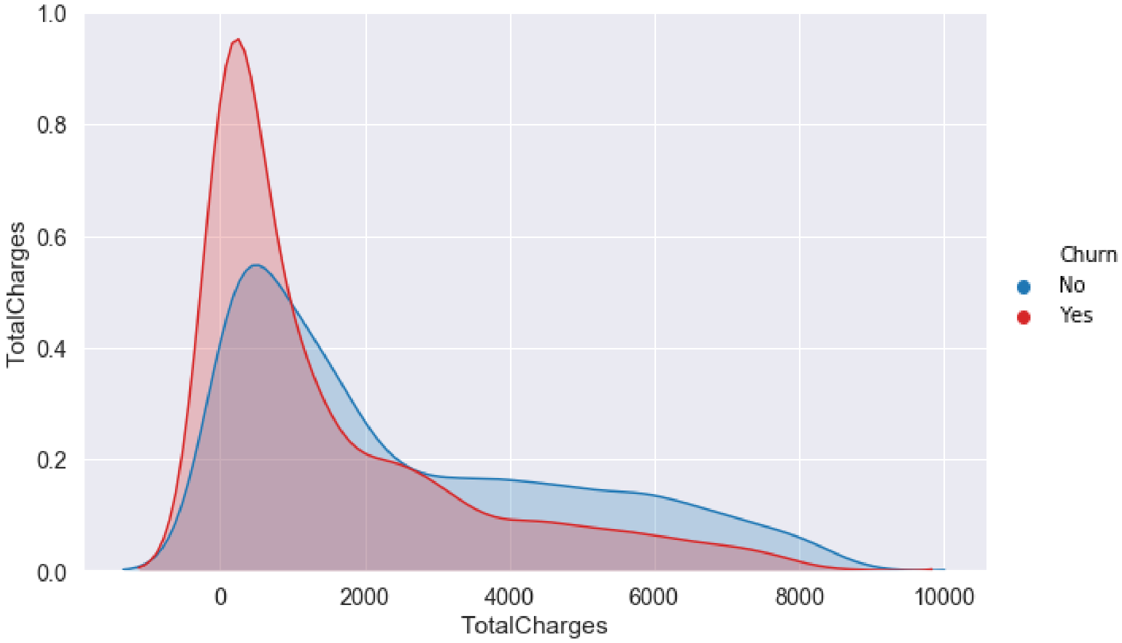
Monthly Charges

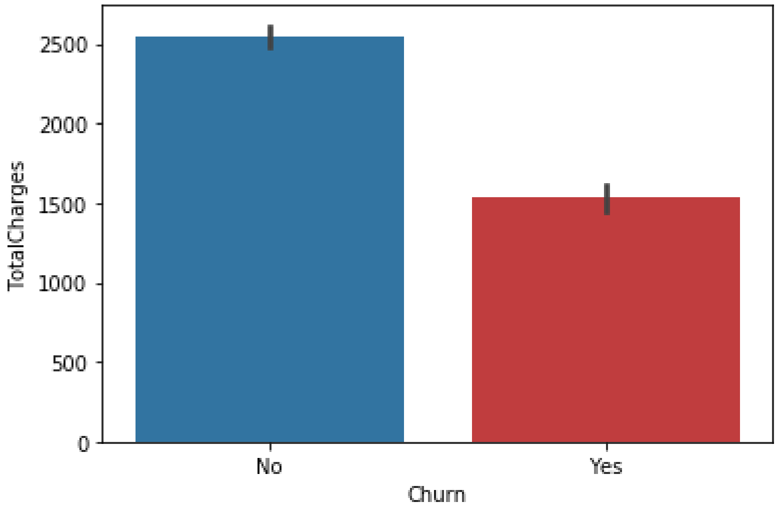




- ➢ Churn rate increases as Monthly Charges increases.

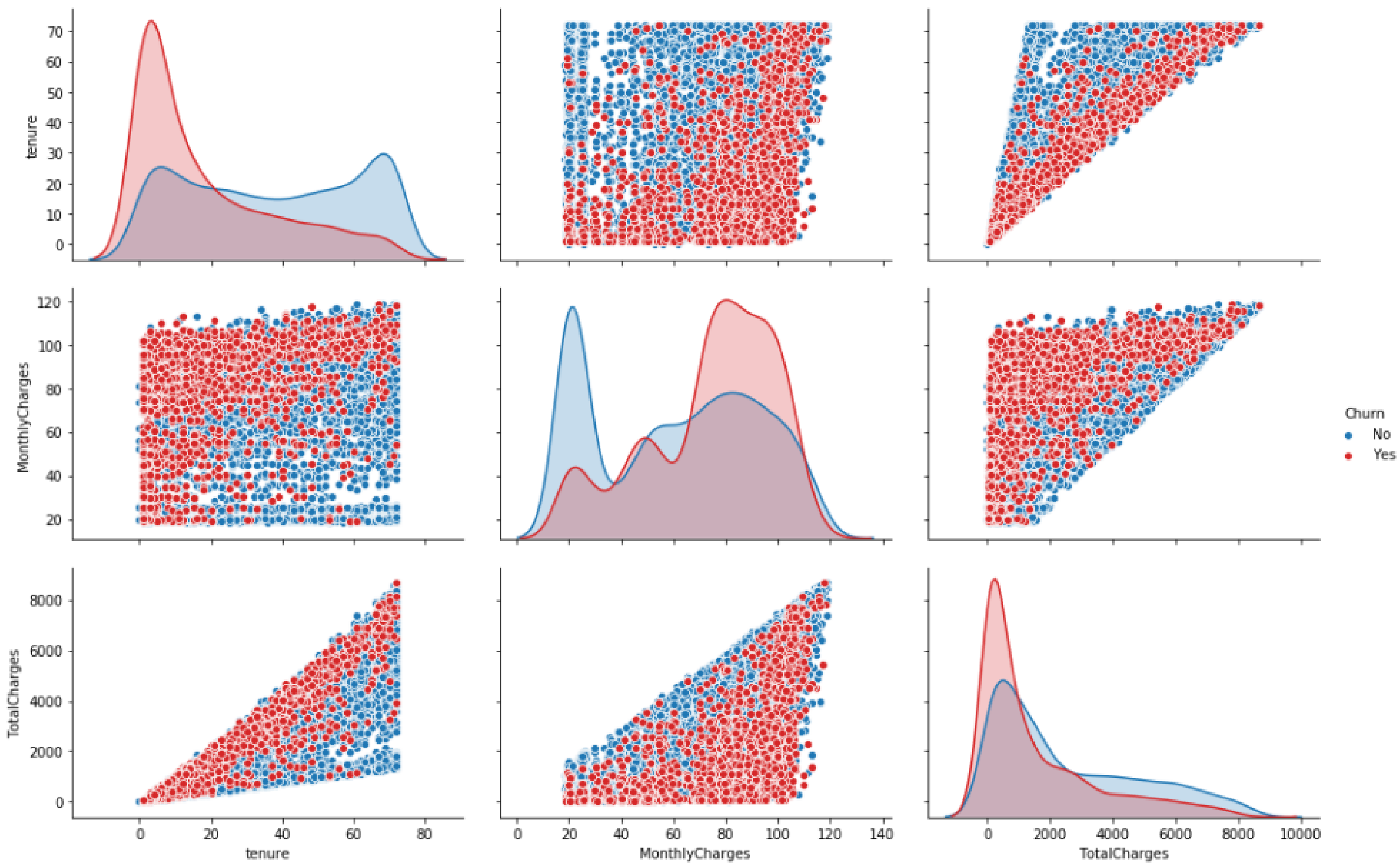
Total Charges



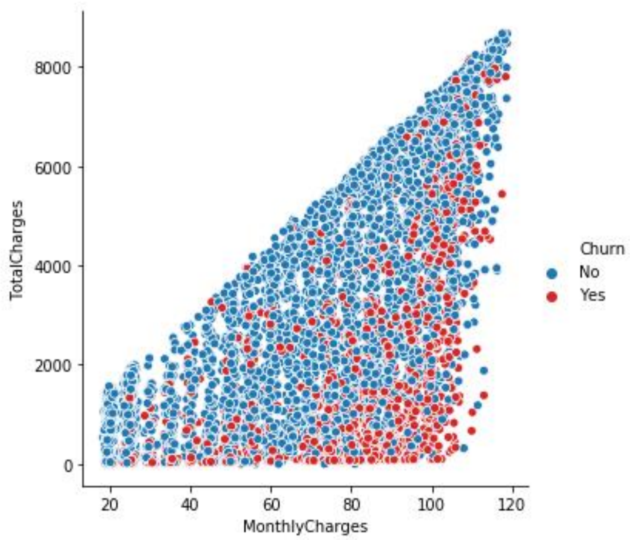


➢ Churn rate decreases as Total Charges increases.

Relationship between the numeric columns with respect to Churn:

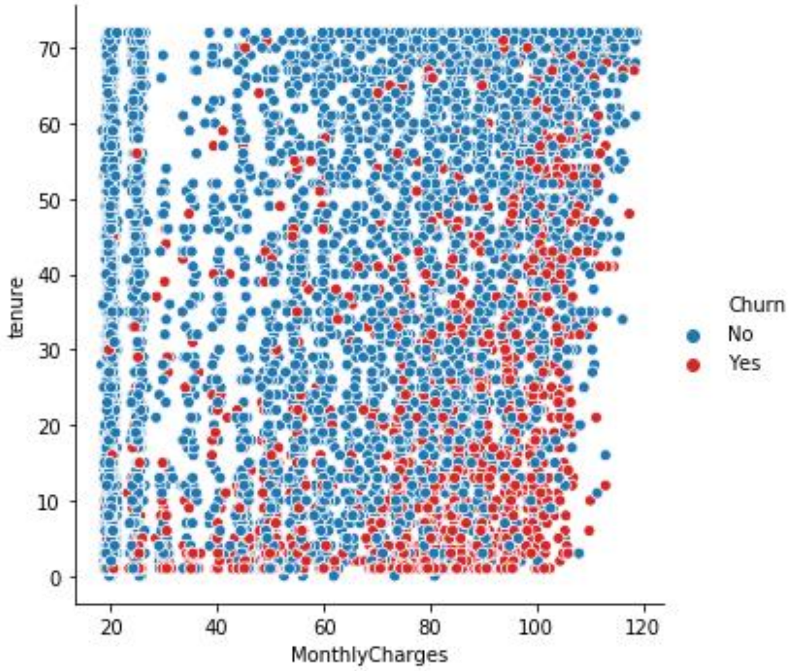


Monthly Charges and Total Charges by Churn -



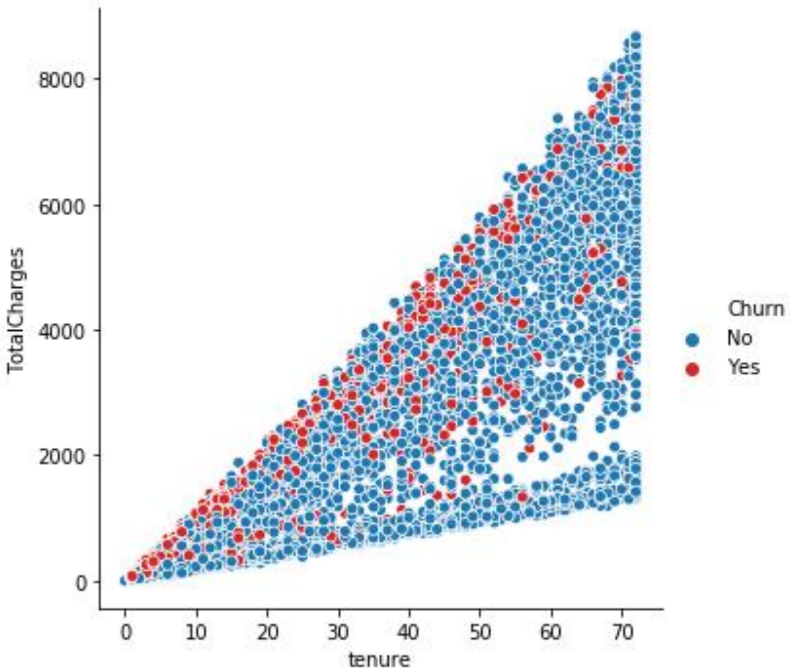
- Total Charge increases as the Monthly Charge increases.
- Churn is mainly towards the bottom which indicates that churn increases with increase in monthly charge.

Monthly Charges and Tenure by Churn -



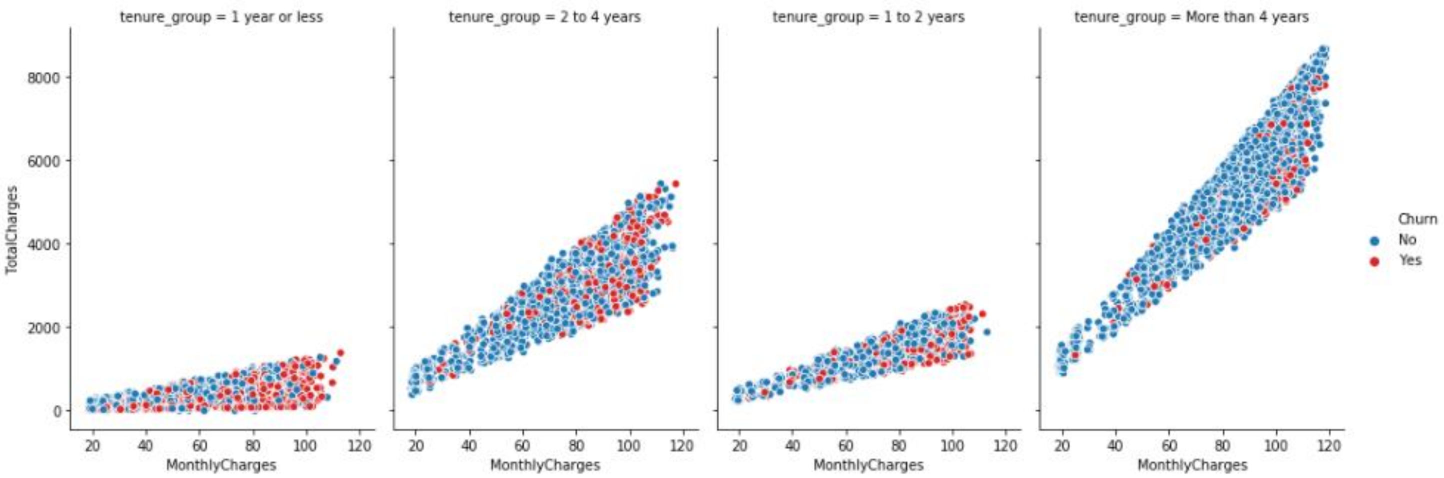
- Monthly charges may or may not increase with tenure.
- Again we can see that churn increases with increase in monthly charge.

Total Charges and Tenure by Churn -



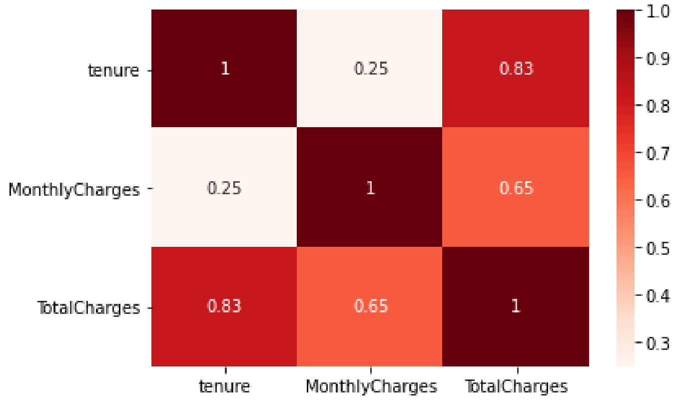
- Total charges increases with tenure.
- Churn rate does not increase so much with increase of either total charge or tenure.

Monthly Charges, Total Charges and Churn by Tenure groups -



- This clearly indicates that the churn rate is high within the 1st year and also it increases with monthly charge.

Correlation between the numeric columns:



- Tenure is highly correlated with total charge.
- Monthly charge is moderately correlated with total charge.

Inferential Statistics

Here we will be applying statistical tools to gain some inferences and insights into the data and discover relationships between various features of our dataset with the target variable by hypothesis testing.

We will use **chi-square test of independence of variables in a contingency table**.

The following results were obtained:

1. Gender has no influence on churn.
2. Senior Citizen has influence on churn.
3. Partner has influence on churn.
4. Dependents has influence on churn.
5. Tenure has influence on churn.
6. Phone Service has no influence on churn.
7. Multiple Lines has influence on churn.
8. Internet Service has influence on churn.
9. Online Security has influence on churn.
10. Online Backup has influence on churn.
11. Device Protection has influence on churn.
12. Tech Support has influence on churn.
13. Streaming TV has influence on churn.
14. Streaming Movies has influence on churn.
15. Contract has influence on churn.
16. Paperless Billing has influence on churn.
17. Payment Method has influence on churn.
18. Monthly Charge has influence on churn.
19. Total Charge has no influence on churn.

Gender, phone service and total charge have no influence on churn according to the chi-square test of independence of variables. 16 out of 19 variables have influence on churn.