Telecom Churn Prediction

ML Capstone Project to predict customer churn in Telecom Industry

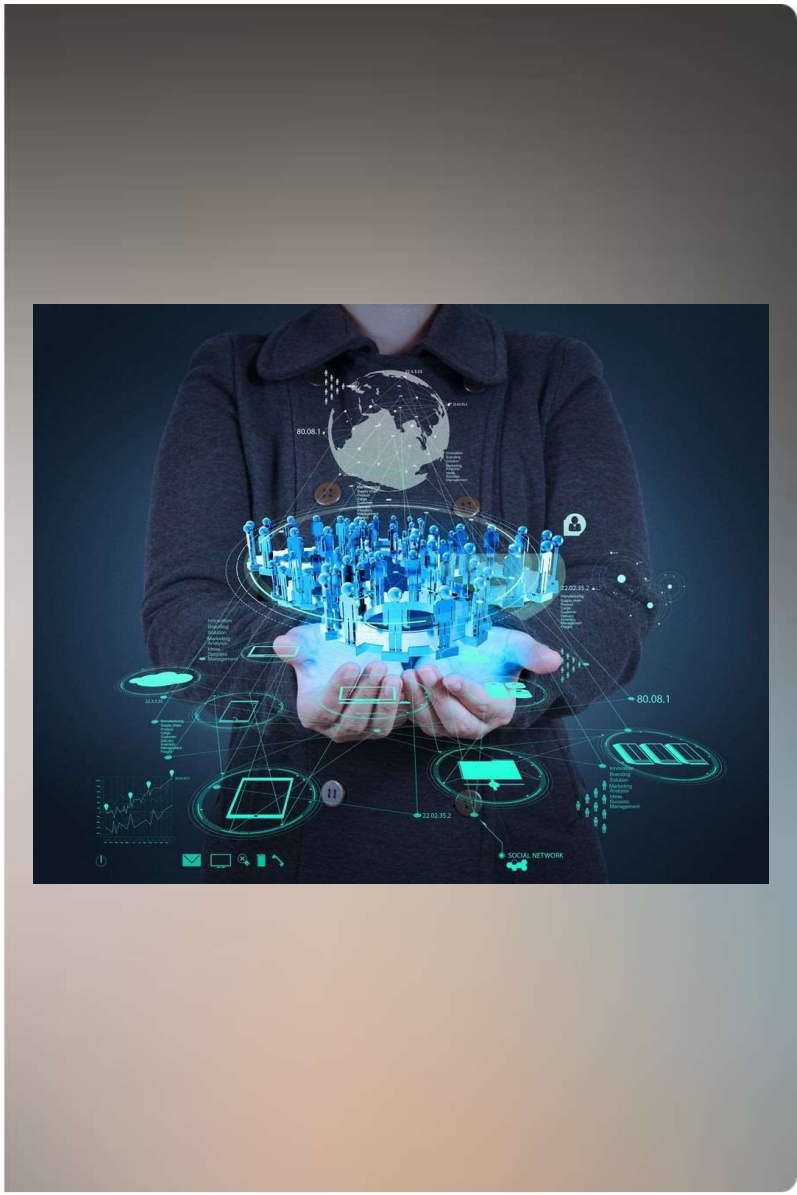
D **by Dheeraj Dua**

Introduction

Countries like India have experienced a significant digital revolution, driven by plummeting data costs and decreasing cellphone prices. As a result, telecom providers now face intense competition and shrinking profit margins. The ability for customers to easily switch providers due to service quality issues has only intensified this challenge.

Customer churn in the telecom sector poses one of the most critical challenges for service providers worldwide. Retaining existing customers is notably more cost-effective than acquiring new ones.

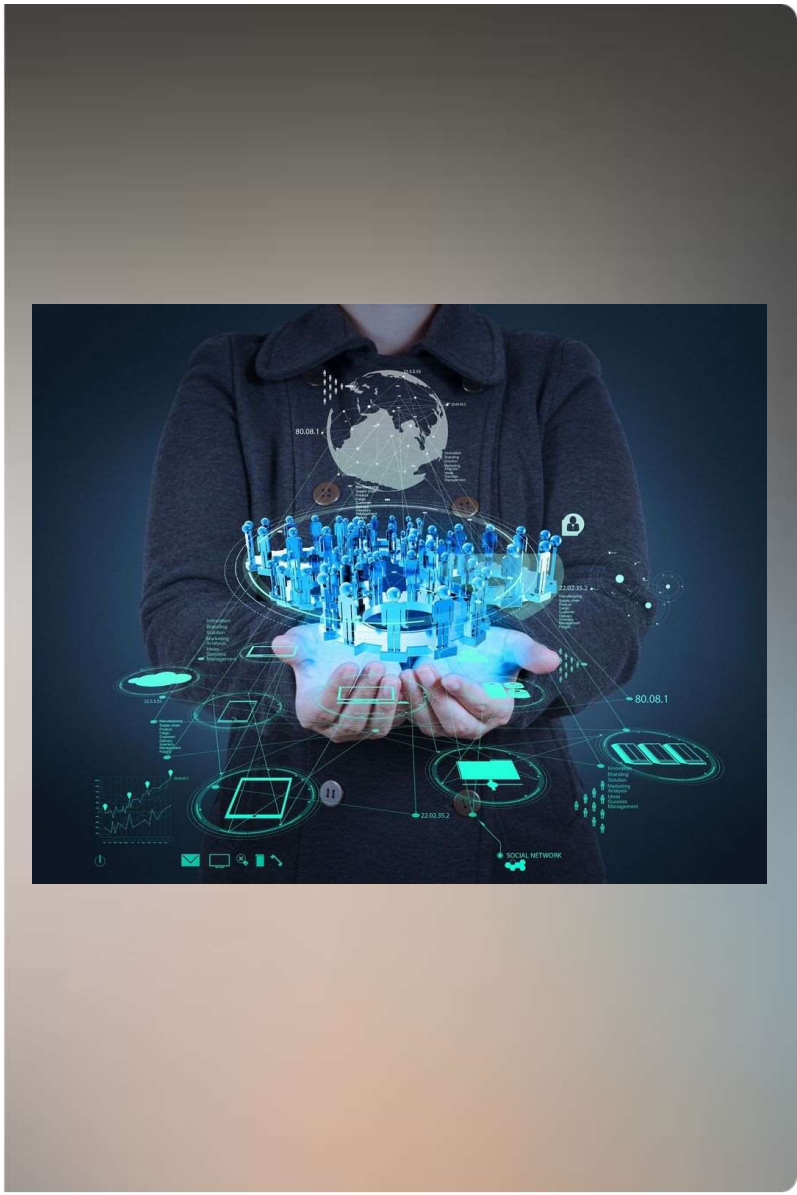
The goal of this case study is to identify the key factors contributing to customer churn within the telecom industry. By understanding the underlying drivers of churn, we aim to develop a predictive model that can forecast customer churn, enabling companies to proactively address these issues. Such a model would empower companies to retain their customers by offering tailored counteroffers, bundled packages, discounts, and other incentives.



Problem Statement

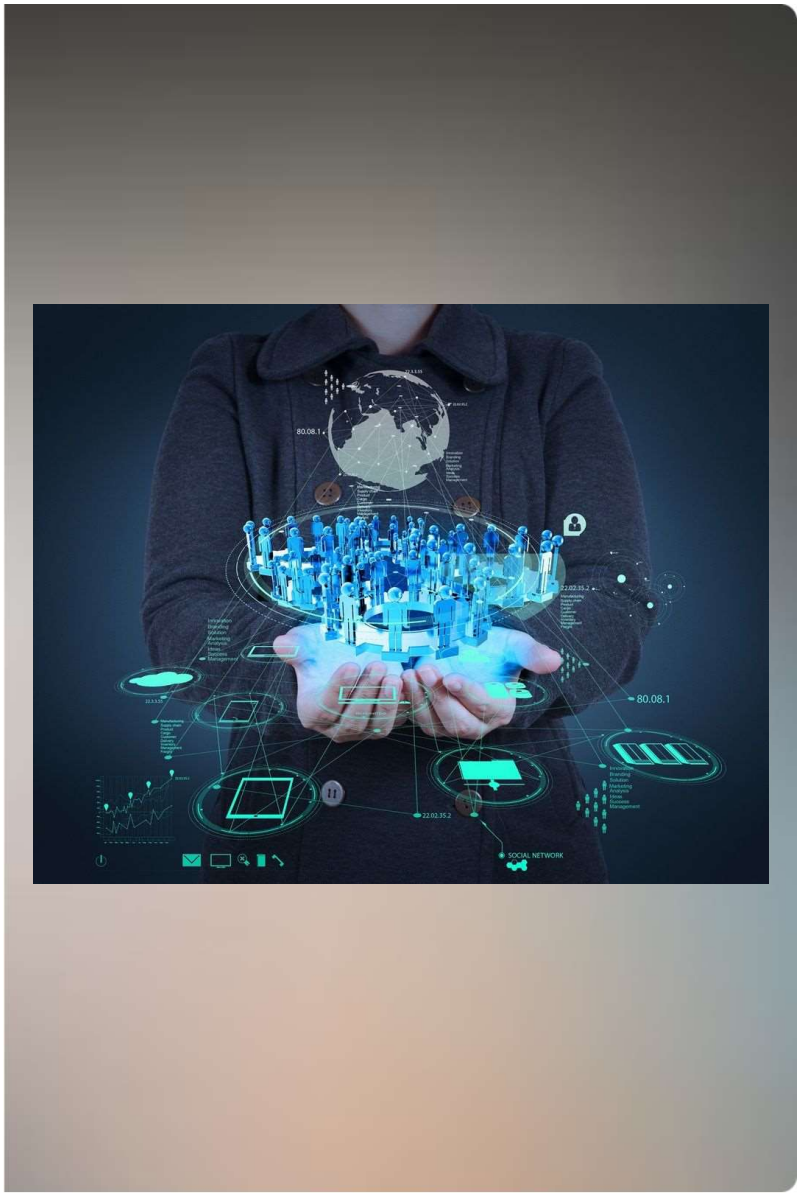
Due to increasing churn rate of customers from last several quarters, the company has decided to dive deep into factors driving it. Past efforts in retaining customers have been reactive where the suggested steps were applied at the point of no return.

The idea is to use machine learning to predict the likelihood of churn for each customer and reasons behind it. These reasons could be different for each customer. This will help with targeted response tailored to customer's preferences thereby helping them stay and improving customer loyalty.



Project Objective

- Leverage given dataset to predict customer churn
- Understanding main variables/features influencing customer churn
- Provide insights to the business on factors impacting customer churn
- Use various ML algorithms to build prediction models, evaluate the accuracy and performance of these models
- Finding the best model and provide recommendations to the Business to improve customer churn



Project Approach

Understanding of the Problem & dataset

Understand the given dataset and its associated features

1

Discovery

Leverage existing data and perform exploratory data analysis(EDA) to Generate actionable insights

2

Modeling

Use different ML algorithms to predict customer churn basis important features

3

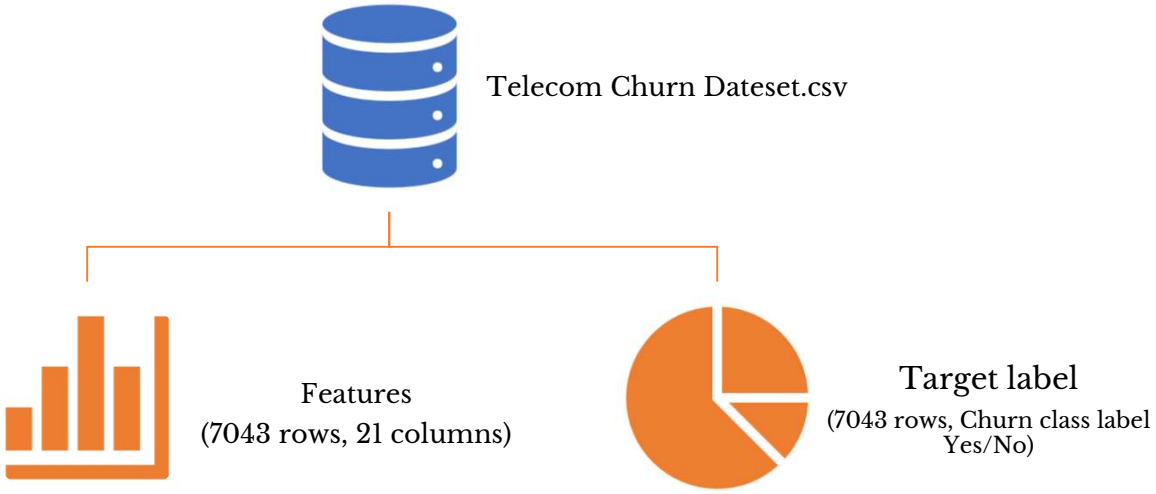
Recommendations

Provide recommendations to the business to improve customer churn

4

Part 1
Understanding of the Problem & Data

Dataset Overview



Name of the Features

Data Summary

skimpy summary

Data Summary

dataframe	Values
Number of rows	7043
Number of columns	21

Data Types

Column Type	Count
string	18
int32	2
float64	1

number

column_name	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
SeniorCitizen	0	0	0.1621	0.3686	0	0	0	0	1	
tenure	0	0	32.37	24.56	0	9	29	55	72	
MonthlyCharges	0	0	64.76	30.09	18.25	35.5	70.35	89.85	118.8	

string

column_name	NA	NA %	words per row	total words
customerID	0	0	1	7043
gender	0	0	1	7043
Partner	0	0	1	7043
Dependents	0	0	1	7043
PhoneService	0	0	1	7043
MultipleLines	0	0	1.2	8407
InternetService	0	0	1.4	10139
OnlineSecurity	0	0	1.4	10095
OnlineBackup	0	0	1.4	10095
DeviceProtection	0	0	1.4	10095
TechSupport	0	0	1.4	10095
StreamingTV	0	0	1.4	10095
StreamingMovies	0	0	1.4	10095
Contract	0	0	1.4	10211
PaperlessBilling	0	0	1	7043
PaymentMethod	0	0	2.4	17152
TotalCharges	0	0	1	7054
Churn	0	0	1	7043

End

Data Preprocessing for Exploratory Data Analysis

```
 #   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
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

Step 1 – Identify missing values

```
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
```

Step 2 –Impute missing values

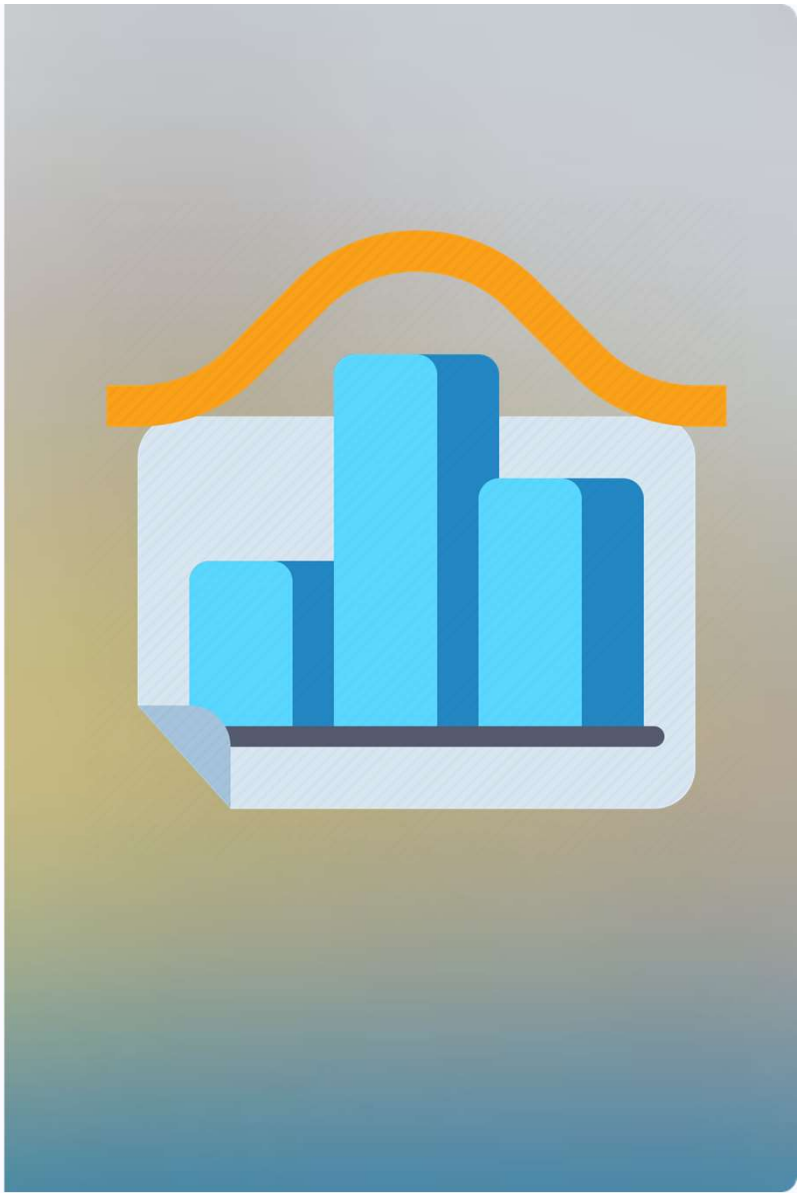
```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV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

Step 3- Remove 'Customer' id column

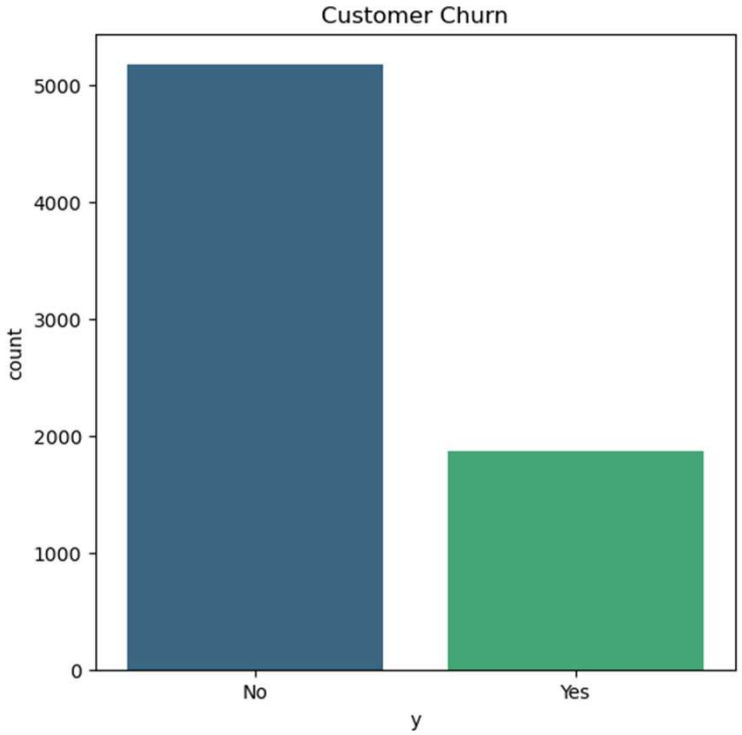
```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7043 non-null   object
 1   SeniorCitizen     7043 non-null   int64
 2   Partner           7043 non-null   object
 3   Dependents        7043 non-null   object
 4   tenure            7043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

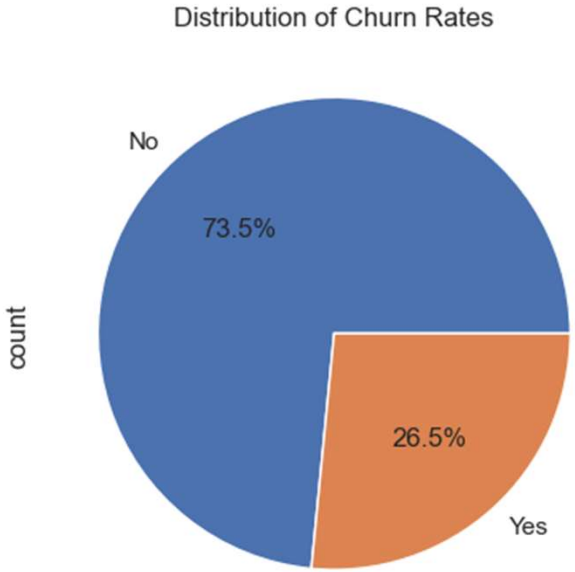
- Converted format of 'Total charges' column into int column since the values are numeric
- 11 missing values were identified in 'Total Charges' column which are replaced by median of that column
- No missing values were identified in other features or target column 'Churn'
- "Customer Id' column is dropped from the dataset since it has no relevance

Part 2- Exploratory Data Analysis



Understanding Target Variable

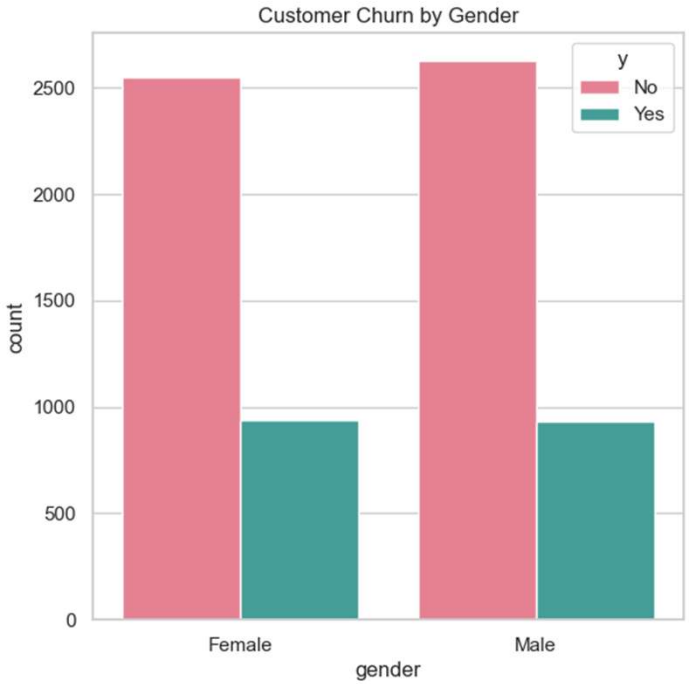


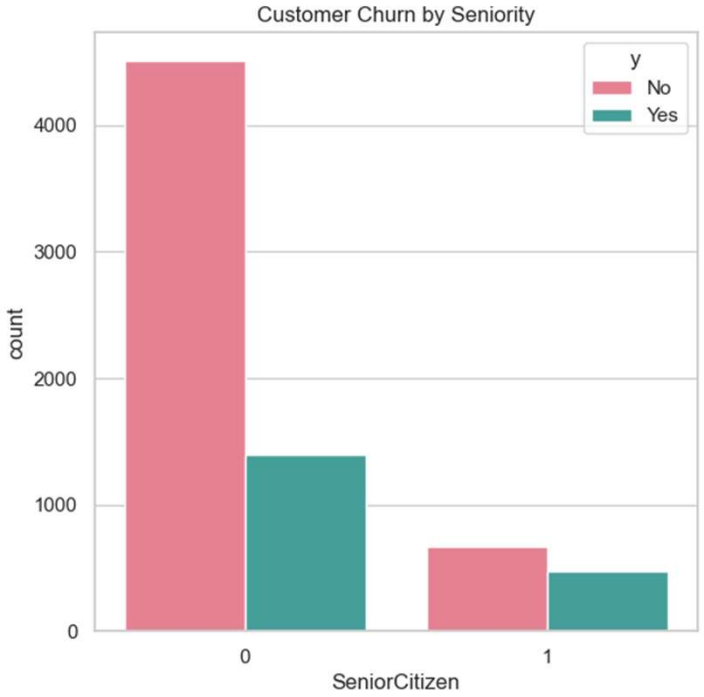


Target label Customer churn(y)

- 73.5% of customers have remained loyal to the telecom company, while 26.5% have switched to a different service provider
- This means that for every 100 customers, approximately 27 are leaving, which can have a substantial impact on revenue and customer acquisition costs.

Bivariate Analysis – Gender and Seniority vs Customer Churn

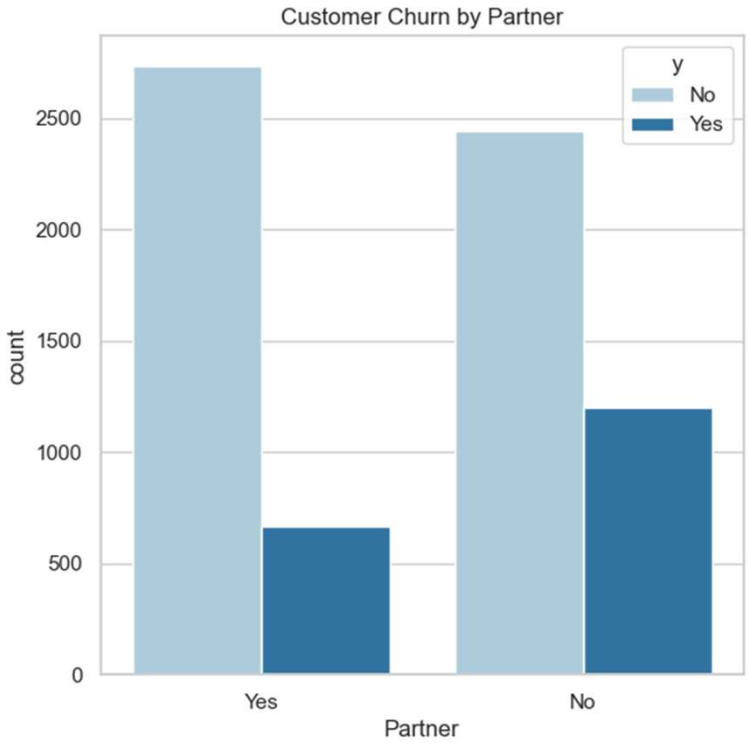


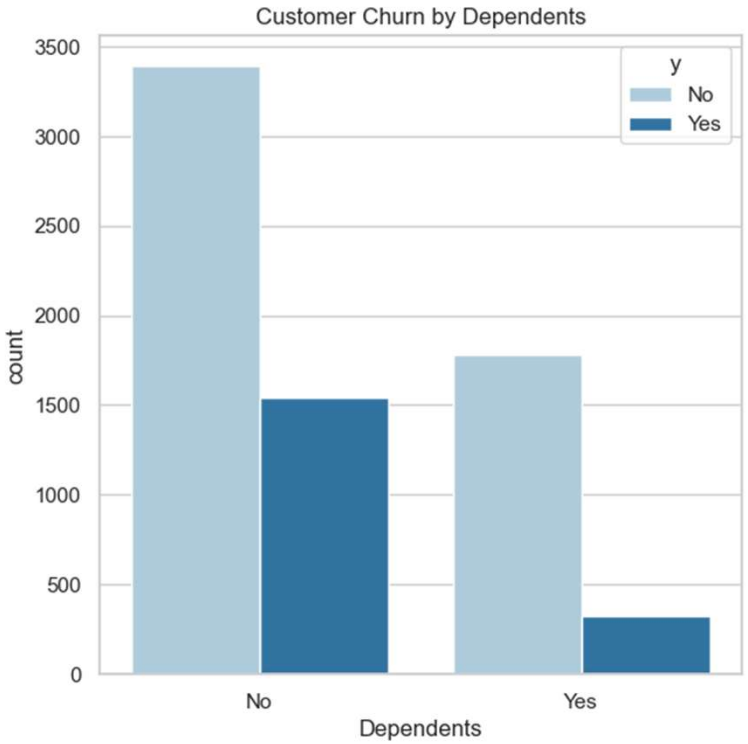


Observations:

- The churn rate is nearly identical for both male and female customers (Fig 1), indicating that gender does not influence customer churn
- Fig 2 shows that senior citizens exhibit a significantly higher churn rate of approximately 40%, compared to younger customers whose churn rate is around 20%.

Bivariate Analysis – Partner and Dependents vs Customer Churn

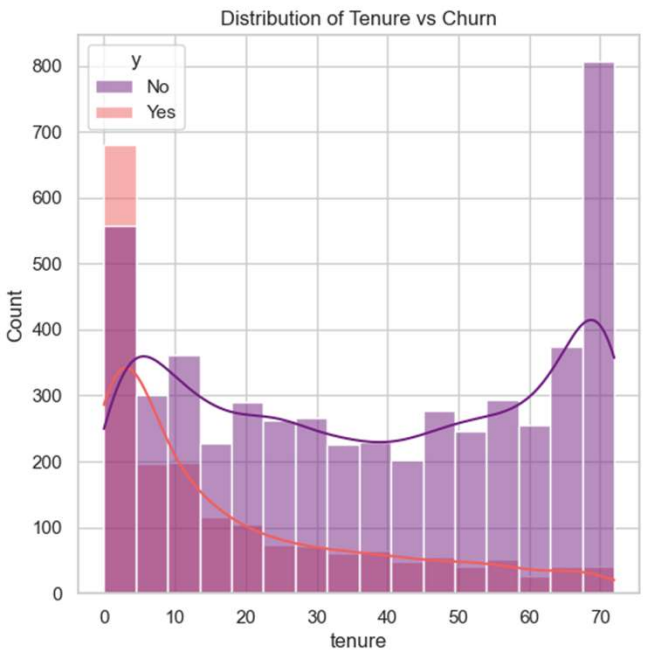


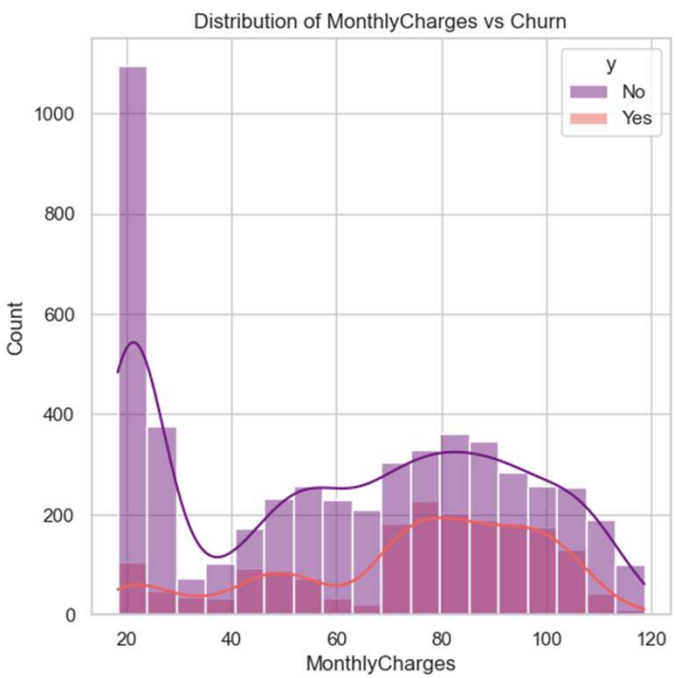


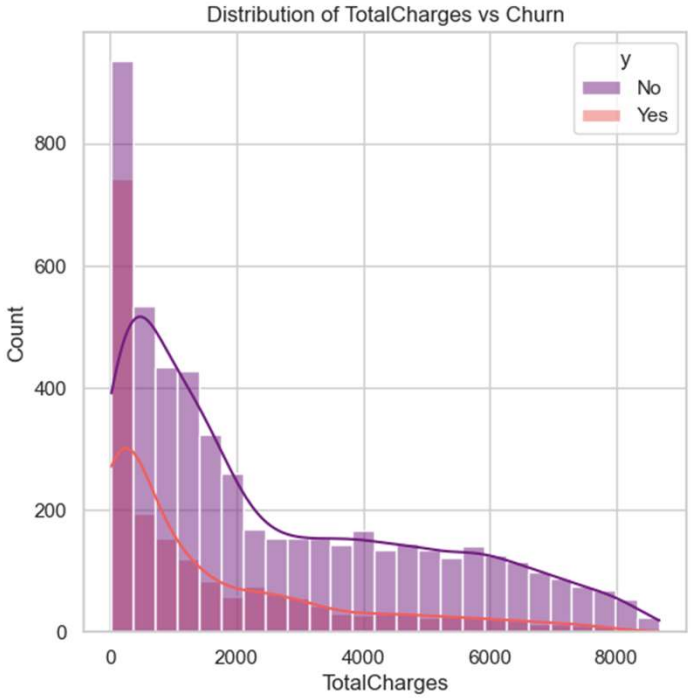
Observations:

- Figure 1 clearly shows that customers without a partner have a higher churn rate, reaching approximately 33% as compared to 20% for customers who have partner
- Figure 2 clearly indicates that customers without dependents have a higher churn rate, reaching around 31%

Distribution of Numerical features vs Churn



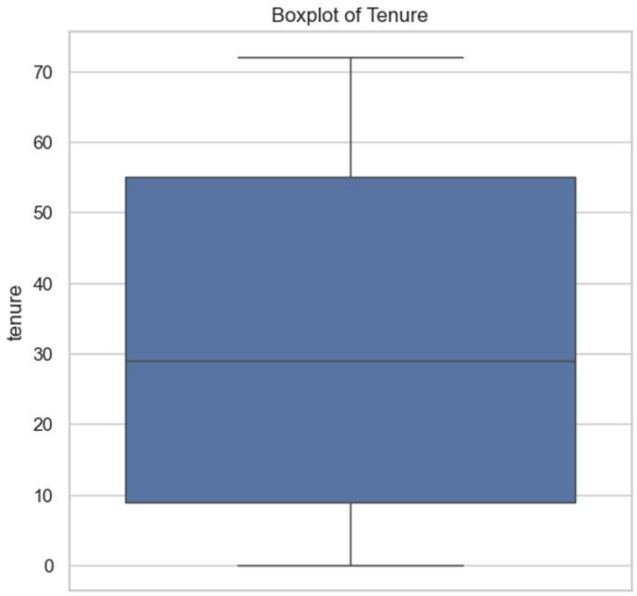


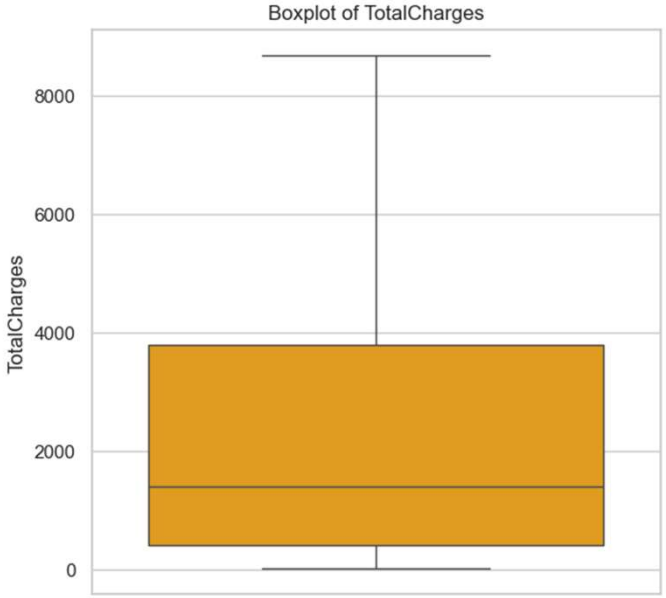


Observations:

- Figure 1 shows that customer churn rate decreases as tenure increases, with the highest churn observed among customers with a tenure of 0–10 months, while those with the longest tenure have the lowest churn rate.
- Figure 2 shows that Customer Churn is also notably higher for customers paying higher monthly charges, particularly in the range of ₹70 to ₹100.
- Figure 3 shows the that the distribution of total charges is right-skewed, and churn tends to decline as total charges increase.
- Exploring numerical features further could provide additional valuable insights.

Box Plot to check Outliers in numerical features

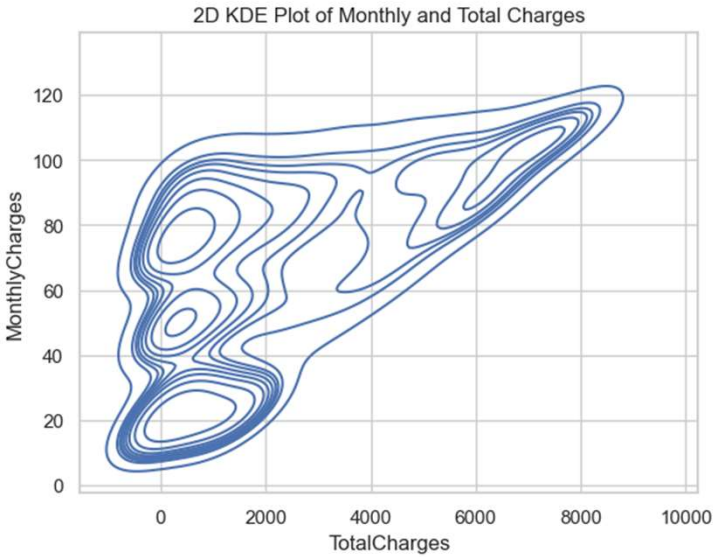


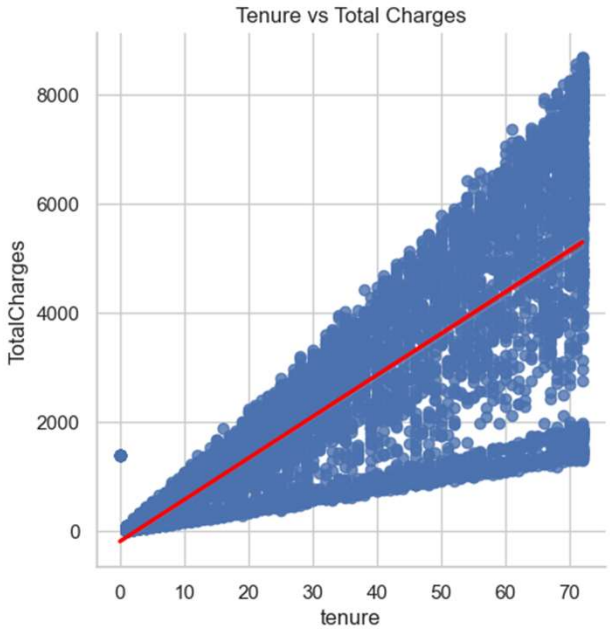


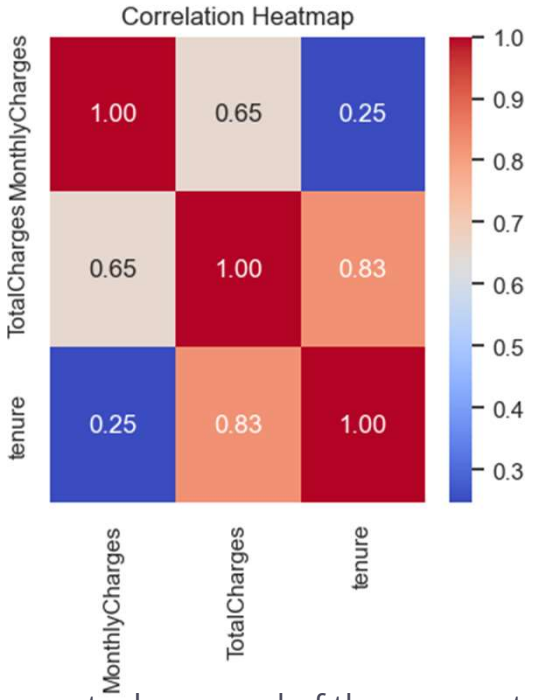
Observations:

- Figure 1 shows, the median tenure is approximately 30 days, and the tenure distribution appears symmetrical.
- In Figure 2, The median monthly charges are around ₹65, aligning with the observation that churn increases as monthly charges rise
- Figure 3 shows boxplot for total charges indicates a skewed distribution with high variance, and the median total charges are approximately ₹1500.

Studying relationship between numerical features



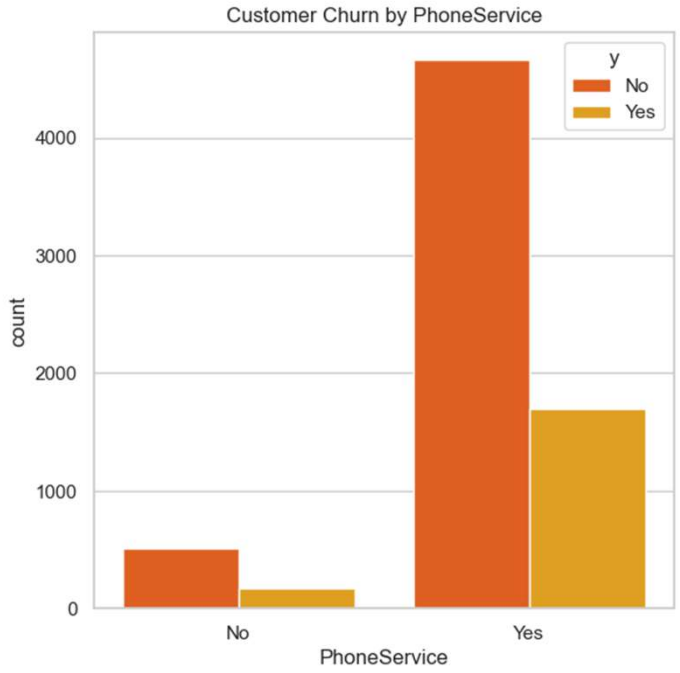


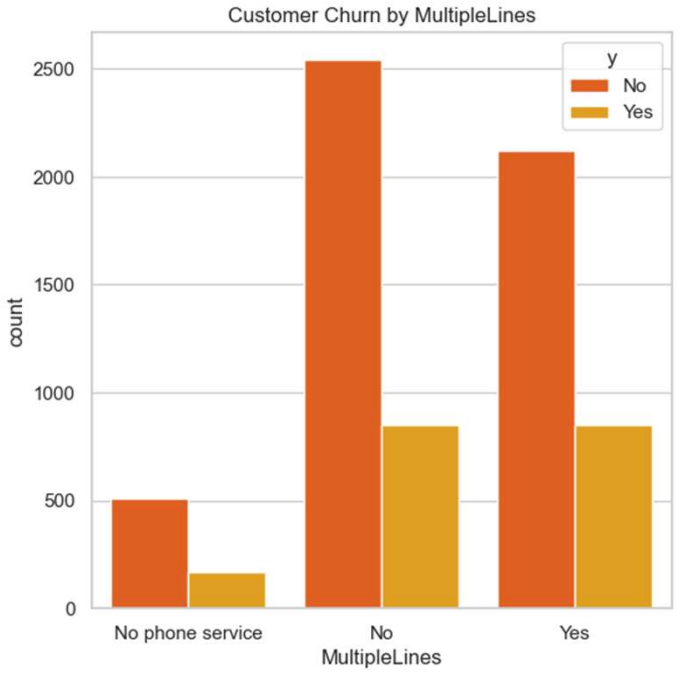


Observations:

- It is evident from above KDE plot in Figure 1 that there are high density of customers who are at a lower end of the segment between 0-2000 total charges and 20-40 monthly charges
- In Figure 1 there's a positive trend, with higher Total Charges associated with higher Monthly Charges
- Figure 2 shows positive correlation between Tenure and Total Charges
- Figure 3 shows correlation summary of 3 numerical columns , since Tenure and Total Charges are highly correlated, we can drop one of the feature before model building preferably Total Charges column since it has higher variance

Customer Churn basis Phone Service

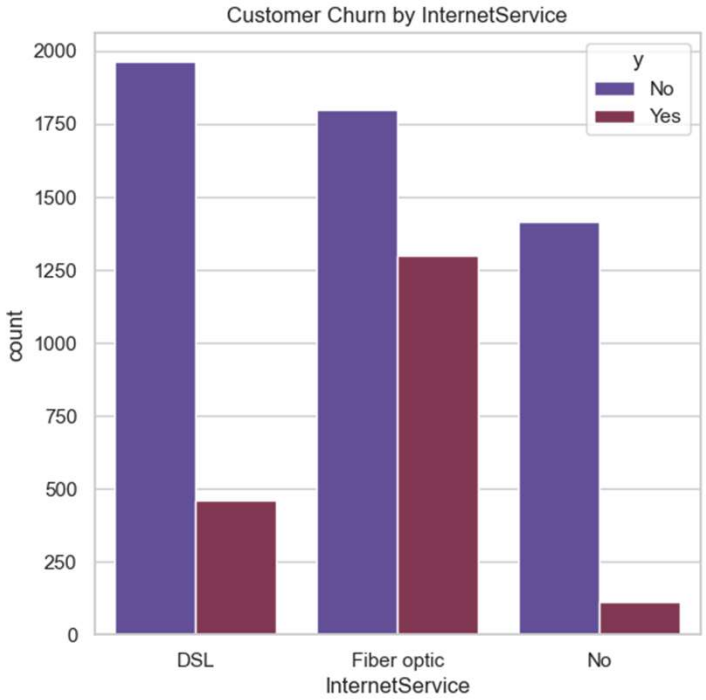


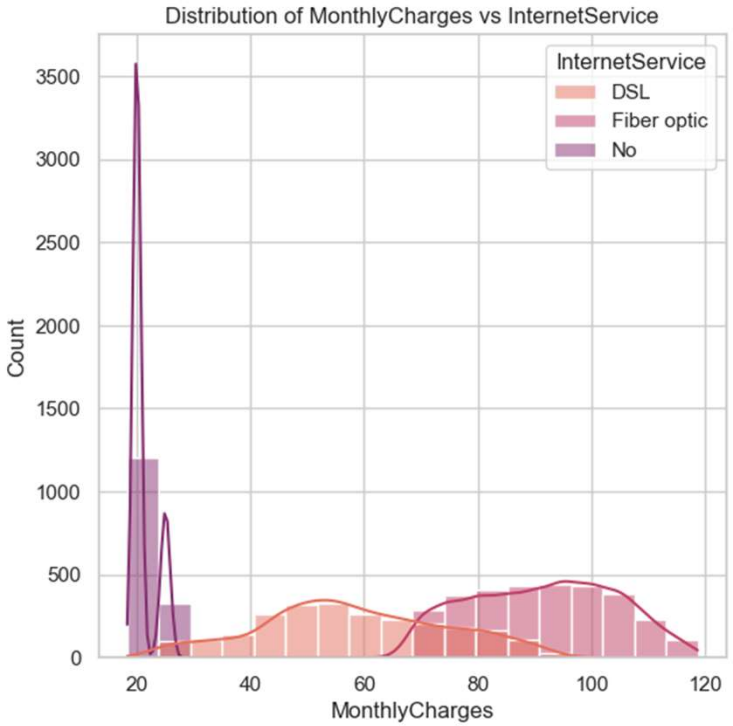


Observations:

- In figure 1, proportion of customer is nearly same for customers who have phone service or who do not have phone service
- Figure 2 shows that, nearly ~29% customers with Multiple lines tend to leave, 25% customers with no phone service leave

Customer Churn basis Internet Service

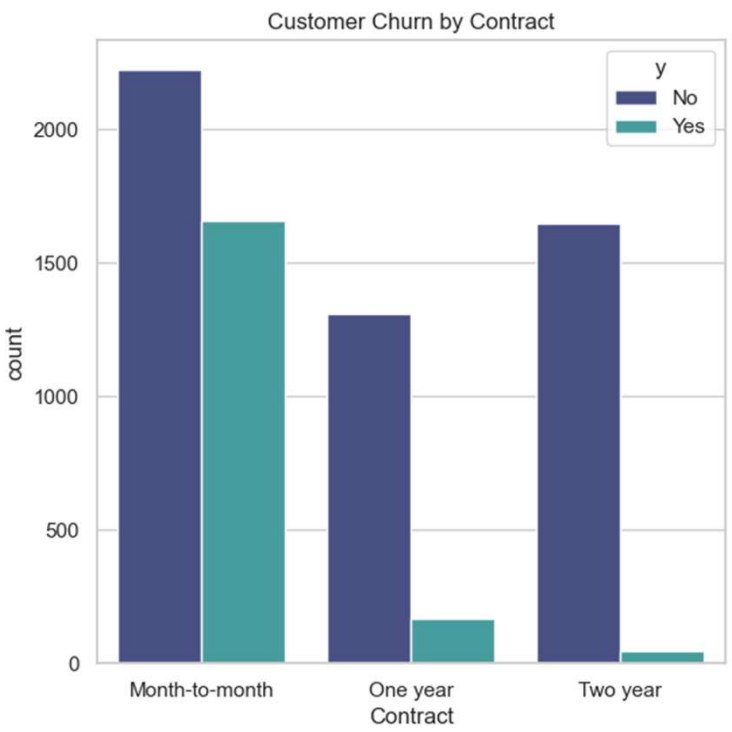


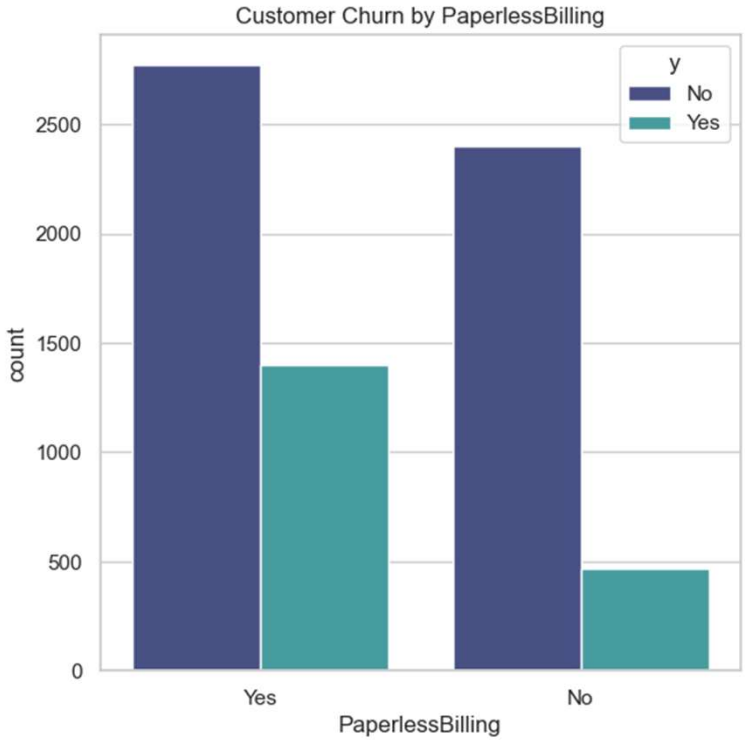


Observations:

- As per figure 1, nearly 40% customers leave who have opted for Fiber Optic internet service as compared to 19% in case of DSL
- From figure, It also implies, that average Monthly charges for Fiber Optic is much higher than customer who have opted for DSL
- Churn rates for customers with paperless billing is nearly 2 times as compared to ones who have not opted for Paperless billing

Customer Churn basis type of contract and Paperless Billing



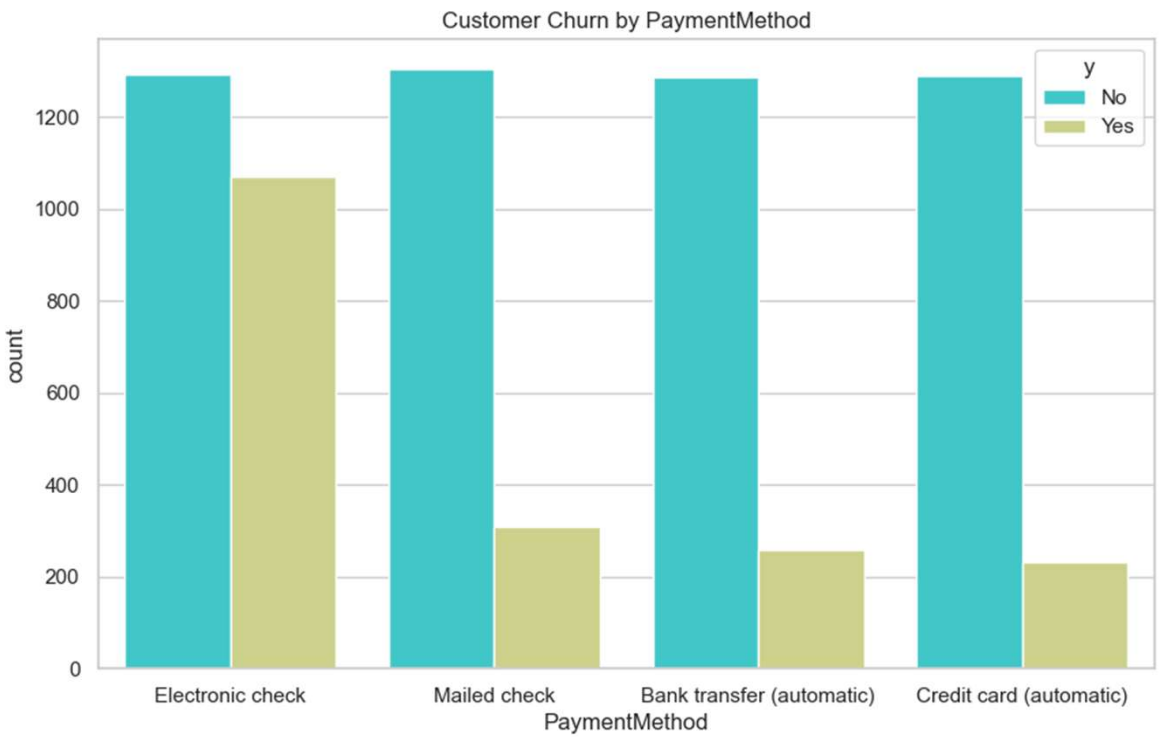


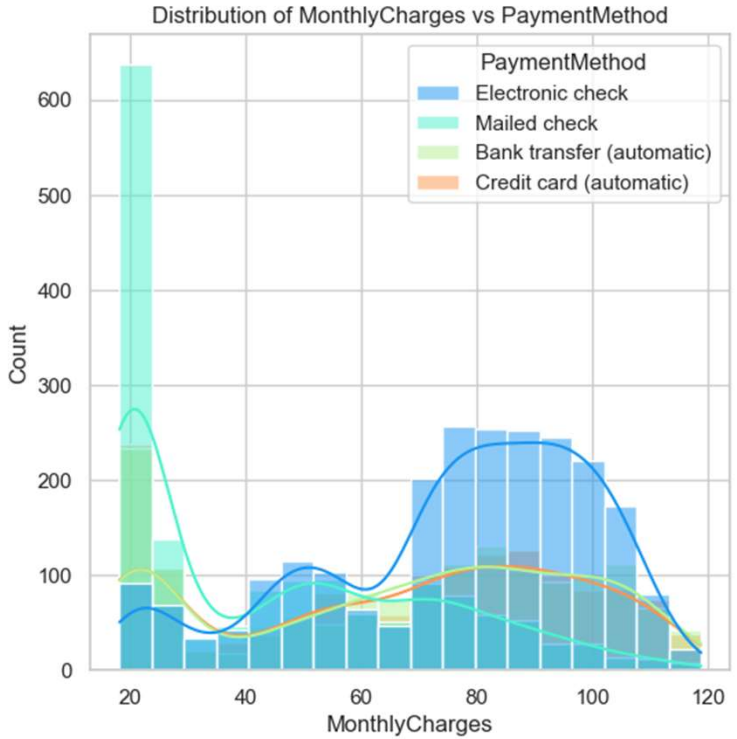
Observations

- Figure 1 shows that 42% customers who have month to month contract are likely to leave, Churn rates are very less for customers who have either one- or 2-year contract; this indicates that company should put efforts to sign up customers for long term contracts like one or two years

- Churn rates for customers with paperless billing is nearly 2 times as compared to ones who have not opted for Paperless billing which means that customer prefer to have paper bills rather than being on paperless billing

Customer Churn basis type payment method



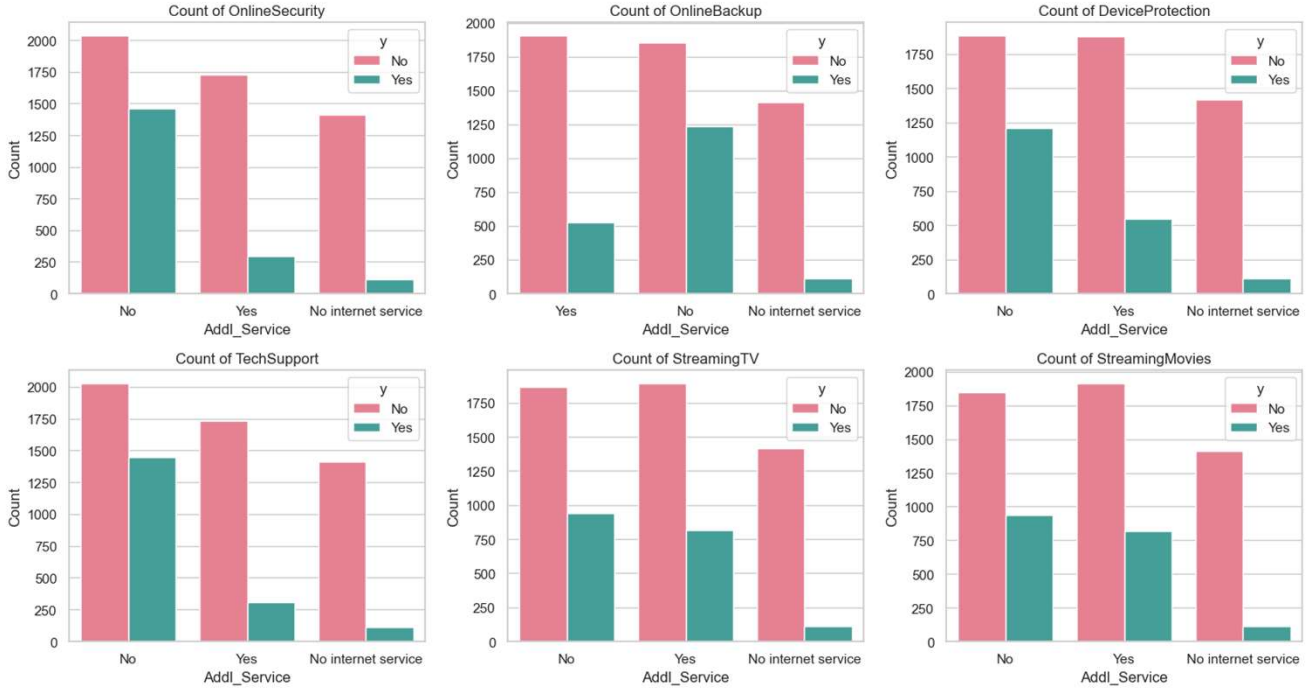


Observations:

- In Figure 1, nearly 45% of the customers who pay by electronic check are likely to leave as compared to other payment methods and the churn rate is lowest for customers who pay by Credit card and are set up on auto pay which is only ~15%

- In Figure 2, customers who pay by electronic check are also paying higher monthly charges, It is important for the company to encourage these customers to opt for other payment methods before they decide to leave
-

Customer Churn basis additional services



Summary of additional services categorical columns

- Figure 1 shows that nearly 40% customers who have no Online security feature is likely to churn as compared to 14% for customers who do have online security
- Figure 2 shows that nearly 40% customers with no online back up are likely to churn as compared to ones who have online back up
- Fig 3 shows that nearly 40% customers who have not opted for device protection are likely to churn as compared to the ones who do have device protection
- Fig 4 shows, nearly 40% customers who have not opted for Tech Support or have not reached out to Tech support are likely to have more churn rate
- In figure 5 & 6, Streaming services or movies do not seem to have impact on churn rates

Customer Churn Analysis Summary

1.Overall Churn Rate:

1. Approximately **27% of customers** leave the company, which means **1 out of 4 customers** switch to a different service provider.

2.Gender Analysis:

1. **Gender** does not impact the churn rate, indicating that churn behavior is consistent across genders.

3.Age Impact:

1. **Senior Citizens** show a high churn rate of **40%**, meaning nearly **2 out of 5 senior customers** are likely to leave.

4.Marital Status:

1. **Single customers** are more prone to churn, with **1 out of every 3** likely to leave the company.

5.Dependents:

1. **Customers with dependents** have a churn rate of **30%**, suggesting that **3 out of 10** such customers may switch services.

6.Customer Tenure:

1. **Retention improves** significantly after customers stay with the company for more than **10 days**, showing a marked reduction in churn rates for longer-tenured customers.
2. The **median tenure is approximately 30 days**, and the distribution of tenure is symmetrical.

7.Monthly Charges:

1. **Churn rate is notably high** for customers with **monthly charges between Rs. 70 to Rs. 100**. This indicates that price sensitivity could be a factor influencing customers' decisions to switch providers.
2. The **median monthly charges** are around Rs. 65.

8.Total Charges:

1. The **distribution of total charges** is skewed to the right, showing that most customers have accumulated lower total charges.
2. **Churn rate decreases** as **total charges increase**, suggesting that customers who spend more over time are more likely to remain loyal.

9. Internet Services:

1. Nearly **40% of customers with Fiber Optic** internet service churn, compared to 19% for DSL users
2. Fiber Optic customers **have higher average monthly charges** than DSL customers

10.Additional Services:

1. **Churn likelihood is close to 40% among customers without online security, online backup, device protection, or tech support**, whereas customers with these services show significantly lower churn rates
2. **Streaming and movie services do not appear** to affect churn rates.

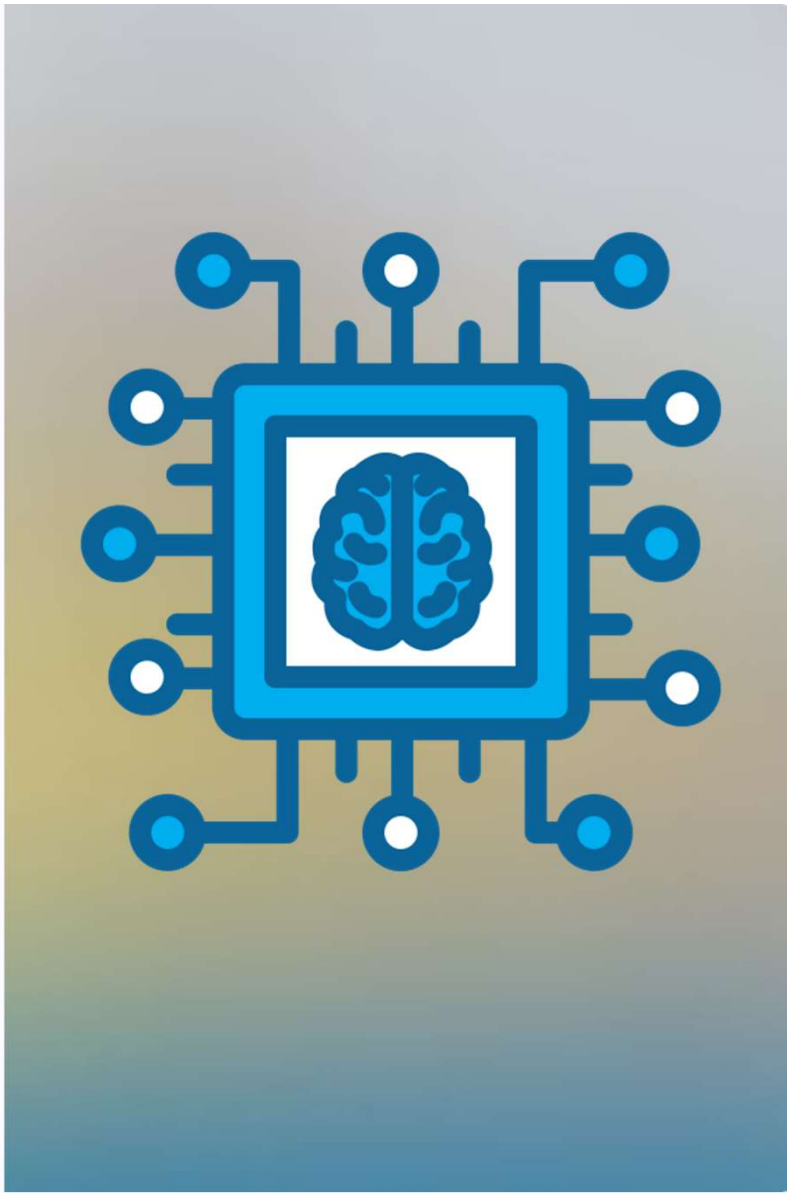
Customer Churn Analysis Summary

11. **Customer Contracts**:
    1. **Nearly 50% of customers are on month-to-month contracts**
    2. **Churn is high (42%) for month-to-month contracts,** while one- and two-year contracts have significantly lower churn rates.
    3. **Customers with two-year contracts have a very low churn rate of only 3%.**

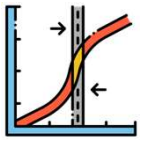
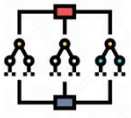
12. **Paperless Billing:**
    1. **60% of customers have chosen paperless billing**
    2. Nearly **50% of paperless billing customers churn**, compared to just 20% of those with paper billing.

13. **Payment Method:**
    1. Nearly **equal number of customers pay by different** payment methods
    2. **Almost 45% of customers who pay by electronic check are likely to churn**, a much higher rate than other payment methods

Part 3- Building ML model



Accuracy Scores basis different ML Models using Cross Validation

	Accuracy Score	Precision	Recall	F1 Score
 Logistic Regression	76.89%	74.00%	80.00%	77.00%
 Decision Tree	79.04%	77.00%	82.00%	79.00%
 Random Forest	84.61%	83.00%	87.00%	85.00%
XGBoost	84.15%	81.00%	88.00%	84.00%

Part 4- Recommendations

Recommendations to improve Customer Churn

- **Focus on Senior Citizen Retention:** Offer discounts, senior-specific support, or loyalty programs to retain senior customers**.**
- **Engage Single Customers:** Provide exclusive bundles, perks, or loyalty programs for single customers to reduce churn**.**
- **Encourage Long-Term Contracts for New Customers:** Offer limited-time discounts or loyalty rewards to encourage new customers to commit beyond the initial period.
- **Address Price Sensitivity in Monthly Charges:** Revisit pricing for customers paying Rs. 70–100 per month, offering slight rate reductions or added perks.
- **Optimize Fiber Optic Service Offerings:** Enhance Fiber Optic value with bundled services, premium support, or adjusted pricing to reduce churn.
- **Promote Essential Add-On Services:** Encourage customers to opt for online security, backup, device protection, and tech support through discounts or free trials.
- **Encourage Long-Term Contracts:** Offer incentives for month-to-month customers to switch to one- or two-year contracts to reduce churn.
- **Improve Paperless Billing Experience:** Enhance paperless billing with reminders and loyalty rewards to reduce churn among these customers.
- **Promote Alternative Payment Methods:** Encourage customers who pay by electronic check to switch to more stable payment methods, offering discounts or rewards.

Thank You