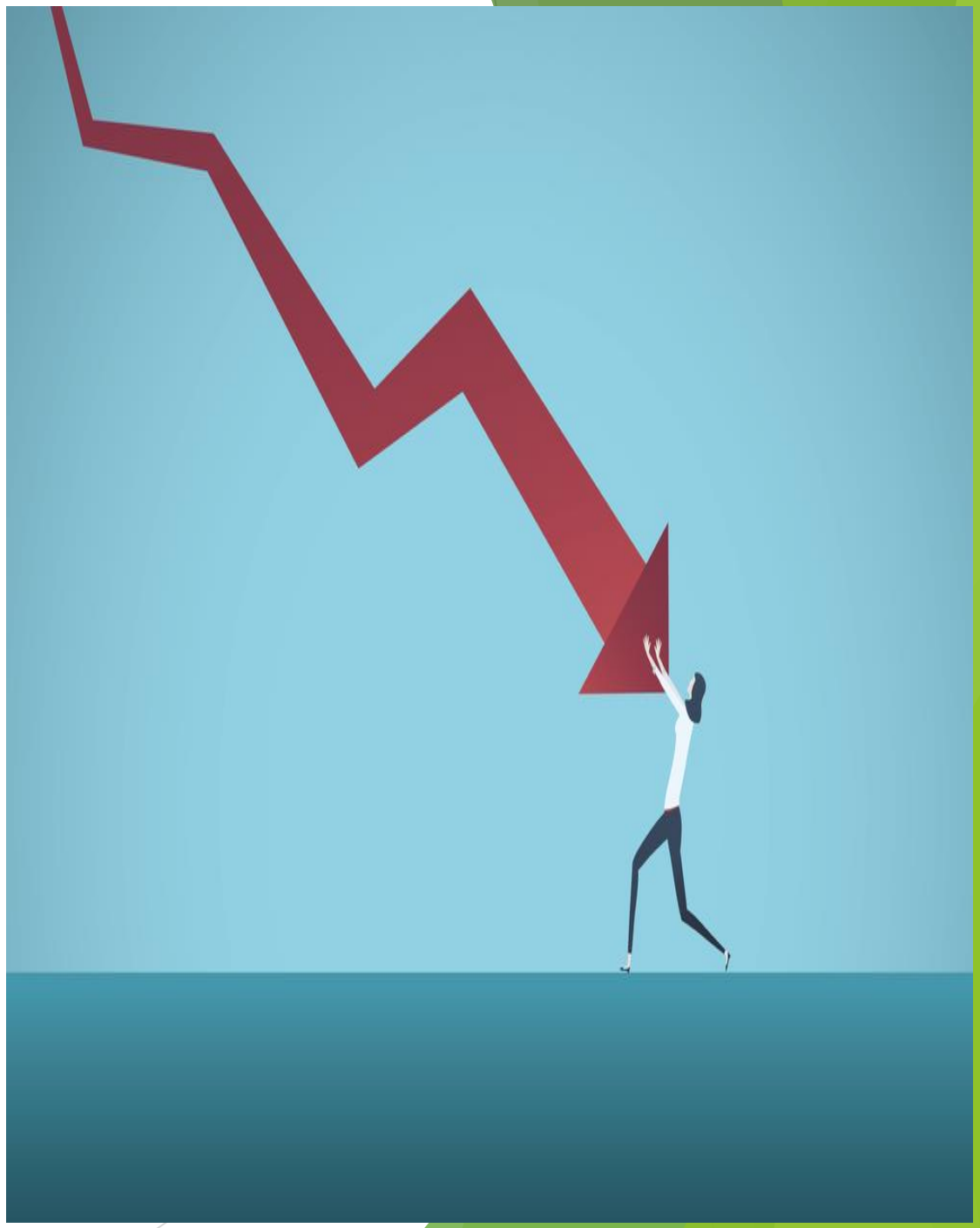
CONNECTTEL TELECOM CUSTOMER CHURN PREDICTION

10Alytics

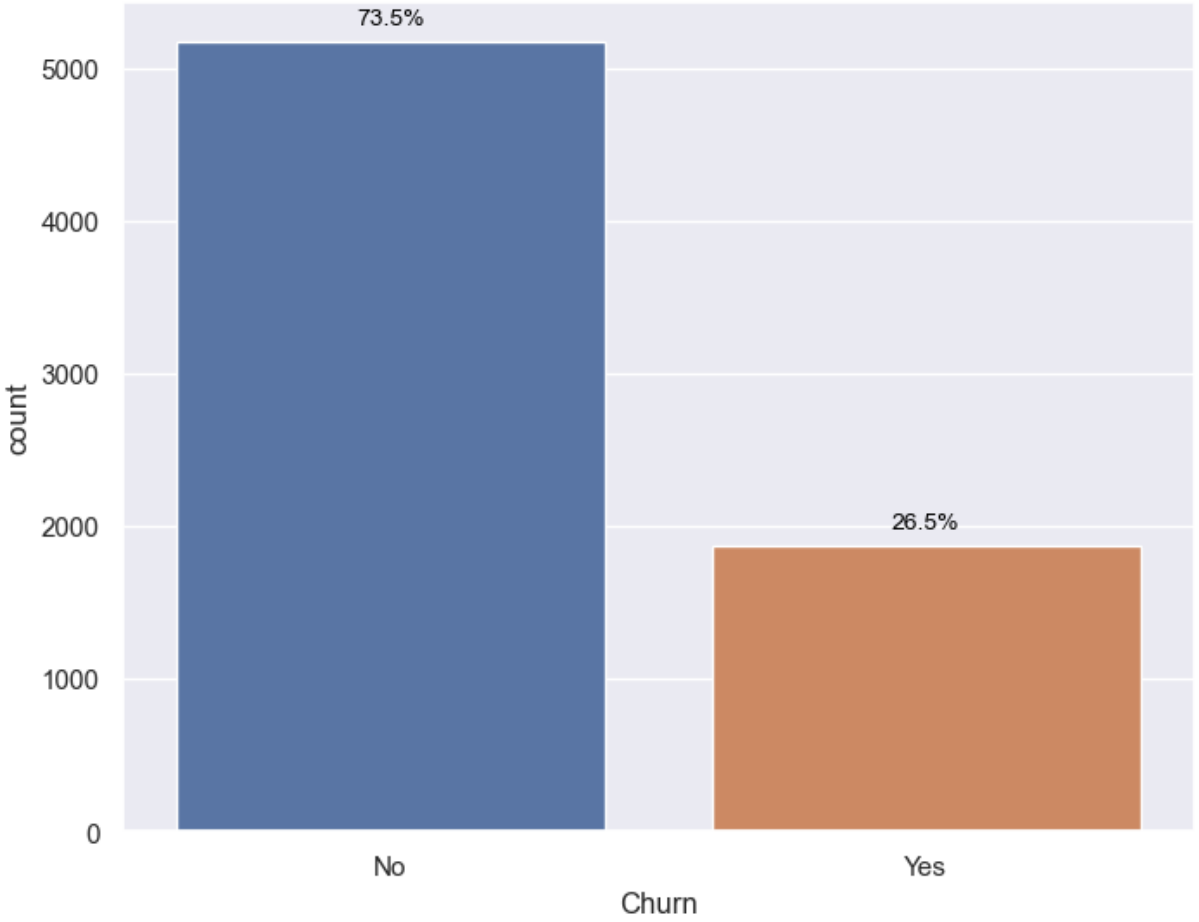
PROBLEM DEFINITON:

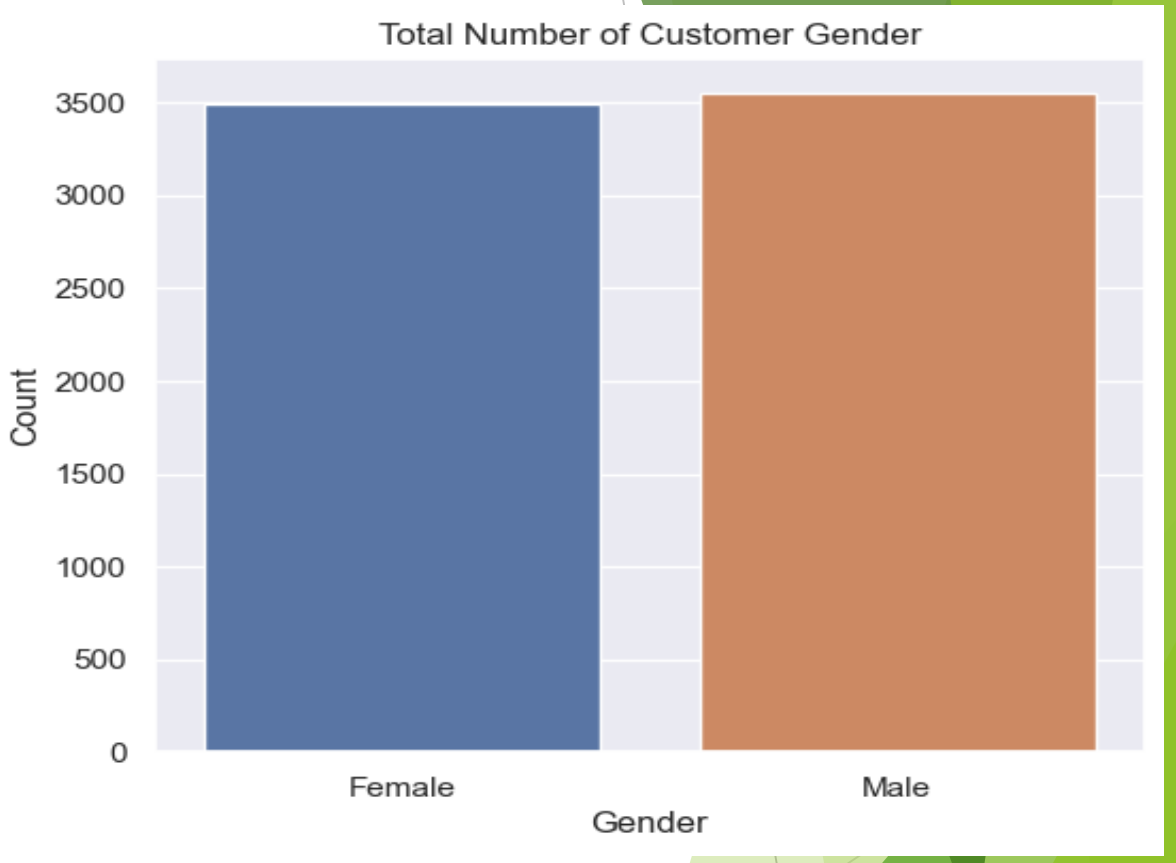
- The problem to be solved with data mining is to develop a predictive model that accurately predicts customer churn for a ConnectTel company.
- Customer churn, or the rate at which customers discontinue their service with the company, is a significant concern for ConnectTel telecommunication company as it directly impacts revenue and profitability.
- By analyzing historical customer data, including demographic information, usage patterns, and customer interactions, the goal is to identify factors that contribute to churn and develop a model that can predict which customers are at risk of churning in the future.



- ConnectTel Churn data includes information about the telecom company that provided home phone services and Internet services to **7,043 customers**. Which customers have left, stayed, or signed up for their service shows?

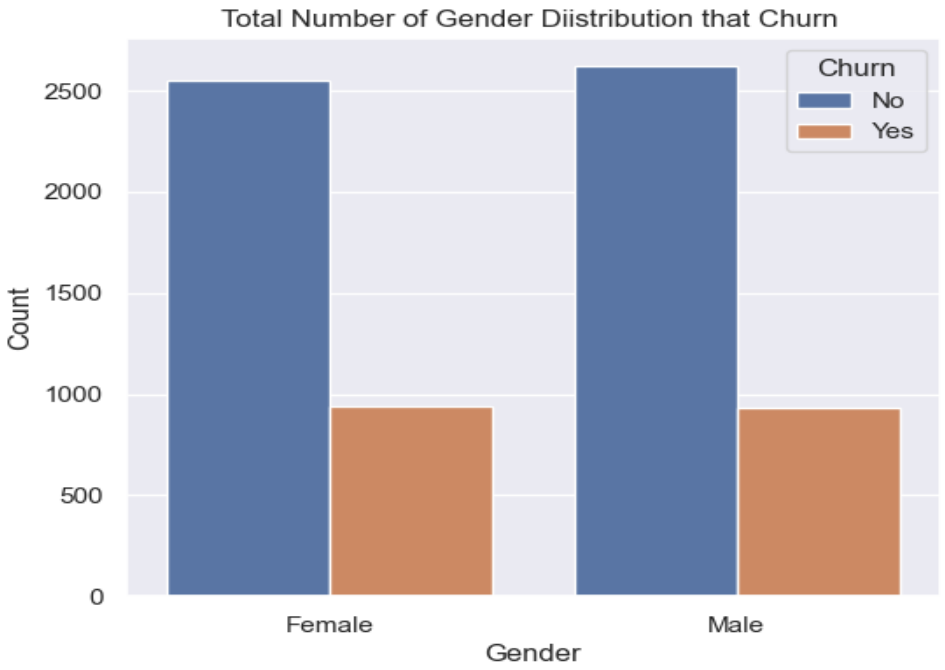
Insight derive from the dataset:





- **The churn rate of customers is 26.5% while non-churn rate is 73.5%**

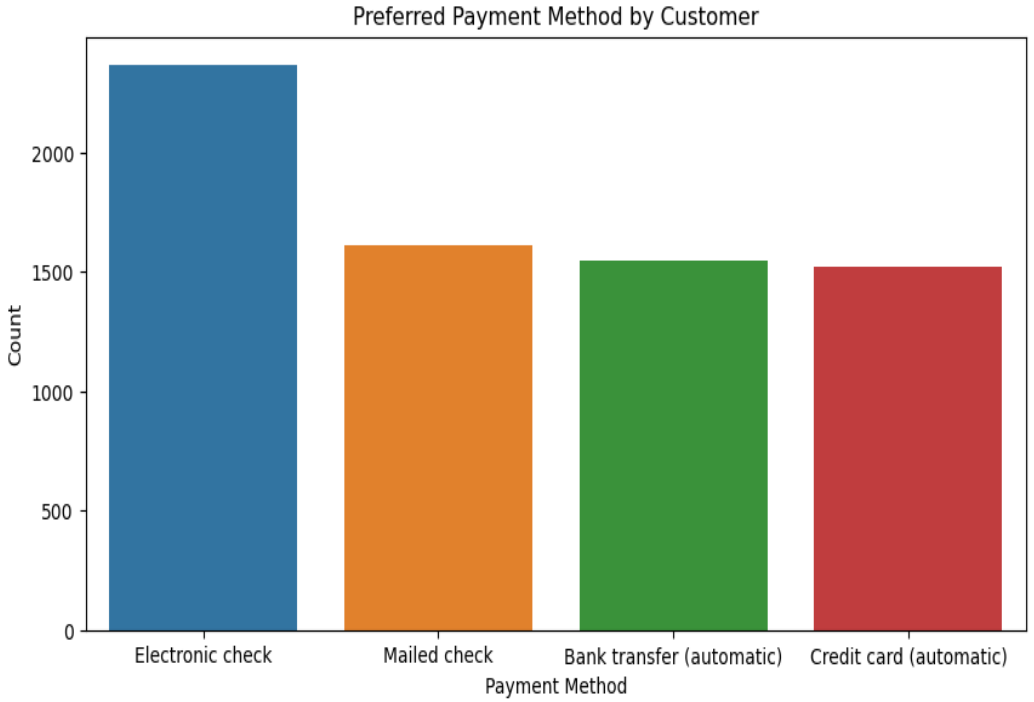
- **So in terms of customer there is no preference, we have balance subscriber level in the sense of male and female customer.**

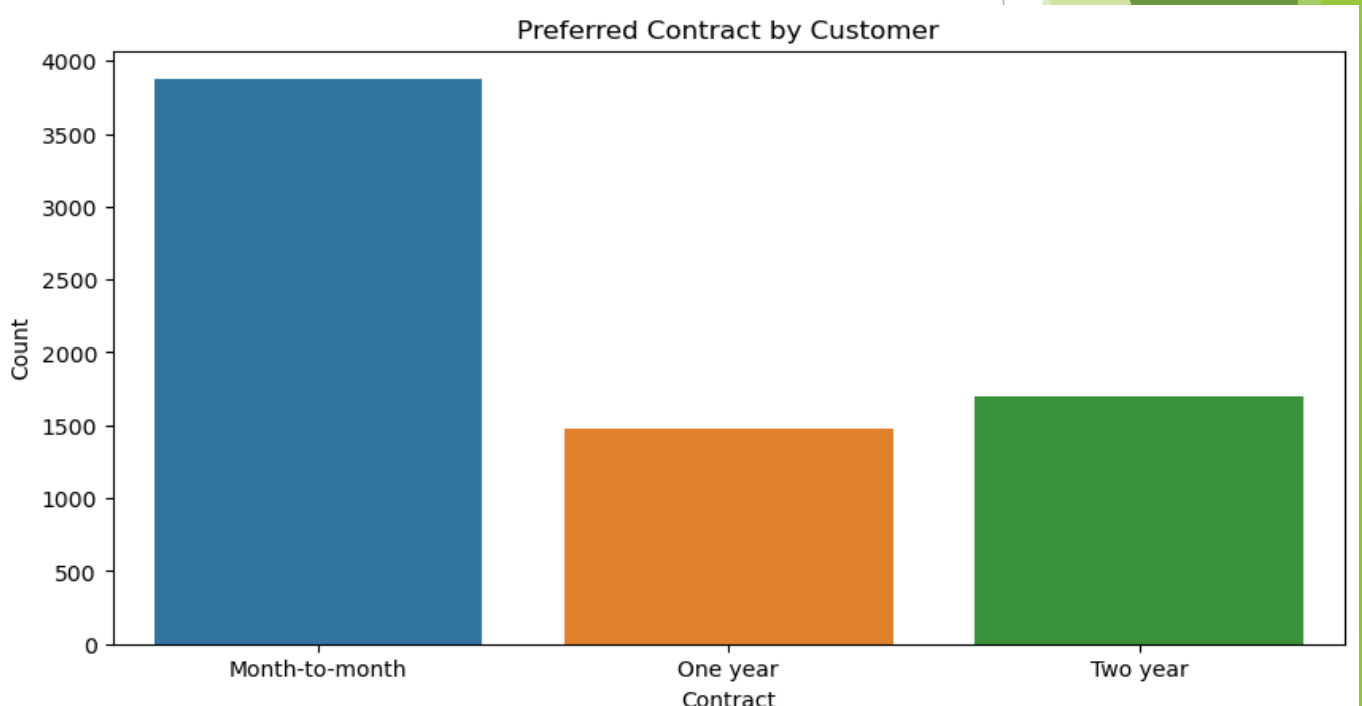


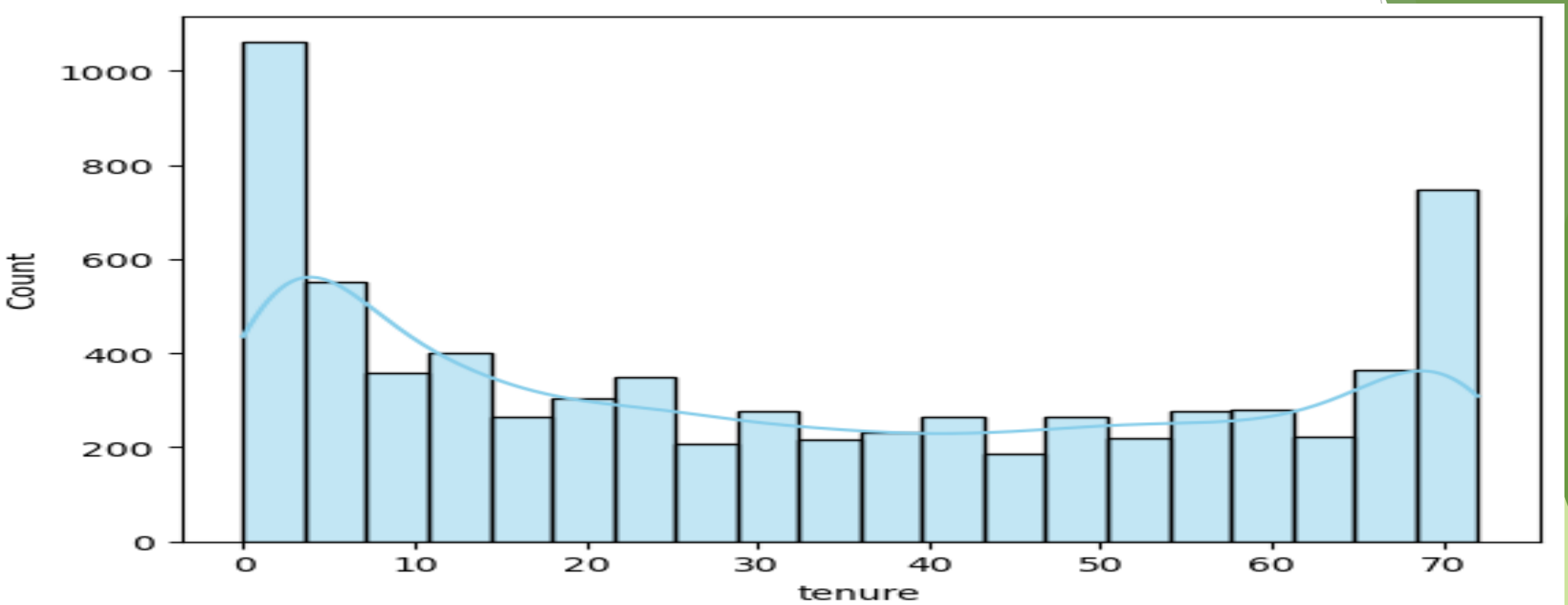
- **There is no significant difference in the number of departures of male and female customers.**

- **In terms of contract, customer prefer to subscribe to Month-to-Month which dominate the majority of while two-year and one year contract follow respectively**

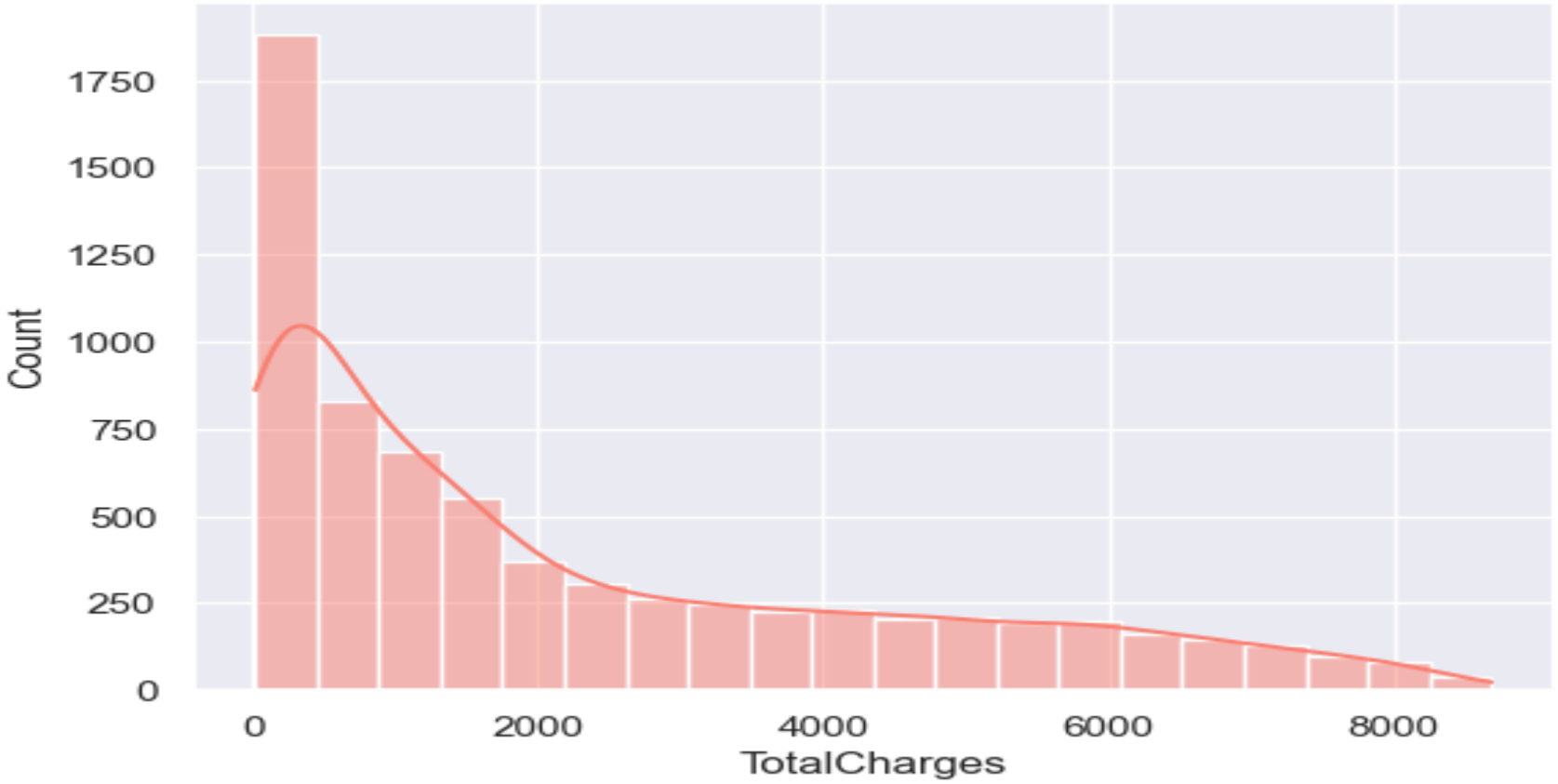
- **In terms of payment method preference selected by the customer, majority of the customer uses Electronic check as payment method; Mailed check, Bank transfer and Credit card which dominate the rest of the customers have balanced number of customers using them as their payment method**



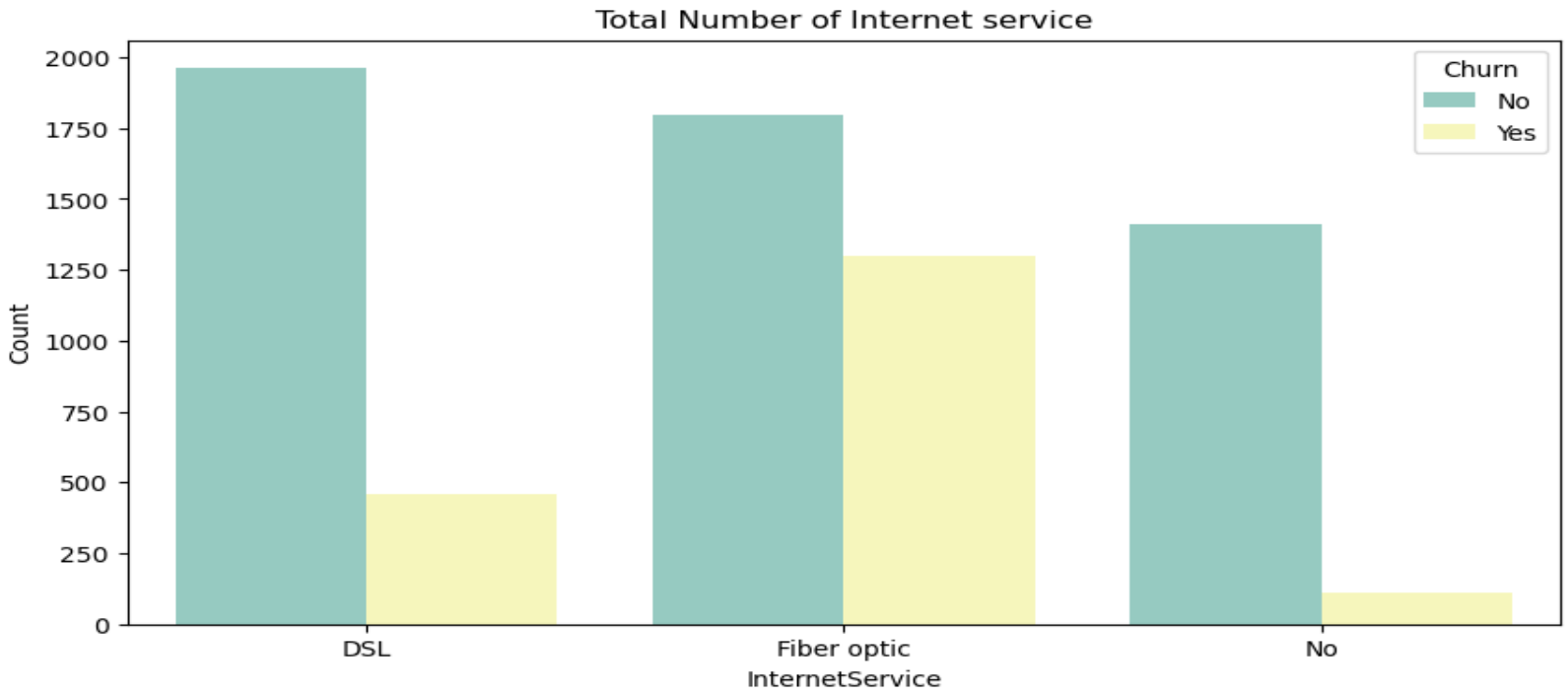




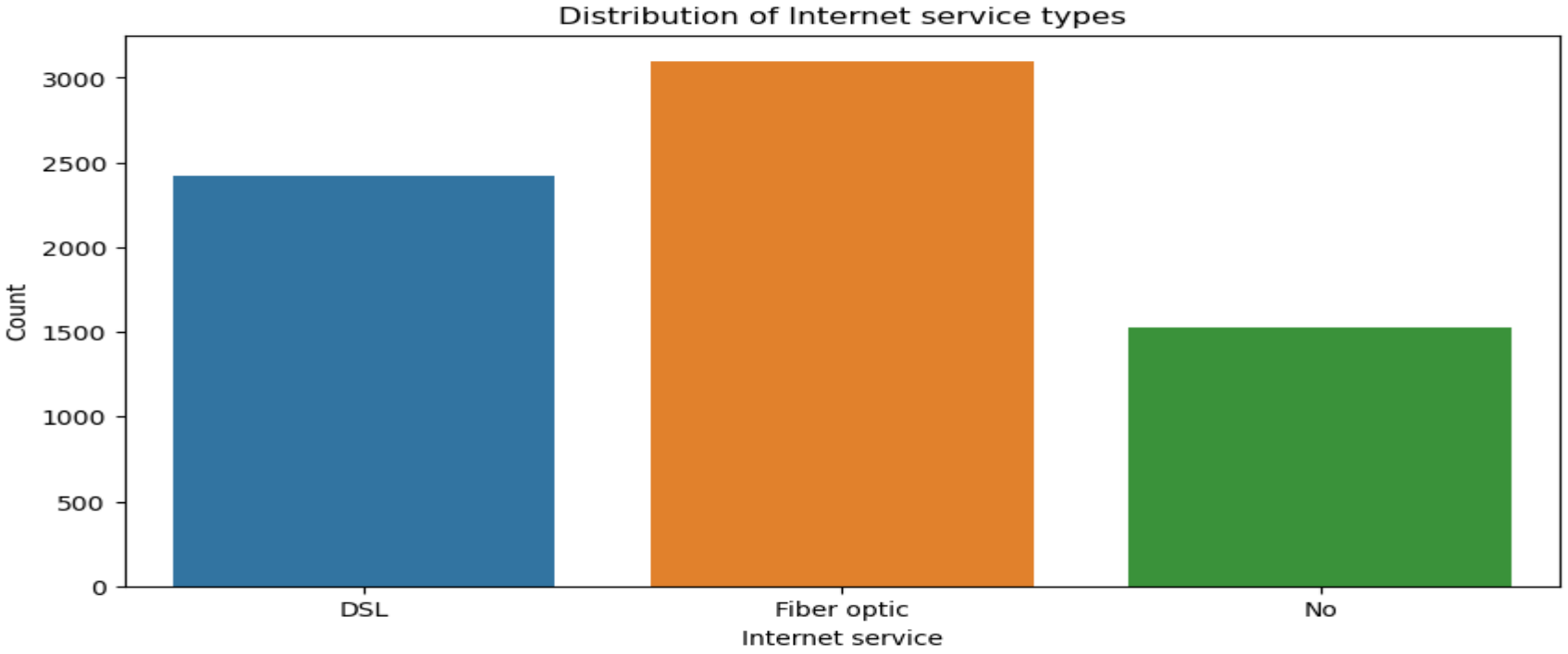
This tenure distributoion shows a bimodal pattern with a peak at the lower and the higher end of the distribution. This entails that there is outstanding numbers of new customer at the beginning point and old customers at the ending point of the distribution



The Total charges distribution is relatively spreadout, indicating a wide range of charges of customer incur. There's a noticeable downward trend which clearly state that there's decrease of total charges from the customer as the price increase

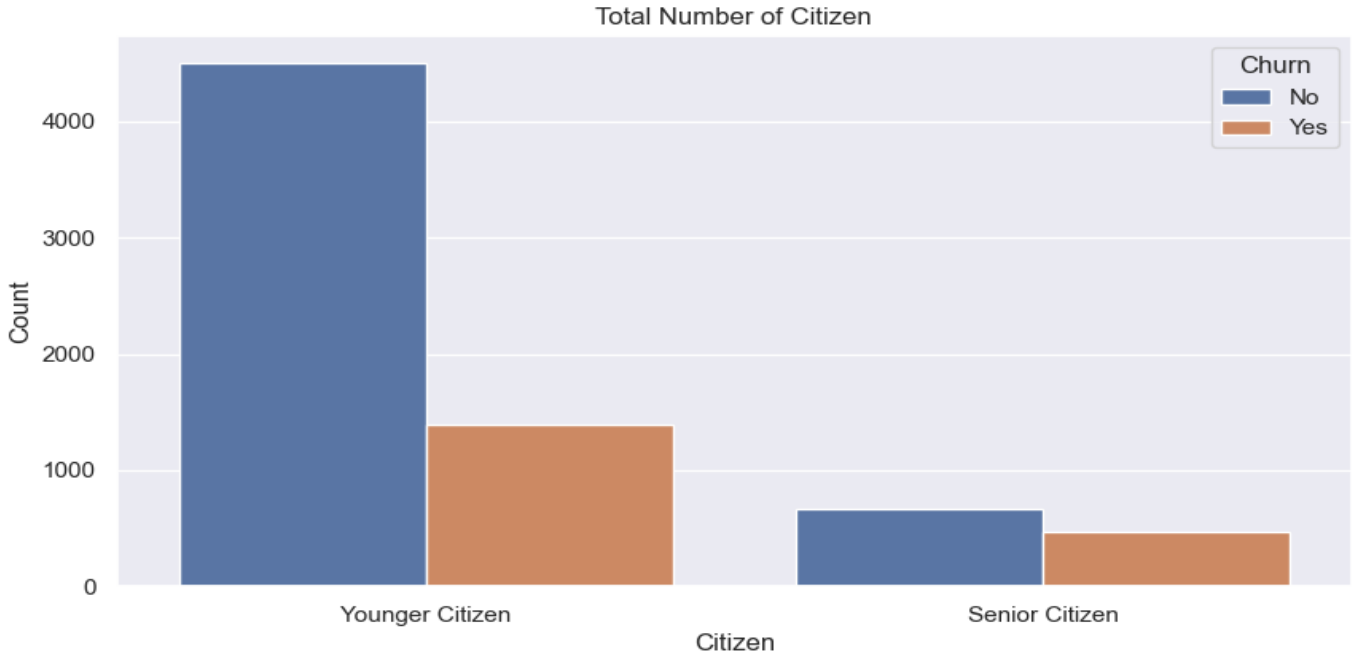


- **Churn rate is much higher in case of Fiber Optic Internet Services**

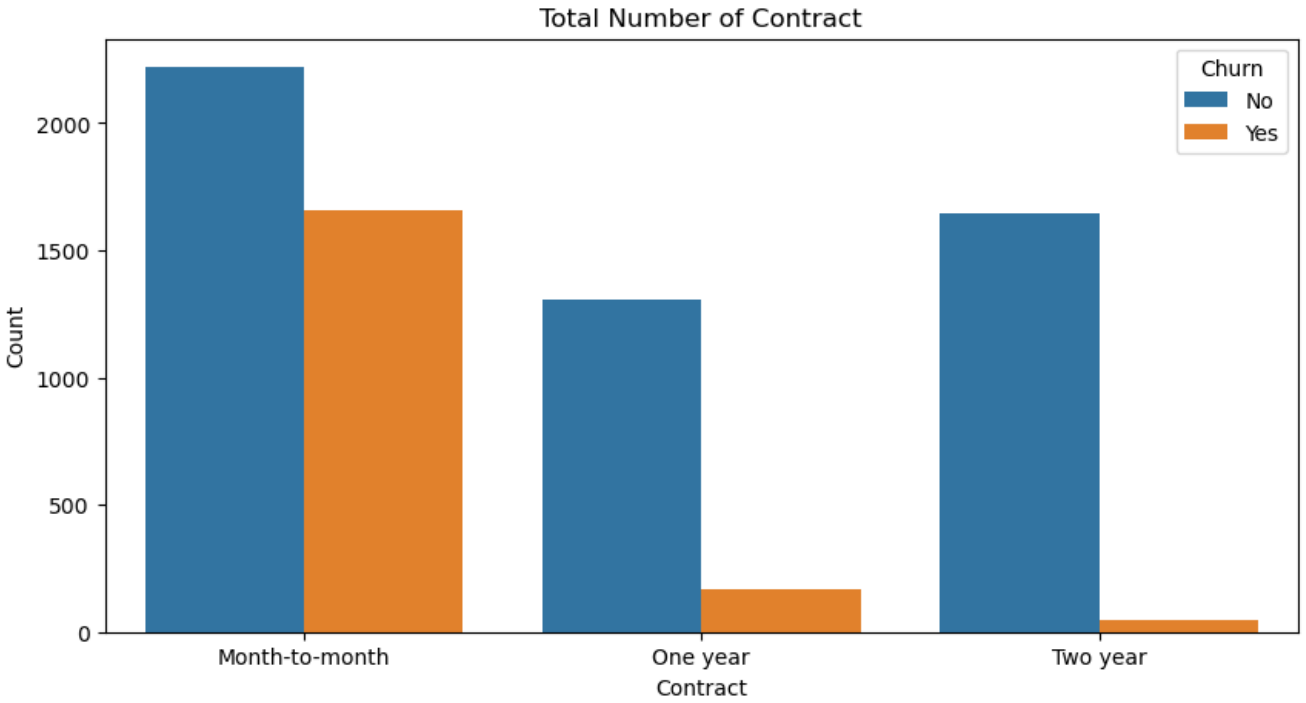


- **In terms of internet service, most subscribers prefer using Fiber optic as their service provide which has the highest number of customer, DSL also has a significant number of subscriber while the rest of the customers don't use internet service at all**

- **After checking the distribution of data in each tenure bin, we found that maximum number of customers have a tenure of either 0-1 years and followed by 5-6 years. Customer who are to churn have a tenure of 0-1 year and there's significant number of loyal customer (5-6 years) which there number of leaving the platform is the least compared to the rest.**

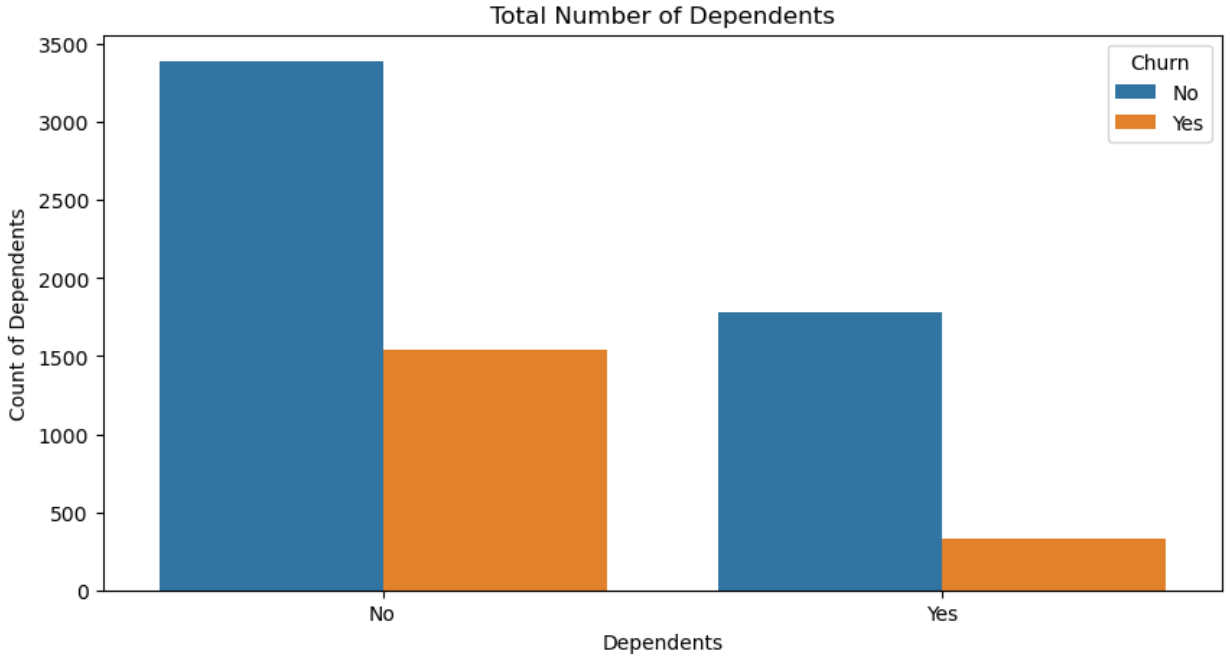


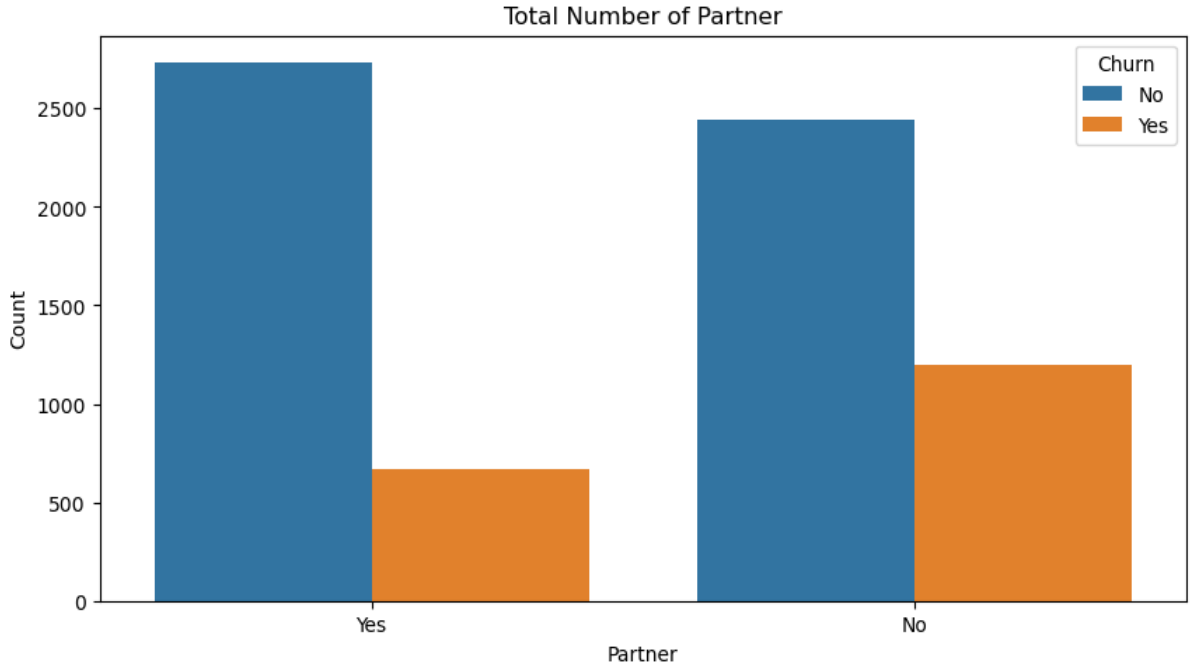
- **The churn rate is higher in terms of younger citizen**

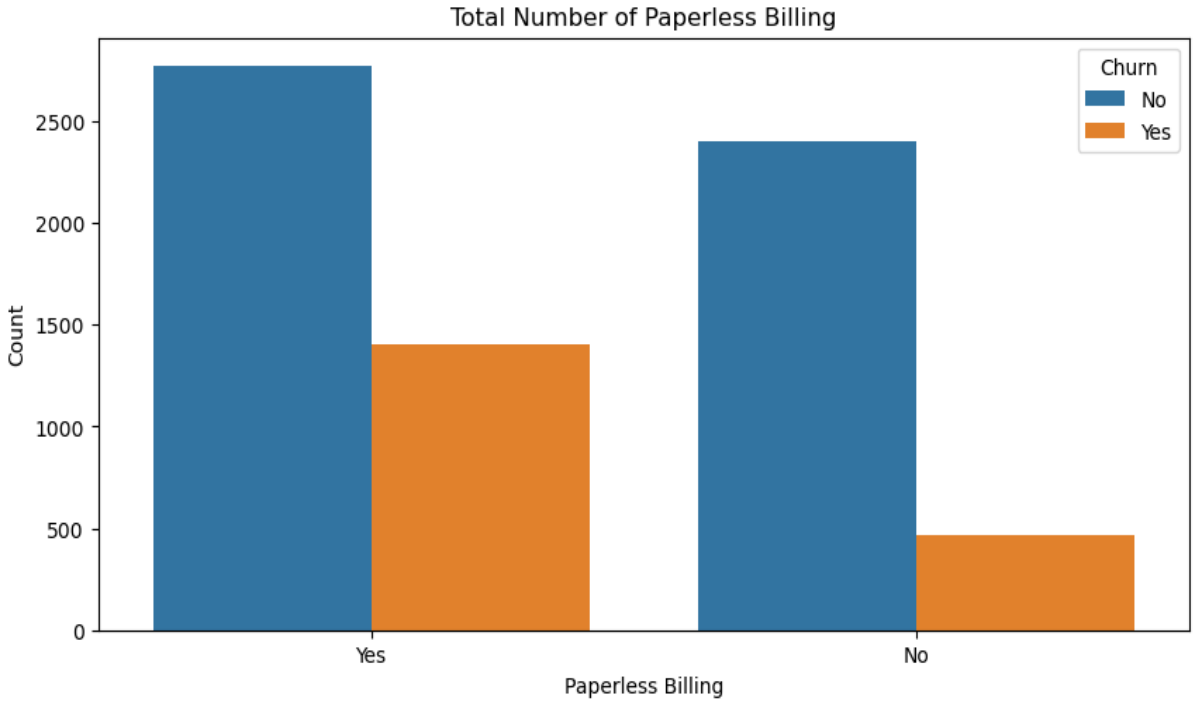


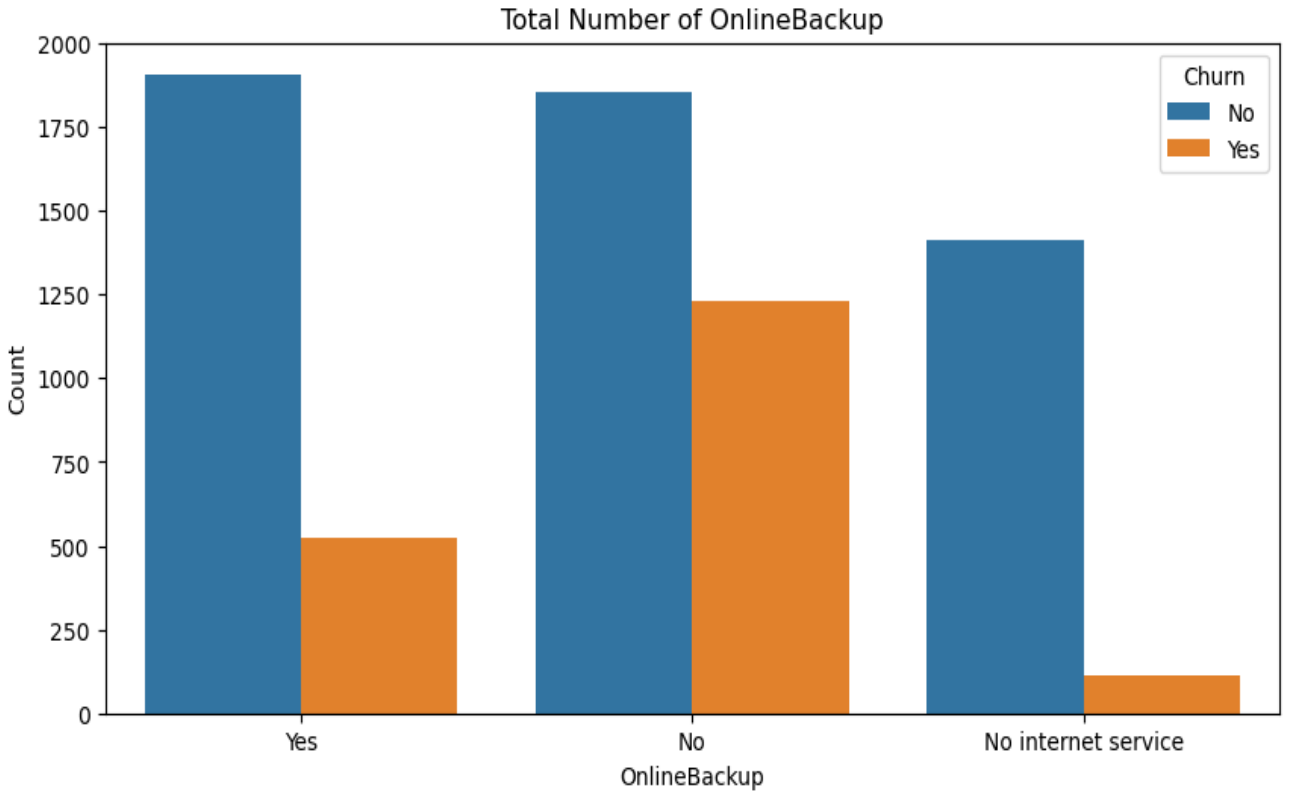
- **A larger percent of Customers with monthly subscription have left when compared to Customers with one or two year contract.**

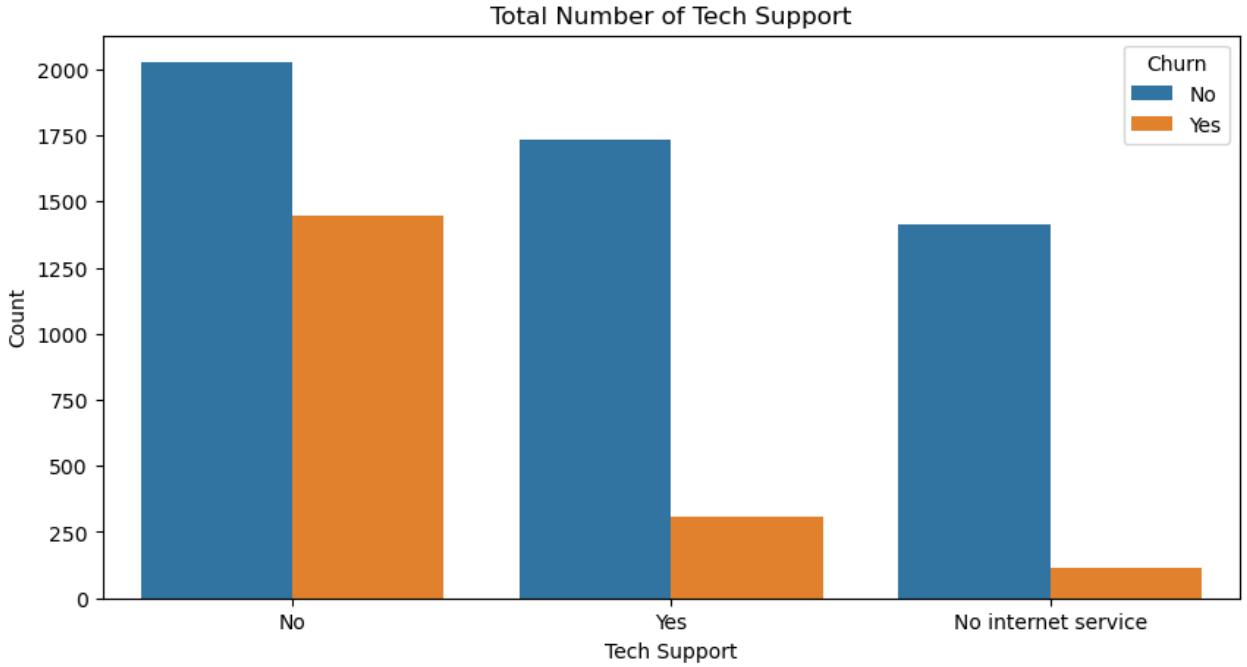
- **Customers with Partners and Dependents have lower churn rate as compared to those who don't have partners & Dependents.**







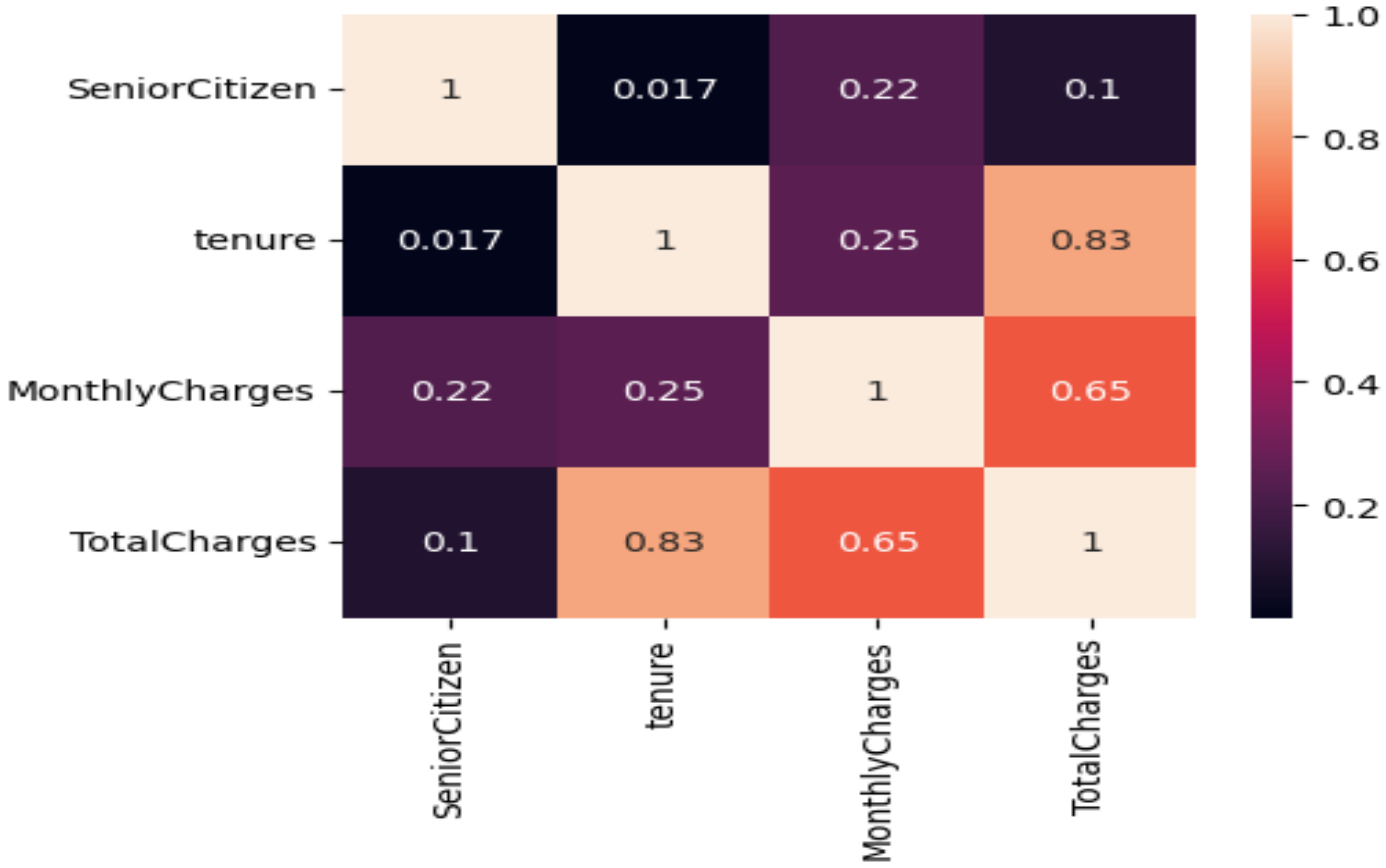




- **Churn rate is much higher in case of customer with Paperless Billing**

- **Customers with Online Backup and Tech Support have lower churn rate as compared to those who don't have Online Backup and Tech Support**

Checking the correlation between continuous variables



In terms of correlation, key things to monitor in your customer for them not to churn, or some of the things you want to priortise in your maintenance check is Tenure on Total charges which has a strong relationship of 83%(percent) correlation and Monthly charges on Total charges which also has a strong relationship of 63%(percent) correlation. It's mandatory to spend more time on the Tenure, Total charges and Monthly charges variables of the customer packages because it's somehow very useful in terms to determine whether or not customer will Churn.

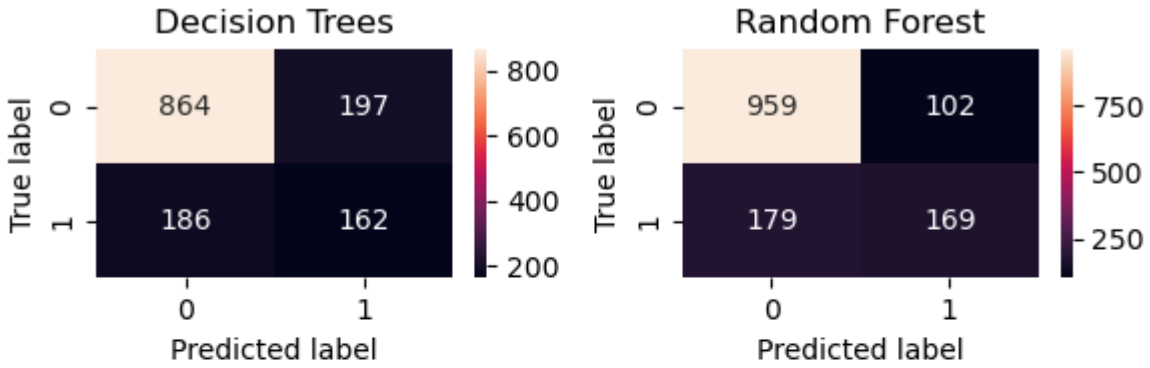
Modeling Approach:
Consider exploring various classification of algorithm(Decision trees, Random Forest, logistic regression) to identify the model that best captures patterns in the data.

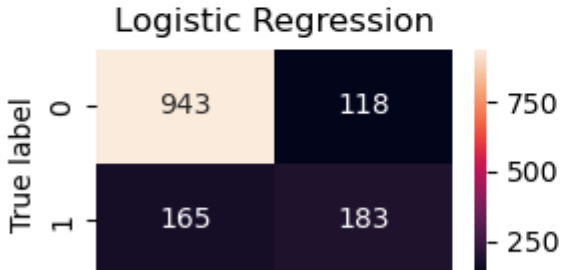
Decision Trees

	PRECISION	RECALL	F1-SCORE
0	0.82	0.81	0.80
1	0.45	0.47	0.47
ACCURACY	0.71		

Random Forest

	PRECISION	RECALL	F1-SCORE
0	0.83	0.90	0.86
1	0.63	0.47	0.64
ACCURACY	0.79		





Logistic Regression

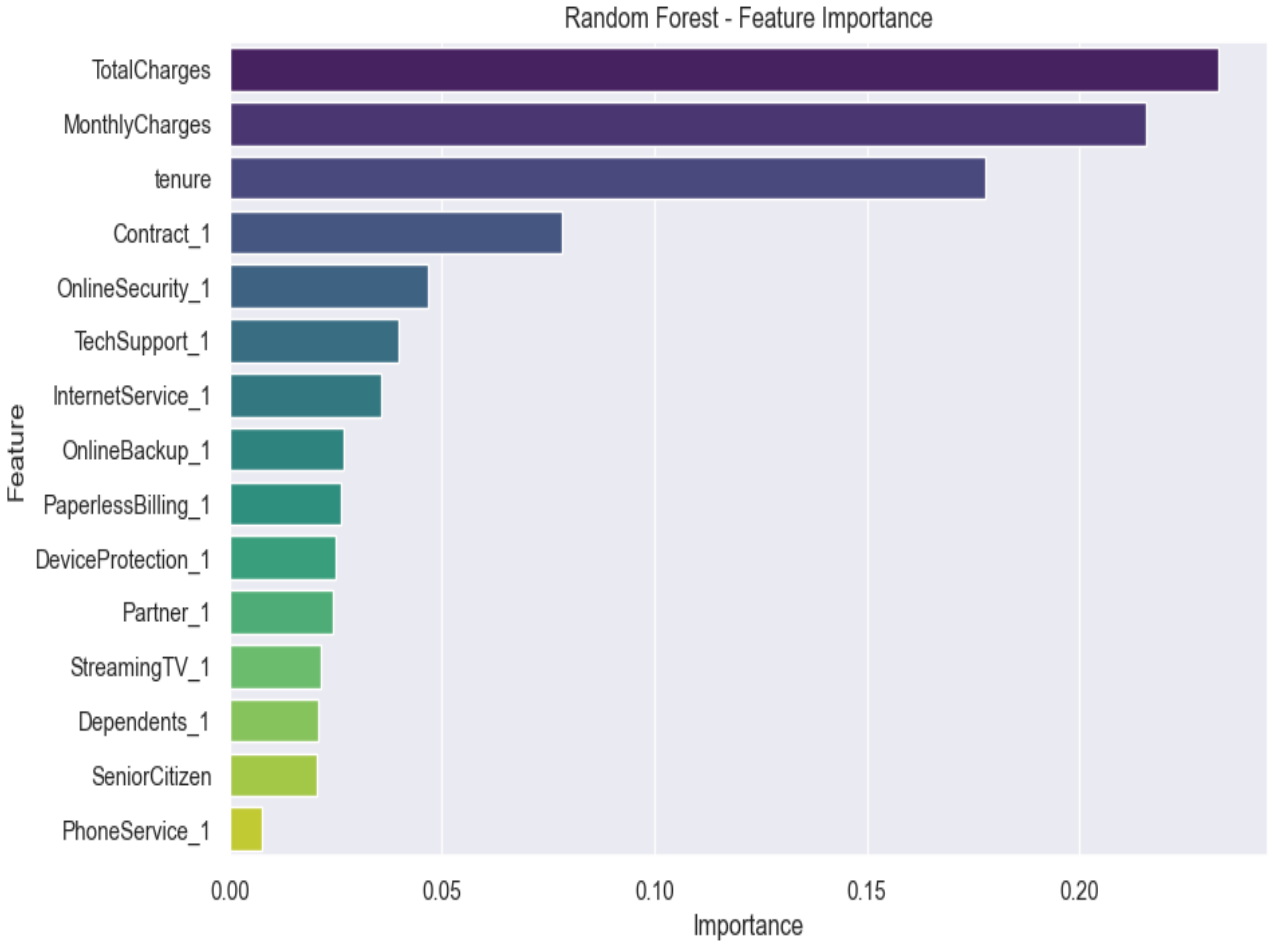
	PRECISION	RECALL	F1-SCORE
0	0.86	0.89	0.86
1	0.67	0.59	0.63
ACCURACY	0.79		

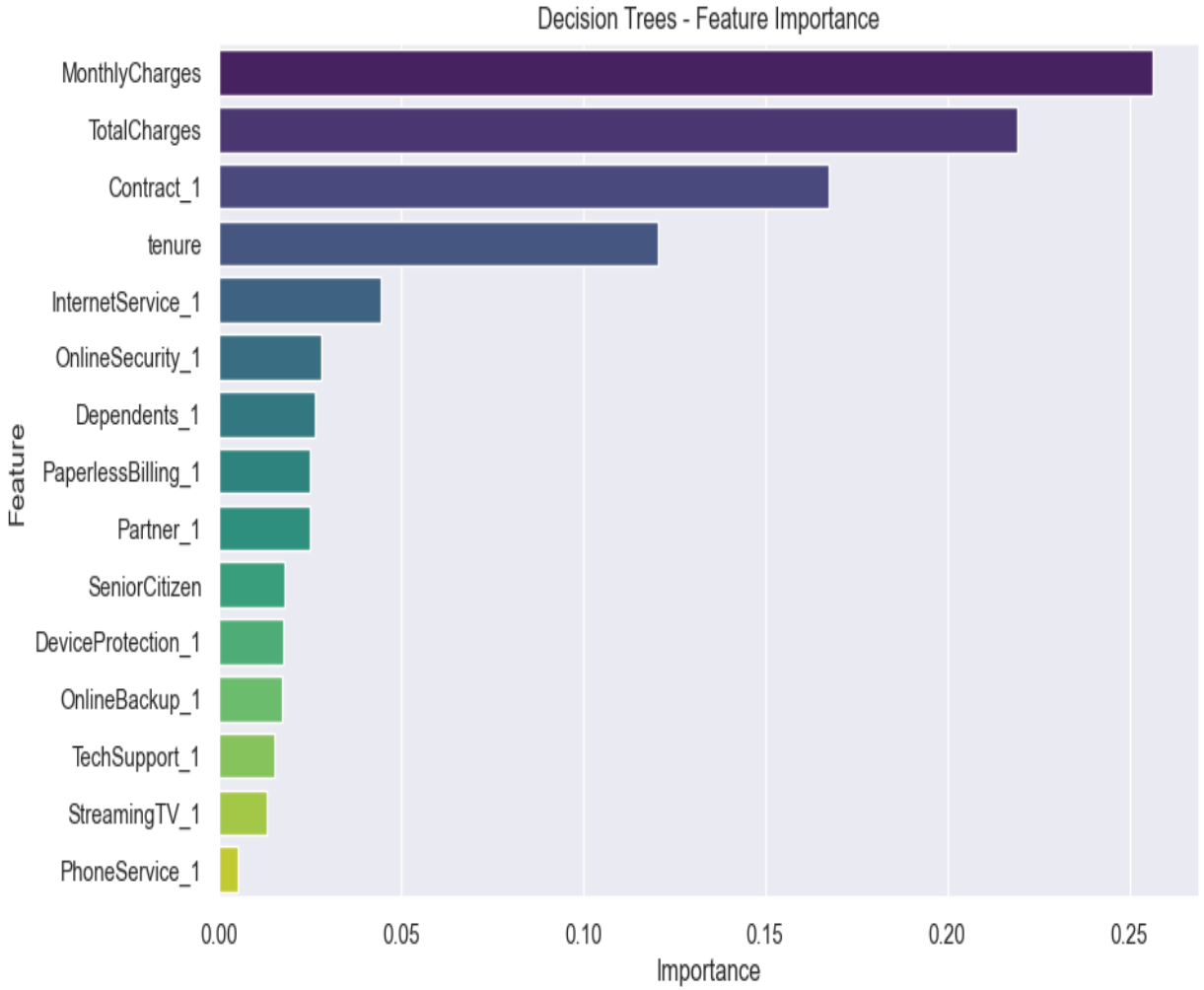
Insight derived from the classification of the algorithm

- The Decision tree model (accuracy - 71%), The overall accuracy of the model is 0.71, indicating that it correctly predicted the class labels for 71% of all instances. For class 0 (non-churn), the recall is 0.80, indicating that the model correctly identified 80% of all actual non-churn instances. For class 1 (churn), the recall is 0.49, indicating that the model correctly identified 49% of all actual churn instances.
- The Random Forest model (accuracy – 79%). The overall accuracy of the model is 0.79, indicating that it correctly predicted the class labels for 79% of all instances. Recall for class 0 (non-churn) is 0.90, indicating that the model correctly identified 90% of all actual non-churn instances. Recall for class 1 (churn) is 0.47, suggesting that the model correctly identified only 47% of all actual churn instances.
- For Logistic Regression model (accuracy – 81%). Accuracy of the model is 0.81, indicating that it correctly predicted the class labels for 81% of all instances. Recall for class 0 (non-churn) is 0.89, indicating that the model correctly identified 89% of all actual non-churn instances. Recall for class 1 (churn) is 0.59, suggesting that the model correctly identified only 59% of all actual churn instances.

- Overall, the Logistic Regression model demonstrates decent performance in predicting both non-churn and churn instances, with higher precision and recall for class 0 (non-churn) compared to class 1 (churn).

Feature Importance





- Total Charges and Monthly Charges are the most important feature(strong indicator) in the models when considering its prediction. Other features are also relevant in both model such as tenure and contract_1. So in conclusion, the features stated earlier are actually the most important indicators to notify whether or not a customer will churn.

Recommendations and Actionable Insights:

- Engage new customers with effectively through targeted communications, value-added services, or rewards programs, they more likely not to churn.
- From the insight derived there's a decrease in total charges as the prices increases. In other words, as prices increase, customers tend to incur lower total charges. This observation could potentially indicate a pricing strategy employed by the company where higher-priced plans offer discounts or incentives that lead to lower overall charges for customers.
- Conduct regular customer satisfaction surveys to gather feedback on service quality, network coverage, pricing, and customer support.
- Offer value-added services, exclusive content, or rewards programs to encourage ongoing engagement and reduce the likelihood of churn.
- Implement proactive customer support strategies such as regular check-ins, personalized assistance, and quick resolution of issues to enhance customer satisfaction and loyalty.