Telco Customer Churn Prediction: A Data Visualization Case Study

1. Introduction

Customer churn is one of the most significant challenges businesses faces. It refers to the phenomenon where customers stop doing business with a company or they just cancel the subscription offered by the company, leading to loss of revenue. Predicting churn early enables companies to implement strategies to retain customers, saving both revenue and resources. Data visualization plays a crucial role in identifying patterns and insights from the data, providing a foundation for effective churn prediction.

This case study focuses on using visualizations to understand customer churn behaviour. The objective is to leverage insights from the data to formulate recommendations for reducing churn.

2. Data Overview

The dataset used in this analysis contains information about a telecommunications company's customers. It includes both demographic and service-related features, such as:

- **CustomerID**: Unique identifier for each customer
- **Gender**: Male or Female
- **SeniorCitizen**: Indicates if the customer is a senior citizen (1) or not (0)
- **Partner**: Whether the customer has a partner (Yes or No)
- **Dependents**: Whether the customer has dependents (Yes or No)
- **Tenure**: Number of months the customer has been with the company
- **PhoneService**: Whether the customer has a phone service (Yes or No)
- **MultipleLines**: If the customer has multiple phone lines (Yes, No, or No phone service)
- **InternetService**: Type of internet service (DSL, Fiber optic, or No)
- **OnlineSecurity**: Whether customer has online security while using internet or not
- **OnlineBackup:** Whether customer has online backup while using internet or not
- **DeviceProtection:** Whether customer has device protection or not
- **TechSupport:** Whether customer has access to tech support or not.
- **StreamingTV:** Whether customer has access to streaming TV or not

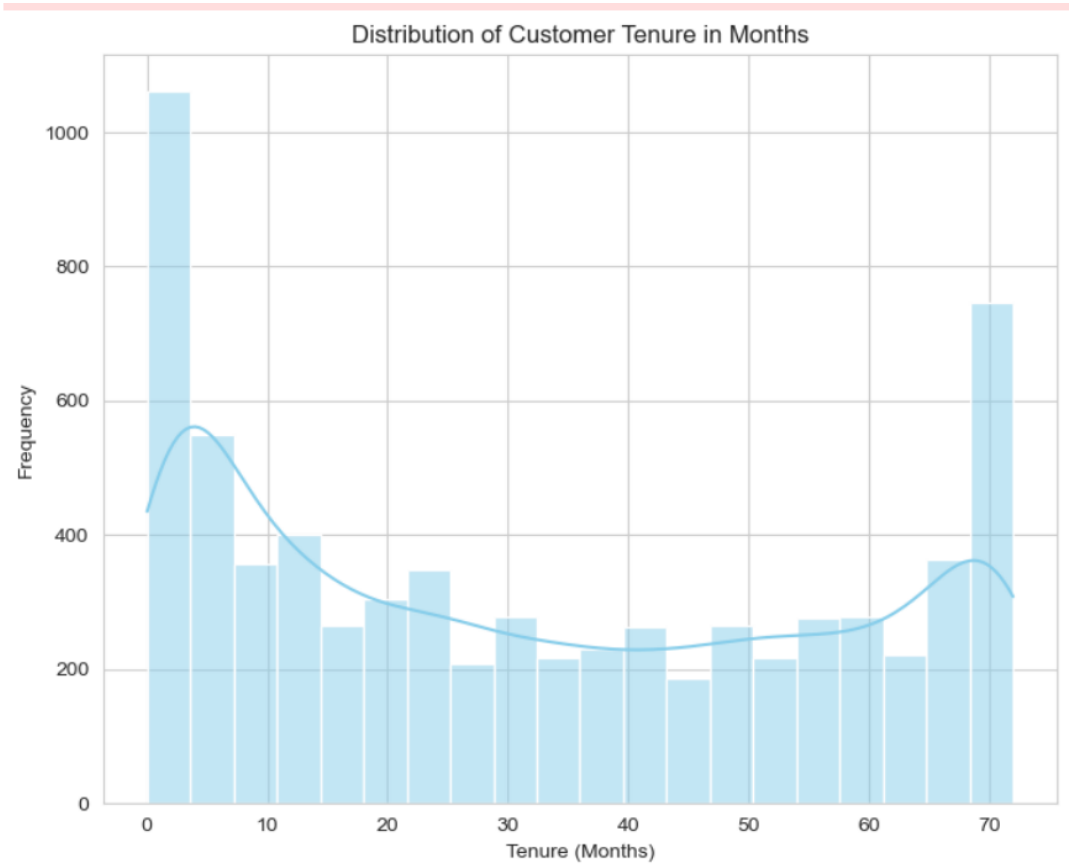
- **StreamingMovies:** Whether customer has access to streaming movies or not
- **Contract**: The customer's contract type (Month-to-month, One year, Two years)
- **PaperlessBilling**: Whether customer has paperless billing or not
- **PaymentMethod:** The preferred payment method by customer(Bank Transfer, Credit Card, Electronic Check,etc)
- **MonthlyCharges**: The amount charged to the customer monthly
- **Churn**: Whether the customer has churned (Yes or No)

The following data values have been created using Python in Jupyter Notebook: -

- **TenureGroup: -** The tenure column has been grouped into categories like "New Customer," "Mid-term Customer," and "Loyal Customer" based on their duration
    - **New Customer**: 0–12 months
    - **Mid-term Customer**: 13–36 months
    - **Loyal Customer**: 37+ months
- **ServiceCategory: -** Define broader categories based on combinations of these services, for example:
    - **Basic Service**: Customers with no internet service or limited additional services.
    - **Standard Service**: Customers with internet service but limited add-ons like online security or tech support.
    - **Premium Service**: Customers with comprehensive services, including multiple add-ons.

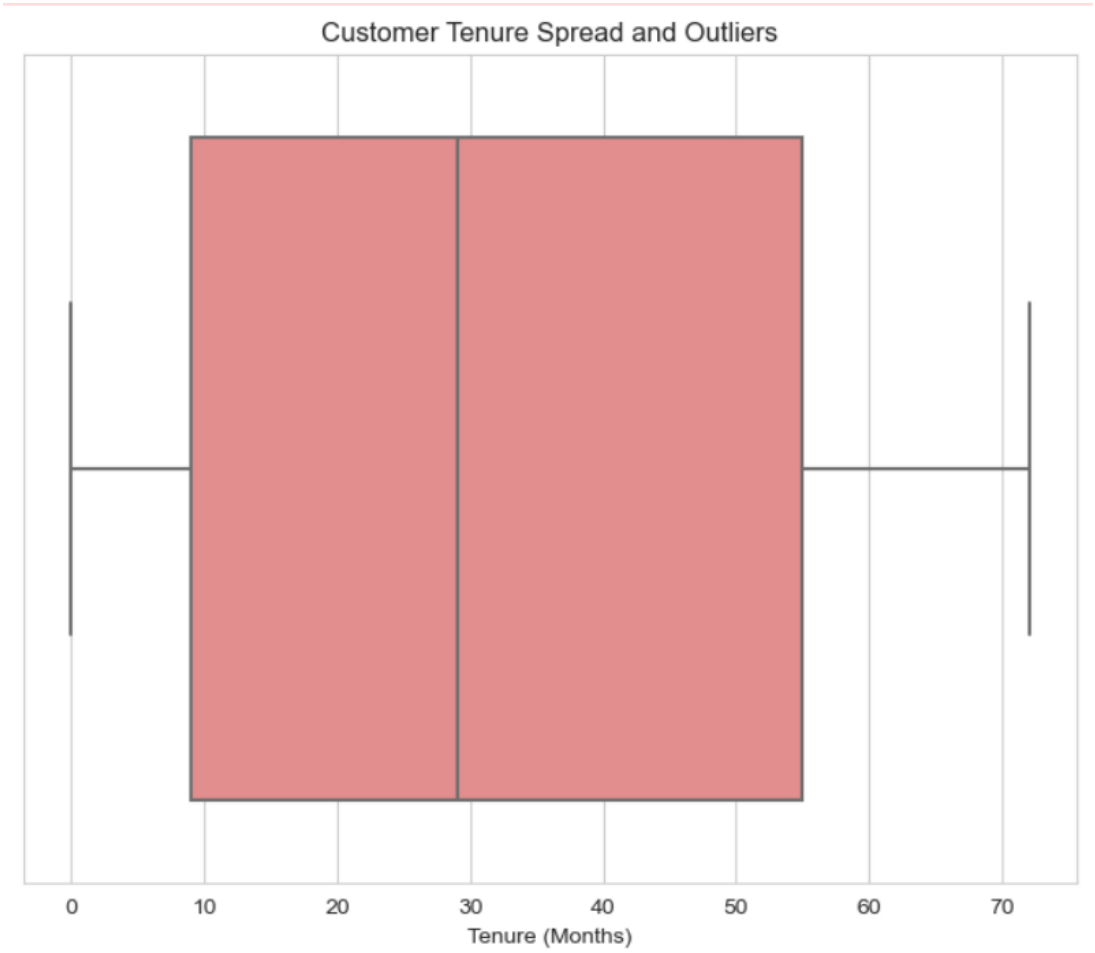
3. Visual Insights

a. Histogram with Density Plot for Customer Tenure Distribution: -



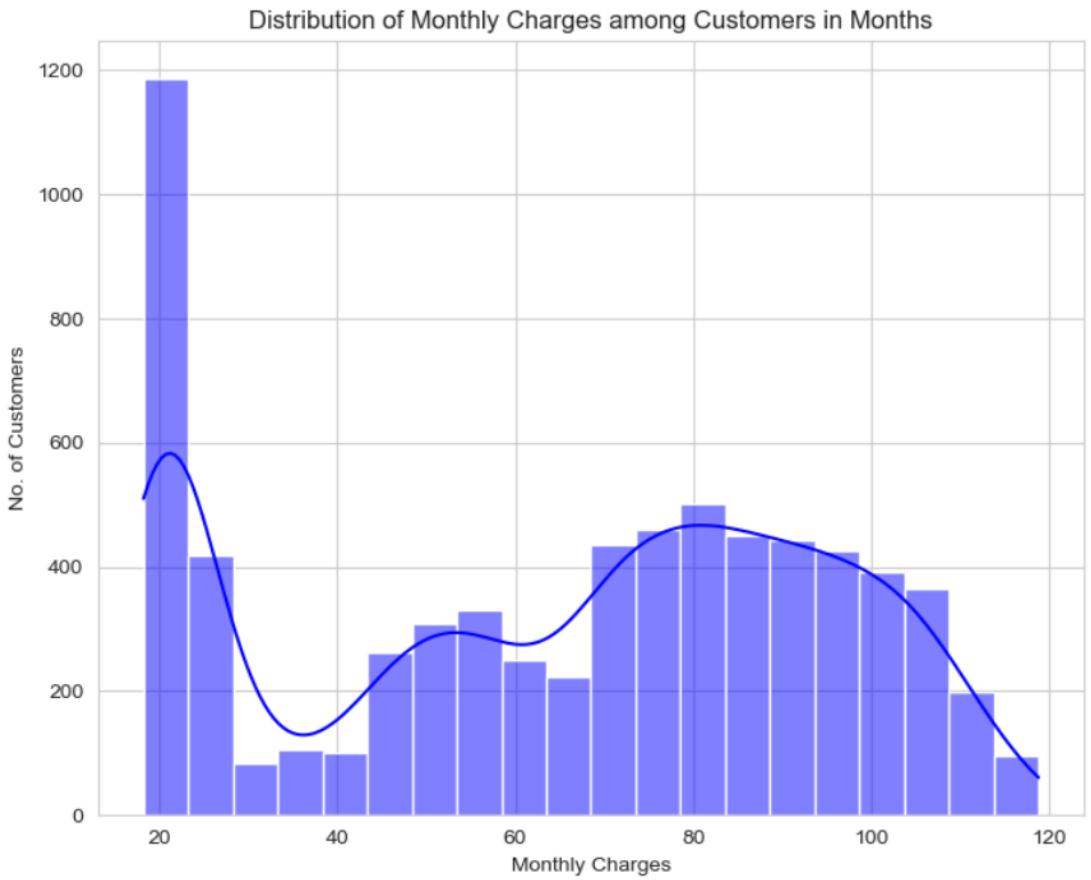
The histogram demonstrates that the customers had a smooth onboarding and the churns have been less likely during first 0 to 10 months. However, the largest number of churns have been seen between 10th month and 60th month after onboarding. The customers who have retained for more than 60 months have less churn rate.

b. Box Plot for Customer Tenure Distribution: -



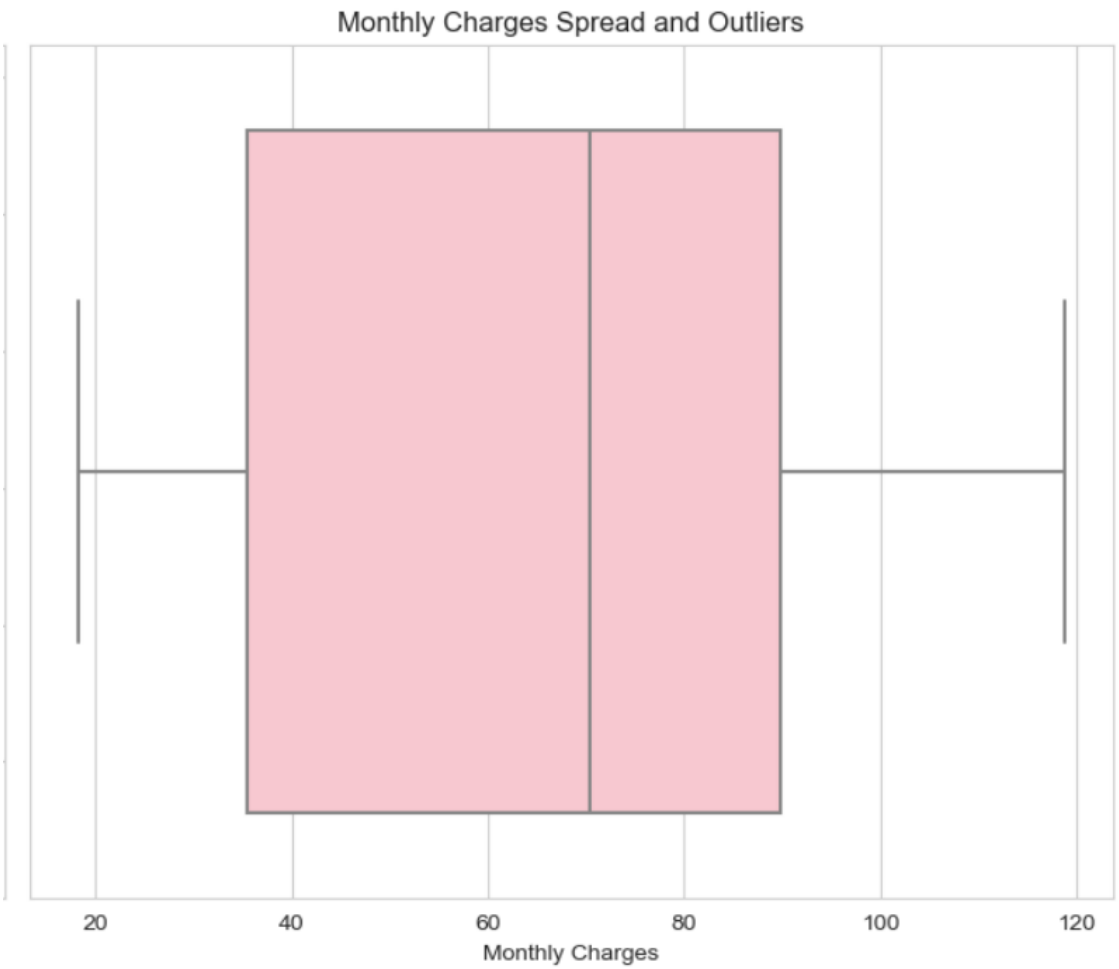
The boxplot demonstrates that median tenure is around 30 to 35 months which means most customers have remained for around 3 years. The IQR (Interquartile Range) shows that most customers have remained between 10 to 50 months indicating either most of them have churned early or have remained for significant amount of time. The number of customers who have stayed for less than 10 months are less than the number of customers who have stayed for 60 to 70 months.

c. Histogram with Density Plot for Monthly Charges Distribution: -



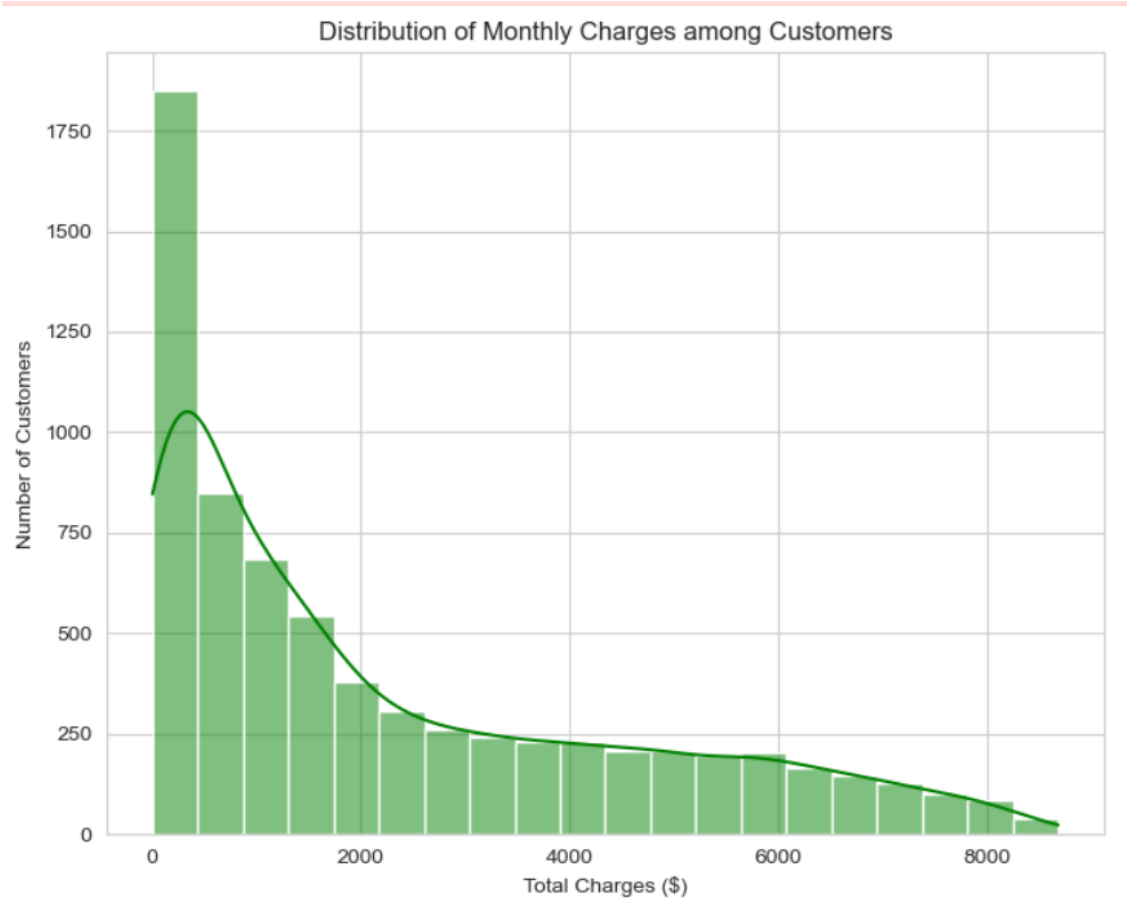
The histogram demonstrates that the large number of customers has retained for basic services which costed around 20 units whereas there are significantly a smaller number of customers paying between 40 to 60 units indicating either customers has churned, or standard service is not attractive towards them. On the other hand, many customers have liked premium services paying around 80 to 90 units but again there is dip in the number customers who are paying for more than 90 units for premium services.

d. Box Plot for Monthly Charges Distribution: -



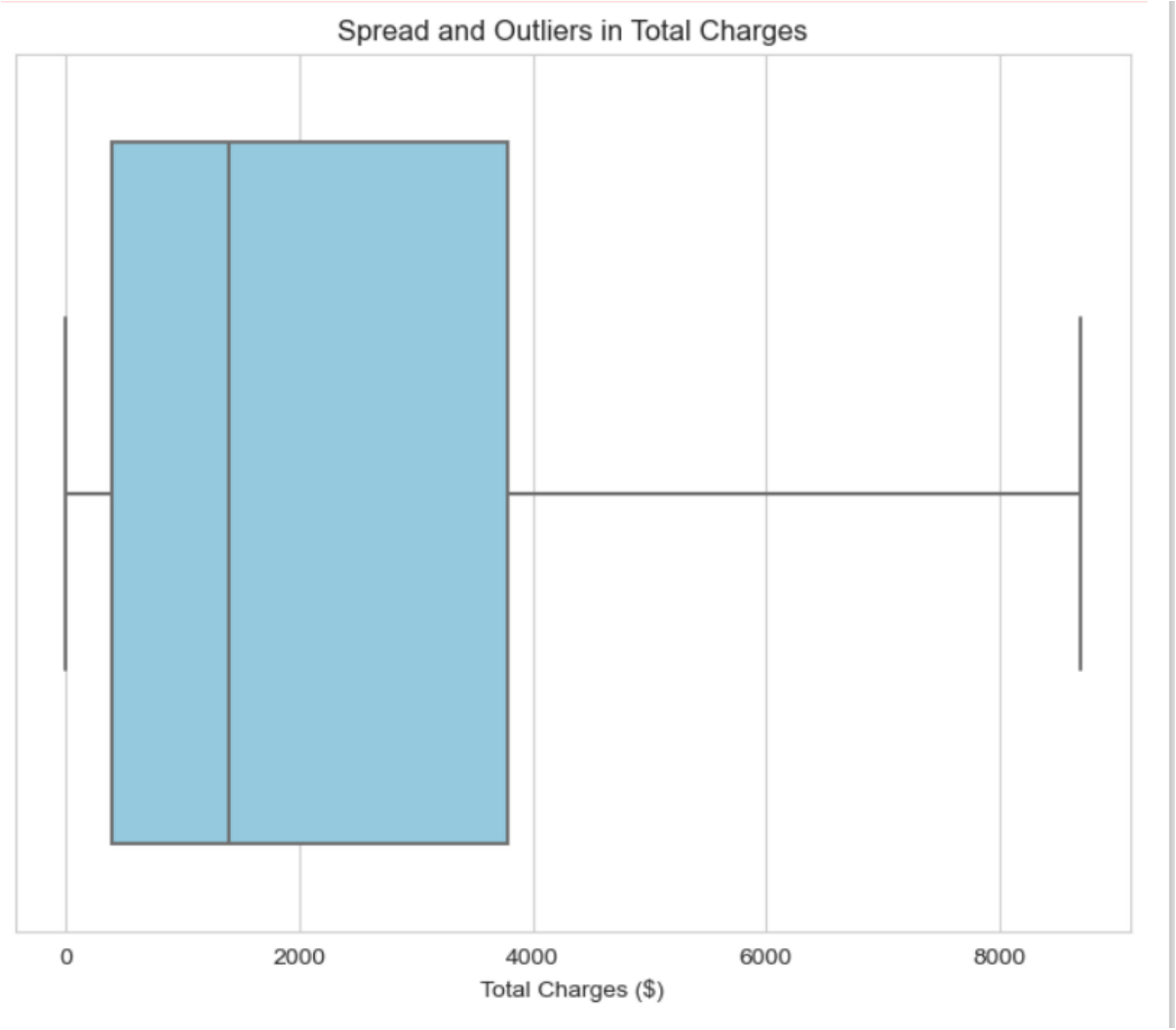
The boxplot demonstrates that the median monthly charges is around 70 units which means most customers are opting for either standard services or premium services. The IQR indicates that most customers are paying between 30 to 90 units indicating that services are spread across the customers. There are limited number of customers paying less than 35 units and the another set of limited customers paying more than 90 units.

e. Histogram with Density Plot of Total Charges Distribution: -



The histogram demonstrates that the maximum number of customers have paid the total charges of less than 1000$ indicating that customers are more interested in basic or minimal services. However, as the total charges increases, there is gradual dip in number of customers indicating very less customers are interested in long term services or premium services.

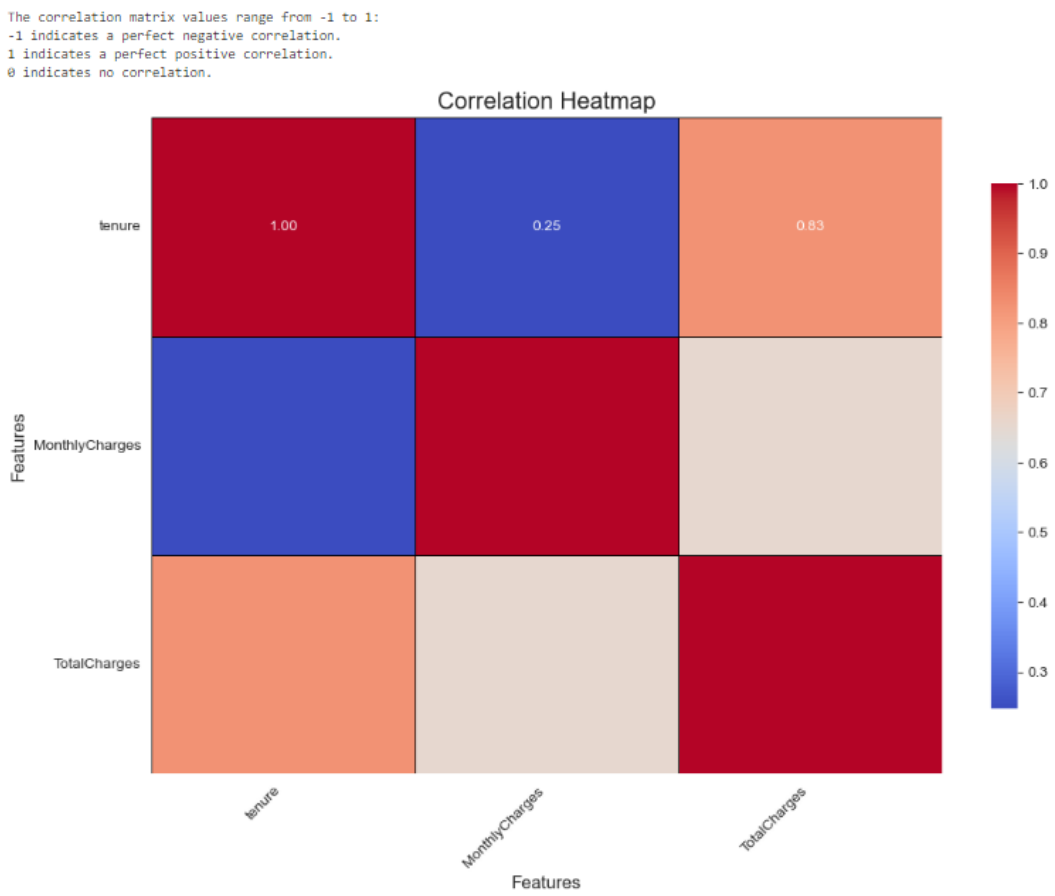
f. Box Plot for Total Charges Distribution: -



The box plot demonstrates that half of the customers are paying less than around $2000 and remaining half above this value indicating most customers are clustered around moderate levels of total spendings. The IQR indicates that majority of the customers have paid around $1000 to $4000 total charges.

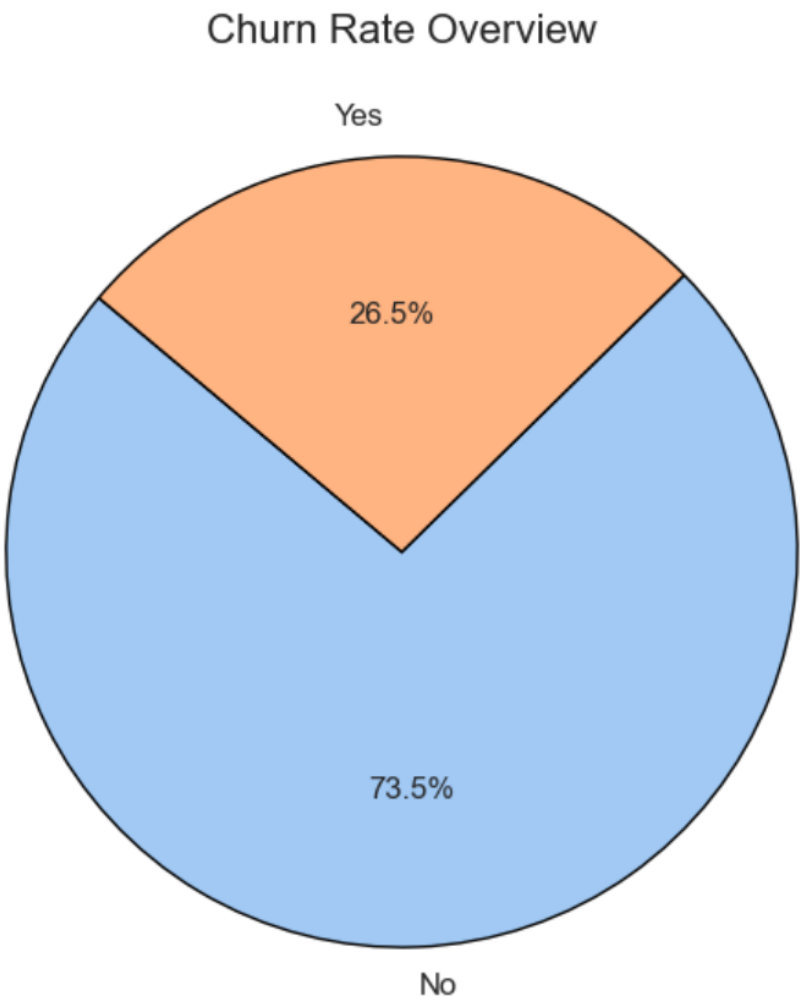
There are some exception cases where customers have opted for premium services and paid upto $8000.

g. Correlation Heatmap between Tenure, MonthlyCharges, TotalCharges



The strong positive correlation (0.83) between tenure and total charges shows that as the tenure increases, the total charges also tend to increase whereas weak correlation (0.25) between tenure and monthly charges indicates that tenure of the customer doesn't significantly affect the monthly payment. The moderately high positive correlation (0.65) between monthly and total charges suggest that high monthly charges could be associated with increasing total charges.

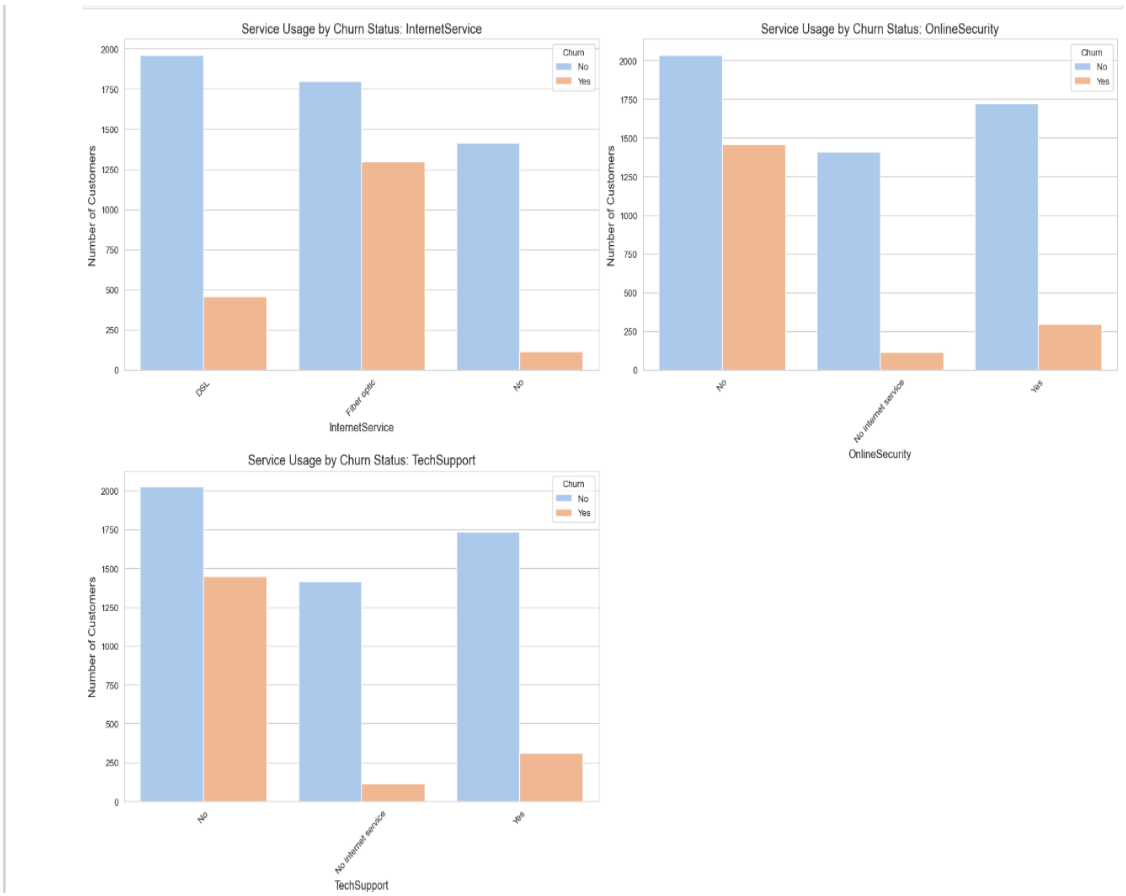
h. Churn Rate Overview



The given chart demonstrates that majority of customers (73.5%) have not churned which means that they are still using the services offered by the company and remaining customers (26.5%) have churned.

So, majority of customers have not churned.

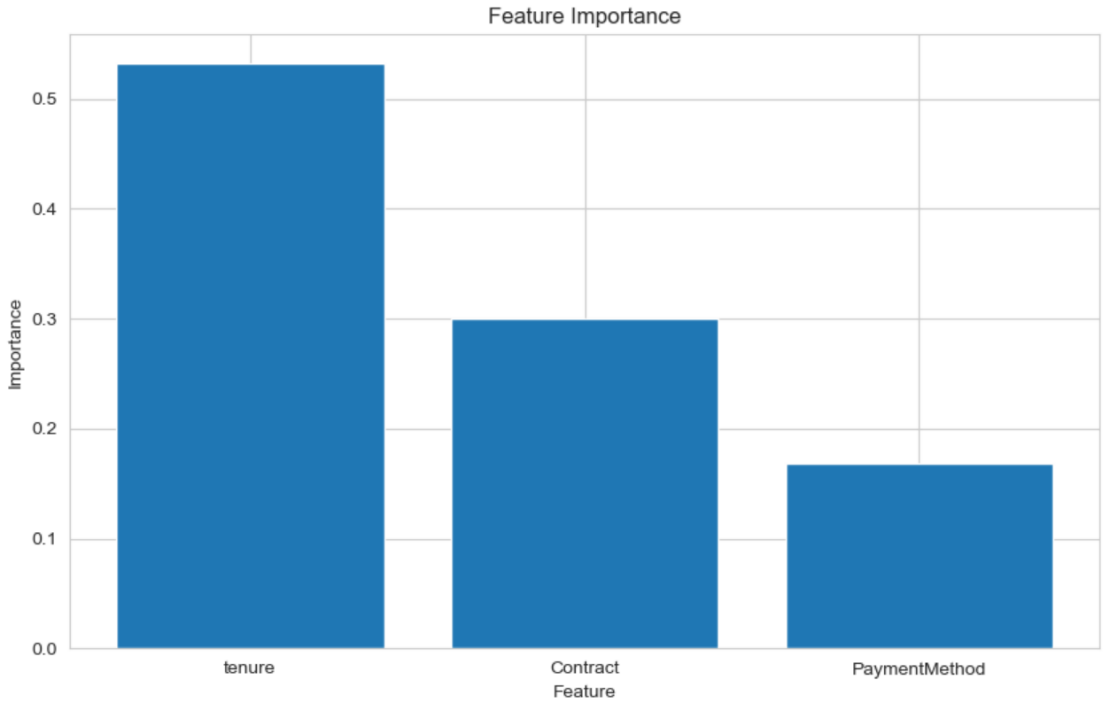
i. Service Usage Patterns



The company provides the following services: - Internet Service, Online Security and Tech Support.

The image demonstrates bar chart for differences in service usage between churned and retained customers. For all the three services, those customers who have not churned have used these services the most compared to the customers who have churned.

j. Feature importance using Random Forest Classifier

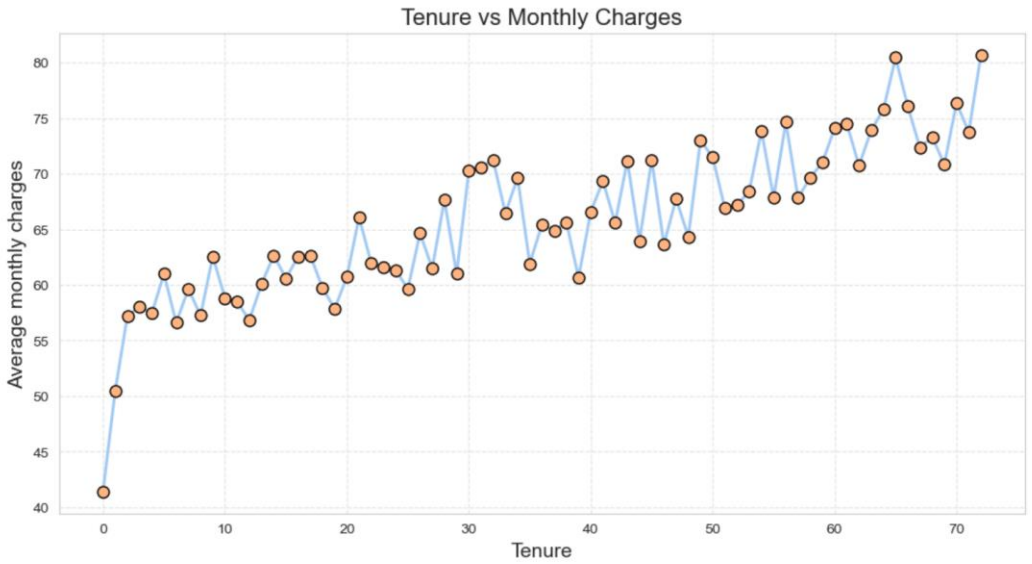


The tenure (over 0.5) has highest churn importance since it helps in identifying how long does a customer stay with the company and how it affects the company revenue, followed by Contract (around 0.25) and then by Payment Method (slightly above 0.1).

So, tenure of the customer will be most helpful for analysing the business improvements and outcomes whereas payment method is least important.

Contract Type is important but not as important as tenure.

k. Tenure vs Monthly Charges Line Chart

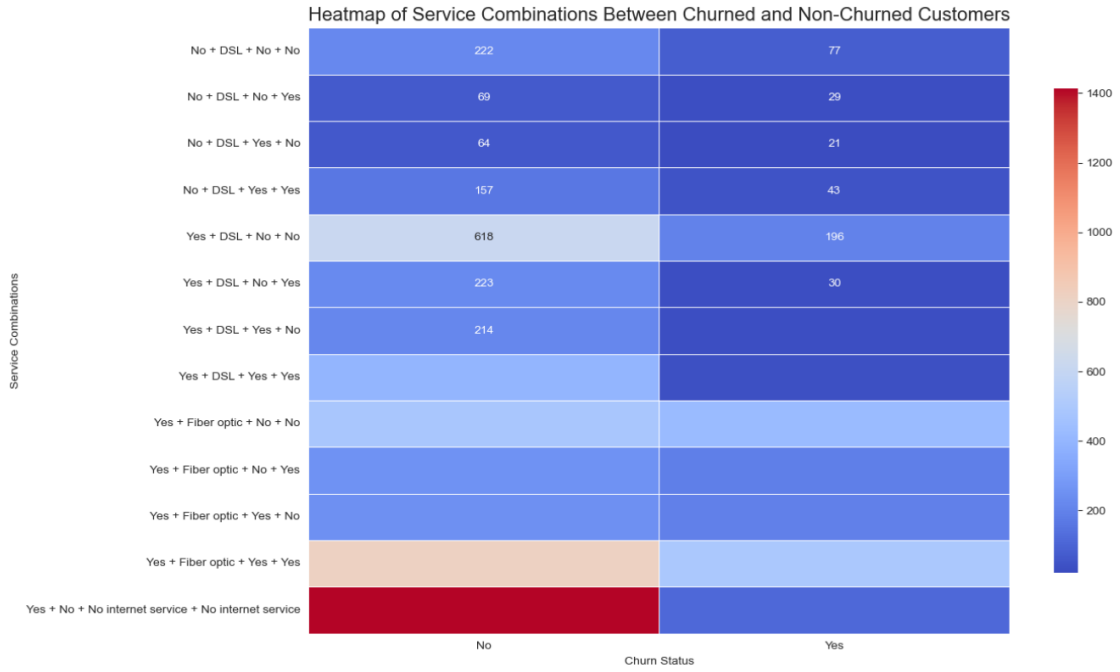


The orange dot represents the average monthly charges with respect to the tenure of customers with Telco services. The blue line indicates the changes in average monthly charges when the tenure increases.

The scatter plot shows that there is a lot of variability, meaning that there isn't a single clear pattern that applies to all the customers. For example, the average monthly charges of the customer with around 33 months tenure are 62 units and the average monthly charges of the customer with around 40 months tenure are 60 units.

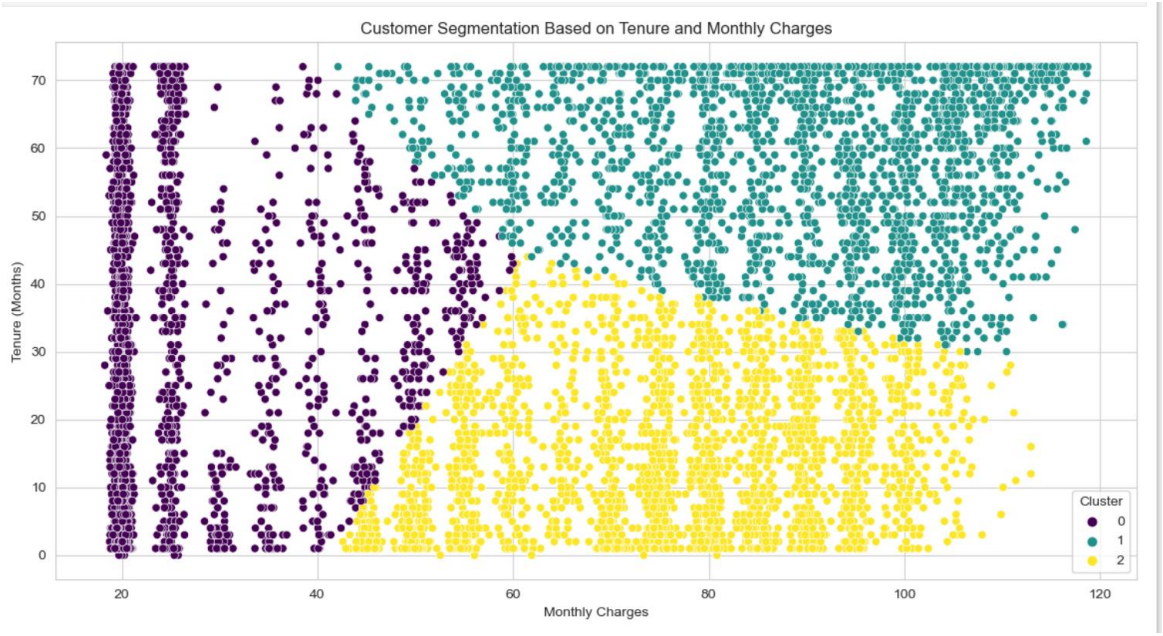
Despite variability, there is a slight upward trend in the blue line indicates that customers with longer tenures have higher monthly charges and reason could be upgrading services to premium ones or increase in prices along with increase in tenure.

l. **Heatmap of Service Combinations with Churned and Non-Churned Customers: -**



The heatmap indicates that if the customer has ‘No DSL + No’ i.e. no DSL have not churned whereas ‘Yes Fiber optic + Yes’ customers have churned. So the certain service combinations might have higher churn rates, for instance if a cell under ‘Yes’ has a higher count, it indicates that customers with that service combination are more likely to churn.

m. Customer Segmentation Based on Tenure and Monthly Charges: -

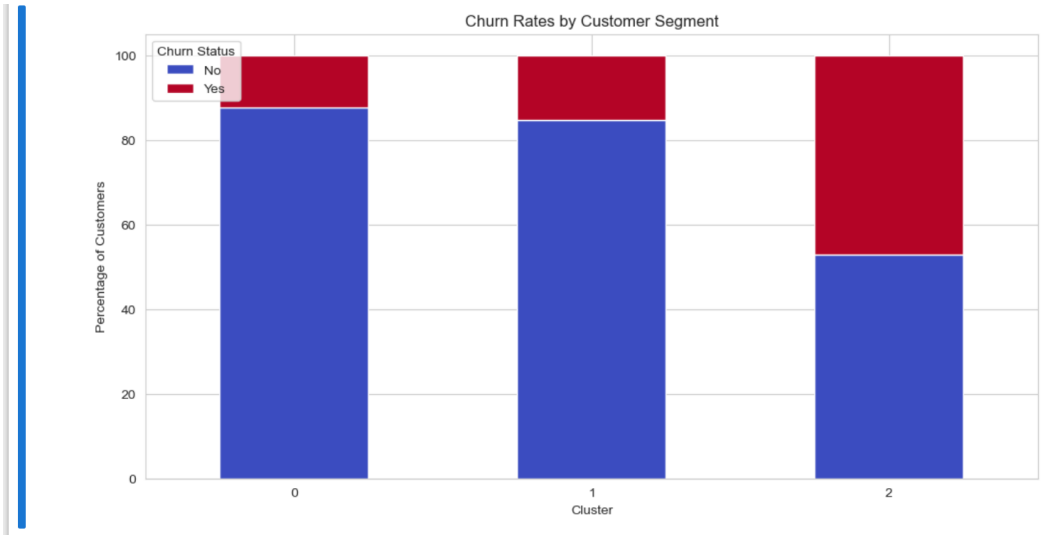


Cluster 0(Purple): - Indicates customers with lower monthly charges and shorter tenure. They are probably the new customers or the one with basic plans.

Cluster 1(Yellow): - Indicates customers with varying monthly charges but generally shorter tenure.

Cluster 2(Teal): - Indicates customers with higher monthly charges and longer tenure. They are probably loyal customers who have been with the service for a long time and are on premium plans.

n. Churn rates by Customer Segment: -

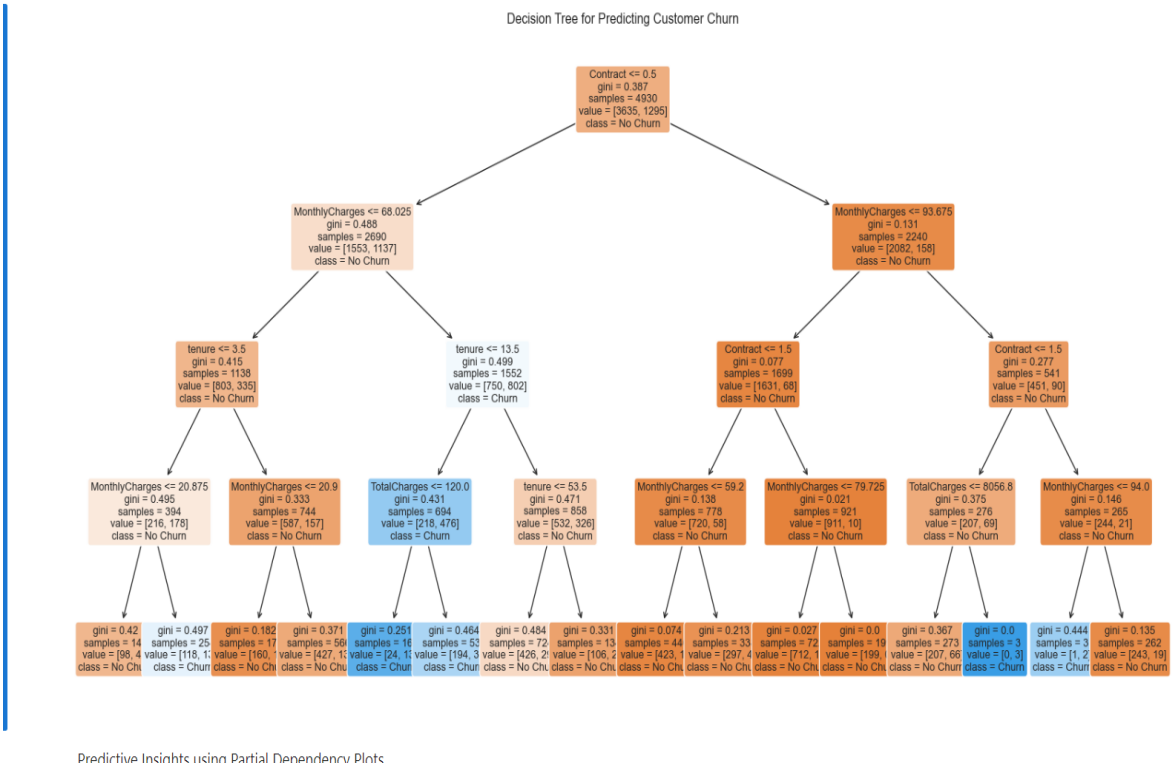


Segment 0: - The red portion of the bar is relatively small, indicating a lower churn rate. The blue portion is larger, showing a higher retention rate. These customers are likely the most satisfied or have the most favourable service conditions.

Segment 1: - This segment has a moderate churn rate. There might be some issues or dissatisfaction among these customers, but it's not as severe as segment 2.

Segment 2: - This segment has the highest churn rate and the lowest retention rate. These customers are the most likely to leave, indicating significant issues or dissatisfaction.

o. Predictive Insights using Decision Tree Visualization: -



Predictive Insights using Partial Dependency Plots

This decision tree is used to predict whether a customer will churn (leave the service) based on various attribute.

Root Node: -

The topmost node is the root node which represents the first decision point. In this case, it is based on the attribute ‘Contract’.

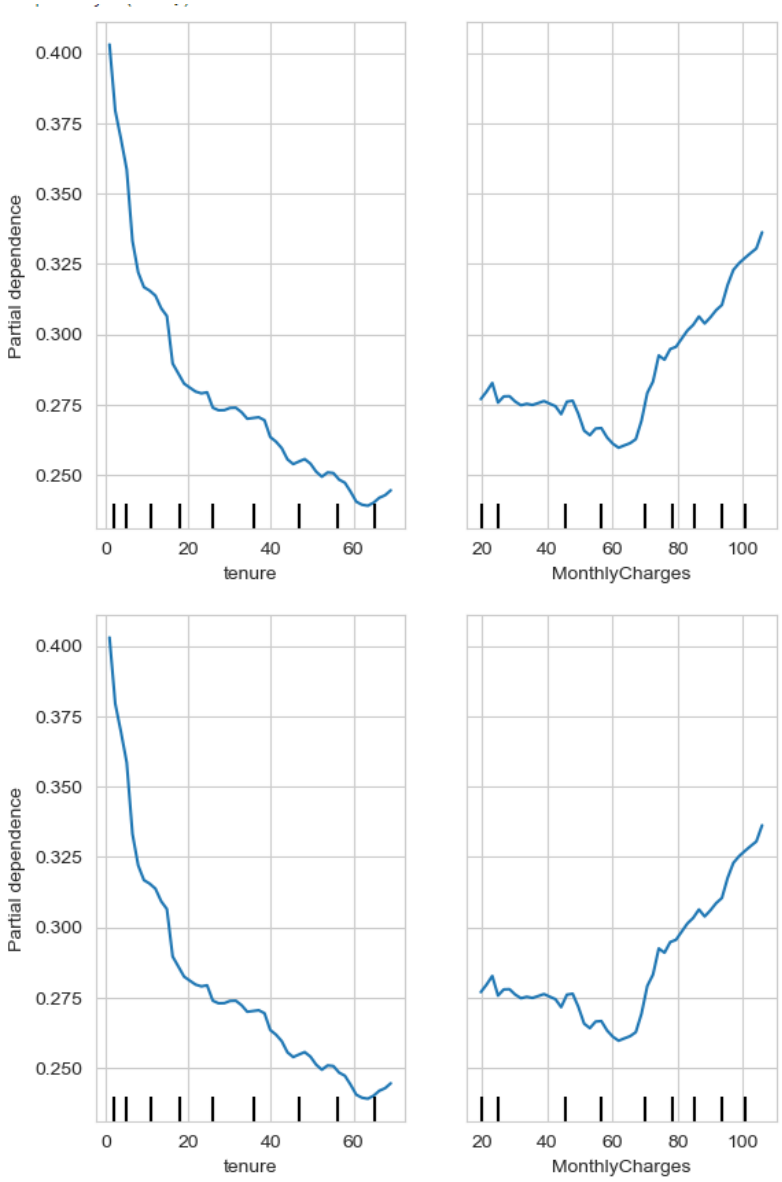
Decision Nodes: -

These nodes represent conditions or attributes that split the data into different branches. For example, a node might check if ‘MonthlyCharges <= 64.75’. Each decision node branches into two paths: one for when the condition is true and one for when it is false.

Leaf Nodes: -

These are the terminal nodes that represent the final outcome. In this case, the outcomes are whether a customer will churn (‘Yes’) or not churn (‘No’).

p. Plot Partial Dependence: -



Partial dependence of tenure: - As tenure increases, the probability of churn decreases, indicating that customers with longer tenure are less likely to churn.

Partial dependence of MonthlyCharges: - As monthly charges increase, the probability of churn increases, suggesting that customers with higher monthly charges are more likely to churn.

4. Solution

Based on the analysis of the Telco Customer Churn case study, the key focus is on identifying patterns in customer behaviour that lead to churn and proposing strategic solutions to reduce churn rates. The following solutions are derived from the insights gathered from the dataset and visualizations:

1. Customer Segmentation for Personalized Interventions:
    - Segmented Customer Groups: By grouping customers into clusters based on their tenure and monthly charges (e.g., new, mid-term, and loyal customers), specific retention strategies can be tailored for each segment. For example:
        - Cluster 0 (New customers with low charges): These customers are at lower risk of churn but may need targeted onboarding and initial service upgrades to boost engagement.
        - Cluster 2 (Loyal customers with higher charges): Since these customers are the most likely to churn, retaining them would require better long-term benefits, discounts on premium services, or loyalty rewards programs to improve satisfaction and retention.
2. Focus on Tenure and Service Enhancement:
    - Tenure was found to be the most significant factor in predicting churn. Therefore, the company should focus on improving customer satisfaction during the mid-tenure phase (10–60 months), where churn rates are highest.
        - Introduce loyalty bonuses or discounts for customers who reach key tenure milestones (e.g., after 12 months, 36 months).
        - Provide personalized service recommendations and offer upgrades, such as premium bundles, to ensure that customers see value in staying long-term.
3. Service Plan Optimization:
    - Customers paying for premium services are more prone to churn, possibly due to dissatisfaction with high costs or service value. The company can:
        - Reassess pricing structures: Adjust the premium plans to provide more flexible pricing options, allowing customers to opt into services they value most.

        - Improve the perceived value of premium services by adding exclusive benefits, personalized customer support, or free trials of additional services.

4. Contract Type and Payment Method:
    - The analysis shows that contract types play an important role in customer retention. Offering longer contracts with built-in incentives (such as discounts or added services) will likely encourage customers to remain with the company longer.
    - Payment Methods: While this feature had lower importance, optimizing customer experience with smooth, automated, and customizable payment options may reduce churn linked to billing issues.
5. Tech Support and Online Security Focus:
    - Customers who use Tech Support and Online Security services were more likely to stay. Increasing the visibility and availability of these services through proactive customer communication, tutorials, or free access for new customers could enhance retention.
    - Customers should be educated on the importance of these features and how they can benefit from them, especially those who have not adopted them yet.
6. Churn Prediction and Proactive Retention:
    - The decision tree analysis suggests that contract type, tenure, and monthly charges are key indicators for churn. The company should use this model to predict customers at risk of churning and implement early interventions such as:
        - Special offers or discounts.
        - Personalized outreach through customer service or support teams.
        - Surveys and feedback loops to address any issues or dissatisfaction early.
7. Service Usage Optimization:
    - Service combinations were found to influence churn, particularly customers with fiber-optic internet service who are more prone to leave. The company can improve the customer experience by ensuring consistent quality and reliability for high-value services like fiber-optic.

5. Conclusion

The analysis of customer churn in this case study reveals several critical insights and potential solutions for improving customer retention:

- Tenure is the most significant factor in predicting churn. Focusing on retaining mid-tenure customers through loyalty programs, value-added services, and improved engagement is essential.
- Service plans and pricing need to be reassessed, particularly for premium customers, to ensure they feel they are receiving adequate value for the cost.
- Contracts and payment methods should be leveraged as tools to encourage long-term customer retention, with longer contracts offering additional incentives.
- Enhancing the availability and visibility of Tech Support and Online Security services can help reduce churn by providing customers with essential service protections.
- A churn prediction model using decision trees can be highly effective in identifying at-risk customers and implementing early intervention strategies.
- Overall, by targeting specific customer segments with personalized retention strategies and optimizing service offerings, the telecommunications company can reduce churn rates and increase customer loyalty.

Implementing these recommendations based on the insights from the case study will likely lead to improved customer satisfaction, reduced churn, and ultimately, better financial performance for the company