**Introduction:**

The objective of this exploratory data analysis (EDA) is to understand the key factors that influence customer churn within a telecommunications company. Customer churn, or the rate at which customers leave a service, is a critical metric for businesses as it directly impacts revenue and growth. In this analysis, we aim to identify the underlying reasons why customers are leaving and determine which factors are most strongly associated with churn.

The dataset used for this analysis contains a variety of features that describe the customers and their interactions with the company. These features include demographic information such as gender and Senior Citizen, relationship status indicators like Partner and Dependents, service usage data such as Phone Service and Internet Service, and customer account details like Contract, Payment Method, Monthly Charges, and Total Charges. The target variable, Churn, indicates whether a customer has left the service.

By analyzing these features, we seek to uncover patterns and insights that will help the company better understand the drivers of customer churn. The ultimate goal is to use these insights to develop strategies for improving customer retention, thereby enhancing the company's overall performance.

**Data Preprocessing**

In the data preprocessing phase, the dataset was carefully examined to ensure its readiness for exploratory analysis. The following steps were undertaken:

1. **Handling Missing Values:**

Upon reviewing the dataset, it was found that there were no missing values in any of the columns. This was confirmed through a thorough check using functions like .isnull().sum(), which indicated that each feature was fully populated. As a result, no imputation or removal of data was necessary, allowing us to proceed directly to the analysis phase.

1. **Outlier Detection**

Outliers can significantly skew analysis results, so it was important to assess the dataset for any anomalies. Box plots and statistical methods were used to detect potential outliers in the numerical features such as MonthlyCharges and TotalCharges. However, the analysis revealed that there were no significant outliers present in the data. This finding allowed us to use the data in its original form without the need for any additional treatment.

1. **Normalization/Standardization**

Normalization or standardization is often necessary when dealing with features on different scales, especially in preparation for certain types of models. However, in this case, the dataset's features were already on comparable scales, and no further normalization or standardization was required. This decision was made to maintain the interpretability of the original data, particularly for features like MonthlyCharges and TotalCharges, where the raw values provide meaningful insights.

With the data being clean, consistent, and ready for analysis, we were able to move forward confidently into the exploratory analysis phase, focusing on uncovering patterns and relationships that contribute to customer churn.

**Exploratory Analysis**

In this section, we explore the key features of the dataset to gain insights into customer behavior and identify potential predictors of churn. The focus will be on analyzing the distribution and relationships between critical variables, including SeniorCitizen, tenure, MonthlyCharges, and TotalCharges.

1. **Descriptive Statistics**

**SeniorCitizen**: The SeniorCitizen feature is a binary variable indicating whether a customer is a senior citizen. The mean value of approximately 0.16 suggests that around 16% of the customers in the dataset are senior citizens. This feature will be explored further to understand its relationship with churn.

**tenure**: The tenure feature represents the number of months a customer has been with the company. The data shows a wide range, with a minimum tenure of 0 months and a maximum of 72 months. The average tenure is approximately 32 months, with a standard deviation of 24.56 months, indicating variability in customer loyalty.

**MonthlyCharges**: This feature represents the monthly charges incurred by the customer. The average monthly charge is around $64.76, with a standard deviation of $30.09. The distribution of MonthlyCharges will be analyzed to identify any significant differences between customers who churn and those who do not.

2. **Feature Distributions**

The distributions of SeniorCitizen, tenure, and MonthlyCharges were visualized using histograms and KDE plots. The analysis of tenure revealed a significant proportion of customers with low tenure, which could indicate higher churn rates among newer customers. The distribution of MonthlyCharges showed a slight skew towards higher charges, which could also be a factor in customer churn.

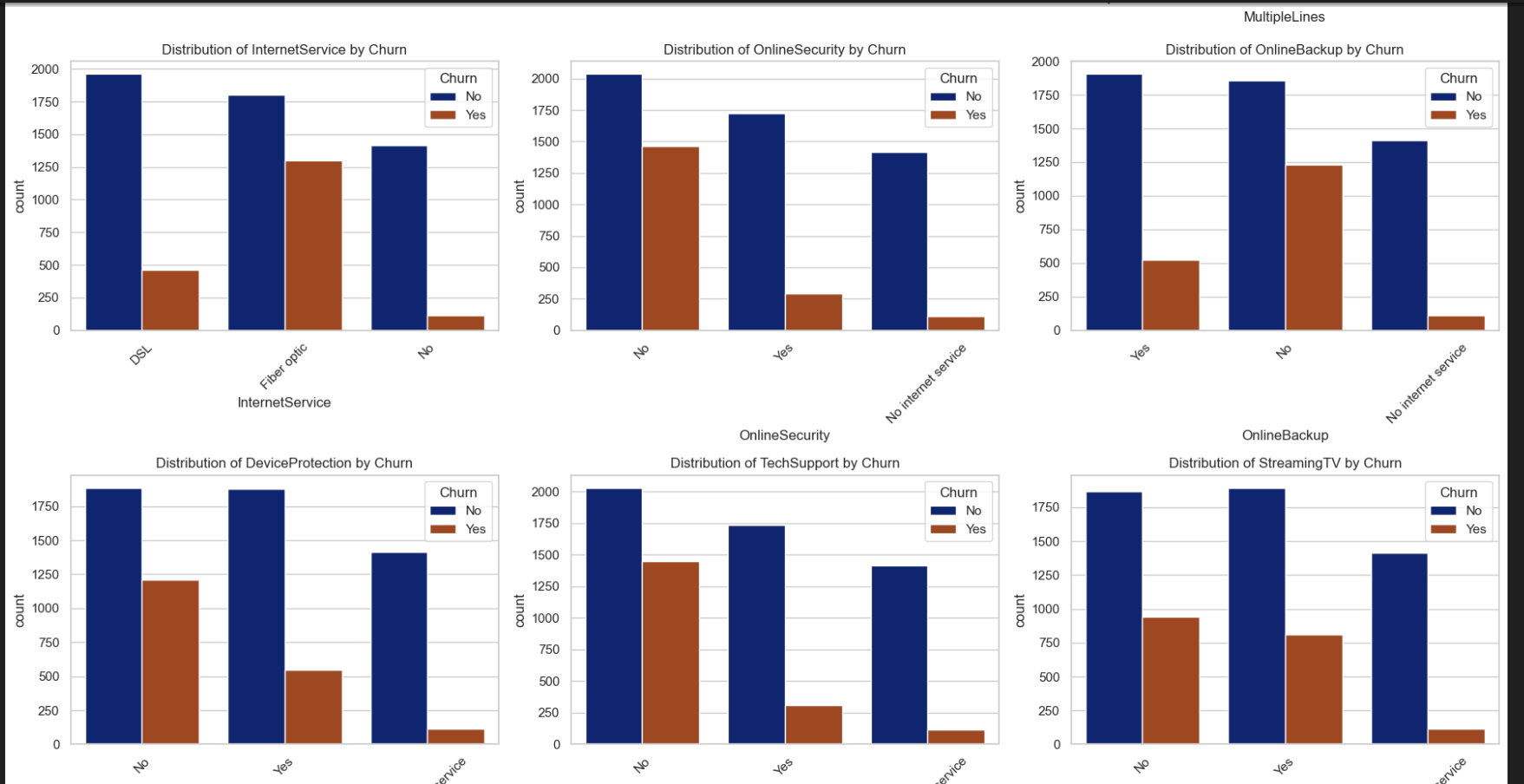
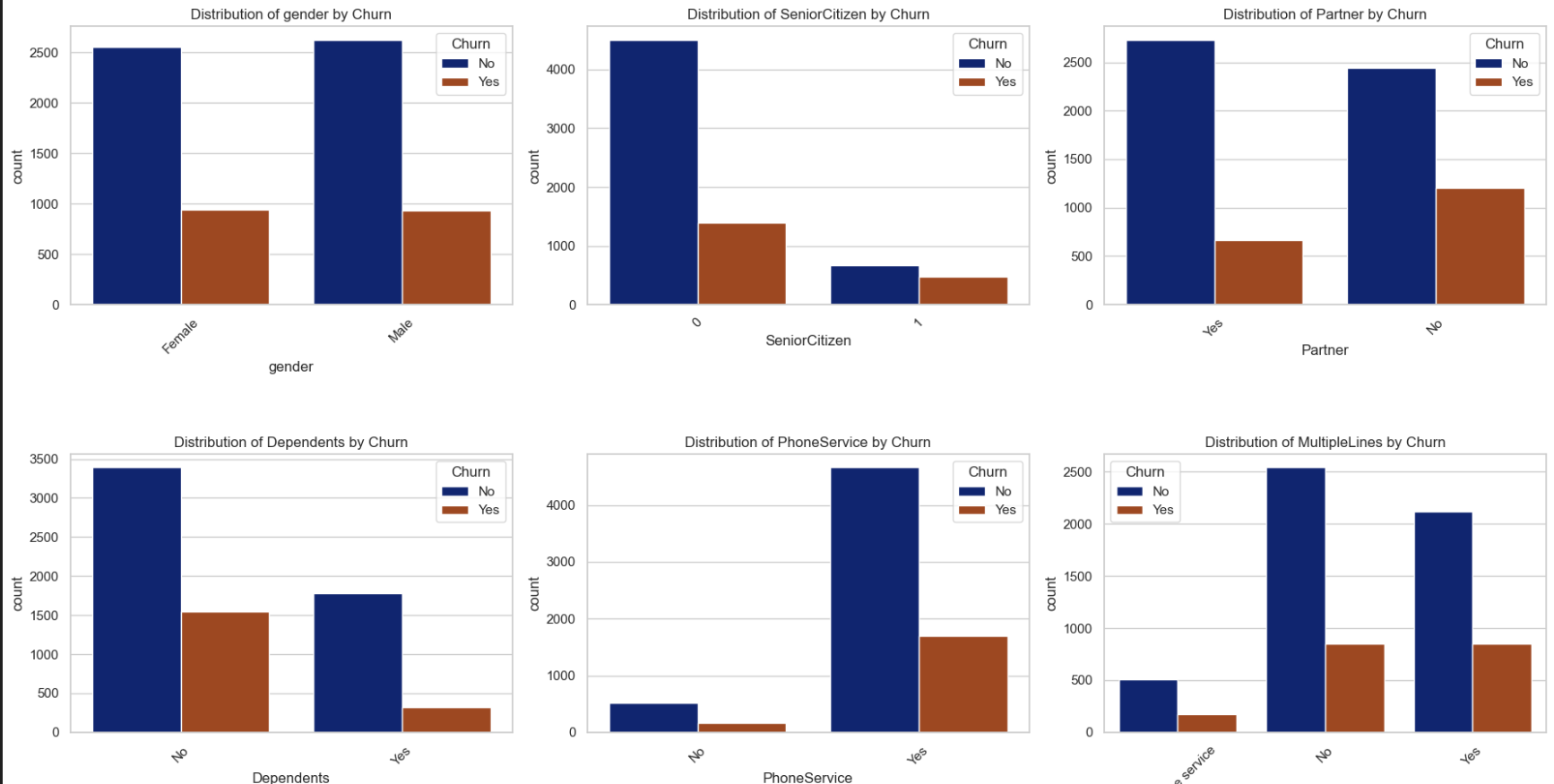
**TotalCharges**: The TotalCharges feature represents the total amount charged to the customer over the entire tenure. Although closely related to MonthlyCharges and tenure, TotalCharges provides additional insights into long-term customer value. Initial exploration indicates a wide range of total charges, with some customers being charged very low amounts (likely due to short tenure) and others accumulating significant charges over time.

3. **Correlation Analysis**

The correlation matrix and corresponding heatmap were used to identify relationships between features and the target variable (Churn). Features such as Contract, tenure, and MonthlyCharges showed notable correlations with churn, indicating that these factors play a significant role in determining whether a customer is likely to leave the service.

**SeniorCitizen**: While the SeniorCitizen feature had a relatively low correlation with churn, it still provides important demographic information that could interact with other features such as Contract or PaymentMethod.

**tenure and MonthlyCharges**: Both tenure and MonthlyCharges displayed interesting correlations. Customers with shorter tenure and higher monthly charges appeared to be more prone to churn, suggesting that new customers with higher service costs are at greater risk of leaving.



4. **Pattern Identification**

The analysis revealed that customers with shorter tenures and higher monthly charges are more likely to churn. Additionally, customers on month-to-month contracts, who have less commitment to the service, also show higher churn rates. These insights suggest that tenure, monthly charges, and contract type are critical factors in predicting customer churn.

The TotalCharges feature, when examined alongside tenure and MonthlyCharges, highlighted that customers with lower cumulative charges (likely due to shorter tenure) are also at higher risk of churn. This reinforces the importance of customer loyalty and the need to address issues early in a customer’s lifecycle to prevent churn.

Conclusion

Through this exploratory analysis, we identified key factors that influence customer churn, including tenure, MonthlyCharges, and Contract type. These insights provide a foundation for further analysis and model development aimed at predicting and reducing customer churn.