Telecom EDA Case Study

Akansha Rai

Business Understanding and Overview:

This dataset contains information about customers of a telecommunications company. Each row represents a single customer, and each column contains different attributes or features about the customers and their services. The goal of analyzing this data is to gain insights into customer behavior, preferences, and factors influencing churn rates.

The analysis of this dataset could include exploring patterns and correlations between different attributes and churn rates. It could help in identifying factors that contribute to customer churn and guide decision-making strategies aimed at reducing churn, improving customer satisfaction, and optimizing business operations.

Understanding the Data:

Churn

Findings

- Churn rate less when the monthly charges are low
- Correlation with Contract_Month-to-Month is Positive correlation ~ 0.40.
- Correlation with Contract_Two year is Negative correlation ~ -0.30.

Missing Data

Findings

- We observed Total charges Dtype is an object where it should be Float.
- After Dtype conversion we could analyze it including missing rows.
- `telco_base_data.info()` after this conversion, it reveals that there are 11 null (NaN) values in the 'TotalCharges' column.

Data cleaning

- As there is no Thumb rule on what criteria do we fill or delete the rows or columns,
- In this case we will drop the NULL rows resulting in the length of dataframe to 7032 from 7043

Percentage of Missing values
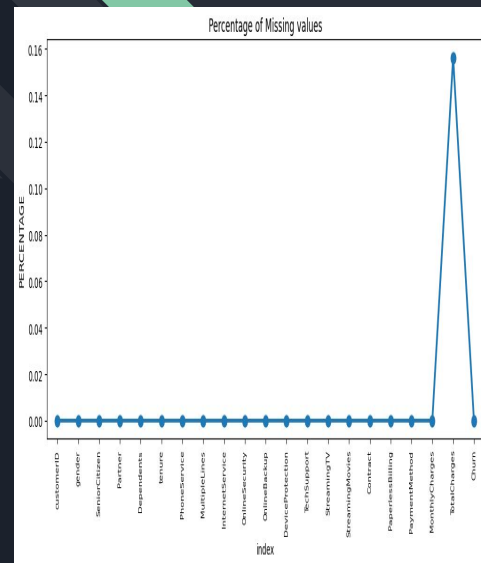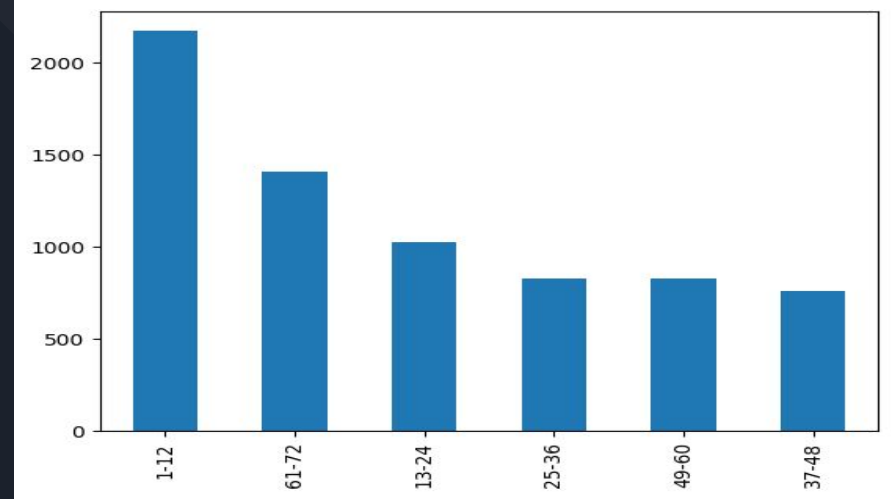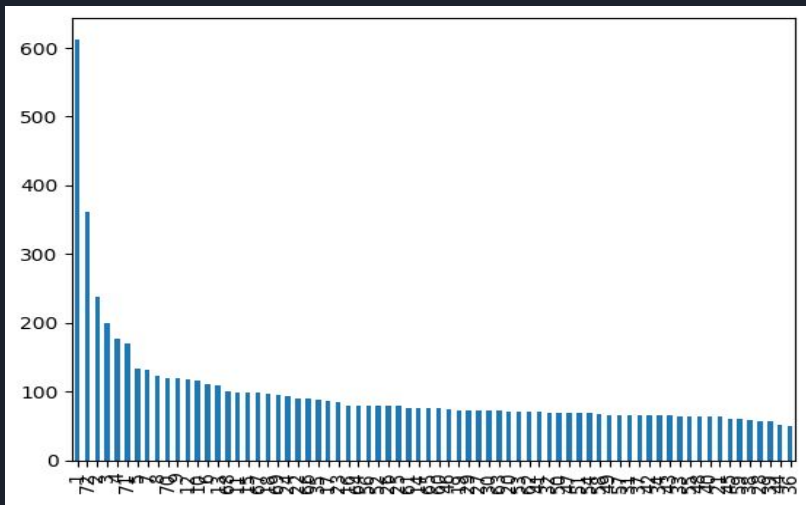
- The data cleaning process results in a new categorical feature 'tenure_group' that provides a more granular representation of customer tenure compared to the original 'tenure' variable.
- This transformation can be useful for subsequent analysis or modeling tasks where the categorical representation of tenure might be more informative or suitable than the continuous tenure values.
- It also ensures that each customer is assigned to an appropriate tenure group based on the specified intervals, facilitating better interpretation and analysis of tenure-related insights.

Initial Intuition From Data

- The mean value is approximately 0.1624, suggesting that about 16.24% of the customers are senior citizens on average.
- The mean monthly charge is approximately $64.80, indicating the average monthly fee paid by customers.
- The standard deviation (std) is $30.09, suggesting a considerable variability in monthly charges among customers.
- The minimum monthly charge is $18.25, while the maximum is $118.75, indicating a wide range of monthly charges.
- The standard deviation (std) is $2266.77, indicating a considerable variability in total charges among customers.
- The minimum total charge is $18.80, while the maximum is $8684.80, indicating a wide range of total charges.

In summary, the initial analysis provides insights into the distribution and variability of key variables such as customer seniority, tenure, monthly charges, and total charges. Further analysis, including visualization and exploration of relationships between variables, can provide more comprehensive insights into the data.

Categorical Analysis

- The code visualizes the distribution of each predictor category with respect to churn status. Each bar in the plot represents the count of customers belonging to a specific category of the predictor variable.
- The hue parameter (`hue='Churn'`) allows for the comparison of churned and non-churned customers within each category of the predictor.
- By observing the count plots, one can analyze how the distribution of predictor categories varies between churned and non-churned customers. This analysis helps in identifying predictors that might have a significant impact on churn behavior.
- For example, if certain categories of a predictor have a higher proportion of churned customers compared to others, it suggests that predictor may be correlated with churn and can be considered important for predictive modeling or further investigation.
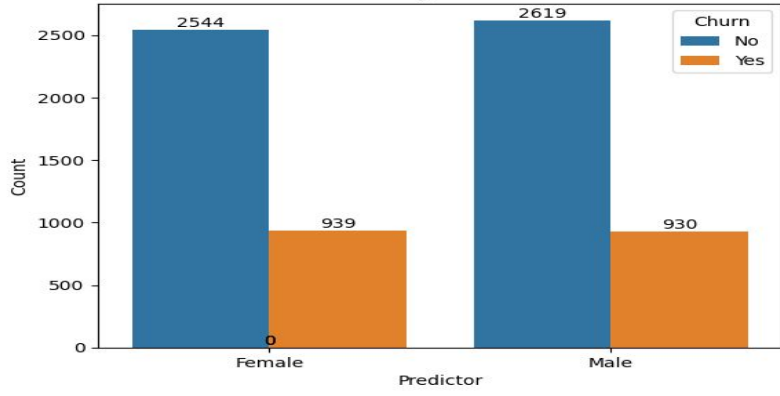
Insights

- The churn rate of Male and female has negligible difference
- The churn rate for female is 26.96% and for Male is 26.20%
- Therefore, it can be inferred that gender does not significantly influence churn behavior in this context.

Count Plot for SeniorCitizen with Churn
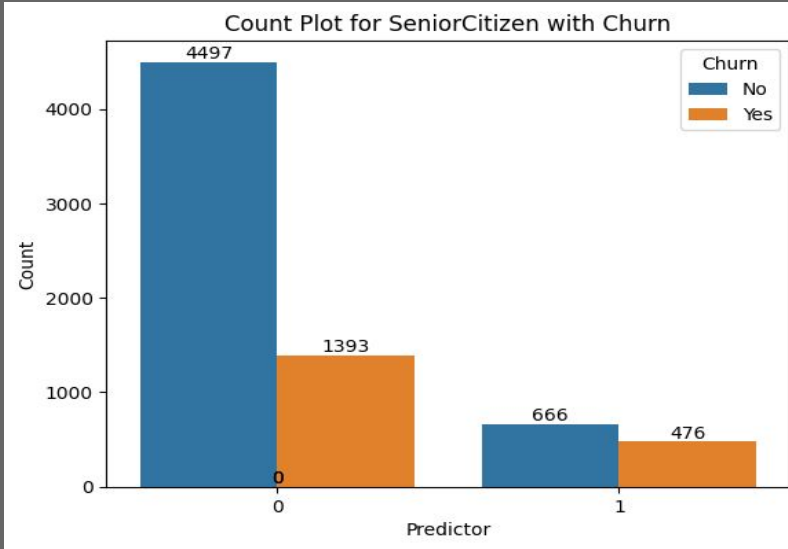
Insights

- The churn rate for senior citizen is 41.68% and for non senior citizen is 6.67%
- It is evident that senior citizens exhibit a substantially higher churn rate despite their smaller population size compared to non-senior citizens.

Insights

- The churn rate for partners is 19.71% and for non partner is 32.97%
- This suggests that being a partner is associated with a lower likelihood of churning compared to non-partners.
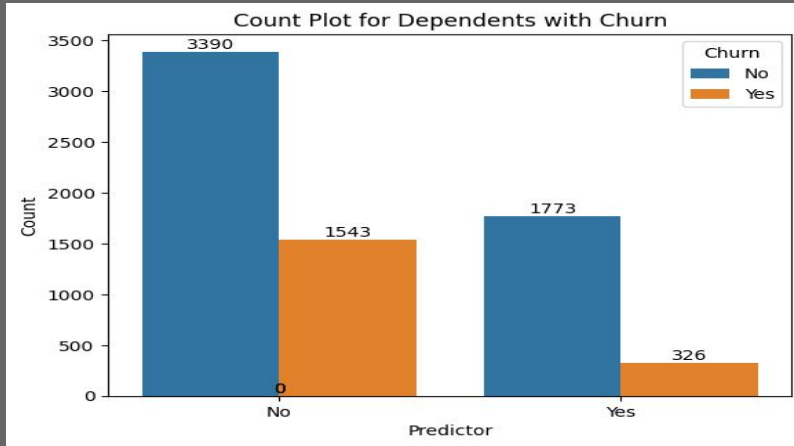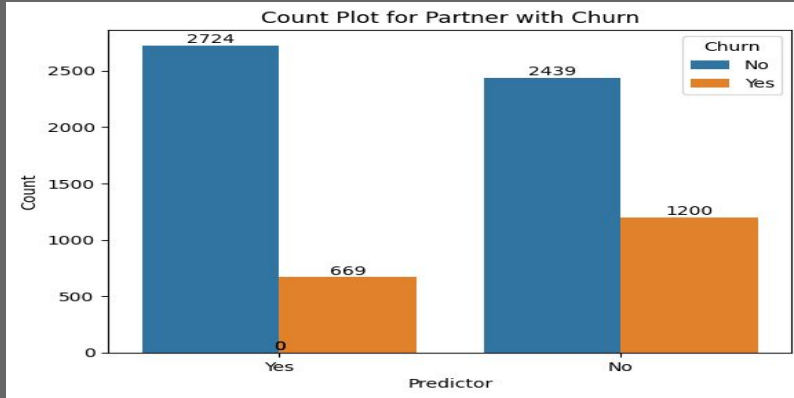
Insights

- The churn rate for customers without dependents is substantially higher at 31.28%, compared to a lower churn rate of 15.53% for customers with dependents.
- Despite the total count of non-dependents being 4933 and dependents being 5890, the significant disparity in churn rates suggests that the presence of dependents is associated with a lower likelihood of churning. Further analysis may be required to understand the underlying factors influencing this difference and to develop targeted retention strategies, particularly focused on non-dependent customers, to mitigate churn and enhance overall customer satisfaction and loyalty.

Numerical Analysis

- Differences in the distributions of monthly charges between churned and non-churned customers may indicate that monthly charges play a role in customer churn.
- Further analysis, such as statistical tests or predictive modeling, can be conducted to understand the relationship between monthly charges and churn more deeply.
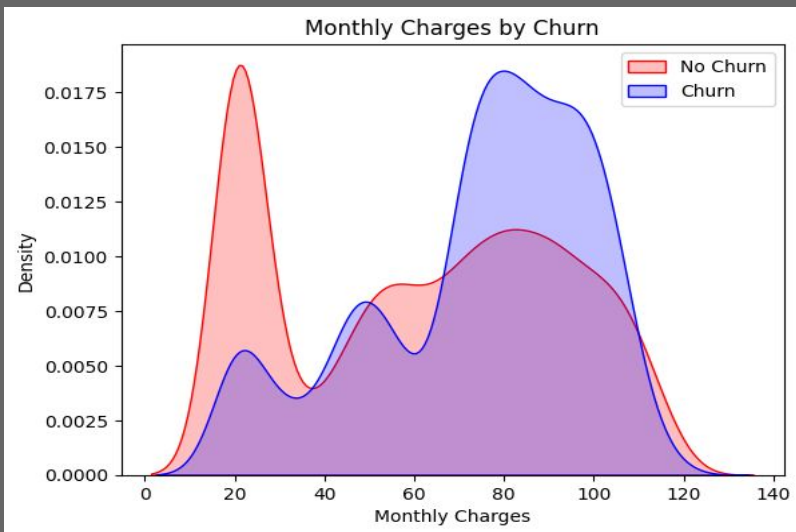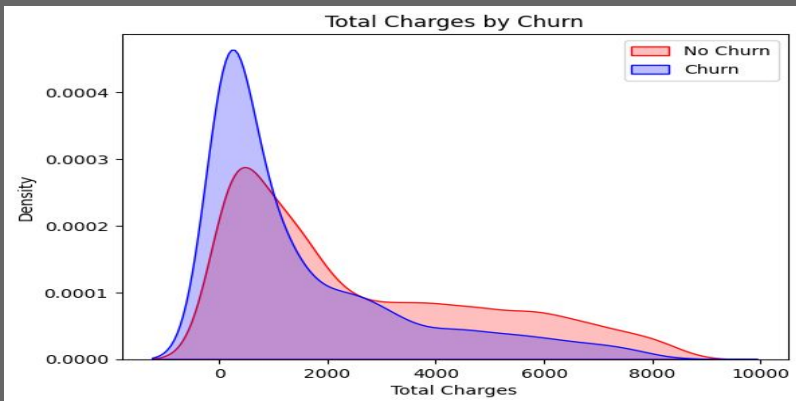
Numerical univariate Analysis

Findings

- People with low monthly charges are less likely to churn than people with high monthly charges.
- Total Charges graph needs future observations as the above graph does explain well

In summary, the code snippet facilitates a numerical analysis of the 'MonthlyCharges' variable by visualizing its distribution stratified by churn status, using KDE plots. This analysis helps in understanding the relationship between monthly charges and customer churn in the telecommunications company's dataset.

Final Thoughts

- The churn rate for Total charges is more when the total charge is low that could have multiple reasons. For instance people churning in initial months, people with more monthly charges etc
- People with High monthly charges are churning more
- Senior Citizens are more likely to churn: ~41.68%
- Customers with higher monthly charges and low total charges are more likely to churn
- People with another phone service are less likely to churn
- People with no internet service are less likely to churn
- People with Fiber optical are more churners ~41.89%