A1:Question for Analysis

Is there a significant difference in customer churn rates between different states?

A2:Benefit from Analysis

Analyzing churn rates across different states could help stakeholders, such as customer service managers and marketing teams, tailor their strategies to reduce churn. For instance, if a particular state shows a higher churn rate, targeted retention efforts could be deployed in that region.

A3:Data identification

To analyze the question, "Is there a significant difference in customer churn rates between different states?" the following columns from the dataset are crucial:

State: This column identifies the state in which each customer resides. It is a categorical variable and will help in grouping customers by geographic location to observe if there are regional differences in churn rates. The state variable will be used to determine if customers in specific states are more likely to churn.

Churn: This binary column indicates whether a customer has churned (`Yes`) or remained with the service (`No`). It is the dependent variable for our analysis and serves as the outcome measure to see if the likelihood of churn varies across different states. Analyzing this variable in relation to the `State` column will allow us to perform the Chi-Square test to identify any significant patterns in customer behavior across regions.

By focusing on these two columns, the analysis will isolate geographic factors as potential drivers of customer churn, enabling the organization to target interventions more effectively. This approach can reveal whether specific states require more attention due to higher churn rates, or if efforts can be evenly distributed across all regions.

B2:Output
Chi-Square: 51.185367617210886
- The Chi-Square statistic measures the difference between the observed and expected frequencies in the contingency table. In this case, a value of 51.185367617210886 indicates that there is some difference between the observed and expected churn rates across different states, but this value alone does not tell us if the difference is statistically significant.

P-value: 0.9999969755238908
- The p-value is used to determine the significance of the results. It tells us the probability that the observed differences in churn rates across states could have occurred by random chance alone. A common threshold for significance is 0.05. If the p-value is less than 0.05, we reject the null hypothesis, meaning the differences are statistically

significant. In this case, the p-value of 0.9999969755238908 is much higher than 0.05, indicating that the observed differences in churn rates across states are not statistically significant.

B3:Justification

The Chi-Square test is ideal for analyzing the relationship between two categorical variables, such as `State` and `Churn`, because it determines whether observed differences in churn rates across various states are statistically significant or simply due to chance. This test compares the observed frequencies of churn in each state with the expected frequencies if no relationship existed between the variables. It tests the null hypothesis that churn rates are independent of the state, helping to assess whether geographic factors influence customer churn. A low p-value would suggest a significant relationship, prompting further investigation into regional factors that may contribute to higher churn rates, whereas a high p-value indicates that any differences in churn rates are not statistically significant. This insight is valuable for stakeholders, guiding them on whether to focus on geographic factors or other aspects such as pricing or customer service in their efforts to reduce churn.
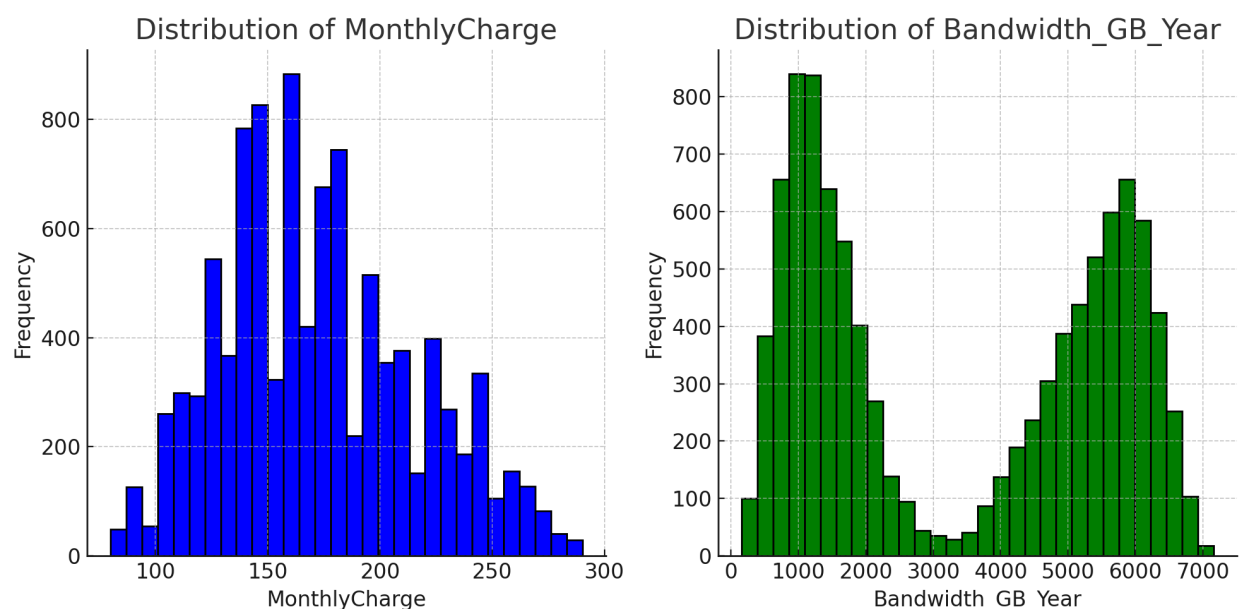
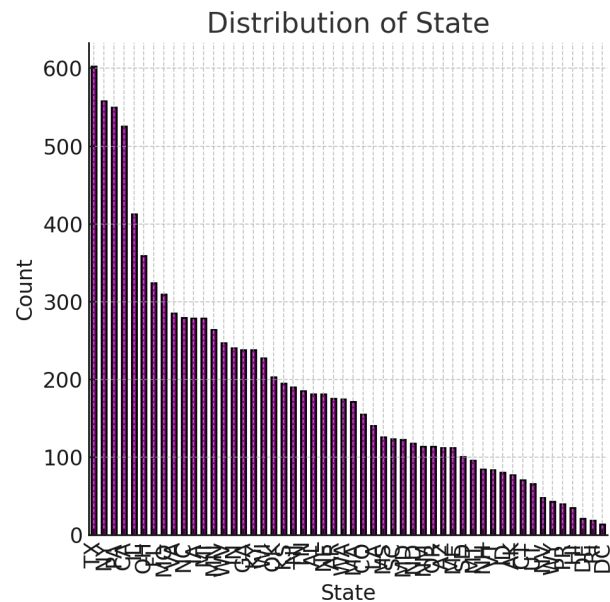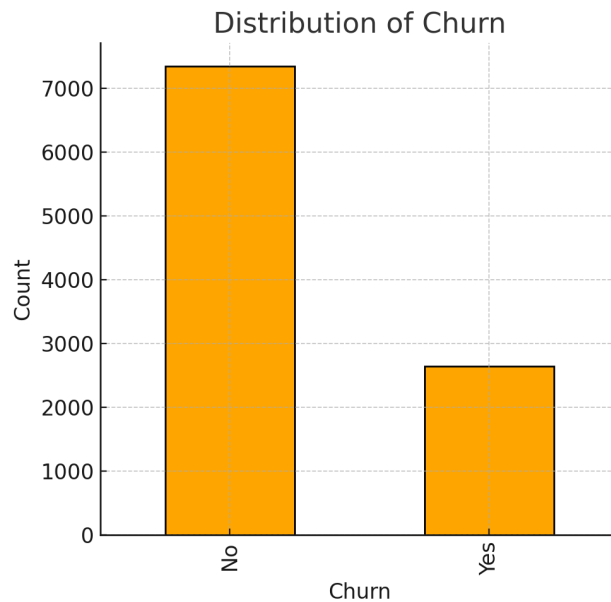## C:Univariate Statistics

Continuous Variables:
- MonthlyCharge: The monthly charge for each customer.
- Bandwidth_GB_Year: The yearly bandwidth usage in GB.

Categorical Variables:
- Churn: Whether the customer churned or not.
- State: The state where the customer is located.

C1:Visual of Findings

Distribution of Churn — Distribution of State

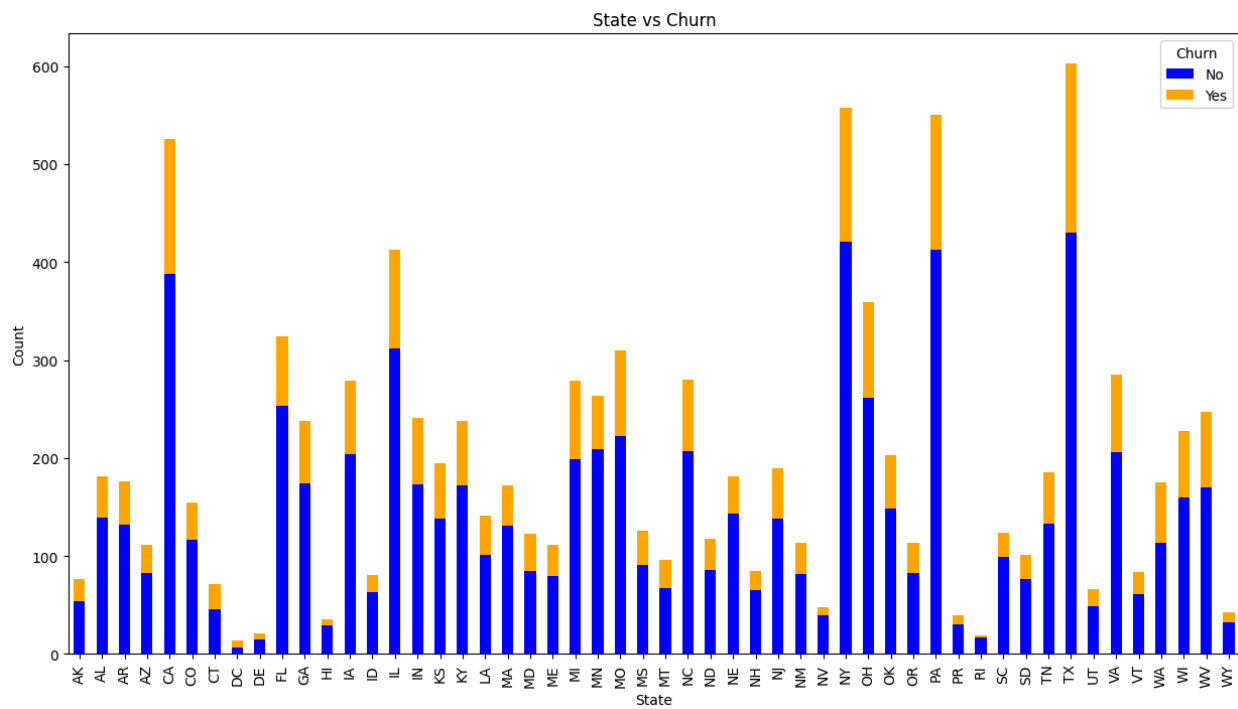**D:Bivariate Statistics**
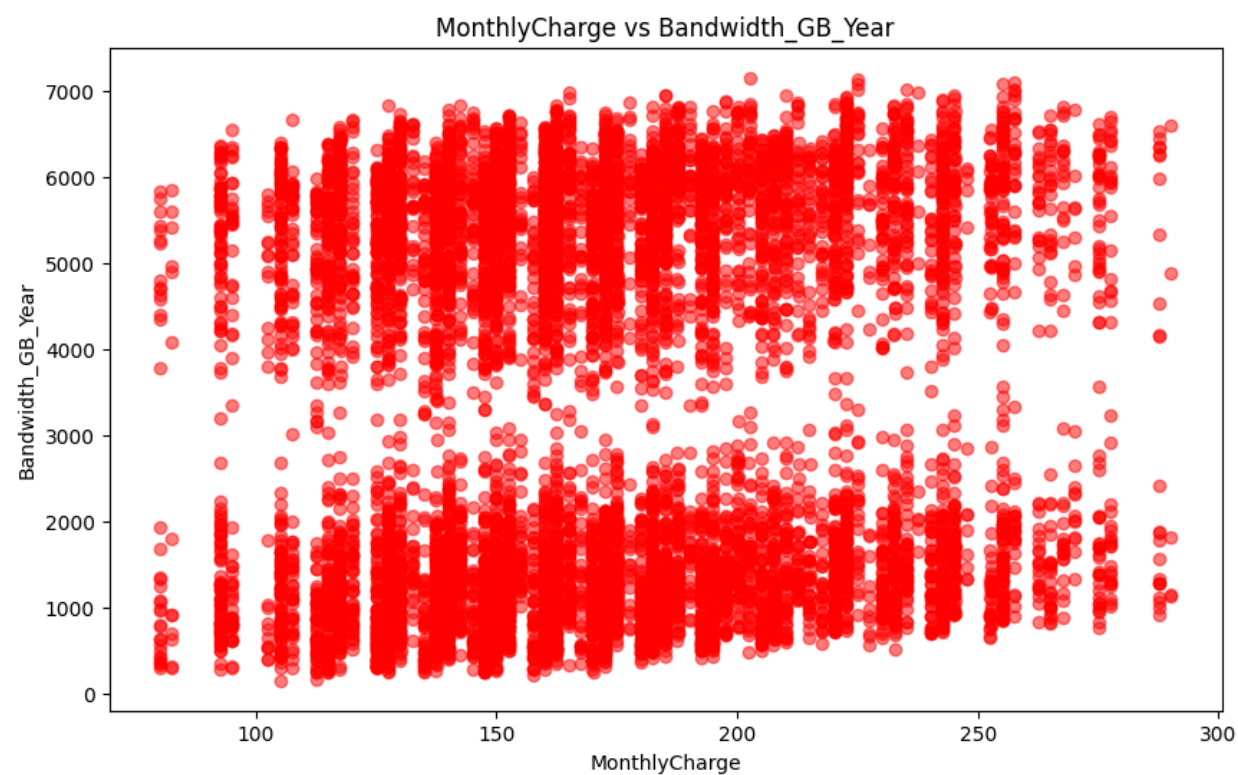
For the bivariate analysis:

Continuous Variables:

- I'll plot a scatter plot to observe the relationship between `MonthlyCharge` and `Bandwidth_GB_Year` .

Categorical Variables:

- I'll plot a stacked bar chart to observe the relationship between `State` and `Churn`.

D1:Visual of Findings

MonthlyCharge vs Bandwidth_GB_Year

State vs Churn

E1:Results of Analysis
The results of the Chi-Square test, which produced a p-value of 0.9999969755238908, strongly suggest that there is no statistically significant difference in churn rates across different states. In the context of hypothesis testing, the p-value is a critical measure that helps determine whether the observed data significantly deviates from the null hypothesis. In this scenario, the null hypothesis posits that there is no association between the state in which a customer resides and their likelihood of churning. If the p-value were less than the commonly accepted significance threshold of 0.05, we would reject the null hypothesis, indicating that state could indeed influence churn rates. However, the exceptionally high p-value in this analysis, much greater than 0.05, leads us to fail to reject the null hypothesis. This result implies that any differences in churn rates between states are most likely due to random variation rather than any real geographic effect on customer behavior. This finding is important because it suggests that the state a customer lives in does not have a significant impact on whether they will churn or not. In other words, geographic location does not appear to be a key factor driving customer decisions to stay with or leave the service. Therefore, efforts to reduce churn should not be geographically targeted but should instead focus on other potential factors, such as customer satisfaction, service quality, or pricing, that might be more relevant across the entire customer base. This analysis helps clarify that location-based strategies may not be effective in addressing churn, and it encourages a broader approach to understanding and mitigating the reasons behind customer attrition.

E2:Limitations of Analysis
The data analysis has several limitations that need to be considered when interpreting the results. First, the dataset may include other confounding variables that could influence customer churn, such as customer service interactions, pricing plans, or customer satisfaction levels, which were not accounted for in this analysis. These factors could potentially have a significant impact on churn rates, but their effects remain unexamined. Furthermore, while the Chi-Square test is useful for determining whether there is a significant association between two categorical variables, it does not provide information about the strength or direction of that relationship. In other words, the test can tell us whether churn rates differ across states, but it cannot explain how much or in what way they differ. This limitation means that even if a significant relationship were found, further analysis would be needed to understand the nature of that relationship. Additionally, the Chi-Square test assumes that the data meets certain conditions, such as expected frequencies being sufficiently large, and any violations of these assumptions could affect the validity of the results. Overall, while the Chi-Square test is a valuable tool, it should be complemented with other methods and analyses to provide a more comprehensive understanding of the factors influencing customer churn.

E3:Recommended Course of Action
Despite the analysis showing no significant difference in churn rates across states, stakeholders should not overlook the possibility that other factors may be driving customer churn. Geographic location alone may not explain why customers are leaving, but variables such as customer satisfaction, service quality, pricing structures, or competition could be key contributors.

Therefore, it is crucial for stakeholders to conduct more targeted analyses that consider these potential influences. For example, gathering and analyzing customer feedback or service interaction data could uncover underlying issues that lead to dissatisfaction and churn. Additionally, examining how different pricing plans or service offerings correlate with churn could help identify areas for improvement. By focusing on these factors, stakeholders can develop more informed strategies to reduce churn, such as enhancing customer support, optimizing pricing, or improving service reliability. Ultimately, a more comprehensive and nuanced approach to analyzing churn will provide actionable insights that can drive more effective customer retention efforts.

G:Sources for Third-Party Code
None

H:Sources

"Matplotlib.Pyplot#." *Matplotlib.Pyplot - Matplotlib 3.5.3 Documentation*, matplotlib.org/3.5.3/api/_as_gen/matplotlib.pyplot.html. Accessed 17 Aug. 2024.