

BY GROUP ONE: MADUABUGHICHI ACHILEFU, OKEKE ONYEDIKACHUKWU, BABATUNDE ODUMUYIWA

1. Introduction

The dataset under examination, "**churn_data.csv**," encompasses a range of customer-related metrics from a telecommunication company. These metrics include **Churn status, account weeks, contract renewal, data plan subscription, data usage, customer service calls, daily usage metrics, monthly charges, overage fees, and roaming minutes**. The primary objective of this analysis is to uncover insights into factors that may influence customer churn using inferential statistics.

2. Questions

The analysis focuses on two key questions:

- Is the average monthly charge for customers who churn higher than for those who do not churn?
- Are there significant differences in the number of customer service calls between customers with and without data plans?

3. Hypotheses

For each question, the following hypotheses are formulated:

Question 1 Hypotheses:

- **Null Hypothesis (H_0):** The mean monthly charge for customers who churn is equal to or less than for those who do not churn.
- **Alternative Hypothesis (H_1):** The mean monthly charge for customers who churn is higher than for those who do not churn.

Question 2 Hypotheses:

- **Null Hypothesis (H_0):** The mean number of customer service calls for customers with data plans is the same as for those without.
- **Alternative Hypothesis (H_1):** The mean number of customer service calls for customers with data plans is different from those without.

4. Methodology

The following statistical tests were employed:

For Question 1, a one-sample t-test was used to compare the mean monthly charge of customers who churned against the overall mean of the dataset at different confidence levels (90%, 95%, and 99%).

For Question 2, a two-sample t-test was applied to compare the mean number of customer service calls between customers with and without data plans at different confidence levels (90%, 95%, and 99%).

5. Results

Question 1:

- **At 90% Confidence Level:** T-Statistic = 5.438, P-Value = $\sim 6.87e-08$, Confidence Interval: 58.25 to 59.98
- **At 95% Confidence Level:** T-Statistic = 5.438, P-Value = $\sim 6.87e-08$, Confidence Interval: 58.09 to 60.14
- **At 99% Confidence Level:** T-Statistic = 5.438, P-Value = $\sim 6.87e-08$, Confidence Interval: 57.76 to 60.47

Question 2:

- **At 90% Confidence Level:** T-Statistic = -1.068, P-Value = ~ 0.285 , Confidence Interval: -0.099 to 0.021
- **At 95% Confidence Level:** T-Statistic = -1.068, P-Value = ~ 0.285 , Confidence Interval: -0.111 to 0.033
- **At 99% Confidence Level:** T-Statistic = -1.068, P-Value = ~ 0.285 , Confidence Interval: -0.134 to 0.055

6. Interpretation

The results from **Question 1** consistently indicate, across all confidence levels, that the average monthly charge for customers who churn is significantly higher than the overall average. This suggests that higher monthly charges could be a contributing factor to customer churn. Consequently, the p-value is significantly lower than the standard alpha level of 0.05. This indicates that we reject the null hypothesis. Therefore, the average monthly charge for customers who churn is statistically significantly higher than the overall average monthly charge.

The findings from **Question 2** demonstrate, at all confidence levels, that there is no significant difference in the number of customer service calls between customers with and without data plans. This implies that the presence of a data plan does not significantly affect the frequency of customer service interactions. Consequently, the p-value is greater than 0.05, indicating that we fail to reject the null hypothesis. This suggests

there is no statistically significant difference in the number of customer service calls between customers with and without data plans.

7. Conclusion

The statistical analysis reveals two key insights: This statistical analysis provides critical insights into key factors influencing customer churn.

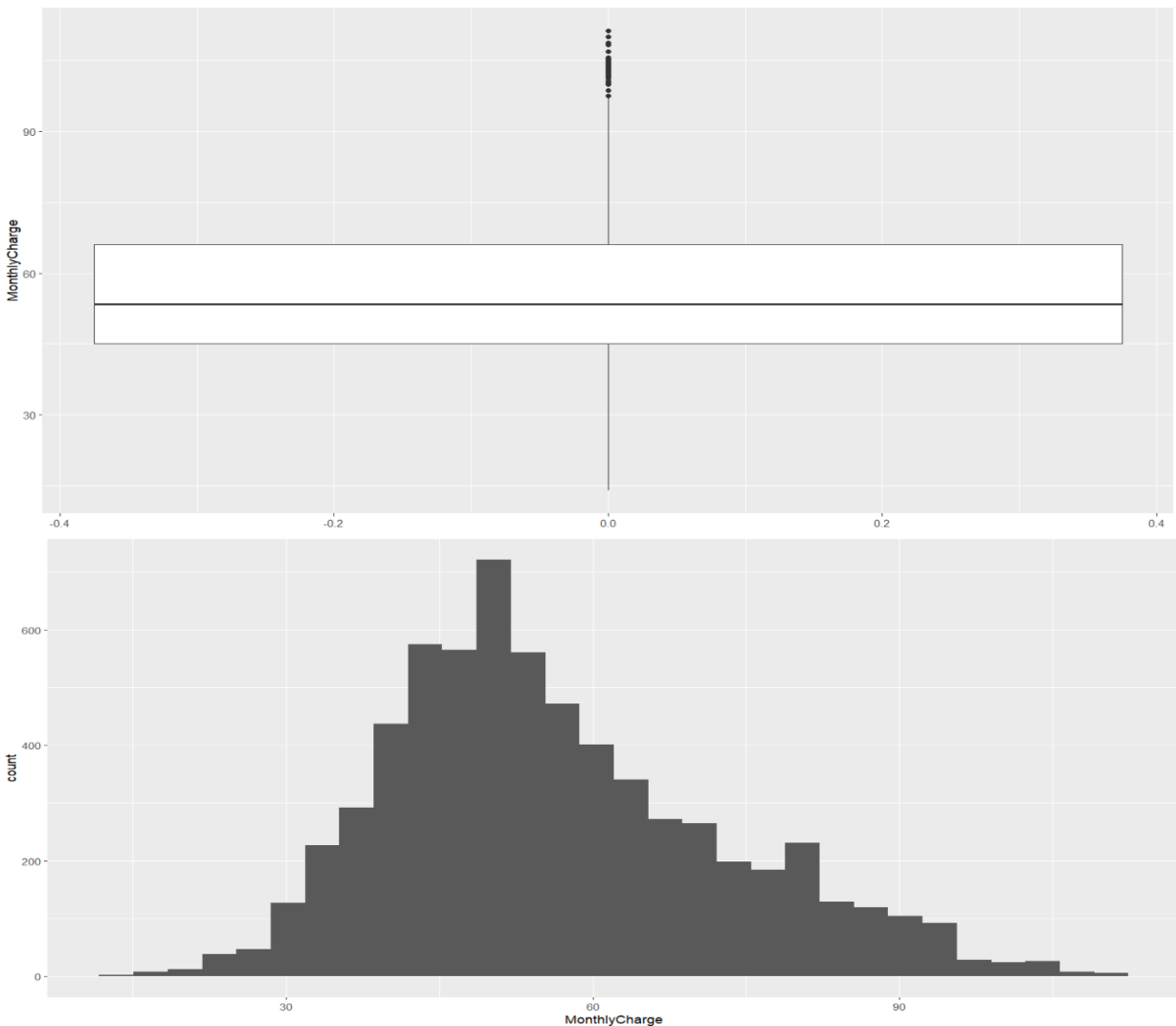
Monthly Charges and Churn: The significant difference in monthly charges for customers who churn, evident at all confidence levels, underscores the importance of pricing strategies in customer retention. For example, customers who churn tend to have higher monthly charges than the general customer base. This finding suggests that pricing strategies might be a critical factor influencing customer churn.

Customer Service Calls and Data Plans: On the other hand, the lack of a significant difference in customer service calls, regardless of data plan status, suggests that factors other than service issues might play a more influential role in customer churn. For example, there is no significant difference in the frequency of customer service calls between customers with and without data plans. This implies that having a data plan does not significantly affect the likelihood of contacting customer service.

These insights are vital for guiding strategic decisions in customer service management and pricing strategies, aiming to enhance customer retention and identify areas for service improvement. For instance, these insights can be instrumental for the telecommunication company in formulating strategies to reduce customer churn and improve overall customer satisfaction. For instance, revisiting pricing structures or enhancing service features for high-charge customers could be potential areas of focus.

ADDITIONAL INFORMATION BASED ON HISTOGRAM AND A BOXPLOT DISTRIBUTION

HISTOGRAM AND A BOXPLOT OF THE MONTHLYCHARGE VARIABLE FROM THE CHURN DATA



The summary statistics provide a more comprehensive view of the **MonthlyCharge** distribution:

- The minimum monthly charge is 14.00, suggesting that the lowest amount a customer is charged per month is 14.
- The first quartile (25th percentile) is 45.00, which means that 25% of the customers pay less than 45 per month.
- The median monthly charge is 53.40, indicating that half of the customers pay less than this amount, and the other half pay more.
- The mean (average) monthly charge is slightly higher than the median at 56.27, implying a right-skewed distribution where higher charges pull the average up.

- The third quartile (75th percentile) is 66.00, showing that 75% of customers pay less than 66 per month.
- The maximum monthly charge is 111.30, which is the highest amount paid by any customer in a month.

Histogram and a boxplot of the MonthlyCharge variable from the churn data

The histogram indicates the distribution of the **MonthlyCharge** across the dataset. It shows how often each range of charges occurs. This particular histogram suggests a unimodal distribution with a peak around the median value. The right skew indicates that there are some customers with very high monthly charges compared to the rest.

The boxplot provides a summary of the distribution of **MonthlyCharge** using five-number summary statistics: **minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum**. It also shows outliers as individual points beyond the whiskers. The boxplot for **MonthlyCharge** suggests that the median is lower than the mean, which is typical for a right-skewed distribution. Outliers on the upper end indicate that there are some values of **MonthlyCharge** that are much higher than the typical range.

Both visualizations help in understanding the spread and central tendency of monthly charges among customers, which can inform the analysis of factors contributing to customer churn. High monthly charges could potentially be a driver for customers to leave the service, especially if these charges are significantly higher than the typical customer's charge.