D207 Performance Assessment

Exploratory Data Analysis

Instructor: Dr. Sewell

Student Name: Nicole Haibach

Student ID: 001260374

Email: nhaibac@wgu.edu

# Describe the Issue

## Research Question

My research question for the assessment will be evaluating what led to customers discontinuing service. Looking at the possible reasons for customers to leave their current service will help the company retain customers and save money overall. Does the overall monthly charge of services affect customer retention?

## Benefits of Analysis

The benefit of analysis would be figuring out the cause of customer turnover. If the company figures out what is a big cause for customers leaving to a different telecommunication company to help reduce the problematic factors, they would overall save money. The cost of acquiring new customers being 10 times more than retaining customers shows that the analysis would help reduce expenditure at the company.

## Identify all relevant data

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Type** | **Description** | **Example** |
| Churn | Categorical | Whether the customer discontinued within the last month | Yes |
| MonthlyCharge | Continuous | The amount charged monthly | 242.63255 |
| Contract | Categorical | The contract term of the customer | One Year |
| Tenure | Continuous | Number of months the customer has stayed | 17.087227 |
| Yearly\_equip\_failure | Continuous | Number of failures or resets in the past year | 2 |
| InternetService | Categorical | Internet service provider | DSL |

# Describe Analysis

## Code for Analysis

t\_test\_results <- t.test(MonthlyCharge ~ Churn, data = CC)

## Provide Output

Welch Two Sample t-test

data: MonthlyCharge by Churn

t = -39.288, df = 4493.5, p-value < 2.2e-16

alternative hypothesis: true difference in means between group No and group Yes is not equal to 0

95 percent confidence interval:

-38.09691 -34.47549

sample estimates:

mean in group No mean in group Yes

163.0090 199.2952

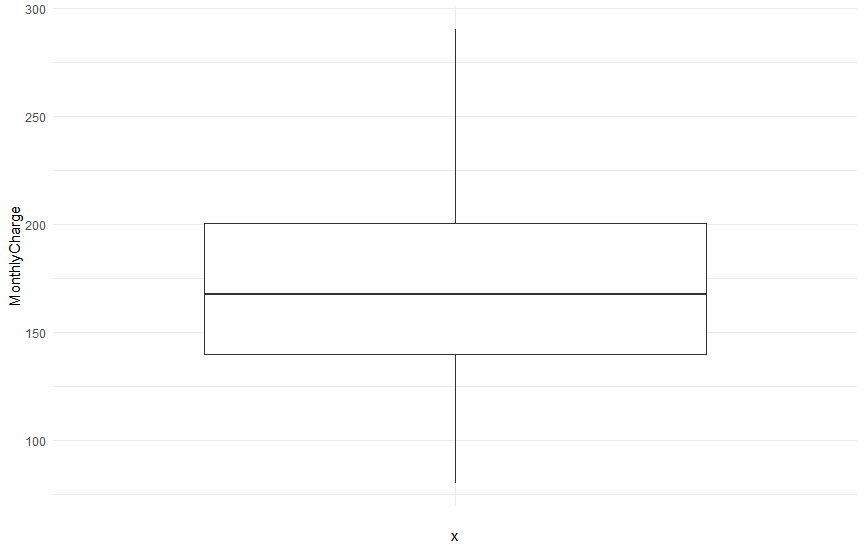
## Justify Technique

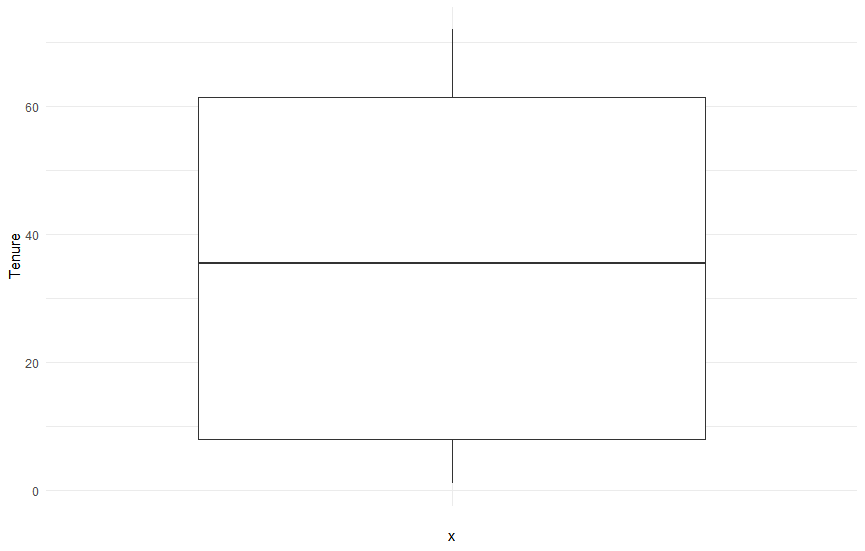
A t-test shows whether the Monthly Charge has anything to do with whether people decided to leave the service provider. “With a t test, the researcher wants to state with some degree of confidence that the obtained difference between the means of the sample groups is too great to be a chance event and that some difference also exists in the population from which the sample was drawn. If our t test produces a t-value that results in a probability of .01, we say that the likelihood of getting the difference we found by chance would be 1 in a 100 times. We could say that it is unlikely that our results occurred by chance and the difference we found in the sample probably exists in the populations from which it was drawn” (Siegle, 2015).

# Identify Distribution using Univariate statistics

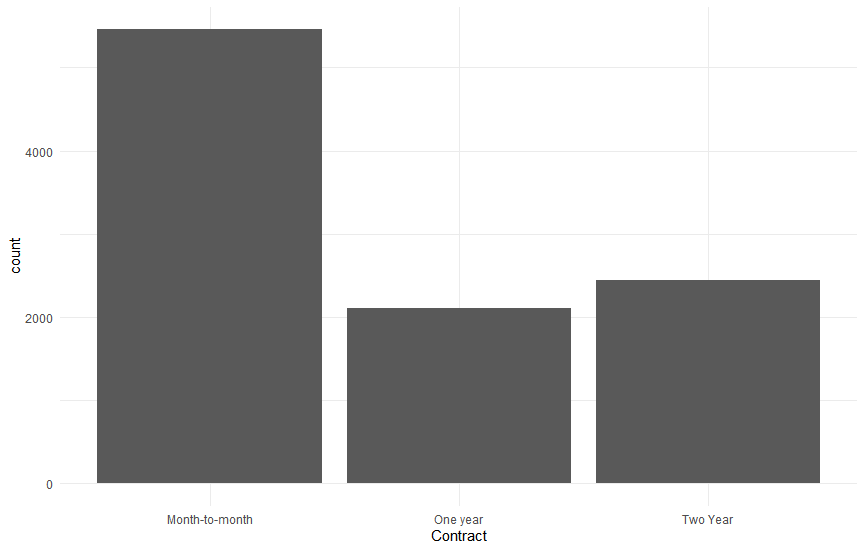
## Represent findings Visually

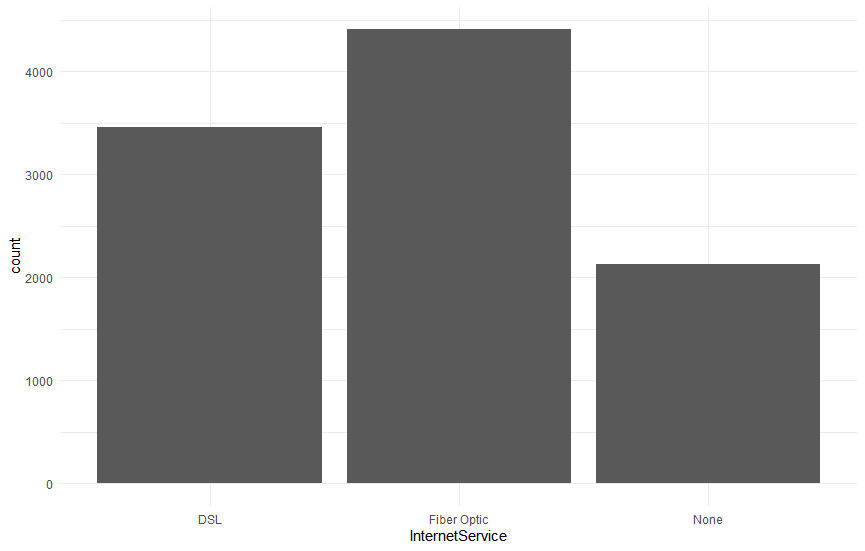
Visual representation of Continuous Variables





Visualization of Categorical Data

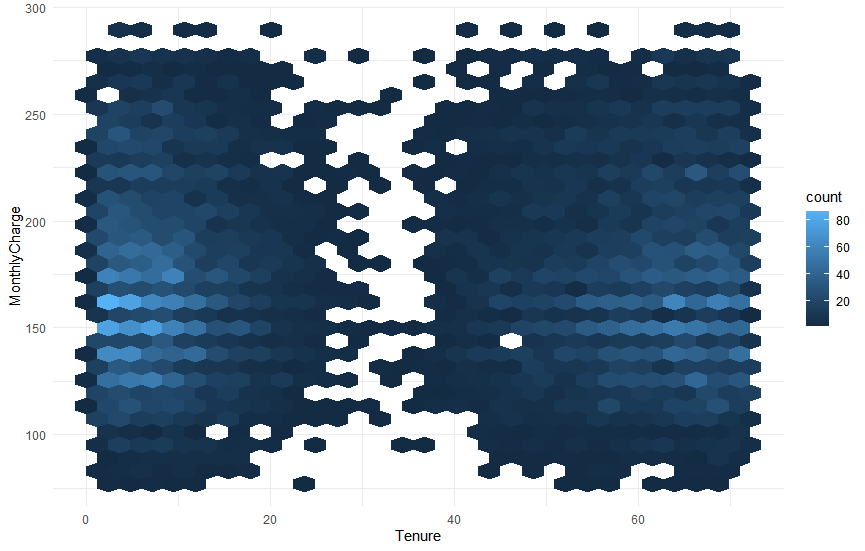




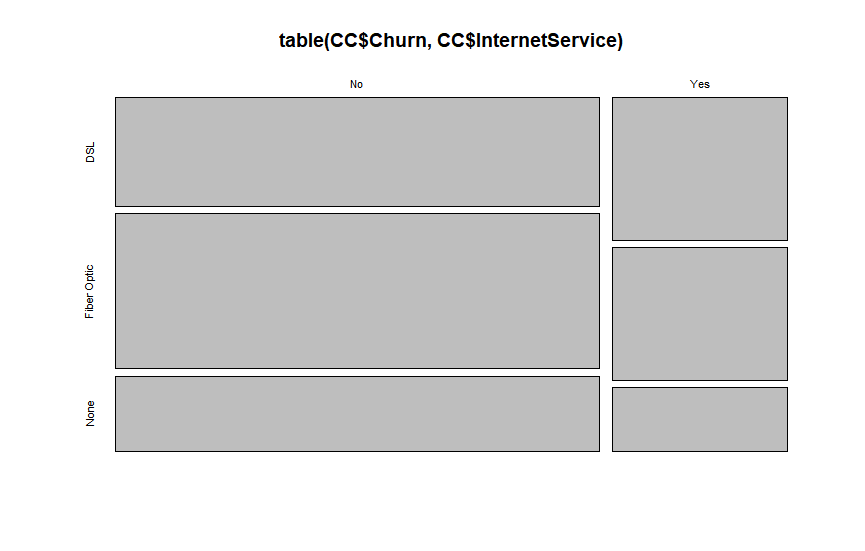
# Identify Distribution using Bivariate Statistics

## Represent findings visually

Visualization of Bivariate Continuous Data



Visualization of Bivariate Continuous Data



# Summarize Implications

## Discuss Results

The results show that the p-value equaling zero means that the observed difference is unlikely due to chance and that the lower the monthly charge the more likely the customer is to stay. The results from the hexagon graph visually show that the lower the monthly charge the more likely people are to join and the more likely they are to stay. Fiber optic is the most likely service to stay. From this information, the company should focus on trying to offer the best deals to people who have long tenure. A good group to focus on would be the ones with fiber optic service to continue growing long-term clientele.

## Discuss Limitations

The limitation of this analysis is the assumption of normal distribution. If your data has unimodal distribution or multiple peaks the t-test may be less reliable also. T-tests are also limited to continuous data, categorical or ordinal data would be better analyzed in other formats. A significant result also does not always guarantee a meaningful difference.

## Recommend a Course of action

The best course of action for the company is to start with the fiber optic customers to make sure they are happy with their plans to keep more of them on long-term. The next course of action is to make calls to the high-paying long-term customers to offer them better deals on their business to keep them around for a longer time since they are susceptible to leaving.

# Code References

*Hexbin function* . (n.d.). Www.rdocumentation.org. Retrieved January 20, 2024, from https://www.rdocumentation.org/packages/hexbin/versions/1.29.0/topics/hexbin

# References

Siegle, D. (2015, May 22). *t Test | Educational Research Basics by Del Siegle*. Uconn.edu; Educational Research Basics by Del Siegle. https://researchbasics.education.uconn.edu/t-test/