**Causes of Internet Churn**

**D207 Exploratory Data Analysis:**

**By Josue Gonzalez**

A.

 **Research Question:**  
Is there a significant association between customer churn and the type of contract they have?

 **Assessing Financial Implications for Stakeholders:**  
Stakeholders can evaluate the financial impact of outages on the organization by calculating potential revenue losses due to customer churn caused by the type of contract the customer has. This insight can help in designing better contract options to reduce churn and improve customer retention.

 **Key Variables in Dataset Analysis:**  
The primary variable of interest is "churn," a Boolean value indicating whether a customer has left (yes) or stayed (no). Important continuous numerical variables to consider include Tenure, MonthlyCharge, Bandwidth\_GB\_Year, and Outages\_per\_second. Additionally, categorical variables such as churn, Contract and internet service, and technical support, will be examined. These variables provide insight into factors contributing to internet service churn.

B

1. We will utilize a Chi-square test and use Python as our programming language.
2. Code and output of calculations-

A screenshot of a computer

Description automatically generated  
A screenshot of a computer program

Description automatically generated

3. The Chi-square test is typically employed to analyze the association between two categorical variables. In this context, we aim to examine the relationship between Contracts, and churn (categorized as "Yes" or "No").

**C.** Univariate statistics

|  |  |
| --- | --- |
| Variable | Detail |
| Categorical | Churn |
|  | Contract |

A screenshot of a computer

Description automatically generated

The second graph shows the distribution of customer churn, categorized into "Yes" and "No".

* **No**: This category has a count of approximately 7300. This indicates that around 7300 customers have not churned and are still with the company.
* **Yes**: This category has a count of approximately 2700. This indicates that around 2700 customers have churned.

The distribution shows that a majority of the customers (around 7300) have not churned, while a smaller portion (around 2700) have left the service

A screenshot of a graph

Description automatically generated

The first graph shows the distribution of three different contract types: "One year", "Month-to-month", and "Two Year".

* **One year**: This category has a count of approximately 2000. This indicates that around 2000 customers have opted for a one-year contract.
* **Month-to-month**: This is the most common contract type with a count of approximately 5500. The high count suggests that a significant portion of customers prefer the flexibility of a month-to-month contract.
* **Two Year**: This category has a count of approximately 2500. This indicates that around 2500 customers have committed to a two-year contract.

The distribution shows a clear preference for month-to-month contracts among customers, followed by two-year contracts, and the least common is the one-year contract.

A screenshot of a computer program

Description automatically generated

|  |  |
| --- | --- |
| Variable | Detail |
| Continuous | MonthlyCharge |
|  | Outage\_sec\_perweek |

Visual Presentations below for continuous variables

A screenshot of a graph

Description automatically generated

In this univariate visual presentation, we are taking a look at Monthly charge the distribution here is slightly right skewed. The histogram shows a concentration of customers around the $70 to $110 range, with a gradual decrease in frequency as charges increase

A screen shot of a graph

Description automatically generated

The distribution of Outage\_sec\_perweek shows a strong right skew, with a high frequency of low outage seconds per week. Most values are concentrated between 0 and 10 seconds, with a long tail extending towards higher outage times, indicating a few instances of significantly higher outages.

Below is a screenshot of the univariate statistics using the df.describe pyton metod.

A screenshot of a computer

Description automatically generated

D. Bivariate Statistics

A screenshot of a computer

Description automatically generated

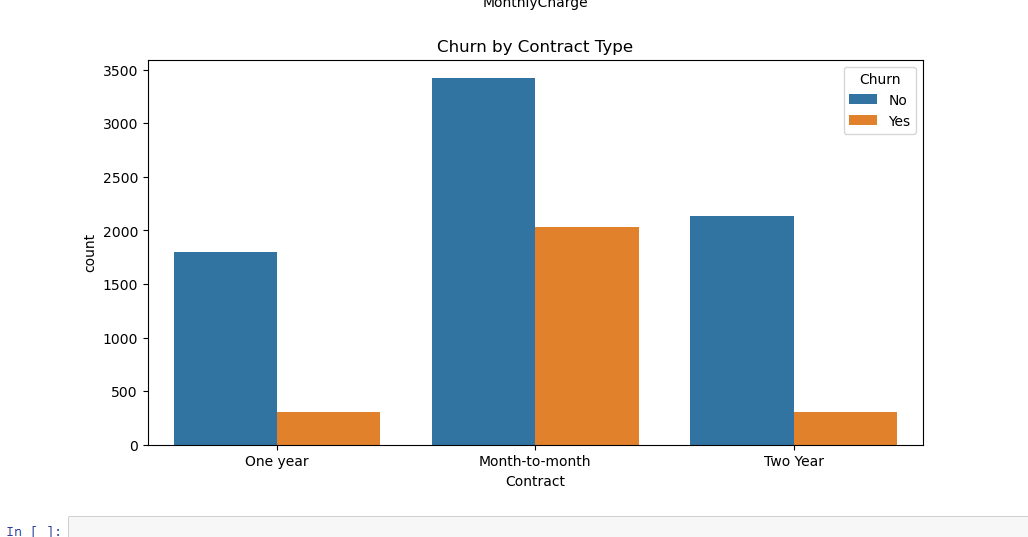
|  |  |
| --- | --- |
| Variable | Detail |
| Continuous | MonthlyCharge |
|  | Bandwidth per year |

A screen shot of a graph

Description automatically generated

the scatter plot shows the relationship between Monthly Charges and Bandwidth usage per year. The data points are widely dispersed, indicating variability in bandwidth usage across different monthly charges. There is no clear linear pattern, but there seems to be a clustering of points in certain regions, suggesting that some bandwidth usages are more common at specific charge levels.

|  |  |
| --- | --- |
| Variable | Detail |
| Catergorical | Churn |
|  | Contract |



The count plot illustrates the distribution of churn within different contract types. It shows that customers with Month-to-Month contracts have a higher frequency of churn compared to those with One Year or Two Year contracts. The plot highlights the significant association between contract type and churn behavior.

F. https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=5a25e64a-2fe4-41a3-ba42-b194003ed399

### E1. Results of Analysis

The chi-square test yielded a p-value of p-value: 0.0000. with this P-value we can reject the null hypothesis at the standard significance level of α = 0.05(James,G &Witten,D & Hastie,T & Tibshirani,R,2021). Consequently, the analysis provides sufficient evidence to suggest a statistically significant relationship between the type of contract and customer churn.

### E2. Limitations of Analysis

Despite the significant p-value, several limitations remain. The dataset might still have biases or inaccuracies that could affect the results. Additionally, the analysis focuses solely on the type of contract without considering other potentially influential factors such as customer demographics or service quality. These limitations highlight the need for a more comprehensive dataset to provide a fuller understanding of the factors influencing customer churn.E3. Recommended Course of Action

Given the significant association between contract type and customer churn, it is recommended to further analyze which contract types are most prone to churn and develop strategies to address this. The company could look into offering a new tier of contract that is three to six months long with a discount % depending on how long the contract is to encourage customers to get longer leading contracts. Also, we could employe qualitative research methods, such as customer interviews and surveys (Naime), which can also offer deeper insights into customer experiences and perceptions related to their contracts and if that would be an affective method of reducing churn.

G.third party code

*with illustrations in Python*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/07/univariate-analysis-visualization-with-illustrations-in-python/

*Seaborn.kdeplot#*. seaborn.kdeplot - seaborn 0.13.2 documentation. (n.d.). https://seaborn.pydata.org/generated/seaborn.kdeplot.html

*Visualizing categorical data#*. Visualizing categorical data - seaborn 0.13.2 documentation. (n.d.). https://seaborn.pydata.org/tutorial/categorical.html

H. Sources

*11. correlation and regression: The BMJ*. The BMJ | The BMJ: leading general medical journal. Research. Education. Comment. (2020, October 28). https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression

*Customer churn prevention: 3 best practices to retain customers*. TTEC. (n.d.). https://www.ttec.com/articles/customer-churn-prevention-3-best-practices-retain-customers

Everitt, B. S., & Skrondal, A. (2011). *The Cambridge Dictionary of Statistics*. Cambridge University Press.

guest\_blog. (2023, July 21). *12 univariate data visualizations with illustrations in Python*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/07/univariate-analysis-

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in python*. Springer.