Performance Assessment

WGU |

D207

Narcisse, Laurie

2023

Contents

[A.  Describe a real-world organizational situation or issue in the Data Dictionary you chose, by doing the following: 2](#_Toc151379764)

[1.  Provide one question that is relevant to your chosen data set. You will answer this question later in the task through an analysis of the cleaned data, using one of the following techniques: chi-square, t-test, or analysis of variance (ANOVA). 2](#_Toc151379765)

[2.  Explain how stakeholders in the organization could benefit from an analysis of the data. 2](#_Toc151379766)

[3.  Identify *all* of the data in your data set that are relevant to answering your question in part A1. 2](#_Toc151379767)

[B.  Describe the data analysis by doing the following: 4](#_Toc151379768)

[1.  Using one of the following techniques, write code (in either Python or R) to run the analysis of the data set: chi-square, t-test, ANOVA 4](#_Toc151379769)

[2.  Provide the output and the results of *any* calculations from the analysis you performed. 5](#_Toc151379770)

[3.  Justify why you chose this analysis technique 5](#_Toc151379771)

[C.  Identify the distribution of two continuous variables and two categorical variables using univariate statistics from your cleaned and prepared data. 6](#_Toc151379772)

[1. Represent your findings in Part C, visually as part of your submission. 6](#_Toc151379773)

[D.  Identify the distribution of two continuous variables and two categorical variables using bivariate statistics from your cleaned and prepared data. 10](#_Toc151379774)

[1. Represent your findings in Part D, visually as part of your submission. 10](#_Toc151379775)

[E.  Summarize the implications of your data analysis by doing the following: 13](#_Toc151379776)

[1.  Discuss the results of the hypothesis test. 13](#_Toc151379777)

[2.  Discuss the limitations of your data analysis. 13](#_Toc151379778)

[3.  Recommend a course of action based on your results. 13](#_Toc151379779)

[F. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the tool(s) used. 14](#_Toc151379780)

[G. Reference the web sources used to acquire segments of third-party code to support the analysis. 14](#_Toc151379781)

# A.  Describe a real-world organizational situation or issue in the Data Dictionary you chose, by doing the following:

## 1.  Provide one question that is relevant to your chosen data set. You will answer this question later in the task through an analysis of the cleaned data, using one of the following techniques: chi-square, t-test, or analysis of variance (ANOVA).

Are there any significant differences in the churn rates based on the contract term (month-to-month, one year, two years)?

## 2.  Explain how stakeholders in the organization could benefit from an analysis of the data.

Stakeholders in the organization could benefit from this analysis by understanding the impact of different contract terms on customer churn. It can help in making informed decisions about contract offerings and potentially reduce churn rates.

## 3.  Identify *all* of the data in your data set that are relevant to answering your question in part A1.

* Churn (Dependent Variable):
  + Description: Churn is a categorical variable that indicates whether the customer discontinued service within the last month.
  + Relevance: Churn is a crucial metric for businesses, especially in the telecommunications industry. It helps assess customer satisfaction and loyalty. Understanding factors influencing churn is vital for customer retention strategies.
  + Ordinal
  + Yes, No
* Contract:
  + Description: Contract is a categorical variable that represents the contract term of the customer, with options like month-to-month, one year, and two years.
  + Relevance: The contract term can significantly impact customer behavior, including their likelihood of churning. Longer contract terms often indicate a commitment from the customer, potentially affecting their decision to switch service providers.
  + Nominal
  + "Month-to-month", "Two year"
* MonthlyCharge:
  + Description: MonthlyCharge represents the amount charged to the customer monthly.
  + Relevance: MonthlyCharge is a key financial metric and can provide insights into customer spending patterns. Higher charges may influence customer satisfaction and, consequently, churn.
  + Continuous
  + 50.5, 75.2
* Bandwidth\_GB\_Year:
  + Description: Bandwidth\_GB\_Year represents the average amount of data used in gigabytes by the customer in a year.
  + Relevance: Bandwidth usage is essential for understanding customer behavior and preferences. It can indicate the level of engagement with services and impact overall satisfaction.
  + Continuous
  + 150.8, 200.3
* Gender:
  + Description: Gender represents the customer's self-identification as male, female, or nonbinary.
  + Relevance: Gender may influence service preferences and usage patterns, impacting churn. Understanding gender-related trends can aid in targeted marketing strategies.
  + Nominal
  + “Male”, “Female”
* InternetService:
  + Description: InternetService represents the customer's internet service provider, with options like DSL, fiber optic, or None.
  + Relevance: Different internet service types may have varying levels of customer satisfaction and reliability. Analyzing this variable can provide insights into the impact of internet service on churn.
  + Nominal
  + “DSL", "None"
* Techie:
  + Description: Techie is a binary variable indicating whether the customer considers themselves technically inclined (yes, no).
  + Relevance: Technically inclined customers may have different expectations and preferences regarding technical support and service features, potentially influencing churn.
  + Ordinal
  + Yes, No
* PaperlessBilling:
  + Description: PaperlessBilling is a binary variable indicating whether the customer has paperless billing (yes, no).
  + Relevance: Billing preferences can affect customer experience. Paperless billing may be associated with modern, convenient services that impact customer satisfaction.
  + Ordinal
  + Yes, No
* PaymentMethod:
  + Description: PaymentMethod represents the customer's payment method, such as electronic check, mailed check, bank transfer, or credit card.
  + Relevance: Different payment methods may be associated with different customer demographics and behaviors, potentially influencing churn.
  + Nominal
  + "Bank Transfer", "Credit Card"

# B.  Describe the data analysis by doing the following:

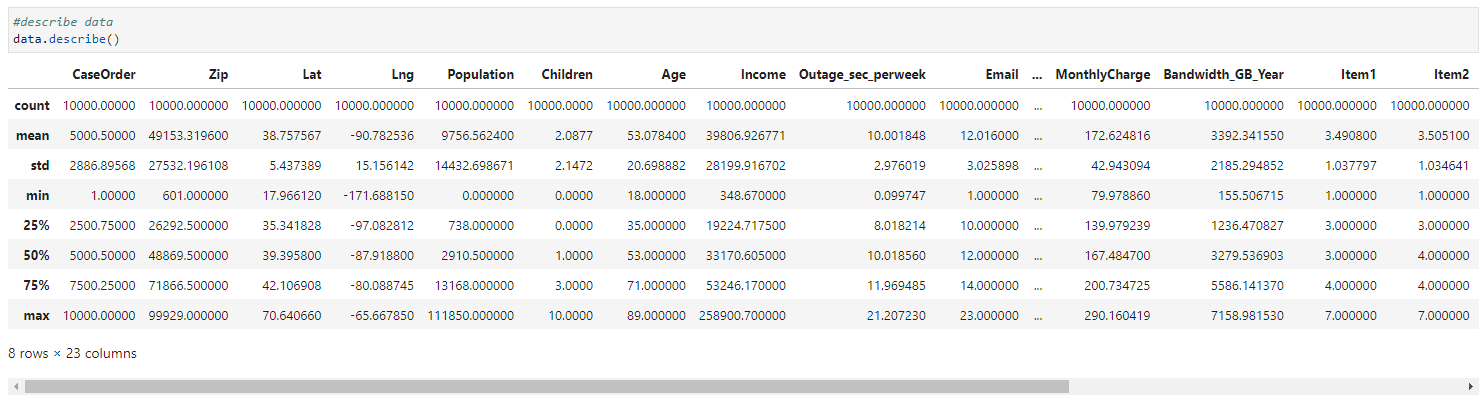
## 1.  Using one of the following techniques, write code (in either Python or R) to run the analysis of the data set: chi-square, t-test, ANOVA

I will use the chi-square test to analyze the relationship between the categorical variables.

A screenshot of a computer program

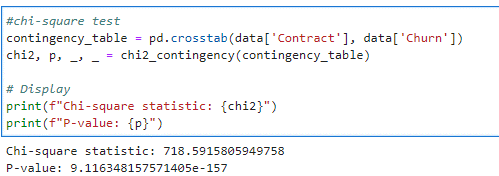
Description automatically generated

## 2.  Provide the output and the results of *any* calculations from the analysis you performed.



A screenshot of a computer

Description automatically generated



3.  Justify why you chose this analysis technique**.**

Chi-square test is suitable for examining the association between two categorical variables, making it appropriate for analyzing the relationship between churn and contract term.

# C.  Identify the distribution of two continuous variables and two categorical variables using univariate statistics from your cleaned and prepared data.

## 1. Represent your findings in Part C, visually as part of your submission.

## 

|  |  |
| --- | --- |
| **Continuous Variables** | **Categorical Variables** |
| **MonthlyCharge** | **Gender** |
| **Bandwidth\_GB\_Year** | **InternetService** |

**Continuous Variables Charts:**

**A screenshot of a computer program

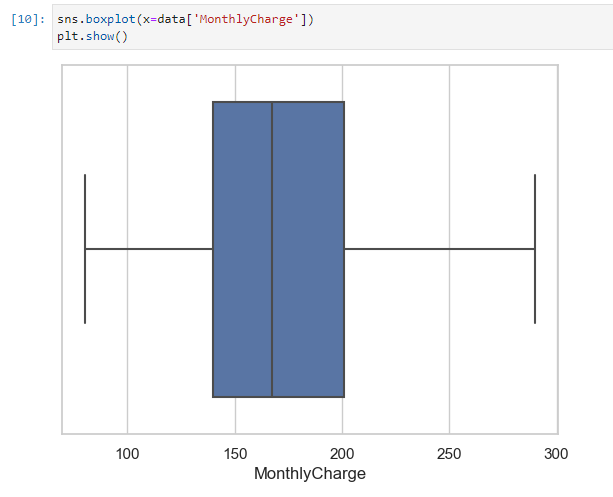
Description automatically generated**

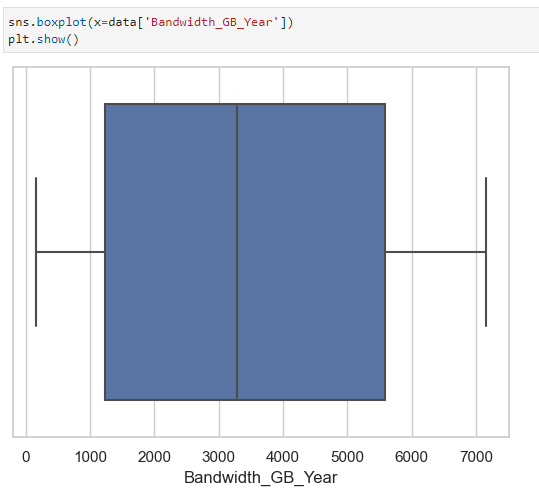
A graph showing a distribution of charge

Description automatically generated

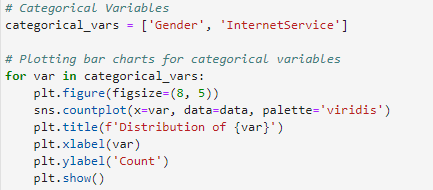
A graph showing a distribution of bandwidth

Description automatically generated

****

****

**Categorical Variables Charts:**

****

A graph of a distribution of gender

Description automatically generated

A graph showing a distribution of internet service

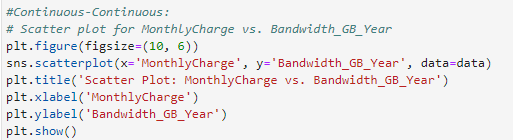
Description automatically generated

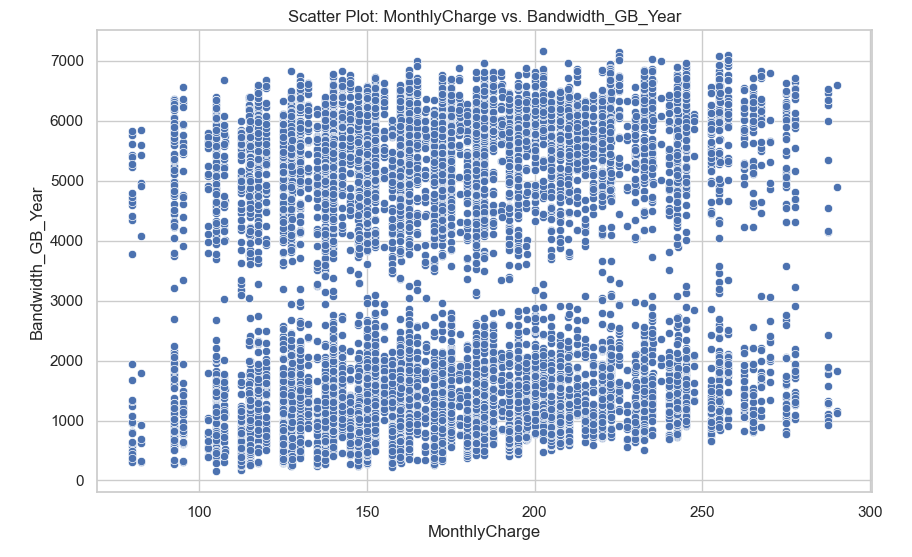
# D.  Identify the distribution of two continuous variables and two categorical variables using bivariate statistics from your cleaned and prepared data.

## 1. Represent your findings in Part D, visually as part of your submission.

## 

**Continuous-Continuous:**

****

****

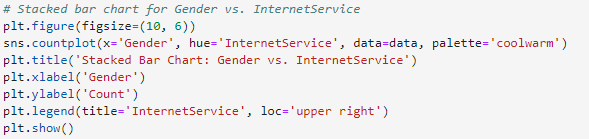
**MonthlyCharge vs. Bandwidth\_GB\_Year (Scatter plot)**

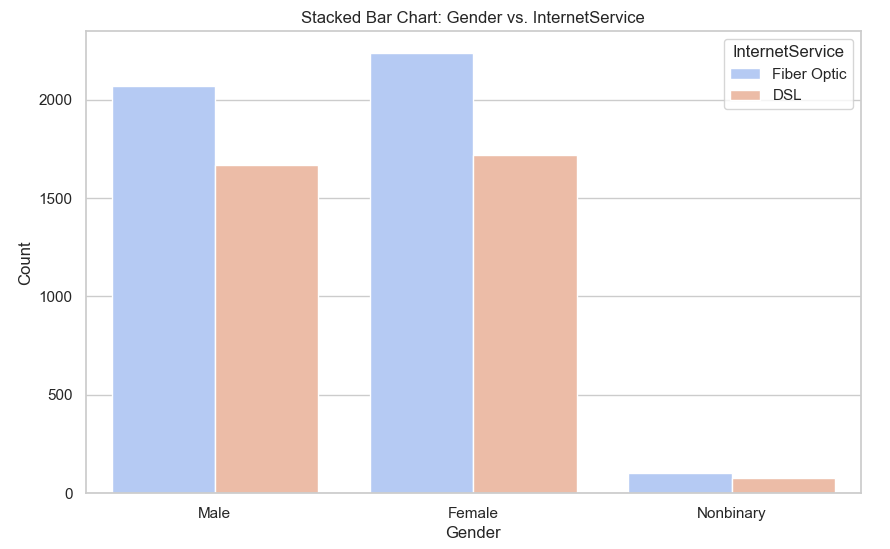
**A screenshot of a computer

Description automatically generated**

**MonthlyCharge vs. Bandwidth\_GB\_Year (Heat Map)**

**Categorical-Categorical:**

****

****

**Gender vs. InternetService (Stacked bar chart)**

# E.  Summarize the implications of your data analysis by doing the following:

## 1.  Discuss the results of the hypothesis test.

The chi-square test of independence conducted on the dataset revealed a statistically significant association between the contract term and churn, χ²(2) = 718.59, p < .001. The chi-square statistic of 718.59 indicates a strong relationship between these variables. The low p-value provides evidence to reject the null hypothesis, suggesting that the likelihood of churn varies significantly across different contract terms. Customers on distinct contract durations exhibit different churn patterns.

## 2.  Discuss the limitations of your data analysis.

While the chi-square test provides valuable insights, it is essential to acknowledge the limitations of this analysis. Firstly, correlation does not imply causation, and external factors not considered in the dataset may influence churn. Additionally, the dataset's cross-sectional nature limits our ability to establish temporal causality. Furthermore, the analysis does not explore potential interaction effects with other variables, which could contribute to a more nuanced understanding of churn dynamics.

## 3.  Recommend a course of action based on your results.

Considering the significant association between contract term and churn, it is recommended that the organization considers tailoring retention strategies based on contract duration. Promoting longer-term contracts, providing exclusive incentives, or customizing communication strategies for specific contract segments may mitigate churn. Moreover, customer engagement initiatives, such as personalized offers and targeted communications, could enhance satisfaction and loyalty, potentially influencing customers on shorter contracts to extend their commitment.

This strategic approach should be complemented by ongoing monitoring and further analysis to adapt strategies based on evolving customer preferences and market dynamics. Additionally, exploring additional factors contributing to churn, such as customer satisfaction surveys or in-depth customer interviews, could provide a more comprehensive understanding and guide more precise retention strategies.

# F. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the tool(s) used.

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c753a305-2c5a-4181-bafe-b0c701023c64>

# G. Reference the web sources used to acquire segments of third-party code to support the analysis.

GeeksforGeeks. (2023, January 12). Univariate, bivariate and Multivariate Data and its analysis. GeeksforGeeks. <https://www.geeksforgeeks.org/univariate-bivariate-and-multivariate-data-and-its-analysis/>

8. the chi squared tests: The BMJ. The BMJ | The BMJ: leading general medical journal. Research. Education. Comment. (2021, April 12). <https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests>

Chi-Square. Python for Data Science. (2018, April 16). <https://pythonfordatascienceorg.wordpress.com/chi-square-python/>

Kelly, T. (n.d.). Univariate Analysis in Python. Statology., from <https://www.statology.org/univariate-analysis-in-python/>

Vehabović, M. (n.d.). Univariate Statistics in Python. Medium., from <https://medium.com/@mehraela/univariate-statistics-in-python-84a293fc3832>

Author not specified. (n.d.). Univariate, Bivariate, and Multivariate Data Analysis in Python. Medium., from <https://medium.com/mlearning-ai/univariate-bivariate-and-multivariate-data-analysis-in-python-341493c3d173>

Kelly, T. (n.d.). Bivariate Analysis in Python. Statology., from <https://www.statology.org/bivariate-analysis-in-python/>

Analytics Vidhya. (2022, February). A Quick Guide to Bivariate Analysis in Python., from <https://www.analyticsvidhya.com/blog/2022/02/a-quick-guide-to-bivariate-analysis-in-python/>