Western Governors University
School of Technology, College of IT
Master of Science, Data Analytics

D207 Performance Assessment

which factors are the highest contributors to customer churn?

Tyson Biegler
Student ID: 012170282
D207 - Exploratory Data Analysis
July 11, 2024

INTRODUCTION

An organization has collected raw data to analyze its operations and support its decision-making processes. The raw data files from the Data Cleaning course have been cleaned and prepared for exploratory analysis.

To compete this assessment, you will access a data dictionary file and its corresponding data flat file. You will analyze the data set using specific techniques, create visualizations, and deliver the results of your analysis.

Note: This assessment requires you to submit pictures, graphics, and/or diagrams. Each file must be a supporting document no larger than 30 MB in size. Diagrams must be original and may be hand-drawn or drawn using a graphics program. Do not use CAD programs because the file will be too large.

SCENARIO

For this task, you will select one of the Data Dictionary and the matching Data Set files from the following link:

[D207 Definition and Data files](#)

You will review the Data Dictionary to understand the needs of the company and to prepare to analyze the data. In this assessment, you will analyze the .csv data file, also referred to as the data set.

You do not need to choose the same data file that you selected for the Data Cleaning course, but it is recommended that you do so.

REQUIREMENTS

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).

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

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

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

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

3. Justify why you chose this analysis technique.

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.

Note: To draw a graph or visualization, you may use one or a combination of the following:

- A spreadsheet program, such as Excel (*.xls)

- A graphics program, such as Paint (*.jpeg, *.gif)

- A word-processing program, such as Word (*.rtf)

- A scanned hand-drawn graph (*.jpeg, *.gif)

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.

Note: To draw a graph or visualization, you may use one or a combination of the following:

- A spreadsheet program, such as Excel (*.xls)

- A graphics program, such as Paint (*.jpeg, *.gif)

- A word-processing program, such as Word (*.rtf)

- A scanned hand-drawn graph (*.jpeg, *.gif)

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

    1. Discuss the results of the hypothesis test.

    2. Discuss the limitations of your data analysis.

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

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.

Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access," and then choose to log in using the "WGU" option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto's website.

To submit your recording, upload it to the Panopto drop box titled "Exploratory Data Analysis – OEM2 \ D207." Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.

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

H. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

I. Demonstrate professional communication in the content and presentation of your submission.