D207_Performance_Assessment

May 18, 2021

1 Performance Assessment | D207 Exploratory Data Analysis Ryan L. Buchanan Student ID: 001826691 Masters Data Analytics (12/01/2020) Program Mentor: Dan Estes (385) 432-9281 (MST) rbuch49@wgu.edu

1.0.1 A. Describe a real-world organizational situation or issue in the Data Dictionary

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.

1.0.2 A1. Question for Analysis:

Can we determine which individual customers are at high risk of churn? And, can we determine which features are most significant to churn?

1.0.3 A2. Benefit from Analysis:

1.0.4 A3. Data Identification:

1.0.5 B. Describe the data analysis

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.

1.0.6 B. Data Analysis:

1. 2. 3.

1.0.7 B1. Code:

1.0.8 B2. Output:

1.0.9 B3. Justification:

1.0.10 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)

1.0.11 C. Univariate Statistics:

1.0.12 C1. Visual of Findings:

1.0.13 Standard imports

[1]:
```
import numpy as np
import pandas as pd
from pandas import DataFrame

import matplotlib.pyplot as plt
%matplotlib inline

import scipy.stats
from sklearn.impute import SimpleImputer
```

1.0.14 Increase Jupyter display cell-width

[2]:
```
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:75% !important; }</style>"))
```

```
<IPython.core.display.HTML object>
```

1.0.15 Load data set into Pandas dataframe

[2]:
```

```

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

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)

1.0.17 D. Bivariate Sttistics

1.0.18 D1. Visual of Findings:

1.0.19 E. Summarize the implications of your data analysis

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.

1.0.20 E1. Results of Analysis

1.0.21 E2. Limitations of Analysis:

1.0.22 E3. Recommended Course of Action:

[2]:

1.0.23 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.

1.0.24 F. Video

Link here

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

1.0.26 G. Sources for Third-Party Code

Format: Author, A. A. (Year of publication). Title of work: Capital letter also for subtitle. Publisher Name.

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

1.0.28 H. Sources

Format: Lastname, F. M. (Year, Month Date). Title of page. Site name. URL

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

1.0.30 I. Professional Communication

[ ]:
```
!wget -nc https://raw.githubusercontent.com/brpy/colab-pdf/master/colab_pdf.py
from colab_pdf import colab_pdf
colab_pdf('D207_Performance_Assessment.ipynb')
```

```
--2021-05-18 23:55:55--  https://raw.githubusercontent.com/brpy/colab-
pdf/master/colab_pdf.py
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
185.199.109.133, 185.199.111.133, 185.199.110.133, ...
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.109.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1865 (1.8K) [text/plain]
Saving to: colab_pdf.py

colab_pdf.py        100%[===================>]   1.82K  --.-KB/s    in 0s

2021-05-18 23:55:55 (39.4 MB/s) - colab_pdf.py saved [1865/1865]

Mounted at /content/drive/

WARNING: apt does not have a stable CLI interface. Use with caution in scripts.


WARNING: apt does not have a stable CLI interface. Use with caution in scripts.

Extracting templates from packages: 100%
```