

OEM2 – OEM2 TASK 1: EDA: EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS – D207

PRFA – OEM2

TASK OVERVIEW

SUBMISSIONS

EVALUATION REPORT

COMPETENCIES

4030.04.01 : Data Interpretation

The graduate interprets central tendency, correlations, and variations to inform organizational decisions.

4030.04.02 : Conducting Parametric Testing

The graduate conducts parametric hypothesis testing.

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

Your submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from sources, even if cited correctly. The originality report that is provided when you submit your task can be used as a guide.

You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect title may contain hyperlinks to relevant portions of the course.

Tasks may **not** be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).

- 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.

File Restrictions

File name may contain only letters, numbers, spaces, and these symbols: ! - _ . * ' ()

File size limit: 400 MB

File types allowed: doc, docx, rtf, xls, xlsx, ppt, pptx, odt, pdf, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

RUBRIC

A1:QUESTION FOR ANALYSIS

NOT EVIDENT

A submission does not provide 1 question to be answered.

APPROACHING COMPETENCE

The submission provides 1 question, but the question is not specific, is not relevant to the chosen data set, or is not addressable through one of the data analysis techniques listed.

COMPETENT

The submission provides a specific question to be addressed through the analysis of the data set using one of the listed analysis techniques. The question is relevant to the chosen data set.

A2:BENEFIT FROM ANALYSIS

NOT EVIDENT

An explanation is not provided.

APPROACHING COMPETENCE

The explanation is not specific to stakeholders in the organization or does not explain how they could benefit from the data analysis. Or the explanation includes incorrect information.

COMPETENT

The explanation correctly addresses how stakeholders in the organization could benefit from the data analysis.

A3:DATA IDENTIFICATION

NOT EVIDENT

The submission does not identify *any* specific data.

APPROACHING COMPETENCE

The submission identifies specific data within the data set, but some or all of the data identified are irrelevant to addressing the question from part A1.

COMPETENT

The submission correctly identifies the specific data within the data set that are relevant to addressing the question from part A1.

B1:CODE

NOT EVIDENT

The submission does not provide *any* code.

APPROACHING COMPETENCE

The submission includes code that has warnings or errors or does not accurately use 1 of the given techniques to analyze the data.

COMPETENT

The submission includes a warning- and error-free code to accurately analyze the data set using 1 of the given techniques.

B2:OUTPUT

NOT EVIDENT

The submission does not include *any* output from running the code or results of calculations.

APPROACHING COMPETENCE

The submission includes either the output from running the code or the results of the calculations but not both. Or the submission includes only *some* of the results of the calculations.

COMPETENT

The submission includes the output from running the code and the results of *all* calculations performed.

B3:JUSTIFICATION

NOT EVIDENT

A justification is not provided.

APPROACHING COMPETENCE

The justification does not address why the chosen method of analysis was selected. Or the justified technique is not one of the three listed in part B1. Or the chosen technique is insufficient or inappropriate for the chosen data set or does not address the question in part A1.

COMPETENT

The justification addresses why the chosen method of analysis was selected. The justified technique is one of the three listed in part B1. The technique is sufficient and appropriate for the chosen data set and addresses the question in part A1.

C:UNIVARIATE STATISTICS

NOT EVIDENT

The submission does not identify the distribution of variables using univariate statistics.

APPROACHING COMPETENCE

The submission identifies the distribution of variables using univariate statistics but does not cover 2 continuous and 2 categorical variables.

COMPETENT

The submission accurately identifies the distribution of 2 continuous and 2 categorical variables using univariate statistics.

C1:VISUAL OF FINDINGS

NOT EVIDENT

A visual representation is not provided.

APPROACHING COMPETENCE

The submission accurately represents the distribution of the 2 continuous variables or the 2 categori-

COMPETENT

The submission accurately represents the distribution of the variables in part C.

cal variables, but not all 4 variables, in part C. Or the visual representation contains inaccuracies.

D:BIVARIATE STATISTICS

NOT EVIDENT

The submission does not identify the distribution of variables using bivariate statistics.

APPROACHING COMPETENCE

The submission identifies the distribution of variables using bivariate statistics but does not cover 2 continuous and 2 categorical variables.

COMPETENT

The submission accurately identifies the distribution of 2 continuous and 2 categorical variables using bivariate statistics.

D1:VISUAL OF FINDINGS

NOT EVIDENT

A visual representation is not provided.

APPROACHING COMPETENCE

The submission accurately represents the distribution of the 2 continuous variables or the 2 categorical variables, but not all 4 variables, in part D. Or the visual representation contains inaccuracies.

COMPETENT

The submission accurately represents the distribution of the variables in part D.

E1:RESULTS OF ANALYSIS

NOT EVIDENT

A discussion of results of the hypothesis test is not provided.

APPROACHING COMPETENCE

The discussion includes the results from the hypothesis test, but they are inaccurate or incomplete.

COMPETENT

The discussion includes accurate and complete results of the hypothesis test.

E2:LIMITATIONS OF ANALYSIS

NOT EVIDENT

An explanation of limitations is not provided.

APPROACHING COMPETENCE

The explanation includes inaccurate limitations of the data analysis or 1 or more of the limitations provided are not applicable to the analysis.

COMPETENT

The explanation includes the limitations of the data analysis and does not include limitations that do not apply to the analysis.

E3:RECOMMENDED COURSE OF ACTION

NOT EVIDENT

A recommended course of action is not provided.

APPROACHING COMPETENCE

The recommendation includes only a response to the question from part

COMPETENT

The recommendation includes *both* a response to the question from part

A1, or the recommendation is missing specific actions that could be taken in response to the analysis. Or the recommendation is irrelevant to the situation or question or would not plausibly address the situation or question.

A1 and specific actions that could be taken in response to the analysis. The recommendation is relevant to the situation and question and would plausibly address the situation and question.

F:VIDEO

NOT EVIDENT

The submission does not provide a Panopto video.

APPROACHING COMPETENCE

The submission provides a Panopto video recording that is missing the demonstration of the functionality of the code, the summary of the tools used, or both. Or the demonstration or summary are inaccurate.

COMPETENT

The submission provides a Panopto video recording that accurately demonstrates the functionality of the code and summarizes the tools used.

G:SOURCES FOR THIRD-PARTY CODE

NOT EVIDENT

The submission does not reference *any* web sources.

APPROACHING COMPETENCE

The submission lists only *some* of the web sources used to acquire third-party code. Or the referenced web sources are not reliable.

COMPETENT

The submission lists *all* web sources used to acquire third-party code, and the referenced web sources are reliable.

H:SOURCES

NOT EVIDENT

The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.

APPROACHING COMPETENCE

The submission includes in-text citations for sources that are quoted, paraphrased, or summarized, and a reference list; however, the citations and/or reference list is incomplete or inaccurate.

COMPETENT

The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available.

I:PROFESSIONAL COMMUNICATION

NOT EVIDENT

Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or grammar. Vocabulary or tone is unprofessional or distracts from the topic.

APPROACHING COMPETENCE

Content is poorly organized, is difficult to follow, or contains errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.

COMPETENT

Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar

WEB LINKS

[Churn Data Dictionary and Data Set](#)

If you have trouble with the link, copy and paste the link directly into your web browser.

[Medical Data Dictionary and Data Set](#)

If you have trouble with the link, copy and paste the link directly into your web browser.

[Panopto Access](#)

Sign in using the "WGU" option. If prompted, log in with your WGU student portal credentials, which should forward you to Panopto's website. If you have any problems accessing Panopto, please contact Assessment Services at assessmentservices@wgu.edu. It may take up to two business days to receive your WGU Panopto recording permissions once you have begun the course.

[Panopto How-To Videos](#)