Faculty of Information Technology

FIT3161 Computer Science Project Part 1
FIT3163 Data Science Project Part 1

Data Analysis Report

(Part of Project Initial Concept and Design/Data Analysis Assignment)

Applicable to Data Science (FIT3163) students only

Introduction

In data science projects, the data analysis component is intended to provide a foundation for your project by providing clarity around your data set. It allows you to document general characteristics, patterns and trends in the data that can lead to the identification of potential problems and the formulation of appropriate solutions. It typically involves the following steps:

1. Cleaning and preparing data

> a. Clean the data of errors and inconsistencies
> b. Deal with missing values and outliers
> c. Transform the data into a format suitable for analysis and processing

2. Conducting exploratory data analysis (EDA)

> a. Explore the data to understand its distribution, the relationships between variables, and note any outliers.

3. Visualising the data to help communicate the findings of your analysis.

4. Interpreting and communicating the results to stakeholders clearly and concisely.

This assignment aims to engage your team in a preliminary analysis of the data available to you. This may reveal to you that further, more complex, analysis is needed later during project execution.

A typical sample of a Full Data Analysis Report template is shown below.

Note: For this assignment, you are not expected to produce a full or thorough report. You only need to include a brief data analysis report in your submission. You are free to use the template as a guide and adapt it. Items marked (*) are required in your submission.

1. **Executive Summary**: Provide a brief overview of the report, summarising the key findings, conclusions, and recommendations.

2. (*) **Introduction**: Introduce the purpose of the data analysis report, including the background information, problem statement, and the significance of the analysis. (An introduction to the project will be already included at the start of your Initial concept and design assignment. You do not need to repeat it here. However, a brief introduction to the data analysis part is required here.)

3. **Objective**: State the objectives and research questions addressed in the project.

4. **Methodology**: Describe the overall methodology used for data collection, analysis, and interpretation. Explain the specific techniques, tools, and software used.

5. (*) **Data Collection**: Provide details about the data sources, including their origin, and format. Discuss any limitations or biases associated with the data if obvious at this early stage of the project.

6. (*) **Data Pre-processing:** Provide a summary of the data that is available to you, including the number of records, the data types.

7. (*) **Data Analysis**: This section may be divided into subsections based on the different analyses you conduct:

    a. (*) Describe inconsistencies, missing values, limitations or biases associated with the data identified at this early stage of the project. Describe briefly how these were checked, any pre-processing steps you may have performed and any challenges encountered.

    b. (*) Descriptive Statistics: Summarise the main characteristics of the data, such as measures of central tendency, variability, and distribution. Present tables, charts, or graphs to support the analysis where possible.

    c. Inferential Statistics: Conduct statistical tests to make inferences about the population based on the sample data. Explain the rationale for selecting specific tests and report the results, including p-values and confidence intervals.

8. (*) **Results and Findings**: Summarise the main results obtained from the data analysis.

9. **Discussion**: Interpret and discuss the results in the context of the objectives. Address any unexpected or interesting insights and provide explanations for observed trends or relationships.

10. **Conclusion**: Summarise the main outcomes of the data analysis and restate the key findings and their implications.

11. **Appendix**: Include any supplementary materials that are relevant to the analysis but not essential to the main report. This may include raw data, detailed statistical tests, additional visualisations, or any other supporting documentation.

Use of Generative AI tools in Projects and Assignments
Policy for FIT316x Units (based on Monash University policy)

Faculty of Information Technology

FIT3161 Computer Science Project Part 1
FIT3163 Data Science Project Part 1

Use of Generative AI Tools is acceptable (unless explicitly forbidden in a particular assignment specification)

● The use of generative AI tools is allowed and is not penalized in marking.

● Students must **acknowledge** when generative AI is used

● Students must clearly indicate which part(s) of the assessment submission contain material where generative AI has been used.

● Students must indicate **how** generative AI was used e.g., what AI tool was used and what questions were asked.

● Students must show **critical thinking** when using generative AI responses. Any errors made by the AI will be assessed as if they were made by the students - i.e. "The AI made a mistake" is not a reason for submission of erroneous work. You will lose marks for this!

To correctly acknowledge the use of Generative AI, please see:

https://www.monash.edu/learning-teaching/TeachHQ/Teaching-practices/artificial-intelligence/policy-and-practice-guidance-around-acceptable-and-responsible-use-of-ai-technologies