



## DATA H3008 – Data Analysis

### Project Details

Dr. Abdelsalam Busalim [Abdelsalam.Busalim@TUDublin.ie](mailto:Abdelsalam.Busalim@TUDublin.ie)

September 2023

### 1. Project Overview

The project is a pivotal component of the Year 3 Data Analysis Module. This assessment is designed to provide you with collaborative and hands-on experience in tackling real-world data analysis challenges. Working in groups, you will apply the knowledge and skills acquired throughout the module to solve complex data problems and present your findings effectively. Every group should find a real dataset related to a specific domain, which could range from business and finance to healthcare, environmental science, or social issues.

You will be expected to apply all the steps included in data analysis life cycle; this may include but limited to:

- **Importing Data:** You must use Python to demonstrate how the data was imported.
- **Data Cleaning:** use various Python libraries to clean and pre-process the data.
- **Data Manipulation:** using various techniques to shape the dataset for data analysis and visualization.
- **Data Visualization:** display the data using plots and graphs.
- **Hypothesis Testing and Statistical Analysis:** Formulate hypotheses based on the data and perform relevant statistical tests. Make data-driven inferences and interpretations.
- **Advanced analysis:** this could include linear regression, multiple regression, etc.

## 2. Project Learning Outcomes:

On successful completion of the project, students should be able to:

- **Practical Application:** to be able to apply the data analysis techniques, Python programming skills, and statistical knowledge acquired during the course to a substantial and authentic dataset.
- **Collaborative Skills:** Will enhance students' ability to collaborate, communicate effectively, and distribute tasks within a group. These skills are essential in professional settings where data analysis is often a collaborative effort.
- **Problem Solving:** The project challenges students to approach a real-world problem, make data-driven decisions, and derive actionable insights, mirroring the types of tasks you may encounter in your future career.
- **Presentation Skills:** Students will gain experience in communicating analysis in a story-telling way effectively through clear and informative data visualizations and presentations.

## 3. Deliverables:

	Deliverables (Submission via Moodle)
1.	Dataset selection/ collection
2.	<b>Technical report:</b> Python notebook that contains all the code.
3.	<b>Report:</b> 2000-3500 words, explains the project steps (e.g., brief introduction to the dataset, problem that the project attempt to solve, results and findings)
4.	Presentations

The remainder of this document includes details on each assignment. Marking Guidelines are provided in the Appendix.

### 3.1 Dataset

Students are required to search for dataset. Selecting an appropriate dataset is a critical first step in any data analysis project. The choice of a dataset should align closely with the objectives of the analysis and the questions you aim to answer. Various links for dataset will be provided via Moodle

It's essential to consider factors:

- Data quality,
- Relevance,
- Size, and accessibility.
- A well-selected dataset should be large enough to yield meaningful insights yet manageable in terms of computational resources.

### 3.2 Technical report

- The technical report should contain all the steps included in the data analysis, from importing the data and preprocessing to analysis and visualization.
- The report must be in notebook format with the *ipynb* format.
- The report should be presented in a professional way, utilizing the text cell to provide description to each step in the analysis.

### 3.3 A report

- The report of the project should be a comprehensive and well-structured document that effectively communicates the entire data analysis process and its outcomes.
- It should follow a logical flow, starting with a clear introduction that outlines the project's objectives, the dataset used, and the research questions addressed.
- The report should describe the data collection and cleaning processes, detailing any challenges encountered and the strategies employed to address them.
- It should also include an exploratory data analysis section, showcasing data visualizations and initial insights.
- The report should present the statistical methods used and the corresponding results, along with clear interpretations of the findings.
- The report should include well-structured and labeled visualizations, tables, and figures to enhance the clarity of the presented information.

### 3.4 Presentation

- The presentation should be structured, with a clear introduction that sets the stage by defining the project's scope, objectives, and the dataset used.
- **Each group member MUST take their turn, presenting their part of the analysis**, whether it's data collection and cleaning, exploratory data analysis, or statistical analysis. Visualizations and key results should be highlighted, and any challenges and solutions encountered during the analysis should be discussed collaboratively.

## 4. Submission and Marking Structure:

Each group is required to submit:

1. An electronic copy (Word or PDF) of the report via Moodle by 08/12/2023 (cut-off at 6pm).
- ✓ The word count for the report must be between 2000- 3500 words excluding references.
2. A technical report in *ipynb* format. (Notebook)
  3. Presentation slides. Presentations will take place in the last week of the semester (12/12/23)

The following marking guidelines for this project will apply subject to the grade descriptors in Appendix A of this document:

Marked component	Weight (%)
Reports Structure and housekeeping	15
Dataset selection and collection	5
Data preprocessing	10
Data visualization	20
Statistical analysis	20
Advanced analysis	5
Presentation	20
Project management	5

## 5. Group Allocation:

- Each group should ideally consist of 3-4 members to ensure a balance between collaboration and manageable group dynamics.
- Students have the option to form their own groups based on their preferences, but it's essential to ensure diversity in terms of skill sets, interests, and expertise to enrich the collective learning experience.
- The contribution of every member within the group must be clearly documented in the project report, specifically under the "Project Management" section.
- This documentation should outline each member's roles and responsibilities, including aspects like data collection, data cleaning, analysis, visualization, and presentation. This ensures transparency and accountability, acknowledging the individual efforts and contributions that led to the project's overall success.

**Appendix A. The lists of grades band**

<b>GRADE</b>	<b>PERCENTAGE BAND</b>	<b>CREDITS AWARDED</b>	<b>INDICATIVE QUALITY OF PERFORMANCE</b>
A	80 – 100	Yes	Excellent
B+	70 < 79	Yes	Very Good
B	60 < 69	Yes	Good
B-	55 < 59	Yes	Above Average
C+	50 < 54	Yes	Fair
C	40 < 49	Yes	Pass
D	35 < 39	Yes	Poor
F	< 34	No	Fail
I	Yes	Deferral of result or Incomplete	
X	yes	Student exempted from the learning of the modules because of previous equivalent learning	
W		Withdrew	
NP		Absent / Not Present	