DSA8040 – Foundations of Data Analytics and Artificial Intelligence – Assignment 1
(30% of Module Mark)

Group Project and Presentation Assessment

The Task
Students will be divided into teams and each team will be **assigned a dataset**. The task for each group is to perform a statistical investigation of their assigned dataset using Python.

In your group, you will need to **agree on two or three research questions** to investigate as part of this project.

Examples of research questions could be…
- ❖ Is there a relationship between the number of hours a full-time student studies during their degree and their grades?
- ❖ What is the average time it takes to recover from heart surgery and what factors impact this?
- ❖ Controlling for other factors, does getting more sleep improve one's emotional state?
- ❖ Does listening to classical music have an effect on one's ability to recall material that has been read?

What research questions your team chooses to investigate is up to your team, though it is recommended that you discuss your plans with the tutor during the practical classes.

Students are recommended to plan their analysis and work together to ensure the different research questions cover different aspects of the data. To showcase your team's understanding of the different methods of analysing data that we cover throughout the module, do try to include a variety of different visualization and hypothesis tests throughout your analysis, where possible and appropriate.

You will then complete the following:
1. Lay out your research questions and provide a justification for their formation and suitability to the dataset.
2. An exploratory data analysis should be carried out to:
> a. assess the structure of the data,
> b. produce summary statistics, and,
> c. create suitable tables and plots.
> d. pre-process the data as necessary.
Be mindful of sample sizes, variable types, missingness, etc.
3. This should be followed by hypothesis testing to confirm or refute some hypothesized relationship(s). At least one research question should be suitable for a hypothesis test. Not all research questions are expected to involve formal hypothesis testing.
4. Finally, draw conclusions and be clear about the limitations of your analysis. This should be presented in layman's terms to aid data-driven decision making around the research question under investigation while remaining technically competent.

One of the research questions you choose **must** be investigated by fitting a regression, in addition to the steps mentioned above. For this, your dependent variable (i.e. the variable you are trying to explain or predict) will be a continuous variable of interest.

For this section of the analysis, you will need to:
- Fit and interpret a suitable simple or multiple linear regression model,
- Test and comment on the normality assumptions of the residuals,
- Test the constant variance (homoscedasticity) assumption (i.e. is the model equally reliable across small and large predictions?), and
- Briefly discuss what these results imply for the reliability and interpretation of the model.

Assessment Guidance
In total, 30% of each student's final module mark is determined through the assessment of this piece of coursework.

Assessment is based on the production of a powerpoint presentation, corresponding Jupyter Notebook, peer assessment forms completed by each member of the group, and delivery of the presentation.

Slides, Notebook & Peer Assessment Form Deadline:
Friday 20th February 2026 at 10am.

The Jupyter notebook file should include Markdown annotation cells indicating which slide the section of code corresponds with which numbered slide in the powerpoint, where appropriate.

These three elements (slides, peer assessment and corresponding notebook) are all part of this assessment and therefore subject to penalization if submitted late.

Presentation Delivery Day:
Wednesday 25th February 2026 at 10am.

Assessment Weightings
The group project will be assessed based on the PowerPoint presentation (40%), the accompanying Jupyter notebook (20%), presentation delivery (20%), and peer assessment of individual contribution (20%).

Peer assessment weightings or scores may be moderated in exceptional circumstances to ensure fairness and consistency.

Peer Assessments
Each team member is expected to contribute throughout the planning, analysis, interpretation and creation of the powerpoint presentation. A student's contribution to the project will be taken into account via peer assessments.

A small element of the assessment acknowledges individual contributions. In order to do this, all students will be asked to allocate a theoretical set of marks amongst individuals in the group, according to the contribution level they perceive for each team member. You may wish to develop group criteria for the allocation of these marks, but each student will be expected to submit their preferred allocation of marks on an individual basis by the deadline above.

Late Submission of Assessed Work

Coursework submitted after the deadline will be penalised at the rate of 5 percentage points deducted from the total mark awarded for that component for each calendar day late up to a maximum of five calendar days, after which a mark of zero shall be awarded. This excludes University closure days.

Exemptions shall be granted to the above only if there are exceptional circumstances. Extensions to deadlines shall be proportionate to the impact of the exceptional circumstances.

Guidelines for Presentation Report Layout
Presentation reports should be created in Powerpoint (or similar software such as Latex) and **<u>submitted as an exported PDF document</u>**.

– The presentation will typically be under 30 slides.
– It should present the findings of the group in a manner that would be comprehensible to a non-statistician (with a university background).
– To ensure the layout of the presentation is professional, please ensure:
  - Slide Headings are underlined and/or in bold (or italic).
  - There are no typographical errors.
  - The font size used is not too small – this will depend on which style of font is used but all text needs to be easily read when in slide show/presentation mode, including text within all figures and captions, when projected to a screen.
  - Ensure table/graphs are clear and easily read.
  - Ensure axes are labelled appropriately
  - Provide a title or brief but descriptive caption for figure/tables e.g. "Figure 1: …"
  - Avoid animations, eye-sore themes, illegible font colours, etc.

The typical format of the presentation is as follows (typical number of slides involved given in brackets):
– Title page (not included in the slide count)
  - Include an appropriate title to describe the analysed dataset.
  - List group members on title page.
– Introduction to/Outline of the dataset (1 – 4 slides)
  - These slides should give an overview of the dataset being analysed. It should provide the reader with sufficient information (e.g. relevant variables within the dataset) to enable the rest of the presentation to be comprehensible.
– Brief description of the problems/questions the team proposed to answer using statistical techniques (1 slide)
– Results of investigating each research question (~ 5 slides per research question)
– Extended analysis for one of the research questions (~3 slides)
– Conclusions and limitations (~1-3 slides)

Plagiarism
Plagiarism is a major offence in submitted coursework. The penalties can be severe. The General Regulations of the University state: "Plagiarism is defined as the presentation of the work of others as the writer's own without appropriate acknowledgement."

Assessment of Written Report (Powerpoint Presentation):
The criteria used to assess written reports in the form of a powerpoint presentation include:
– Structure of presentation
– The introduction

- Use of language including technical terms
- Are tables/figures easily read and clear?
- Appropriateness of statistical analyses performed
- Demonstration of understanding of statistical results
- Demonstration of a clear presentation of conclusions to a non-statistician
- Content of final conclusions section