Assignment 2: Data Analysis Project (Group)

Overview

This is a group assessment task where students will work collaboratively in small teams (4-6 members) to undertake an analysis of a given data set and to report on their findings in the context of a real-world decision problem. The group will produce a written report and give a brief in-class presentation of their findings.

Students will apply selected techniques to the analysis of large-scale data sets, to undertake exploratory and confirmatory data analysis. They will use the results of their analysis to draw conclusions and to support specified decision outcomes.

Percentage Value of Task: 25% (20% report/content/presentation, 5% individual contribution).

Timelines

- Forming your group and dataset should be completed by 5:00PM Friday of **Week 4**.
  - Note: *the same team will undertake Assignment 3 (ML challenge) on the same dataset*
- Final report submission is due no later than 8:00 pm (AEST) Friday of **Week 6** (23 April 2021).
- Group presentations will be conducted during class **Week 7**.
  - Note: *all members must present a section of the final oral presentation*

Submission

- You should submit your assignment as a zipped file via the dropbox on the unit site. Each submission should have 3 subfolders:
  - Final Report (Only PDF file, use the template "Report Template - Assignment 2.docx")
  - Source code (Only Ipynb file)
  - Dataset (Only CSV file)
- Your assignment should be submitted by each member of the group individually/separately.
- Late submission by a group member will incur a 5% late penalty per day (including weekends, public holidays, etc.) for that member. Submissions more than five days late will not be marked.

Assessment Details

✓ Dataset

The list of some suitable datasets is provided on the unit site; refer to "Datasets". Each group will be given a dataset from this list, but you can choose your own dataset (different from others) *by submitting a new request*. ***Using a dataset without approval will lose an immediate 20% of marks.***

✓ Forming your group/team

Groups will be formed during the pracs; it is expected that all members in a team attend the same practical class. Please form your group, choose your data and fill the information in the spreadsheets uploaded into Microsoft Teams SIT307 General Channel / Files:

- Cloud Group List
- On-Campus Group List

- Students who did not group-up by the end of Week 2, will be randomly assigned to different groups.

Important Note: *To avoid plagiarism, cross-pollination of solutions and project reuse and to ensure all presentations are different;* ***only one team will be permitted to use a particular dataset****. You can change your data at any time by submitting a new request. However if that dataset/problem has since been taken then you will be refused.*

✓ Team work

It is up to your group as to how the tasks are to be divided up and where you should work together.

- One approach would be to have one person prepare the data, a second performs the experiments, and the third does the analysis. Then all three write the final report together.
- A second approach might be to all work on the activity stage together then split up the report writing into sections – one for each.

Whichever approach is taken all members **must submit the (same) group assignment individually and present a section** of the final oral presentation.

✓ Report

Once you have performed your experiments you will prepare a report using a suitable format. You must not exceed 10 pages (this includes your reference section) and should include marks for the contribution of each team member. Please use the template "Report Template - Assignment 2.docx".

The report should be written in ***formal English*** and use an appropriate referencing style such as APA or Harvard (https://www.deakin.edu.au/students/studying/study-support/referencing).

Your report should contain a description of your dataset and a discussion of the final data preparation used; exploratory analysis (missing values, etc.) and what was found; the methodology used in experiments; and, a discussion. Your report should contain headings and a clear structure. As a guide the following structure is suggested as a possibility (different datasets/problems may require a different structure – you should change your structure where appropriately required):

- Cover page (including students names and individual contribution marks)
- Introduction
- Exploratory analysis: Methodology, Results and Discussion. This may include:
  - Data cleansing, like missing data or measurement error, data type of columns, etc.
  - Establishing the main structure and key variables.
  - Investigation interesting patterns, like identifying relationships between variables that are particularly interesting or unexpected.
  - Checking assumptions/hypothesis in relation to a specific pattern.
  - Using effective visualizations to demonstrate findings/results.
- Conclusion
- References

✓ Oral Presentation (to be delivered in the Week 7)

You whole group will present together. Your presentation should be for 8 minutes to allow for 2 minutes of questions. Each member of the team should speak for about two minutes of the presentation.

Your presentation should use a formal presentation medium such as PowerPoint or other software. It should completely cover all the topics covered in your final report.

- **On Campus students:** A schedule for groups will be given in week 6.
- **Cloud students** must submit a video presentation of their project edited so that all members of the group share the presentation. There must be no edits during each speaker's presentation (only between presentations).