

Assignment Specification

| | |
|---|--|
| Continuous Assessment 1 Class Group: TU059 / TU060 | |
| Worth: 50% of the overall mark for the module | Due date: 09/11/2022 (Wednesday), 23:59pm |
| Penalty for late submission: $4^d\%$ where d is the number of days late. Note after 3 days the penalty is set to 100%. | |
| Objective: To demonstrate what you have learned so far in terms of importing data, exploring and understanding data, and answering possible questions of interest. | |
| Description of Tasks Find data in publicly accessible online sources e.g. UCI , data.world , Kaggle , webpages, APIs etc. Part 1 Import the piece/s of data and perform any cleaning and merging to produce a final dataframe. (33 marks) Part 2 Carry out exploration of this dataframe to develop an overall understanding of the data. (33 marks) Part 3 Focus on a particular subset of the dataframe and drill down into it extracting details to answer a series of questions that are of interest to you as an analyst - Ideally the motivation for such questions would be framed within the context of a hypothetical use case scenario. (34 marks) | |
| Submission You should submit a Jupyter or RStudio notebook. Make sure you explain each code cell using markdown cells. Focus on what was done, flow of reasoning (motivation for steps), depth of analysis and conclusions reached. In relation to code, ensure that variable names are informative and appropriate and meaningful comments are placed. The code should work without issue when run on the lecturer's machine. | |
| Plagiarism Code sourced from elsewhere must be clearly cited (bear in mind your Python and/or R skills are also being examined). Plagiarism will result in a zero mark (0%). You should make yourself familiar with the plagiarism policy of Technological University Dublin. | |

Grading Rubric

| | Part 1 | Part 2 | Part 3 |
|--------------------------|--|---|--|
| Challenging (70+) | Work is complex involving data in multiple formats and/or files. Exhibits elements of creativity e.g. problem solving and self-learned concepts. | Exploration is in-depth and insightful, demonstrating an excellent understanding of key characteristics and patterns in the data. | Complex manipulations of the data in pursuit of latent trends that answer a series of interesting questions. |
| Very good (60-69) | Work involves data from more than one format and/or file. There are some challenging aspects to the work. | Exploration covers a variety of relevant views on the data. | Some complexity in terms of manipulations with a coherent basis for operations. |
| Good (50-59) | Work involves data from more than one format and/or file. | Exploration is sufficient but somewhat repetitive | Good manipulation and analysis, but feels a little contrived in places. |
| Ok (40-49) | Work based on a single piece of data. Did not require much cleaning and/or merging. | Exploration is limited and/or misleading. | Basic manipulations of the data, work missing a clear focus and/or contains some inaccuracies. |
| Needs work (0-40) | Significant departures from assignment brief and/or heavily incomplete. | Significant departures from assignment brief and/or heavily incomplete. | Significant departures from assignment brief and/or heavily incomplete. |

A good answer should be insightful. It should address the assignment in a way that indicates your comprehension of and control over the assignment itself as well as an understanding of the underlying issues. The message should be communicated clearly, concisely, and directly.

General guidelines

- Start from the beginning. Find a dataset that interests you and is complex enough for allowing an interesting analysis.
- Once a problem has been solved, test and document it before moving to the next part.
- We have gone over many notebooks so far. Review then if you feel lost or don't know how to proceed.
- An excellent assignment will likely require some self-learned concepts and the use of functions/parameters not seen in class. Still, try to complete a basic version of the assignment first and add complexities as much as possible later on.