

Data Science for Business

Assignment 1: Data Collection & Preparation (20 Marks)

Marking Guidelines (100%):

Data collection & storage: 20%

Marks awarded for:

- Correct use of one or more Web API(s), and storage in an appropriate format. Takes into account the complexity of the API, the use of multiple collection runs (where required), and the use of multiple related APIs.

Data pre-processing & analysis: 60%

Marks awarded for:

- Loading the data into an appropriate representation for analysis.
- Application of any necessary pre-processing steps to improve the representation of the data (e.g. missing value imputation, filtering rows, filtering features, merging/grouping features, deriving new features, integrating data from multiple APIs if required).
- Exploration, summarization, and characterization of the cleaned dataset. This should include the use of data visualization and tables where appropriate.
- Interpretations and insights made about the dataset, based on the above. This includes suggestions for further analysis.

Code quality & explanation text: 20%

Marks awarded for:

- Use of Markdown cells to explain each step of the process, from data collection to interpretation.
- Readability and clarity of the code. This includes use of comments, though they should not be used excessively.
- Complexity of the code, and appropriate use of Python packages. Note that code performance/speed will not be taken into account

Penalties for Late Submission:

- 1-5 days late: 20% deduction from overall mark
- 6-10 days late: 40% deduction from overall mark
- Assignments will not be accepted after 10 days without an extenuating circumstances form and/or a medical certificate.