

INFO3406 Project Stage 1

Explore, Clean, Define



Overview

The objective of stage 1 is to explore a data set and define a research question based on a research/business requirement. Activities include: (1) selecting a data set; (2) exploring, summarising and preparing the data; and (3) defining the problem and project requirements.

Report (13 marks)

The two-page report (not counting title page and references or appendix) should describe the problem, proposed approach and your data:

Section 1: Problem

- Describe the problem from a general perspective, highlighting the business/research need.
- List the research question(s) you will answer in stage 2 of the project.

Section 2: Approach

- Describe the approach you will take to solving the problem and any requirements. This is your plan for stage 2

Section 3: Data

- Describe the data from a general perspective e.g. source, size, fields of interest.
- How did you acquire the data?
- Describe any exploratory analysis you have done to refine your understanding of the data and research question. One or two supporting figures/tables would be great.
- Describe any data preparation e.g. transformation, sampling, cleaning.
- Which tools did you use to clean and explore the data set?

The report is worth 13% of the overall mark for INFO3406. It will be marked on data relevance, problem definition, data preparation, data summarisation and report completeness. The report should use the high-level headings above (problem, approach, data). It should use line spacing of at least 1.15 and body font size of at least 10pt.

The goal is to convey the problem clearly and concisely.

INFO3406 Project Stage 1

Explore, Clean, Define



Marking Criteria	Marks	
Problem Definition	0 – 2	
Problem Approach	0 – 1	
Acquisition of relevant Dataset	0 – 2	
Data Summarisation	0 – 4 (Explorative Analysis)	
Tools Appropriate	0 – 2	Python, SQL and Unix
Report Structure and Style	0 – 2	TODO at end
TOTAL	13	