

Advice on Doing Your Project Assignment

Paul Wohlfarth

March 3, 2021

1 Instructions

With this project you should demonstrate that you can do statistical analysis that is relevant for business by yourself. Find some data and perform a segmentation analysis using kmeans-clustering. The data you collect must be quantitative (i.e. not categorical) and contain at least 6 variables. You must address the following points:

1. **Motivation:** Why did you choose your data? Which relationship(s) are you aiming to investigate? What's your research question? What do you expect to see?
2. **Methods:** Explain how you specified clusters. How many clusters did you choose and why?
3. **Results and conclusions:** Describe and present your results in an appropriate way. What do you conclude from your findings? Is there anything your analysis cannot address?

2 Submission Format

Your final submission should be one report of no more than 2000 words. The report must be submitted in Jupyter notebook (.ipynb-files) **as well as in either HTML or PDF format (this means you must submit two copies of your report)** and contain all code with appropriate comments to replicate results. Code cells in Jupyter notebooks must be executable, i.e. all required packages must be imported, comments must be formatted appropriately, where used in code cells, and data must be submitted in a separate file. Use comments only to describe your code in code cells (be concise!) and use markdown text for any other text that is not directly related to your code (i.e. your motivation, description of methods, results and conclusions should be written in markdown text).

3 Deadlines

There are two deadlines:

1. Proposal: Monday, 3rd May 0:00
2. Main Report: Monday, 31st May 0:00

Proposal: This should only be a few lines telling me what you want to do and what data you'd like to use. The proposal won't be marked but I will deduct five percentage points if you don't submit it on time (you could then at most get 75% for the assignment).

4 Grading Criteria

Your grades will be informed by a number of criteria. These are not marked separately, so you will receive one grade. But the below should give you some idea of what I particularly pay attention to when I'm marking your projects.

4.1 Form and presentation.

This includes the complete submission of all required documents and the overall presentation of results.

- Is there: a clear structure overall; clarity in individual paragraphs and sentences; logical arguments; and careful use of evidence? Are spelling and grammar correct?
- Are any technical terms or abbreviations explained? The word limit is short so make every word count.
- Does your code run (apart from having to specify file paths) and can it be replicated easily?

4.2 Motivation.

How well do you motivate your project? Is it relevant? Is it original?

4.3 Analysis.

- Can you demonstrate a good understanding of the relevant context and methods?
- Is there a logical argument and effective use of evidence to support the argument?
- Did it answer the question posed?

4.4 Data collection/presentation

- Has the appropriate data been collected (given time limitations)?
- Have the data been checked?
- Does the work show understanding of what the data actually measure and the limitations of the data?
- Does the work demonstrate the ability to summarize and present data in a clear and effective way?

4.5 Statistical Methods.

Do you use the appropriate specification? Can you explain why you apply methods in the way you did? Does your report show understanding of the methods.

4.6 Interpretation.

How does your report answer your research question? How convincing are your conclusions? Are they backed by the evidence you present?