

Lab Report

186.868 Visual Data Science 2024W

11912007 - Yahya Jabary

Contents

Discover Stage	1
Wrangle Stage	1
Profile Stage	1
Model Stage	2
Report Stage	2

This report contains the contents of all tasks to be completed in the laboratory section of this course. The individual subtasks are described in separate chapters.

Discover Stage

The tasks of this section are:

- Discover your topic
 - Either select one of the suggested topics, or come up with your own idea. → Our custom topic was confirmed by the lecturer via email.
 - The topic should be possible to easily understand the data context, even if you are not an expert.
- Describe the two datasets you selected
 - Use at least two independent datasets.
 - The selected datasets must be multidimensional: they have to contain more than 5 variables.
 - The selected datasets must contain a sufficient amount of data rows.
- Implicit: define questions to answer with the data
- Keep the length to 1-2 A4 pages

Wrangle Stage

The tasks of this section are:

- Join the two or more datasets you selected into one big data table
 - How did you join the datasets? Which keys did you use to join the data? Did all keys match? Did you have to introduce new keys?
- Solve issues like formatting issues, missing data, faulty values, and non-matching keys.
 - Which data cleaning steps have been necessary? Did you experience data issues, and if so, which ones? How did you solve them? Did you use automated methods? Did you use visualization to inspect data issues?
- Visually show and explain the data quality of your dataset (for example, before and after cleaning steps). Come up with your own, creative, solution here.
- Use a charting library, not fully-featured applications.
- Keep the length to 3/4 to 1 A4 page.

Profile Stage

In this section we explore the data in detail, to completely understand its structure, and to discover any interesting patterns that can be found in there.

The tasks of this section are:

- Find at least 3 informative insights in your dataset. For each one add a short text describing the insights plus one visualization.

- Use a charting library, not fully-featured applications.
- Keep the length to 3/4 to 1 A4 page per insight.

Model Stage

The tasks of this section are:

- Build a model of the data you studied, to find answers to the questions you selected (possible models: linear regression, clustering, pca, anomaly detection, etc.)
- describe the modeling process
- Create one or more visualization(s) that describe the results of your model
 - How would you increase trust of your customers/colleagues in your modeling approach by using data visualization?
- Use a charting library, not fully-featured applications.
- Keep the length to 1-2 A4 pages.

Report Stage

The tasks of this section are:

- Show your findings in an interactive dashboard to a broader audience. The findings may be related to the Model/Wrangle/Profile stages.
 - Use appropriate charts, visual encodings (e.g., color).
 - Use at least four different types of visualizations / charts.
 - Include interaction (e.g., filters, zoom, not a jupyter notebook), brushing & linking (changes in one view affect others, but not global filters)
- Use a library, not fully-featured applications.
- See examples: <https://tuwel.tuwien.ac.at/mod/page/view.php?id=2433356>