



Lisbon School
of Economics
& Management
Universidade de Lisboa



LISBOA
UNIVERSIDADE
DE LISBOA

Course: Programming & Data Sciences

Project

Each team performs a project. In this project, the group must identify a research question and objective and support it in a research paper (conference paper or journal paper). The group must also specify a data source and a possible solution. Study of movies (IMDB database) or Airbnb are viable alternatives. Databases obtained from Kaggle are not allowed.

The team must collect data, use Python to explore data, create a model, and deliver the results. Jupyter Notebook supports the project. Students must choose a subject and obtain data using open data datasets (e.g., Pordata, European Commission, Lisbon Municipal Government Site, among others). Students must apply one (or several) techniques. Students may use the following techniques: regression, logistic regression, cluster analysis, and network science. Bagging and boosting algorithms, as well as neural networks, may be used. Note that some of the topics may require more self-study than others.

Deliveries include:

- Report (students must submit the report in **both** formats: .docx filetype and PDF filetype)
- Jupyter Notebook
- Dataset(s) (including the source of those dataset(s)).
- Eventually, students may also deliver results in the form of a web app developed in Flask

Pay attention to plagiarism. ISEG Rules will be strictly followed.

1. Report

The report should have the following structure:

I. Introduction

In the introduction, students must provide the context of the project. It is also essential to identify the generic problem the students expect to solve and the main objective of the empirical work.

II. Literature review

The group should identify a small group of papers that may help to identify similar work developed by other authors.

III. Empirical Work

To make the empirical work, students must follow a data science life cycle (like CRISP-DM or POST-DS) as follows:

1. Data context

2. Data collection
3. Data Preparation
4. Exploring Data
5. Data Modelling
6. Evaluation
7. Deployment

IV. Results and discussion

The team should compare the results obtained from the empirical work with those of other authors from the literature review.

V. Conclusions

What was the purpose of the work? What do you conclude from the literature and the empirical work?

References

According to the format

2. Jupyter Notebook

The report must explain clearly the data analysis steps.

The team is encouraged to explore other techniques different from those presented in Class. However, to be valued, the group members must master the techniques.

3. Datasets

Obtaining original, curious, and valuable data is hard work. So, it is a task that is valued in the project. Students must identify the data sources. The process of data collection should be well described.

The easiest way of obtaining datasets is by selecting datasets already used in published papers (sometimes published with the articles). Obviously, students must avoid if they want to have a good grade. But it is possible.

4. Web App Using Flask

Web app using Flask is not mandatory task.

5. Presentation dates:

According to the scheduling. All the members of the team must present the project.