Project Statement

This project uses data science concepts and Python to analyze data and produce knowledge. The context is the information about Lisbon, Portugal. You may obtain data from several sources, like the following:

- Lisboa Aberta (https://dados.cm-lisboa.pt/)
- INE
- Pordata
- Airbnb
- Dados Gov (https://dados.gov.pt/pt/)
- Geodados (https://geodados-cml.hub.arcgis.com/)
- Kaggle is not allowed!!!!.

Other sources may be used as long as they are credible. Scraping is also allowed.

The group must start by defining a question they must answer, e.g., the impact of Airbnb on real estate prices. It may be related to economic, social, cultural, sports, or any other question considered relevant by the group.

Then the group must share this information in the appropriate place. It is not possible to have repeated questions.

The group must use the CRISP-DM approach presented in class. It should also use as many techniques as possible, like network science, non-supervised learning algorithms, supervised learning algorithms, time series, or NLT.

The result will include:

- The Jupyter Notebook/Python application. Students may develop a small application in Flask. It is not mandatory.

- Report PDF (and DOC/DOCX or LateX)

All the information must be available to the lecturer of the course.

Deadline:

- Jupyter notebook draft: Saturday, May 13th
- Final Report: 13th May Saturday.

Presentation:

- Presentation: between 5 and 10
- Maximum presentation time: 15
- All group members must present the project (May 16th or May 18th).

Submission place:

- MS Teams group channel

The Report has a template like one research paper. The structure of the paper is the following:

Title

Abstract

Keywords:

1. Introduction
What is the problem solved in the research reported in this paper?
The objectives of the paper must be described.
What contributions did you have?

2. Literature review
Who already has already studied the subject analyzed here?

3. Methodology
The methodology followed is the CRIST-DM/POST-DS [1]: business understanding, data understanding, data preparation, modeling, model validation, and deployment

4. Results
Presentation of the main results, identifying.

5. Discussion

6. Conclusions

References
[1] C. J. Costa and J. T. Aparicio, "POST-DS: A Methodology to Boost Data Science," 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), Seville, Spain, 2020, pp. 1-6, doi: 10.23919/CISTI49556.2020.9140932.**(This reference is mandatory)**

The following items will be considered in the evaluation of the Teamwork:

Code/Programming
Data collection
Data analysis and exploration
Modeling
Data presentation
The data collected interest.
Quality and presentation of Report
Quality and presentation in class
References quality