

## Project Statement

The purpose of this Project is applying data science concepts and Python language to analyse data and produce knowledge. The context may use the following data, that will be supplied by the course lecturer:

- Information about movies imdb dataset
- Information about movies exhibited in Portugal
- Wikipedia information by language
- other (must be interesting information obtained from open data data or available) Kaggle is not an alternative.

All the data supplied must be supplemented by data from other sources.

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.

---

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 research reported in this paper?

The objectives of the paper must be described.

What contributions did you have?

#### 2. Literature review

Who already have already studied the subject analysed here?

#### 3. Methodology

The methodology followed is the CRIST-DM/POST-DS [1]: business understanding, data understanding, data preparation, modelling, 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.

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

Code/Programming:

Data collection

Data analysis and exploration

Modelling

Data presentation

Interest of the data collected.

Quality and presentation of Report

Quality and presentation in class