Programming for Data Science
Final Project

niccolo.marastoni@univr.it

February 2026

The project consists of an analysis performed on a dataset chosen by each student. The dataset needs to have a decent quantity of data points (rows), and each data point needs to have some features (columns). Once an appropriate dataset has been found, it has to be submitted to the professor along with an outline of the ideas for the project. **Wait for it to be approved before starting to work on it**.

What follows is a rough outline of how a project can be structured and evaluated. These steps serve as a guideline for anyone that appreciates a more structure approach to the project. You are free to take these steps in any order and focus on any of them as you see fit in your project.

The project is worth: **3 points**.

1 Data Exploration + Data Wrangling

1. Explore the dataset Your first task is to explore the data as seen in class, finding correlations between attributes and finding some interesting aspects that justify the next parts of the analysis.

2. Clean up the dataset This step encompasses the replacement of null values with appropriate data or their outright removal. During this step you are encouraged to modify the data according to the considerations matured during the exploration.

3. Show some interesting plots An essential skill of a data scientist is being able to show the important information by using easily understandable graphs. Use the libraries introduced in class to showcase some interesting aspects of the dataset.

2 Find a model that explains the data

For this part you can use any tool you are familiar with, be it statistical methods, machine learning or even deep learning.

Possible ideas:

- regression model if you are trying to predict continuous values
- classification model when you have discrete and finite labels
- clustering if you want to find underlying structures of the data, without prior information

3 Build a presentation

This ties in nicely with the 3rd step in part 1: you need to be able to present the data in an intuitive way. A nice way to simplify the presentation is removing the underlying code from the view. In order to get full marks in this part you need to submit a presentation built with streamlit (as seen in class) or similar libraries (Flask, Django etc.).

The presentation needs to showcase all the parts outlined in the project so far.

4 Track progress through Git (optional but recommended)

You should create a public GitHub repository for your project, or you can create a private repository and change it to public when you are ready to submit.

Using GitHub as a showcase for your portfolio is only a small part of the project, it is important that you also show the progress of your project in the commit history.

5 Project Submission

Once your project is finished, you can send me the link to your public repository on GitHub via email. You can also send your project in a .zip attachment if you decided to forgo the GitHub part (try to avoid this) and make sure to **include your colab notebook**. The email subject should have the following structure: "<name> <surname> - <studentID> - Final Project Submission".

For example: "Niccolò Marastoni - VR364254 - Final Project Submission".