EBI02156 - Technical Test

Please follow the instructions below and **submit your work to** recruitment@ebi.ac.uk **before the 9th October 2023 (i.e. please complete the test by the 8th)**.

We would like you to provide the code you wrote to analyse the data, and the visualisation of the results. Please note that it is important that you provide plots of your analysis as part of your answer.

You may email your answers (including source code) as a zip file, although we would prefer you to host the code on a personal GitHub project space and email the link, while sending visualisation of the results to the recruitment team. Please label plots clearly so it is obvious which task they address.

Please provide adequate documentation on how to run your code and an indication of how long it took you to complete the test.

Background

Two datasets are provided that relate to red and white variants of the Portuguese "Vinho Verde" wine. The inputs include objective tests (e.g. pH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

Input variables (based on physicochemical tests):

1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 - density
9 - pH
10 - sulphates
11 – alcohol

Output variable (based on sensory data):

12 - quality (score between 0 and 10)

Tasks

Please write a Python script to

1. Load the datasets into a Pandas DataFrame

2. Perform exploratory data analysis (EDA) to get a sense of the data. Please include summary statistics, data visualization (e.g., scatter plots, histograms)

.

3. Compute the correlation matrix for all numerical variables.

4. Create a heatmap to visualize the correlation matrix.

5. Perform a multiple linear regression analysis to predict the wine quality based on the other variables.

6. Interpret the regression results, including coefficients, p-values, and R-squared.

7. Provide recommendations based on your analysis.

Resources

The datasets will be provided as a zip file.