Data Analyst Exam Project (EP)

Welcome to your exam project. In this assessment, you will be required to perform a few tasks that consist of:

- Analysing data from the provided dataset.
- Writing a report to illustrate your findings.
- Creating visualisations.
- Selecting and editing images.
- Constructing a Microsoft PowerPoint presentation.

Scenario

Understanding the chemical properties of beverages is critical to understanding and assessing their quality. As an assistant to a restaurant chain manager, you are asked to prepare a report on the quality of red variants of Portuguese Vinho Verde. You need to find a scientific way to rate the wines based on their chemical characteristics and present the results in a user-friendly manner. Based on your Technical report, the chain manager can choose the best type of wine from the winery's offer.

Working in Microsoft Excel, using all the knowledge acquired during the year, use the winequality-red.csv to demonstrate your skills and understanding. The dataset is based on a study conducted by Cortez et al. (2009).

A description of the dataset can be found below:

Feature name	**Feature description**
fixed acidity	Most acids involved with wine or fixed or non-volatile (do not evaporate readily). This feature indicated the level of such acids
volatile acidity	The amount of acetic acid in wine, which at too high of levels can lead to an unpleasant vinegar taste
citric acid	In small quantities, citric acid can add 'freshness' and flavour to wines. This feature describes the level of citric acid in the wine
residual sugar	The amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/litre and wines with greater than 45 grams/litre are considered sweet
chlorides	The amount of salt in the wine
free sulfur dioxide	The free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulphite ion; it prevents microbial growth and the oxidation of wine
total sulfur dioxide	Amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2 concentrations over 50 ppm, SO2 becomes evident in the nose and taste of wine
density	The density of the wine is close to that of water, depending on the percentage of alcohol and sugar content
pH	Describes how acidic or basic the wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale
sulphates	A wine additive which can contribute to sulphur dioxide gas (S02) levels, which acts as an antimicrobial and antioxidant
alcohol	The percentage of the alcohol content of the wine
quality	Output variable (based on sensory data, a score between 0 and 10)

You must analyse the provided dataset to attempt to predict a final quality for a given wine based on its physicochemical properties and report it. Managers will then use the report as insight to order large quantities of wine. Thus, big money is at stake! Consider a pallet of wine to be used to transport 500 bottles, meaning that with an average price for a bottle of wine of 20 EUR, the average cost of a pallet can be around 10,000 EUR. To serve the manager as well as possible, prepare a submission that will at least include the following:

- A demonstration of your ability to perform an EDA on a dataset and demonstrate the ability to solve practical tasks and ensure data is of a high enough standard.
- Break down the dataset to demonstrate an understanding of trends, patterns, and anomalies in the data.
- The ability to test hypotheses and validate any assumptions about the data.
- The ability to support any conclusions and provide logical and clear arguments.

Deliverables

The mandatory deliverables for the EP consist of two separate documents:

1. A report, submitted as a .pdf file, detailing your analysis. This report is targeted at the restaurant chain manager.
2. A presentation, submitted as a .pptx file, that the restaurant chain manager can use to present your findings to the board.

Details of the report

Your report needs to be in the form of a Microsoft Word document, exported to PDF before submission, with appropriate sections and headings used:

- **Introduction**

  Telling the reader what the report is about, why it is necessary, and how you will address the report's needs.

- **Body**

  - **Initial Assumptions and Hypotheses**

    Read through the Red Wine Quality dataset (Cortez et al., 2009) to better understand how the data can be used. Then, use this section to write up your initial assumptions and hypotheses.

  - **Exploratory Data Analysis (EDA)**

    This section will be relatively extensive, detailing your chosen approach in performing exploratory data analysis. Break this section up into sub-sections, using sub-headings that best fit the individual tasks or sub-steps you had to perform during your data analysis approach.

- **Trends, Patterns, and Anomalies**

  Use this section to discuss trends, patterns, and anomalies you identified during your EDA. Where applicable, use sub-headings to highlight important topics to your reader.

- **Discussion**

  Use this section to discuss your initial assumptions and hypotheses and whether they hold. If not, discuss why.

- **Conclusion**

  A summary of your findings and whether you deem your study successful. End the section with a bit of personal reflection on your journey in compiling the report and how you feel about your abilities regarding performing EDA.

Your report should follow the various guidelines for creating a report, as covered in the second semester of the year. A few tips are mentioned below:

- Select the appropriate type of report and language style.
- Include a title.
- Use styles for titles, sub-titles, and headings.
- Include an automatically generated table of contents.
- Use pagination.
- Ensure your body text is at a set font type and size.
- Set your paragraph spacing to 1.5.
- Set the justification of your document to “justify”.
- Perform proofreading of your report.
- Use document formatting (theme, style, or/and colours).
- Refer to experts’ opinions and use citations (Harvard style).
- Insert images and Excel graphs to support your arguments. Use appropriate wrapping style.

You *may* include a single Microsoft Excel spreadsheet saved as a .xlsx file as a supplementary/optional submission. If your analysis requires multiple sheets, these sheets should be included as different worksheets inside the overall workbook. This spreadsheet will be marked ONLY in particular circumstances (should your .pdf be unclear). We recommended that you ensure your .pdf is fully comprehensive and your .xlsx workbook (should you decide to include it) is as concise as possible.

Details of the presentation

Your manager must present a formal report in the form of a Microsoft PowerPoint presentation to the board of directors at their annual board meeting. Using the dataset and your findings in the report in the above Scenario, make the necessary changes in the format and data visualisation, allowing your manager to present to the board. Bear in mind that the board members do not require all the technical details of managing the restaurant chain nor the specifics of wines. Thus, the presentation needs to be geared towards a less technical audience who may need some background to help them understand the problem.

The presentation must showcase...

- Tables and Charts or any other visualisations techniques you mastered during the course.
- Insert images, graphics or media that you edited.
- Be creative and include transitions and/or animations.
- Insert a slide explaining your design thinking and the design principles you implemented in your presentation.

Summary

To summarise, your complete submission MUST include the following:

- A report in the form of a .pdf file named Exam Project Report.pdf
- A presentation in the form of a Microsoft PowerPoint presentation called Exam Project Scenario.pptx

And *MAY* include the following:

- A spreadsheet in the form of a .xlsx file

References

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, *47*(4), 547–553. https://doi.org/10.1016/j.dss.2009.05.016

UCI MACHINE LEARNING (n.d.). *Red Wine Quality*. Kaggle. Retrieved March 19, 2023, from https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009