



UNIVERSITY OF GHANA

(All rights reserved)

FIRST SEMESTER: 2022/2023

SENG 207: PROGRAMMING FOR ENGINEERS (3 Credits)

PROJECT - 1

Scenario:

Assume you work in a Beverage Manufacturing Company as **Data Analyst**. The **Production Department** has provided your team with a dataset of all inventories of red wine. The dataset is saved in a **csv** file format as **data.csv**. Your task as a data analyst is to perform exploratory data analysis on the dataset and make a presentation to your investors in the end of year meeting. Use this scenario and instructions below to develop your insight. Note that the information below is only to serve as a guide; students are encouraged to research on innovative ways to make their work interesting and insightful.

Task:

- Pre-process data to get rid of all null values and duplicates.
- In the event of null values, explain your method and the reason for adopting such approach.
- Show statistical inferences of your entire dataset
- Show statistical inferences of individual columns in your dataset.
- Show some hidden inferences
- Show data correlation among columns

Hint: Where necessary, use graphs, charts and images to explain the numbers or data. Use markdown to put in text where needed. Research on how to use heatmaps, histogram plots, pie charts, bar graphs, subplots, etc. Use good colors and maintain consistency in your color choices and fonts.

Submission: Each member of the team must keep a copy of the codes and other resources and submit same onto a cloud repository such as *github*. Send URL of your cloud repository to Ernest who will compile them as a *text file* for onward submission.

Submission date: 27-March-2023 at 23:59hrs GMT.

Presentation: Live presentation in class. Submit ppt to Ernest a day before presentation. No changes to ppt slides or jupyter notebook file will be entertained on the day of presentation.

Presentation date: 28-March-2023 at each class' lab section period. The presentation time for each team will be 10 minutes.