**Project Report: Analyzing the Wine Quality Dataset**

Name: Harshit Gupta

Course: Python Data science(offline)

In this task for the internship provided by EICT IITK , I have to use the given data and visualise it to solve the given question where I have used different libraries of Python to calculate, analyse and draft graphs for better visualisation and understand.

In order to understand the variables that affect wine quality, the goal of this study was to analyse the Wine Quality of red and white dataset. In order to determine the most crucial elements influencing wine quality, the project included data loading, data cleaning, investigation through visualisations and statistical analysis, and the implementation of a machine learning model.

Learningoutcome:

* Data Loading and Preprocessing Outcome: I mastered the skills necessary to use pandas to load a dataset into a Python environment and carry out necessary data preprocessing operations.
* df = pd.read\_csv('winequality-red.csv',delimiter=';')
* Data Exploration Outcome: By using statistical analysis and visualisations, I improved my data exploration skills.We used histograms to visualise the distribution of wine quality ratings, which helped us understand how the scores were distributed across different quality levels. With the use of scatter plots and a correlation matrix, we also investigated the connection between other characteristics (such as alcohol concentration) and wine quality.
* Details: We successfully imported the dataset from an external source, checked for missing values, and verified that the data types were suitable for analysis. Although there were no missing values in this dataset, we showed how to manage them if there were.
* Feature Importance Analysis Outcome: I discovered how to employ machine learning models, particularly Random Forest, to pinpoint crucial elements in determining wine quality.

To illustrate feature significance ratings and make it simpler to comprehend and discuss the results, we generated visualisations. The main attributes were shown in a bar chart, and the report's findings were presented in an understandable manner. I successfully analysed the red & white Wine Quality dataset for this project, determined the key variables determining wine quality, and presented the results using code and visualisations.

Identifying the most impactful characteristics in a dataset using a machine learning model (Random Forest) is a skill I gained that is important for making data-driven decisions.

I became better at producing educational visualisations that successfully convey information.