First Assignment's description:

This dataset describes the amount of various chemicals present in wine and their effect on it's quality. The datasets can be viewed as a regression task. Your task is to predict the quality of wine using the given data. A simple yet challenging project, to anticipate the quality of wine. The complexity arises due to the fact that the dataset has fewer samples and is highly imbalanced.

Task 1: collect the data from the database using any SQL engine

Task 2: clean up the data if needed (handle duplicates, missing values, add, delete or modify features, correct the data types).

Task 3: use proper statistical and visualization techniques for analyzing the data and interpretation of the relations among the features.

Task 1:

Since you've already provided the dataset as a CSV file(WineQT), in this part, I'm demonstrating how to load a CSV dataset into a Python environment, connect to a SQL database (specifically SQLite in this example), and then store the data in a SQL table. After the data is

stored, I show how you can execute SQL queries directly from Python to retrieve and work with the data.

Task 2:

Check for Duplicates:

checks for duplicate rows and removes them if any are found.

Handle Missing Values:

it checks for missing values and fills them with the mean value of their respective columns.

Correct Data Types:

check data types, ensuring that.

Task 3:

Exploratory Data Analysis:

Quality Distribution: we visualized how the wine quality ratings are distributed to understand class imbalance.

Correlation Matrix: we created a heatmap to explore linear relationships between chemical properties.

Feature Distributions: we used histograms to see the spread of each chemical property and identify any potential outliers.

Feature Distributions: we used histograms to see the spread of each chemical property and identify any potential outliers.

Sajjad Vosough HM_1