

Wine Quality dataset

The Wine Quality dataset from the UCI Machine Learning Repository contains red and white variants of the Portuguese "Vinho Verde" wine. It includes 11 physicochemical features (e.g., acidity levels, sugar content) and one quality score for each wine. The quality score is based on sensory data and ranges from 0 (very bad) to 10 (very excellent). The dataset is commonly used for regression and classification tasks, such as predicting the quality of wine based on its chemical properties. We have used Red Wine Quality dataset to perform all the operations .

Operations Performed:

Calculate the average alcohol content: Compute the mean alcohol content across all wines in the dataset .

Find the maximum and minimum pH values: Determine the highest and lowest pH values in the dataset .

Calculate the standard deviation of residual sugar: Compute the standard deviation of the residual sugar content in the dataset and save it to a CSV file.

Find the maximum and minimum values of each column: Identify the maximum and minimum values for each column in the dataset and save them to a CSV file.

Select only numeric columns for correlation calculation: Select the numeric columns in the dataset to calculate the correlation matrix.

Plot the correlation heatmap: Visualize the correlation matrix as a heatmap to identify relationships between variables.

Find the number of wines with quality greater than 6: Count the number of wines with a quality rating higher than 6 and save it to a CSV file.

Plot the histogram for 'fixed acidity': Create a histogram to visualize the distribution of fixed acidity in the dataset.

Plot the relationship between 'alcohol' and 'quality': Generate a scatter plot to explore the relationship between alcohol content and wine quality.

Plot a boxplot of 'alcohol' content by 'quality': Create a boxplot to compare the distribution of alcohol content across different quality ratings.

Bar plot of average quality by alcohol content levels: Plot the average quality of wines based on different levels of alcohol content.

Density plot for sulphates across different qualities: Visualize the density distribution of sulphates in wines based on quality ratings.

Perform ANOVA on quality by alcohol content: Conduct an analysis of variance (ANOVA) to examine the relationship between alcohol content and wine quality.

Perform a t-test on alcohol content between high and low-quality wines: Conduct a t-test to compare the alcohol content of high-quality wines (quality > 6) and low-quality wines (quality <= 6).

Perform K-means clustering: Apply K-means clustering to group wines into clusters based on their chemical properties, specifically focusing on columns related to acidity, sugar, chlorides, sulfur dioxide, density, pH, sulphates, and alcohol.

