**REPORT**

**ON**

**WINE QUALITY ANALYSIS**

**BY: SHIVANI VERMA**

**TABLE OF CONTENT**

**SUMMER TRAINING ASSIGNMENT REPORT**

* Declaration
* Introduction
* Project Details
* Clean and prepare the data.
* Explore the data using visualizations and statistical analyses.
* distribution of the wine quality scores
* the relationship between the different features and wine quality
* the most important factors that influence the quality of wine

# 

**DECLARATION**

I , SHIVANI VERMA bearing university roll no 2000560100100 , student of Bachelor of Technology ,Computer Science and Engineering, Babu Banarasi Das Northern Indian Institute of technology , Lucknow, hereby declare that the Summer Training assignment entitled “WINE QUALITY ANALYSIS” submitted by me, is a record of bonafide work carried out independently by me under the guidance of Mr Santosh Kumar sir, trainer at IITK Foundation for Advanced Continuing Education & Training (IFACET).

I assert that all the information presented in this report is factual and accurate to the best of my knowledge and is based on my own work and experiences during the summer training.

Any external sources of information, including data, literature, and references, have been duly cited and acknowledged in accordance with the prescribed citation style.

I have not submitted this report, or any part thereof, for any other purpose, including academic evaluation, and it has not been previously published.

**Shivani Verma**

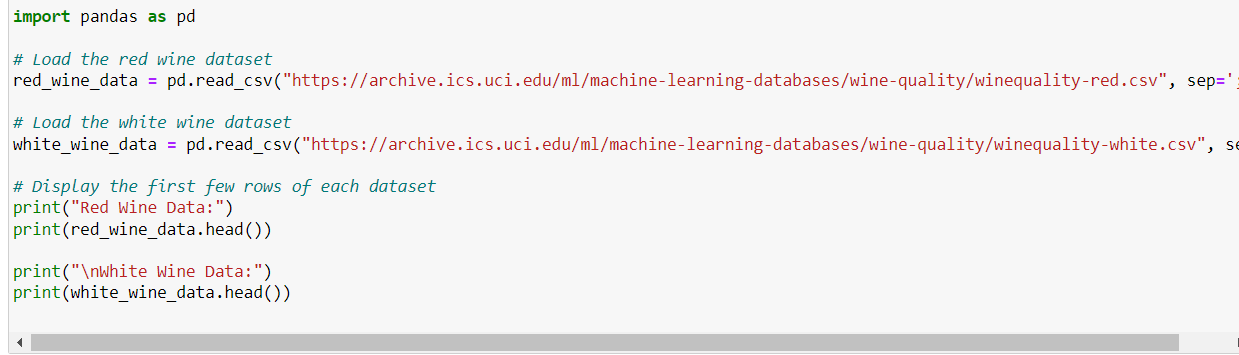
**Wine Quality Analysis Report**

Introduction

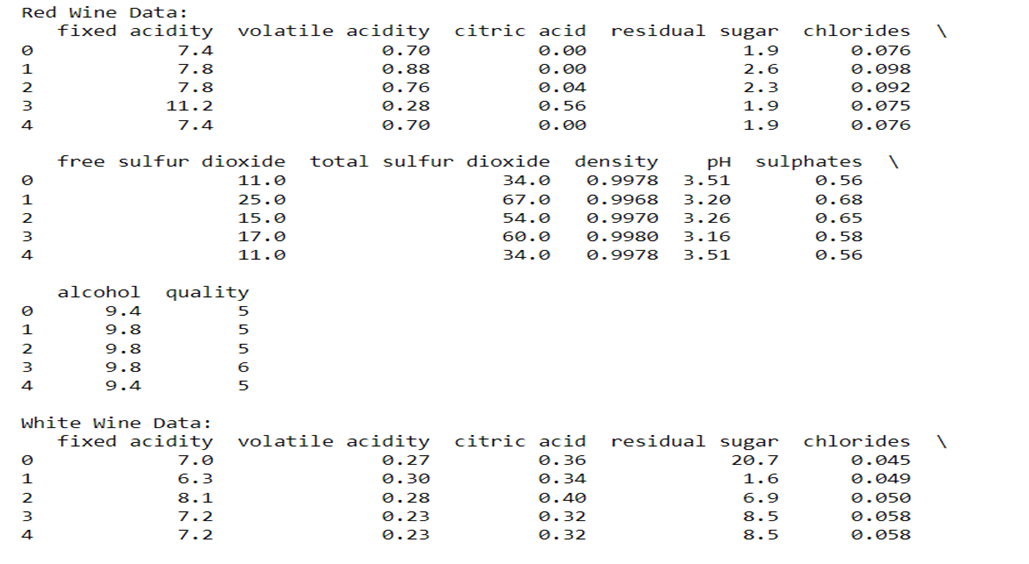
In this report, I present an analysis of the Wine Quality dataset, which comprises data on red and white wines. I have performed data exploration and visualization using Python in a Jupyter Notebook. The analysis aims to provide insights into the distribution of wine quality scores, the relationship between different features and wine quality, and the most important factors that influence the quality of wine.

Data Loading and Cleaning

I began by loading both the red and white wine datasets into my Python environment using the Pandas library. The data was obtained from the UCI Machine Learning Repository. I ensured that there were no duplicate entries in the datasets.



OUTPUT:

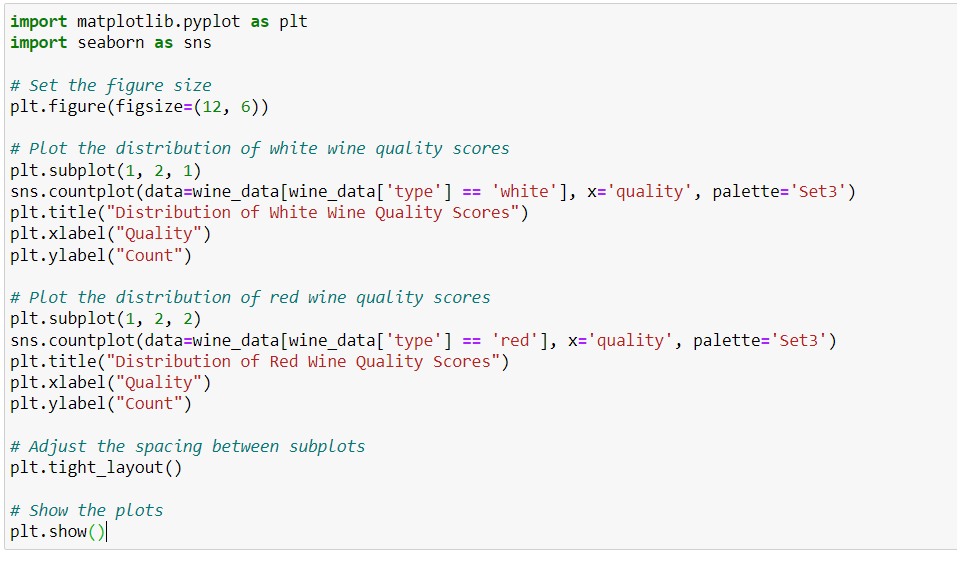


EXPLORING DATA USING VISUALIZATIONS

AND

STATISTICAL ANALYSIS

Distribution of wine quality scores:



This code generates a side-by-side comparison of the distribution of wine quality scores for white and red wines. It creates a single figure divided into two subplots. In the first subplot, it displays a countplot representing the distribution of quality scores for white wines, using a color palette defined by 'Set3'. The second subplot does the same for red wines. The plots show how the quality scores are distributed for each wine type, helping to visually compare the quality distributions between white and red wines. The 'plt.tight\_layout()' function adjusts the spacing between the subplots for better readability, and 'plt.show()' displays the resulting plots. This code allows for a quick assessment of the distribution of quality scores across different wine types.

OUTPUT:

A comparison of a bar chart

Description automatically generated

In the left subplot, a countplot is displayed for white wines, showing the distribution of quality scores. The x-axis represents different quality scores, while the y-axis indicates the count of wines with each quality rating. The color palette 'Set3' is applied to distinguish the bars representing various quality levels.

The right subplot provides a similar countplot for red wines, allowing for a direct comparison between the two wine types. Each subplot is titled to indicate whether it represents white or red wine quality scores.

These visualizations are essential for understanding the distribution of wine quality scores and how they vary between white and red wines. By comparing the two plots side by side, it becomes evident whether there are any significant differences in the quality ratings for the two wine types, providing valuable insights for further analysis and decision-making in the wine industry.

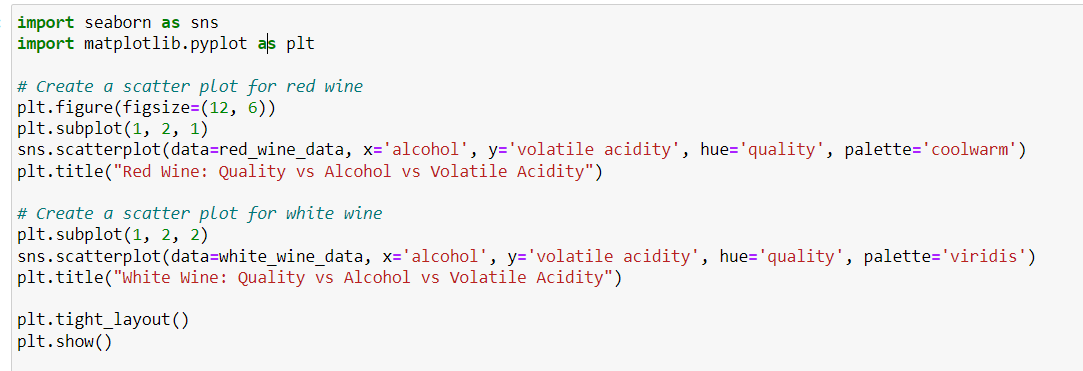
RELASHIONSHIP BETWEEN DIFFERENT FEATURES

AND

WINE QUALITY

Alcohol Content And Volatile Acidity:

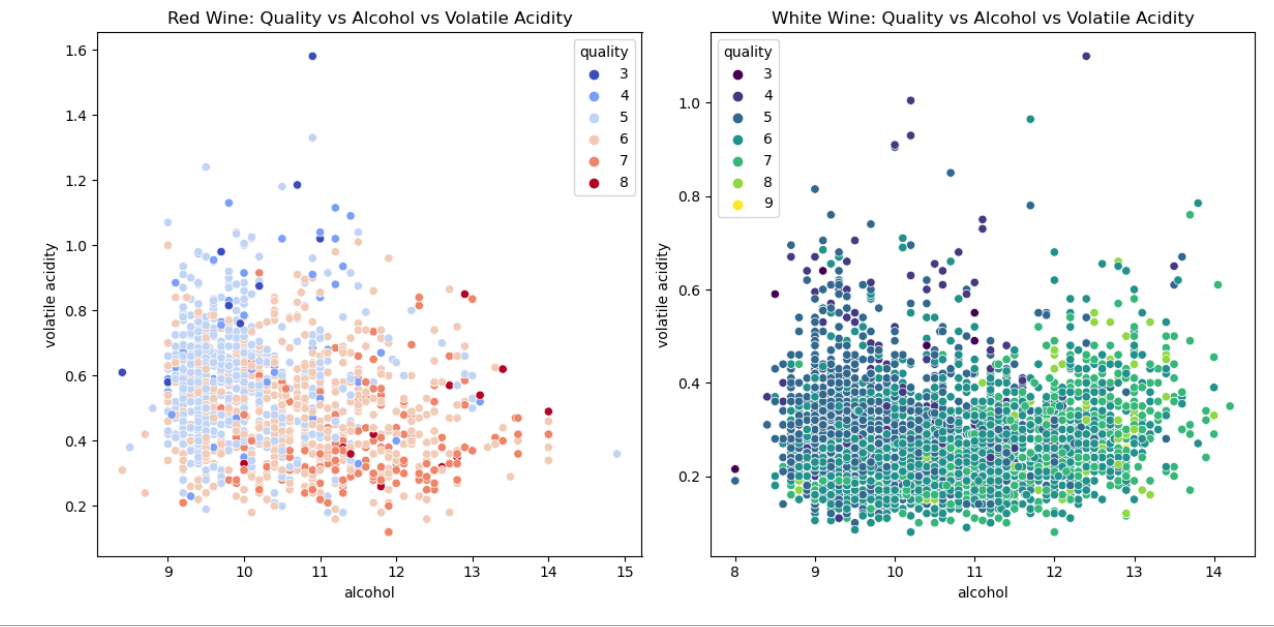
The provided code generates a pair of scatter plots that visually represent the relationship between wine quality and two key features: alcohol content and volatile acidity. These scatter plots are divided into two subplots, one for red wine and the other for white wine, enabling a direct comparison of these wine types



**Red Wine Scatter Plot (Left Subplot):**

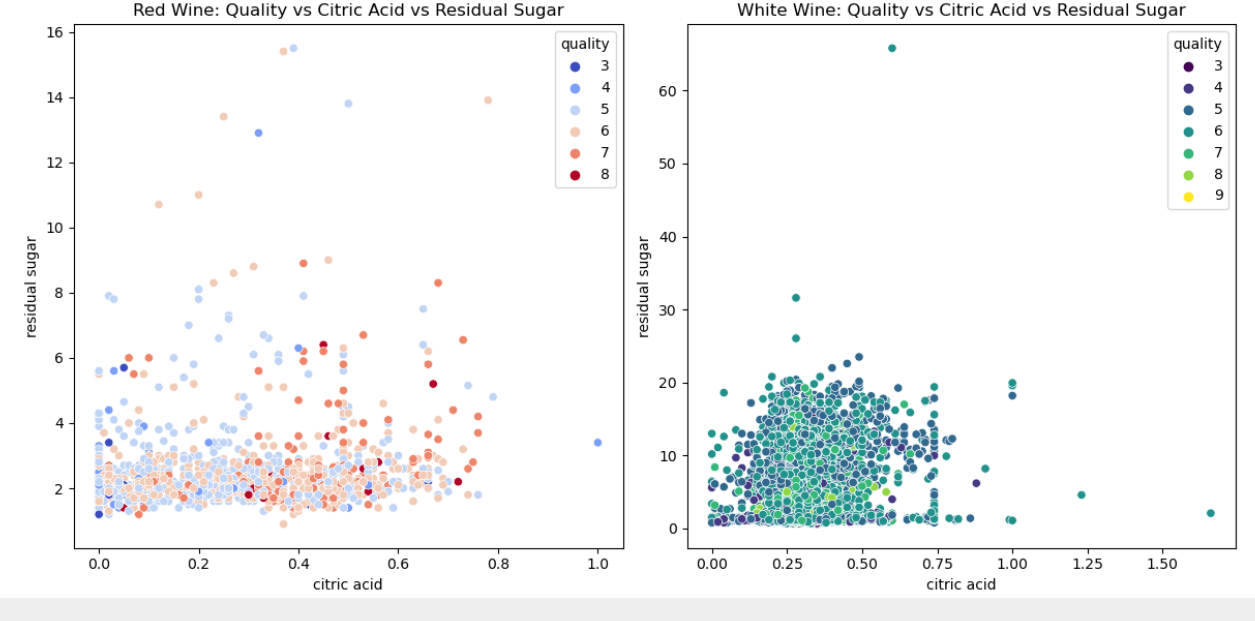
* + The left subplot shows a scatter plot for red wine.
  + The x-axis (horizontal) represents alcohol content, and the y-axis (vertical) represents volatile acidity.
  + Each data point in the scatter plot corresponds to a red wine sample.
  + The color of each data point is determined by the wine's quality, using a color palette defined by 'coolwarm.' Higher quality wines are represented by warmer colors (e.g., red or yellow), while lower quality wines are represented by cooler colors (e.g., blue or green).
  + The title of the subplot is "Red Wine: Quality vs Alcohol vs Volatile Acidity."

1. **White Wine Scatter Plot (Right Subplot):**
   * The right subplot shows a scatter plot for white wine.
   * Similar to the red wine plot, the x-axis represents alcohol content, and the y-axis represents volatile acidity.
   * Each data point in the scatter plot corresponds to a white wine sample.
   * The color of each data point is determined by the wine's quality, using a color palette defined by 'viridis.'
   * The title of the subplot is "White Wine: Quality vs Alcohol vs Volatile Acidity."



By observing these scatter plots, one can visually assess whether there are discernible patterns or trends in the relationship between wine quality and the combination of alcohol content and volatile acidity. The color differentiation allows for a quick identification of high-quality and low-quality wines based on these two features.

SIMILARY:



A screenshot of a graph

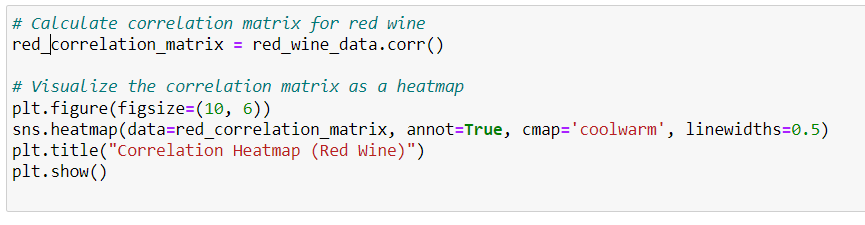
Description automatically generated

A graph of a quality and chloridant

Description automatically generated with medium confidence

**Correlation Matrix Calculation:**

First, the below code calculates a correlation matrix specifically for the attributes of red wine. This matrix quantifies the degree and direction of relationships between pairs of attributes. Each cell in the matrix represents the correlation coefficient, which ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation. Positive values suggest a positive relationship, while negative values suggest a negative relationship.

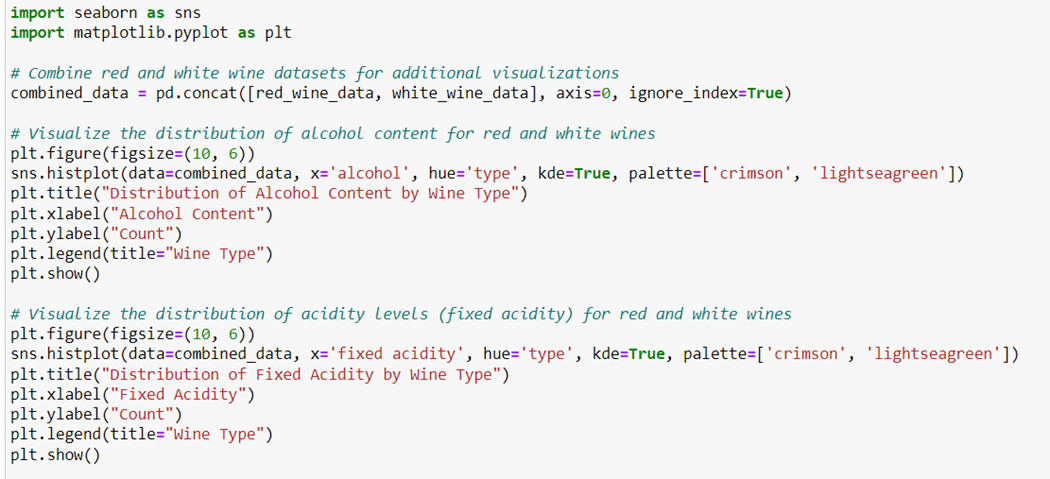


A screenshot of a data analysis

Description automatically generated

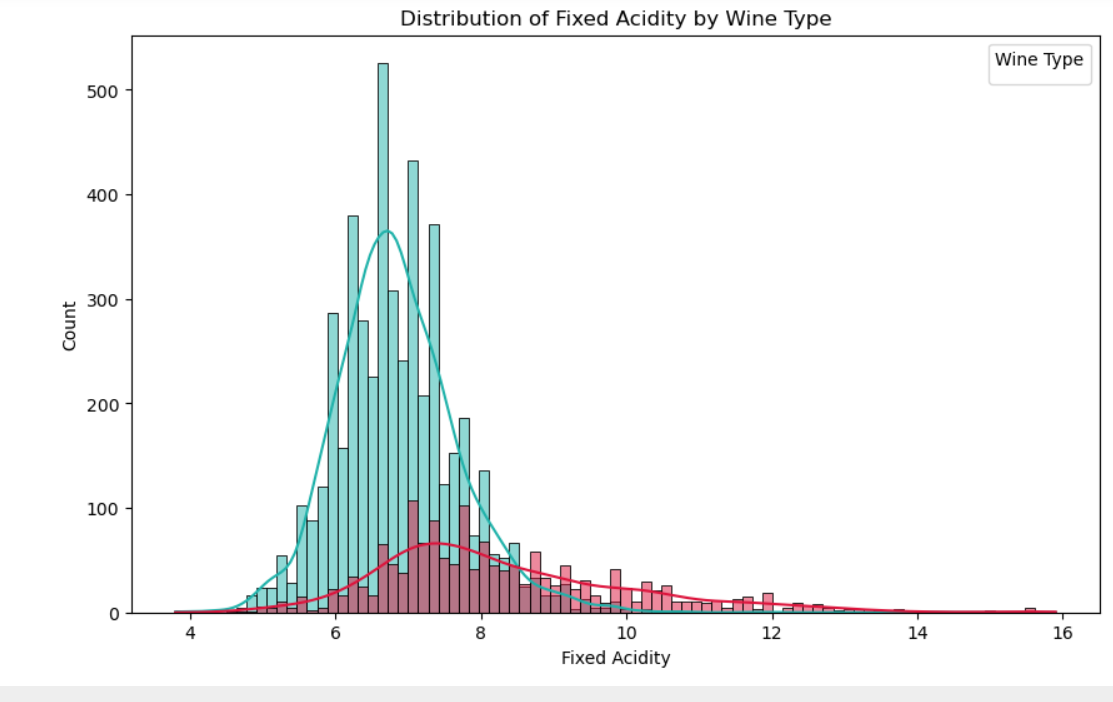
ALCOHOL CONTENT AND FIXED ACIDITY

The provided code generates two separate visualizations, each comparing characteristics of red and white wines, specifically focusing on alcohol content and fixed acidity levels:



A graph of alcohol content

Description automatically generated



1. **Distribution of Alcohol Content by Wine Type:**
   * The first visualization consists of two histograms, one for red wine and the other for white wine, shown side by side. The x-axis represents alcohol content, and the y-axis shows the count of wines with specific alcohol levels.
   * The histograms are color-coded, with 'crimson' representing red wine and 'lightseagreen' representing white wine. This color distinction makes it easy to differentiate between the two wine types.
   * Additionally, kernel density estimates (KDE) are superimposed on the histograms, providing smoother representations of the alcohol content distributions.
   * The title of this visualization is "Distribution of Alcohol Content by Wine Type."
2. **Distribution of Fixed Acidity by Wine Type:**
   * The second visualization follows a similar format but focuses on fixed acidity levels for red and white wines.
   * It presents two histograms side by side, with the x-axis representing fixed acidity and the y-axis displaying the count of wines with specific acidity levels.
   * Like the previous visualization, the histograms are color-coded, with 'crimson' for red wine and 'lightseagreen' for white wine. Kernel density estimates are also included.
   * The title of this visualization is "Distribution of Fixed Acidity by Wine Type."

These visualizations provide an insightful comparison of alcohol content and fixed acidity levels between red and white wines. By visualizing these distributions, you can quickly identify any differences or similarities in these key characteristics, helping in the assessment of each wine type's unique properties. The color differentiation and KDEs aid in the interpretation of these distributions.

MOST IMPORTANT FACTORS

THAT INFLUENCE

QUALITY OF WINE

A bar graph with different colored bars

Description automatically generated