**Clustering and Fitting Analysis of Wine Quality Data**

**Name:** Sreeya kannamala  
**Student ID:** 23094968  
**GitHub Repository:** [GitHub Link](https://github.com/your-repo)

**Introduction**

This report analyzes a wine quality dataset by applying clustering and linear regression techniques. The dataset contains various chemical properties of wines, such as fixed acidity, volatile acidity, alcohol, and quality scores. The objectives of this study are:

1. Grouping wines into clusters based on their chemical properties.
2. Predicting wine quality using alcohol levels through linear regression.
3. Visualizing the relationships and findings using multiple plots.

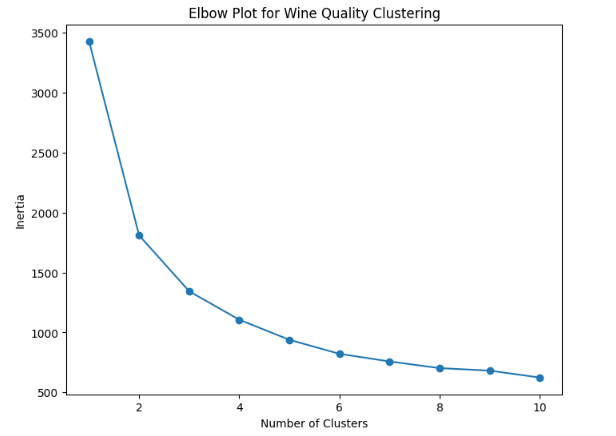
This analysis provides insights into the chemical factors influencing wine quality and clusters wines with similar characteristics.

**Elbow Plot (K-means Clustering)**

The Elbow Plot is used to determine the optimal number of clusters for the K-means algorithm. It displays the inertia (sum of squared distances) as the number of clusters increases.

**Plot Description:**

The elbow point is observed at 3 clusters, indicating the optimal number of clusters. Beyond this point, the reduction in inertia becomes insignificant, suggesting diminishing returns from additional clusters.



**Scatter Plot (Cluster Visualization)**

K-means clustering is performed using fixed acidity, volatile acidity, and citric acid. The resulting clusters are visualized in a scatter plot.

**Plot Description:**

The scatter plot highlights three distinct wine clusters:

* **Cluster 1:** Wines with higher fixed acidity and lower volatile acidity.
* **Cluster 2:** Wines with moderate acidity levels.
* **Cluster 3:** Wines with low fixed acidity and higher volatile acidity.

These clusters provide meaningful groupings based on chemical compositions.

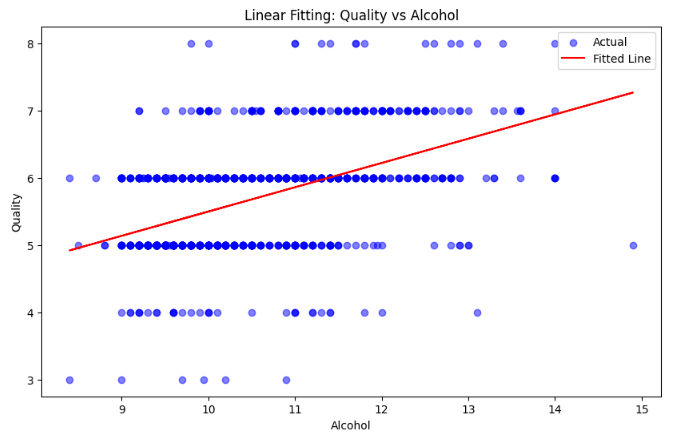


**Linear Regression (Quality vs. Alcohol)**

Linear regression is applied to study the relationship between alcohol levels and wine quality. The fitted regression line indicates the trend.

**Plot Description:**

* **Trend**: A positive correlation is observed between alcohol and quality; higher alcohol levels generally correspond to better quality scores.
* **Insights**: The regression line effectively captures this trend, supporting alcohol as a significant predictor of wine quality.



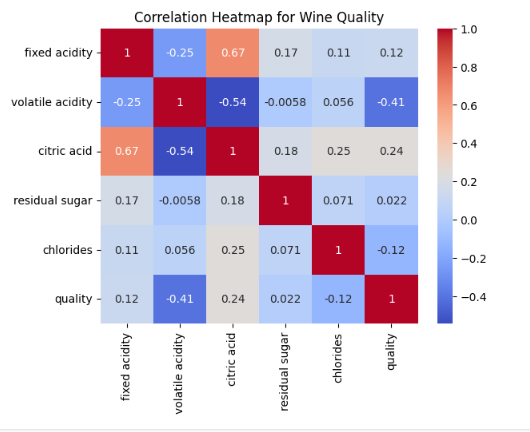
**Heatmap (Correlation Analysis)**

A heatmap is generated to visualize the correlations between key features, such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, and quality.

**Plot Description:**

* **Key Findings**:
  + Quality has a moderate positive correlation with alcohol.
  + Fixed acidity shows a weak correlation with quality but is positively correlated with citric acid.
  + Volatile acidity negatively correlates with quality, indicating its potential role in reducing wine scores.

This heatmap highlights the most influential features affecting wine quality.



**Conclusion**

1. **Clustering**: K-means clustering segmented wines into three groups based on their chemical properties, providing insights into wine characteristics.
2. **Regression Analysis**: Alcohol levels were found to be a strong predictor of wine quality, with a clear positive relationship.
3. **Correlation Analysis**: The heatmap revealed additional insights into the relationships between chemical properties and quality.

These findings help understand wine quality determinants and group wines with similar profiles for better categorization.

**GitHub Repository:** [GitHub Link](https://github.com/your-repo)