**QUESTION AND ANSWER:**

**1. What is the distribution of the wine quality scores?**

- The distribution of wine quality scores is right-skewed or positively skewed. This means that there are more data points to the left (lower scores) of the distribution than to the right (higher scores). In a right-skewed distribution, the mean is typically greater than the median.

**2. What are the relationships between the different features?**

- Through data visualization using histograms, scatter plots, and line graphs, we've observed that the data exhibits a common relationship among all the features. The majority of wines fall within the quality range of 5-7, and their feature values are concentrated in the range of 0-0.4. There are relatively few data points outside these ranges, indicating a consistent relationship among the features.

**3. Are there any outliers in the data?**

- Yes, there are outliers in the data. We identified outliers using box plot visualization. Out of 1599 rows, 141 rows contained extreme or outlier values. After data cleaning, we were left with 1458 rows that were free from outliers.

**4. What is the accuracy of the linear regression model?**

- The accuracy of the linear regression model is 78%.

**5. What are the most important features for the linear regression model?**

- The most important features for the linear regression model are Volatile Acidity, Alcohol, and Sulphates.

**6. What is the Mean Squared Error (MSE) of the linear regression model?**

- The Mean Squared Error (MSE) of the linear regression model is 0.003589.

**7. What is the R-squared (R2) score of the linear regression model?**

- The R-squared (R2) score of the linear regression model is 0.786592.

**8. How can you improve the performance of the linear regression mo**del?

- To enhance the performance of the linear regression model, we recommend the following steps:

1. Data Cleaning: Remove null values and outliers.
2. Data Normalization: Normalize the data.
3. Feature Scaling: Scale the features.
4. Feature Selection: Select the most relevant features.

**9. What are the limitations of the linear regression model?**

- The limitations of the linear regression model include:

1. Limited Expressiveness: Linear regression may not capture complex relationships, such as nonlinear or interactive effects.
2. Non-Normality of Residuals: It assumes normally distributed residuals; deviations from normality can affect hypothesis testing.
3. Homoscedasticity Assumption: It assumes constant variance of residuals; heteroscedasticity can lead to incorrect standard errors.

**10. What are the implications of your findings for the real-world problem?**

- The findings suggest that linear regression can be a useful tool for modeling wine quality based on selected features. However, it's important to consider the limitations and explore other modeling techniques for more complex relationships or improved accuracy in predicting wine quality. Further analysis and domain knowledge can provide valuable insights for real-world applications in the wine industry.