

P2_Visualization

April 17, 2023

1 VISUALIZATION

```
[1]: import pandas as pd
import matplotlib.pyplot as plt

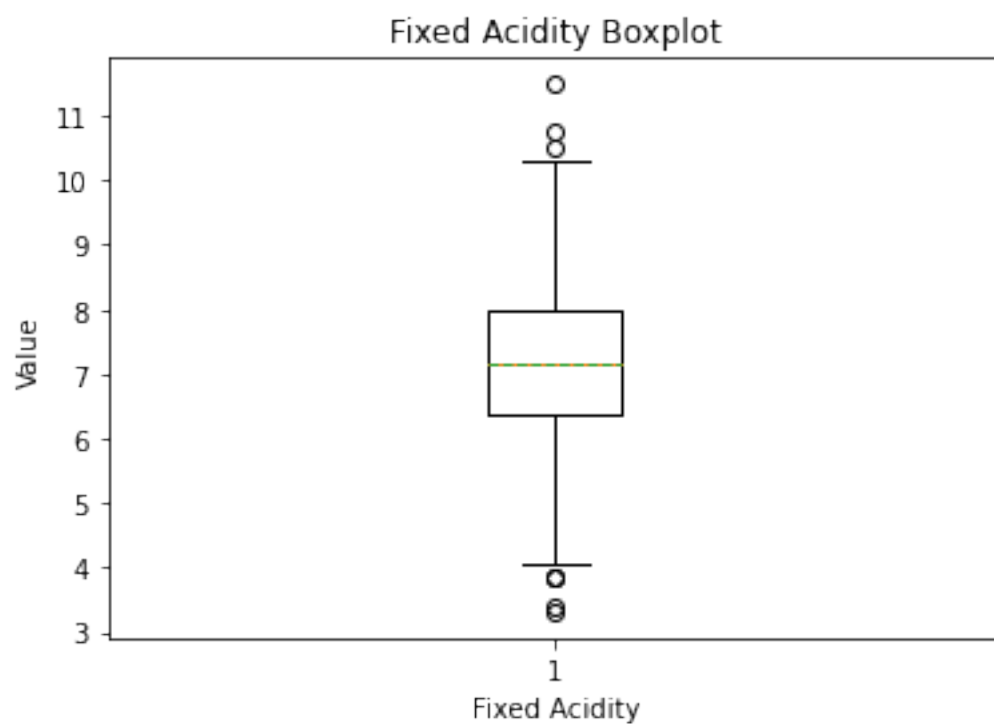
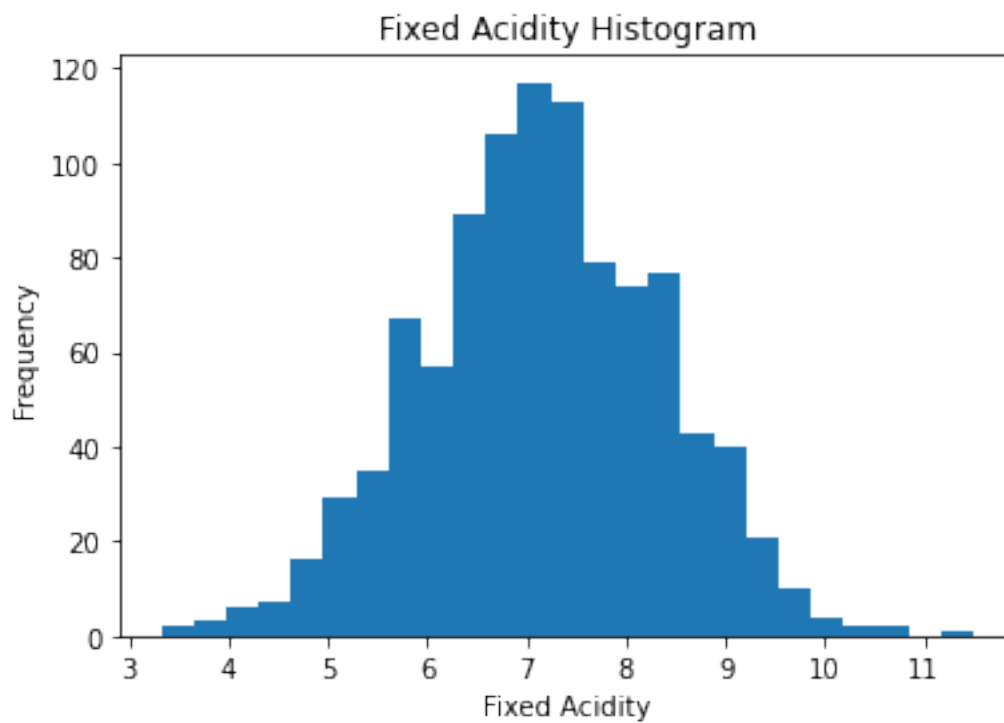
# Read csv file
df = pd.read_csv("../data/anggur.csv")
```

1.1 Kolom Fixed Acidity

```
[2]: df_fixed_acidity = df["fixed acidity"]

# Histogram
plt.hist(df_fixed_acidity, bins = 25)
plt.title('Fixed Acidity Histogram')
plt.xlabel('Fixed Acidity')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_fixed_acidity, showmeans = True, meanline = True)
plt.title('Fixed Acidity Boxplot')
plt.xlabel('Fixed Acidity')
plt.ylabel('Value')
plt.show()
```



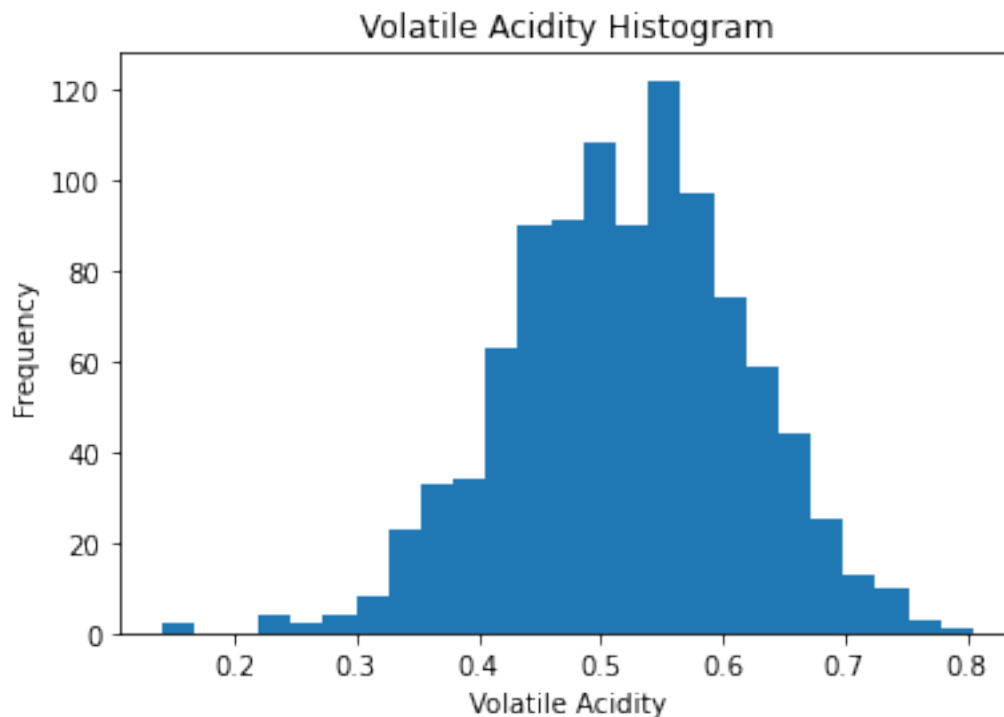
Berdasarkan histogram tersebut, terlihat bahwa distribusi Fixed Acidity cenderung condong ke kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 3-4 dan 10-12. Terlihat juga bahwa kuartil bawahnya berada di sekitar 6.2, kuartil tengahnya berada di sekitar 7.1, dan kuartil atasnya berada di sekitar 8. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 7.1 (sedikit di atas median).

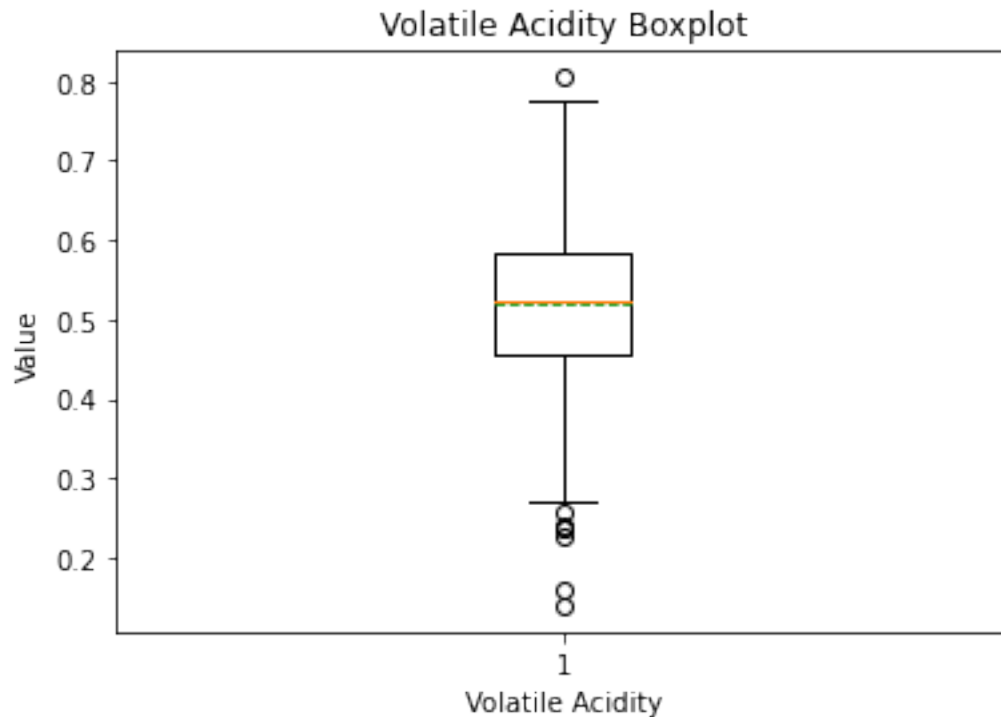
1.2 Kolom Volatile Acidity

```
[3]: df_volatile_acidity = df["volatile acidity"]

# Histogram
plt.hist(df_volatile_acidity, bins = 25)
plt.title('Volatile Acidity Histogram')
plt.xlabel('Volatile Acidity')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_volatile_acidity, showmeans = True, meanline = True)
plt.title('Volatile Acidity Boxplot')
plt.xlabel('Volatile Acidity')
plt.ylabel('Value')
plt.show()
```





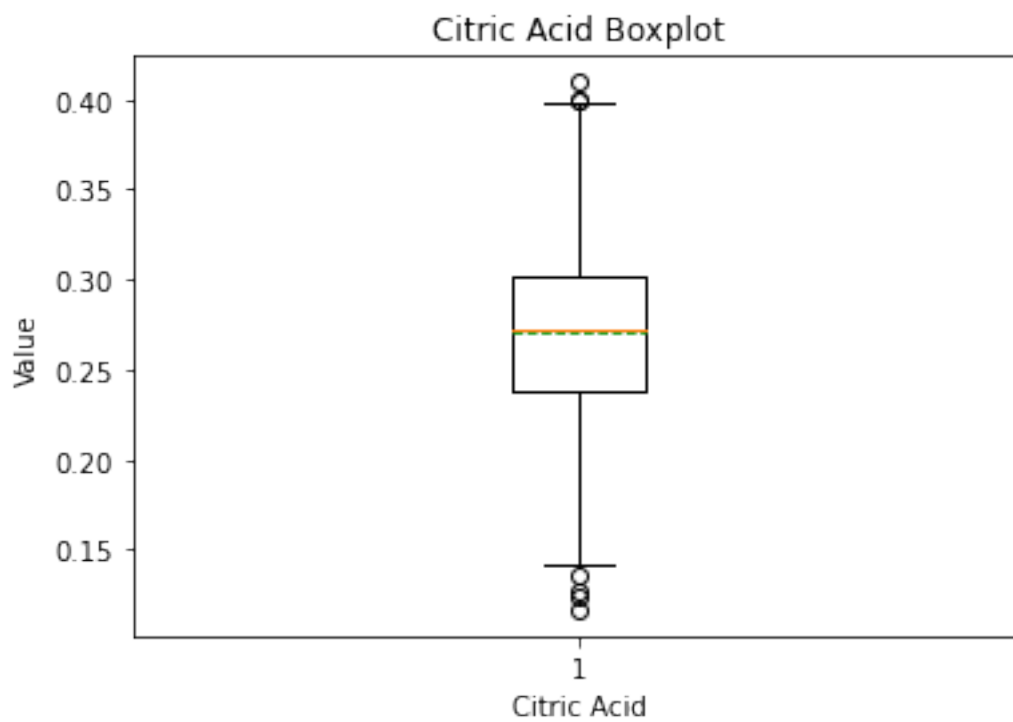
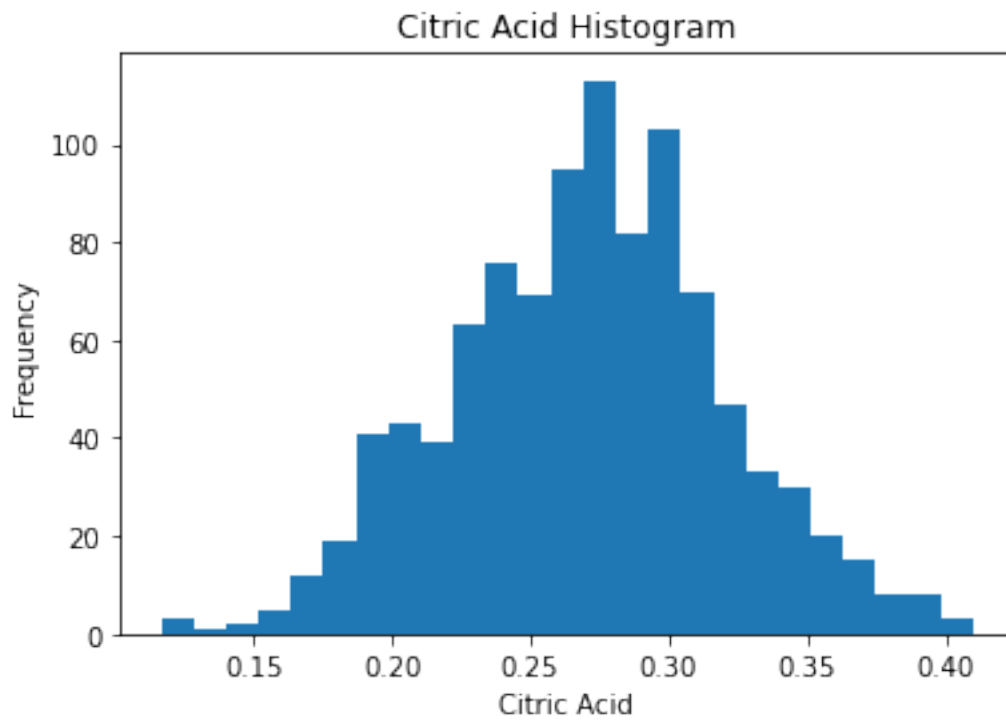
Berdasarkan histogram tersebut, terlihat bahwa distribusi Volatile Acidity cenderung condong ke kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 0.1-0.28 dan 0.78-0.83. Terlihat juga bahwa kuartil bawahnya berada di sekitar 0.46, kuartil tengahnya berada di sekitar 0.53, dan kuartil atasnya berada di sekitar 0.58. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 0.53 (sedikit di bawah median).

1.3 Kolom Citric Acid

```
[4]: df_citric_acid = df["citric acid"]

# Histogram
plt.hist(df_citric_acid, bins = 25)
plt.title('Citric Acid Histogram')
plt.xlabel('Citric Acid')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_citric_acid, showmeans = True, meanline = True)
plt.title('Citric Acid Boxplot')
plt.xlabel('Citric Acid')
plt.ylabel('Value')
plt.show()
```



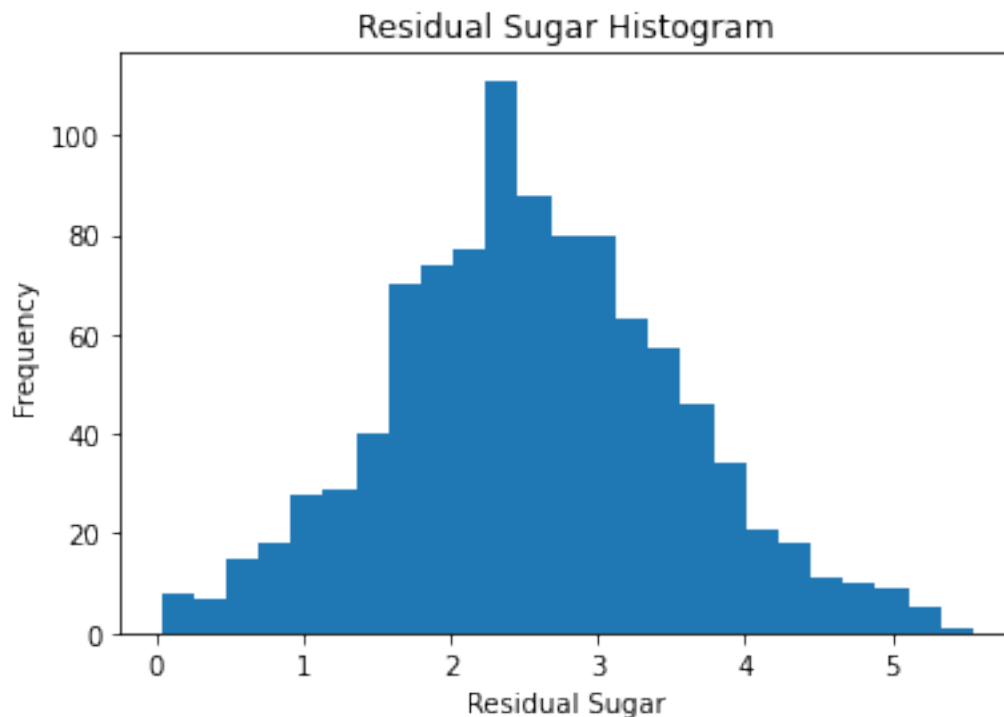
Berdasarkan histogram tersebut, terlihat bahwa distribusi Citric Acid cenderung condong ke arah kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 0.11-0.15 dan 0.4-0.43. Terlihat juga bahwa kuartil bawahnya berada di sekitar 0.24, kuartil tengahnya berada di sekitar 0.27, dan kuartil atasnya berada di sekitar 0.30. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 0.27 (sedikit di bawah median).

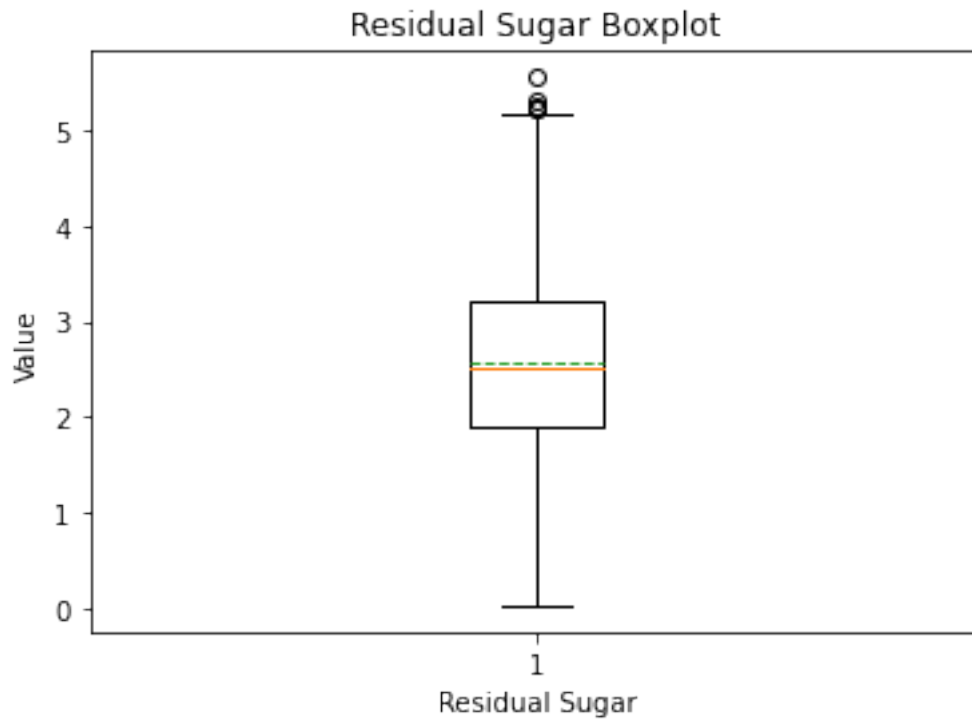
1.4 Kolom Residual Sugar

```
[5]: df_residual_sugar = df["residual sugar"]

# Histogram
plt.hist(df_residual_sugar, bins = 25)
plt.title('Residual Sugar Histogram')
plt.xlabel('Residual Sugar')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_residual_sugar, showmeans = True, meanline = True)
plt.title('Residual Sugar Boxplot')
plt.xlabel('Residual Sugar')
plt.ylabel('Value')
plt.show()
```





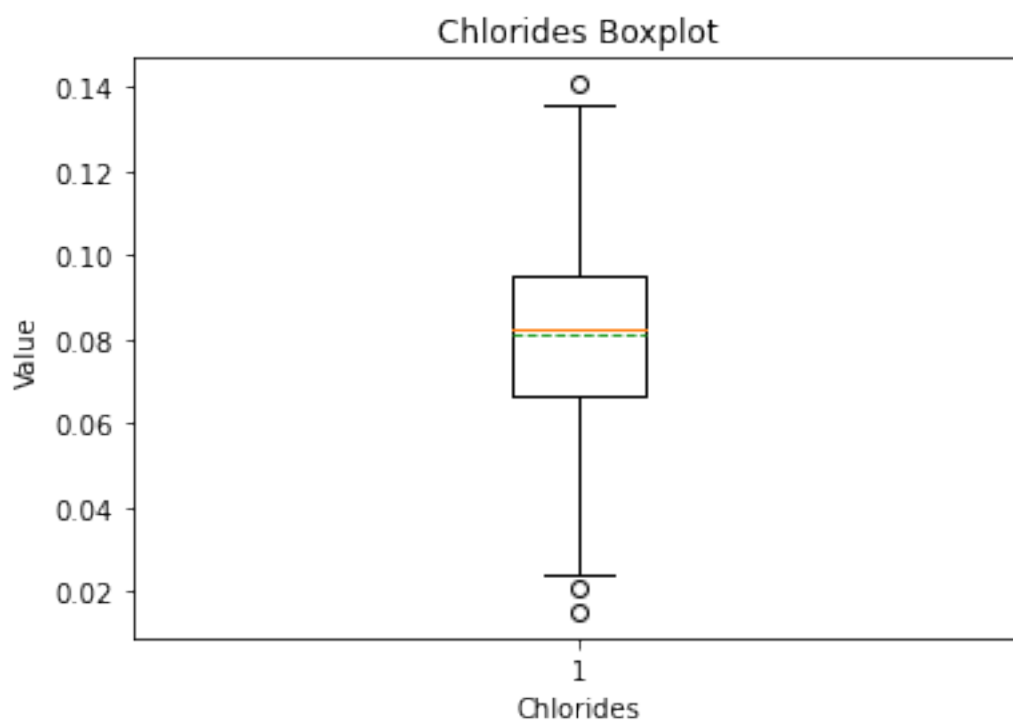
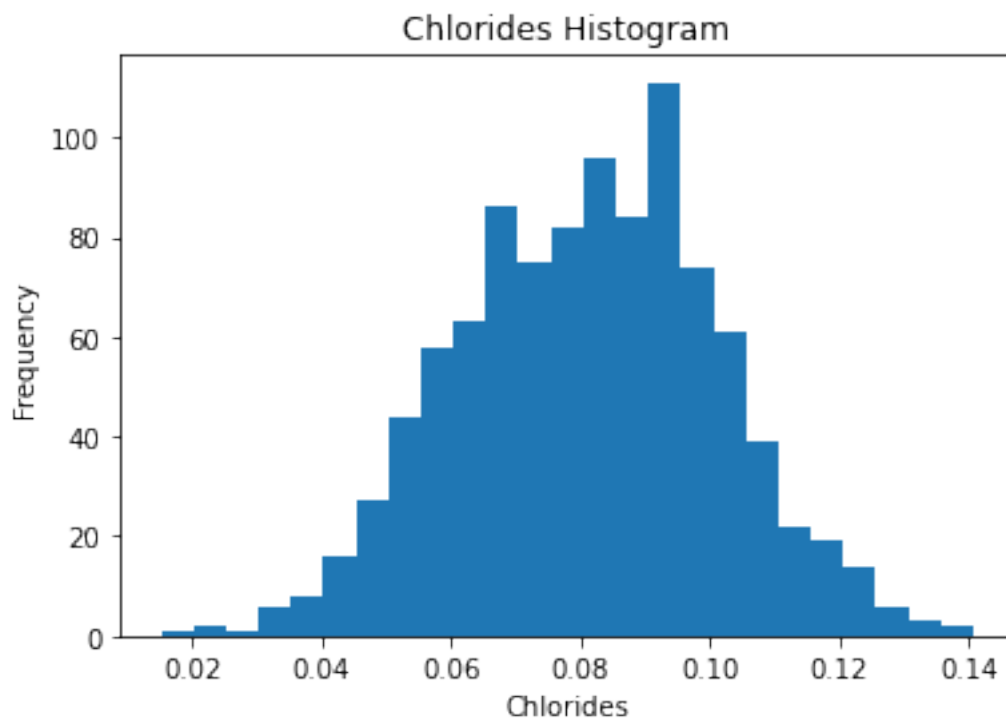
Berdasarkan histogram tersebut, terlihat bahwa distribusi Residual Sugar cenderung condong ke arah kanan (positively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 5-6. Terlihat juga bahwa kuartil bawahnya berada di sekitar 1.9, kuartil tengahnya berada di sekitar 2.51, dan kuartil atasnya berada di sekitar 3.2. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 2.55 (sedikit di atas median).

1.5 Kolom Chlorides

```
[6]: df_chlorides = df["chlorides"]

# Histogram
plt.hist(df_chlorides, bins = 25)
plt.title('Chlorides Histogram')
plt.xlabel('Chlorides')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_chlorides, showmeans = True, meanline = True)
plt.title('Chlorides Boxplot')
plt.xlabel('Chlorides')
plt.ylabel('Value')
plt.show()
```



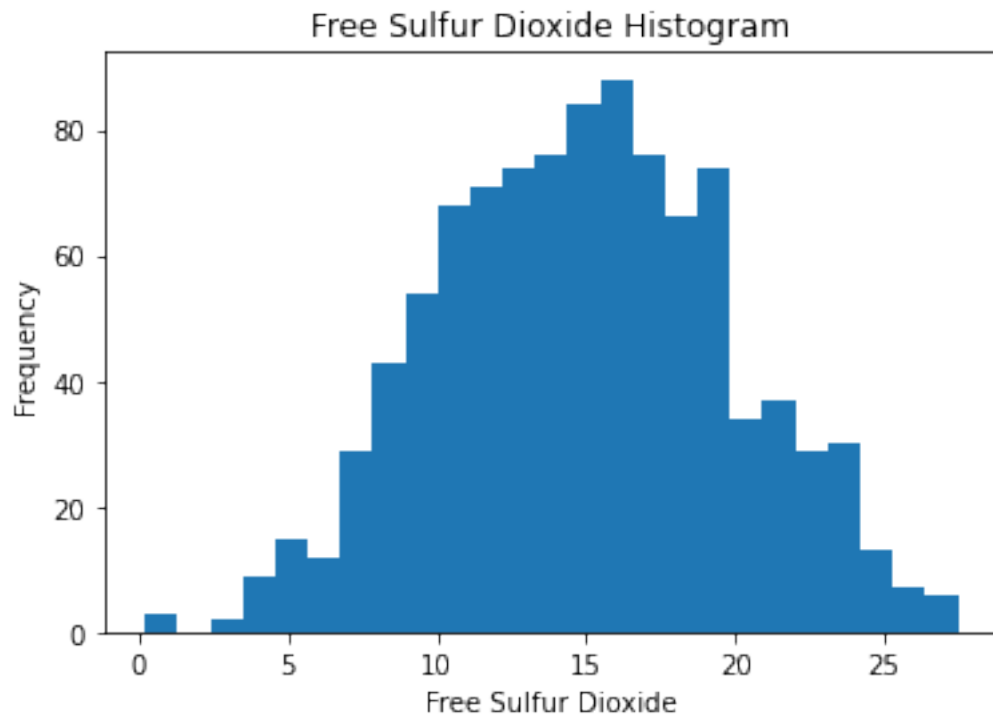
Berdasarkan histogram tersebut, terlihat bahwa distribusi Chlorides cenderung condong ke arah kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 0.01-0.03 dan 0.13-0.141. Terlihat juga bahwa kuartil bawahnya berada di sekitar 0.067, kuartil tengahnya berada di sekitar 0.082, dan kuartil atasnya berada di sekitar 0.096. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 0.081 (sedikit di bawah median).

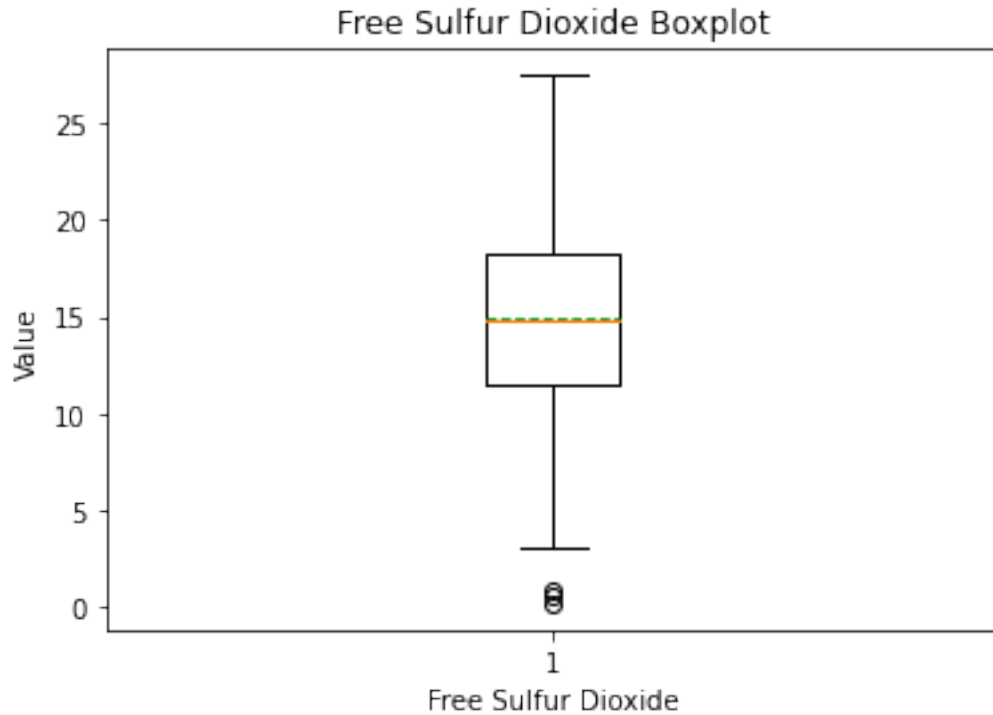
1.6 Kolom Free Sulfur Dioxide

```
[7]: df_free_sulfur_dioxide = df["free sulfur dioxide"]

# Histogram
plt.hist(df_free_sulfur_dioxide, bins = 25)
plt.title('Free Sulfur Dioxide Histogram')
plt.xlabel('Free Sulfur Dioxide')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_free_sulfur_dioxide, showmeans = True, meanline = True)
plt.title('Free Sulfur Dioxide Boxplot')
plt.xlabel('Free Sulfur Dioxide')
plt.ylabel('Value')
plt.show()
```





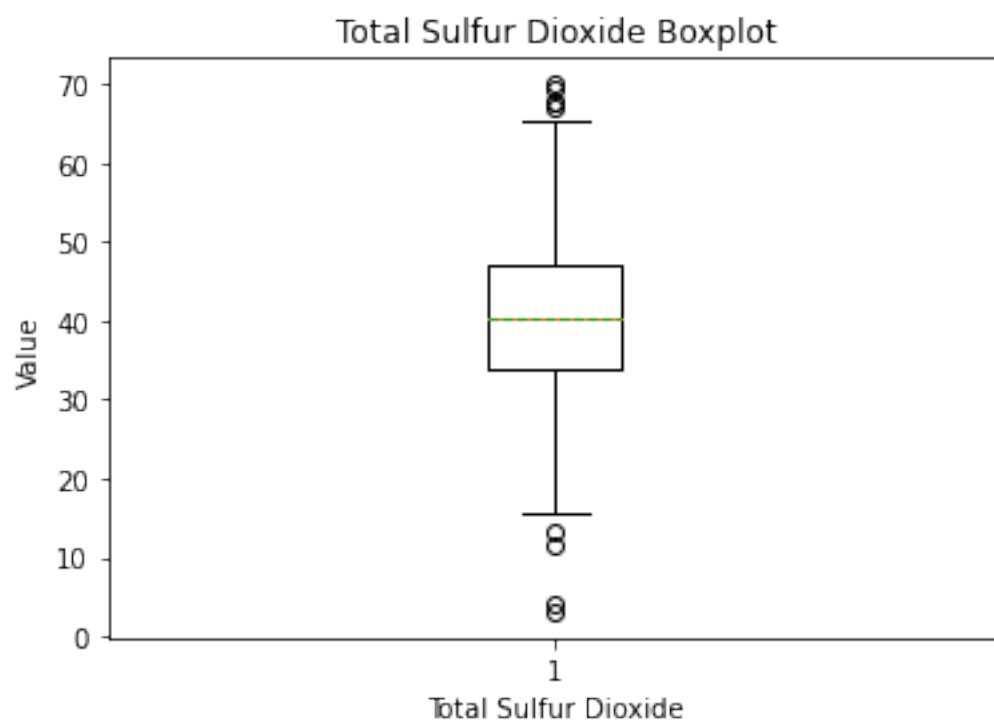
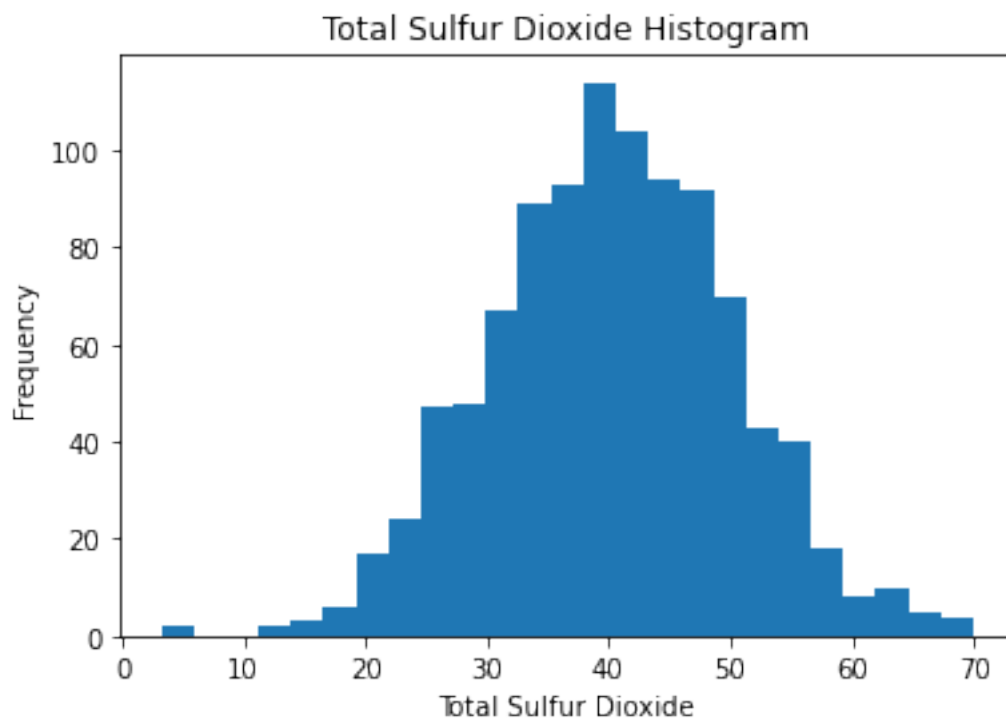
Berdasarkan histogram tersebut, terlihat bahwa distribusi Free Sulfur Dioxide cenderung mendekati distribusi normal karena bentuknya hampir simetris. Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 0-2.5. Terlihat juga bahwa kuartil bawahnya berada di sekitar 11.4, kuartil tengahnya berada di sekitar 14.85, dan kuartil atasnya berada di sekitar 18.3. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 14.9 (sedikit di atas median).

1.7 Kolom Total Sulfur Dioxide

```
[8]: df_total_sulfur_dioxide = df["total sulfur dioxide"]

# Histogram
plt.hist(df_total_sulfur_dioxide, bins = 25)
plt.title('Total Sulfur Dioxide Histogram')
plt.xlabel('Total Sulfur Dioxide')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_total_sulfur_dioxide, showmeans = True, meanline = True)
plt.title('Total Sulfur Dioxide Boxplot')
plt.xlabel('Total Sulfur Dioxide')
plt.ylabel('Value')
plt.show()
```



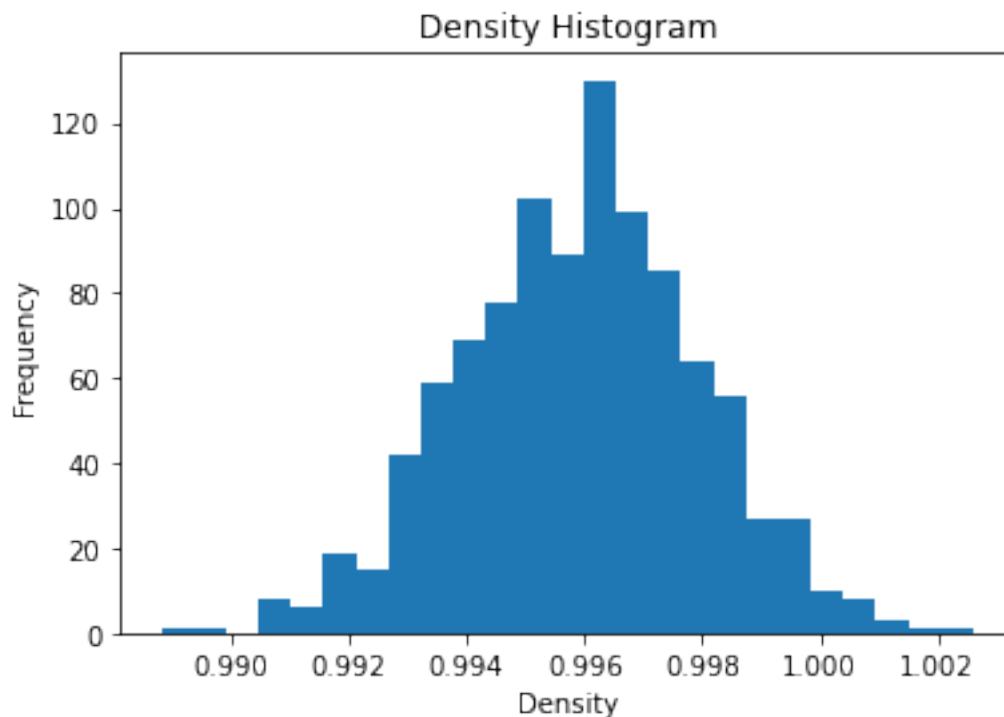
Berdasarkan histogram tersebut, terlihat bahwa distribusi Total Sulfur Dioxide cenderung condong ke kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 3-15 dan 65-70. Terlihat juga bahwa kuartil bawahnya berada di sekitar 33.5, kuartil tengahnya berada di sekitar 40.2, dan kuartil atasnya berada di sekitar 47. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 40.2 (sedikit di atas median).

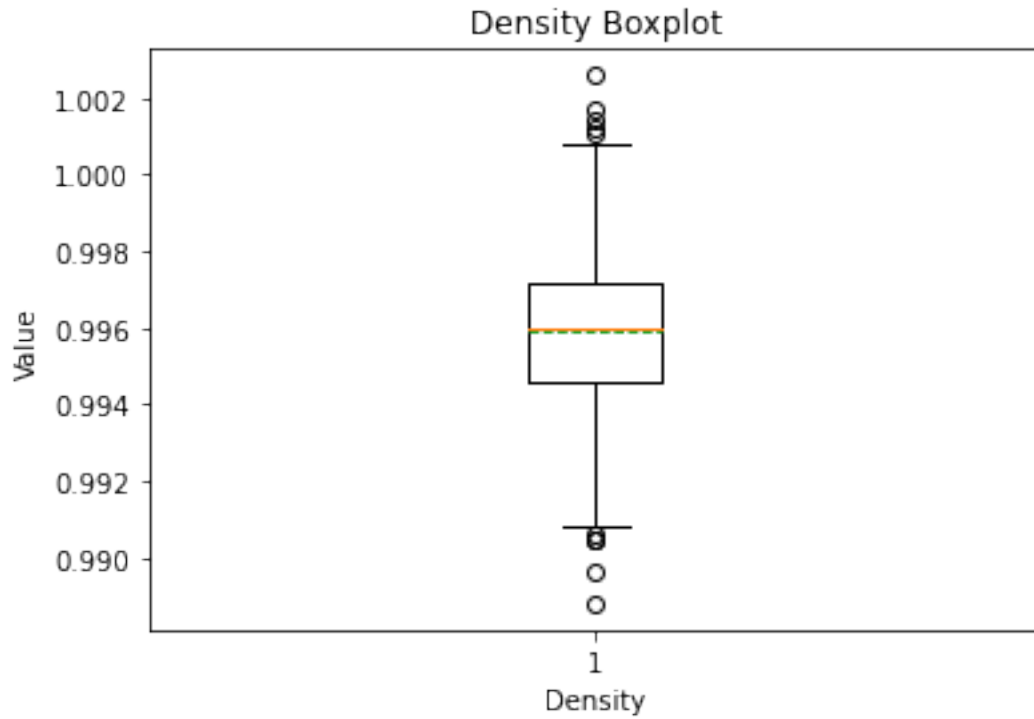
1.8 Kolom Density

```
[9]: df_density = df["density"]

# Histogram
plt.hist(df_density, bins = 25)
plt.title('Density Histogram')
plt.xlabel('Density')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_density, showmeans = True, meanline = True)
plt.title('Density Boxplot')
plt.xlabel('Density')
plt.ylabel('Value')
plt.show()
```





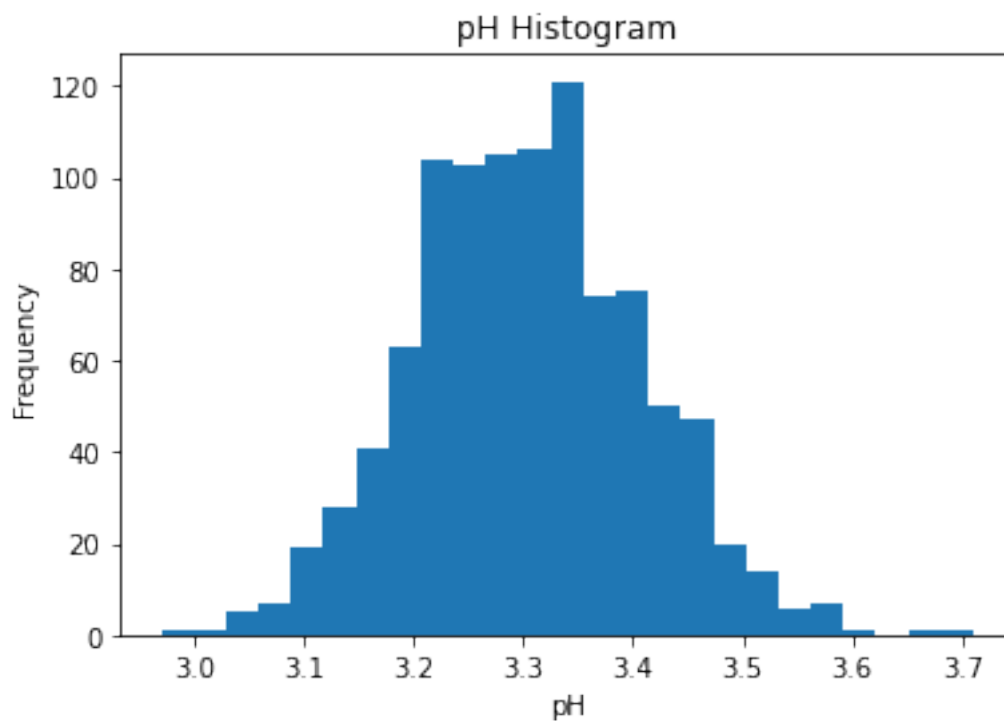
Berdasarkan histogram tersebut, terlihat bahwa distribusi Density cenderung condong ke kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 0.989-0.991 dan 1.001-1.003. Terlihat juga bahwa kuartil bawahnya berada di sekitar 0.995, kuartil tengahnya berada di sekitar 0.996, dan kuartil atasnya berada di sekitar 0.997. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 0.996 (sedikit di bawah median).

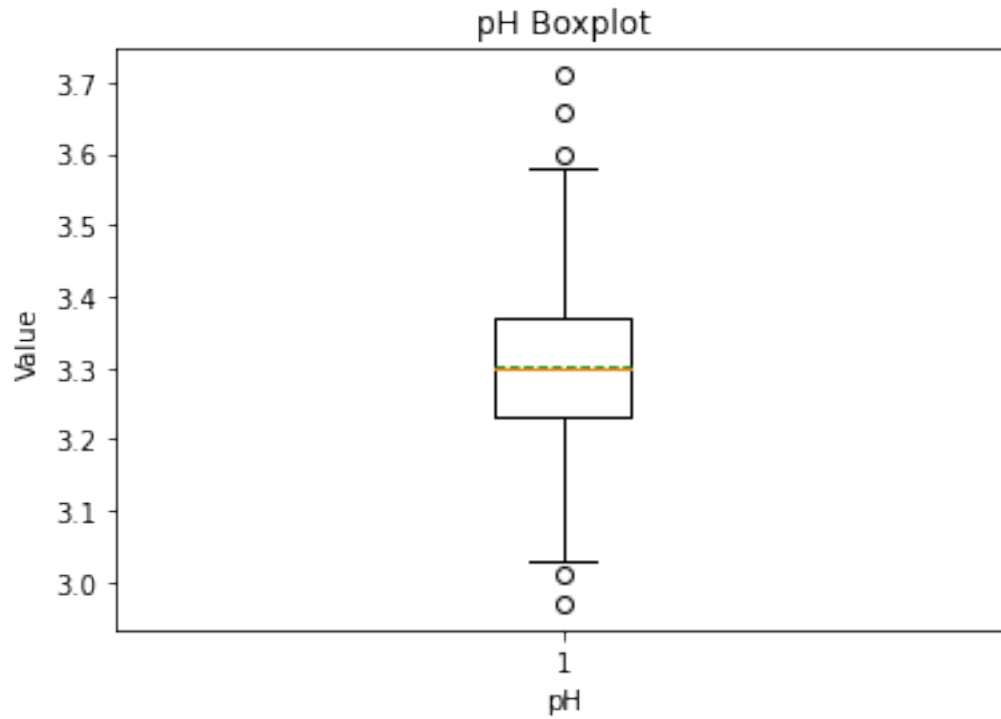
1.9 Kolom pH

```
[10]: df_pH = df["pH"]

# Histogram
plt.hist(df_pH, bins = 25)
plt.title('pH Histogram')
plt.xlabel('pH')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_pH, showmeans = True, meanline = True)
plt.title('pH Boxplot')
plt.xlabel('pH')
plt.ylabel('Value')
plt.show()
```





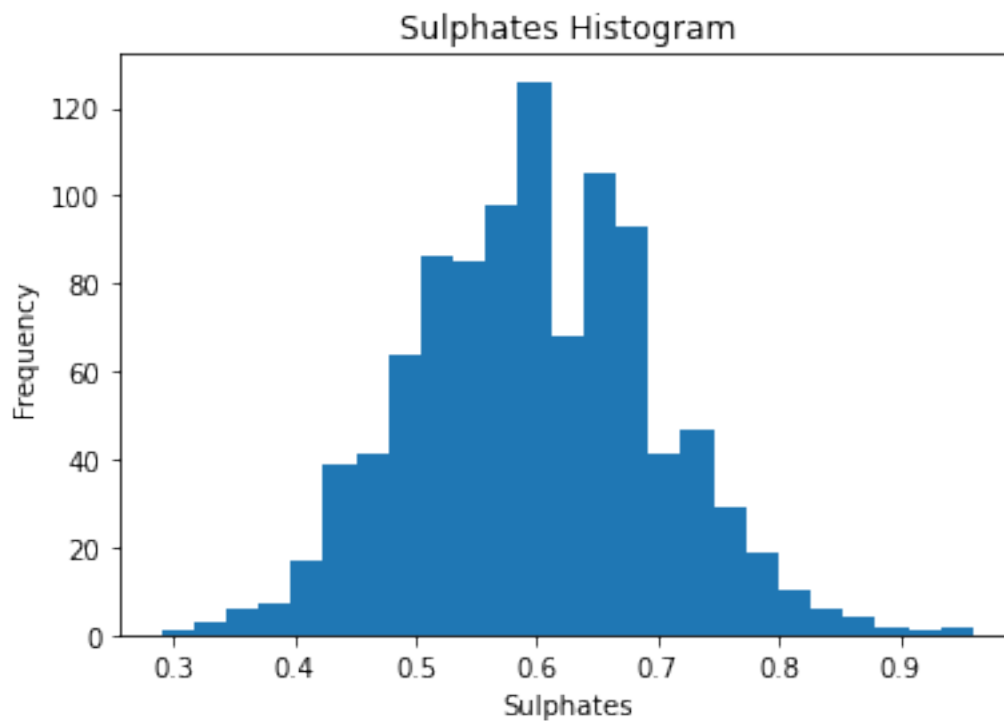
Berdasarkan histogram tersebut, terlihat bahwa distribusi pH cenderung condong ke kanan (positively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 2.95-3.05 dan 3.58-3.73. Terlihat juga bahwa kuartil bawahnya berada di sekitar 3.24, kuartil tengahnya berada di sekitar 3.3, dan kuartil atasnya berada di sekitar 3.37. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 3.3 (sedikit di atas median).

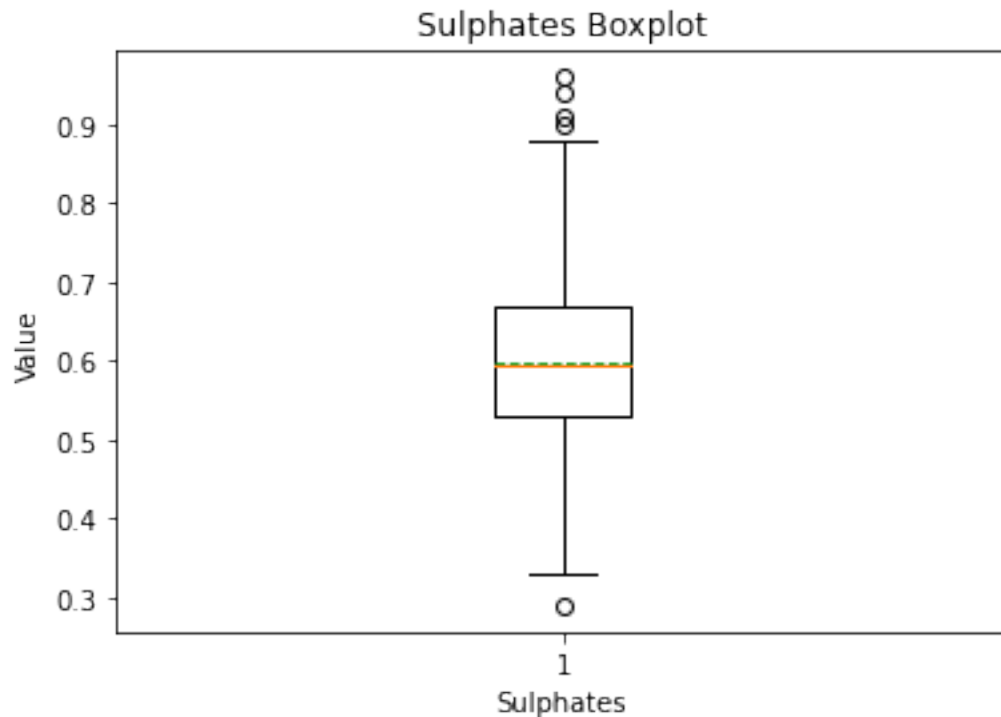
1.10 Kolom Sulphates

```
[11]: df_sulphates = df["sulphates"]

# Histogram
plt.hist(df_sulphates, bins = 25)
plt.title('Sulphates Histogram')
plt.xlabel('Sulphates')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_sulphates, showmeans = True, meanline = True)
plt.title('Sulphates Boxplot')
plt.xlabel('Sulphates')
plt.ylabel('Value')
plt.show()
```





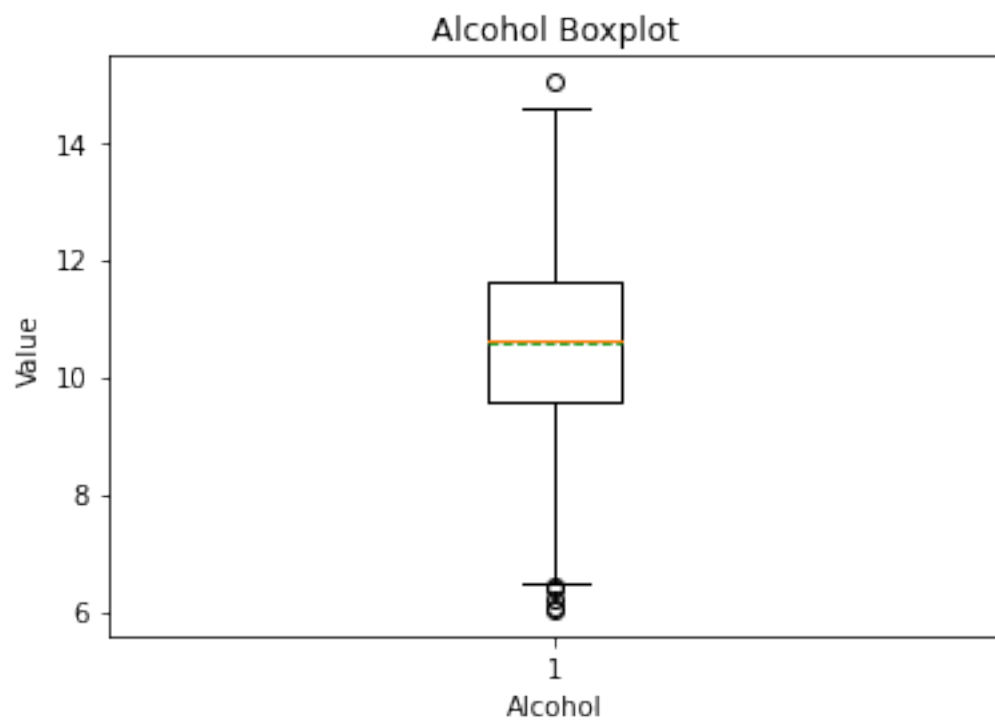
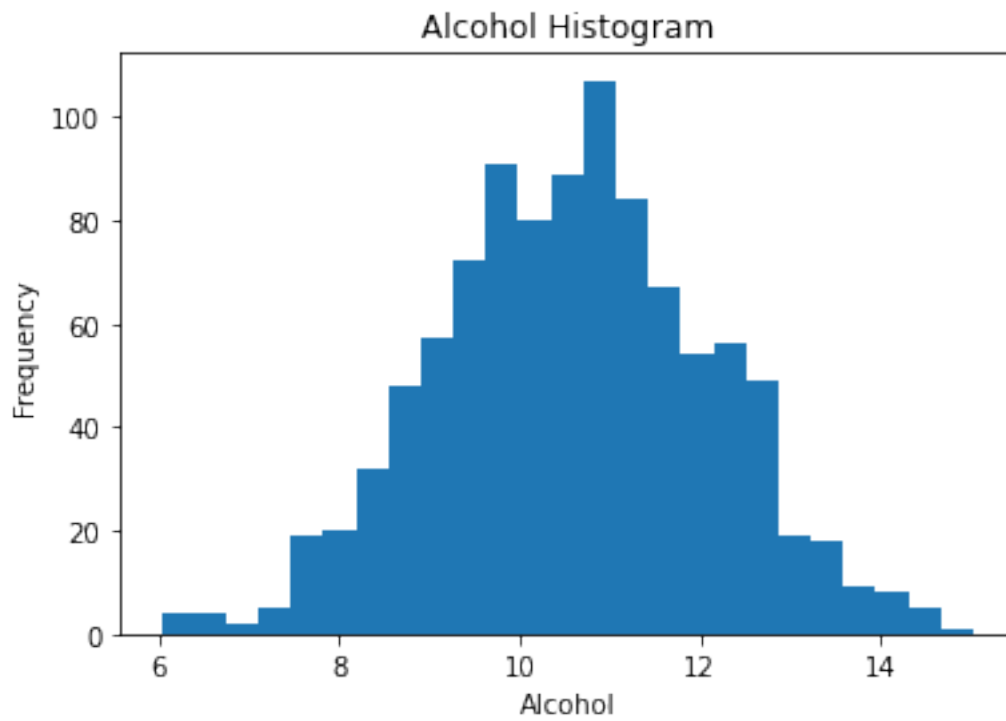
Berdasarkan histogram tersebut, terlihat bahwa distribusi Sulphates cenderung condong ke kanan (positively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 0.29-0.31 dan 0.87-0.98. Terlihat juga bahwa kuartil bawahnya berada di sekitar 0.53, kuartil tengahnya berada di sekitar 0.6, dan kuartil atasnya berada di sekitar 0.67. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 0.6 (sedikit di atas median).

1.11 Kolom Alcohol

```
[12]: df_alcohol = df["alcohol"]

# Histogram
plt.hist(df_alcohol, bins = 25)
plt.title('Alcohol Histogram')
plt.xlabel('Alcohol')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_alcohol, showmeans = True, meanline = True)
plt.title('Alcohol Boxplot')
plt.xlabel('Alcohol')
plt.ylabel('Value')
plt.show()
```



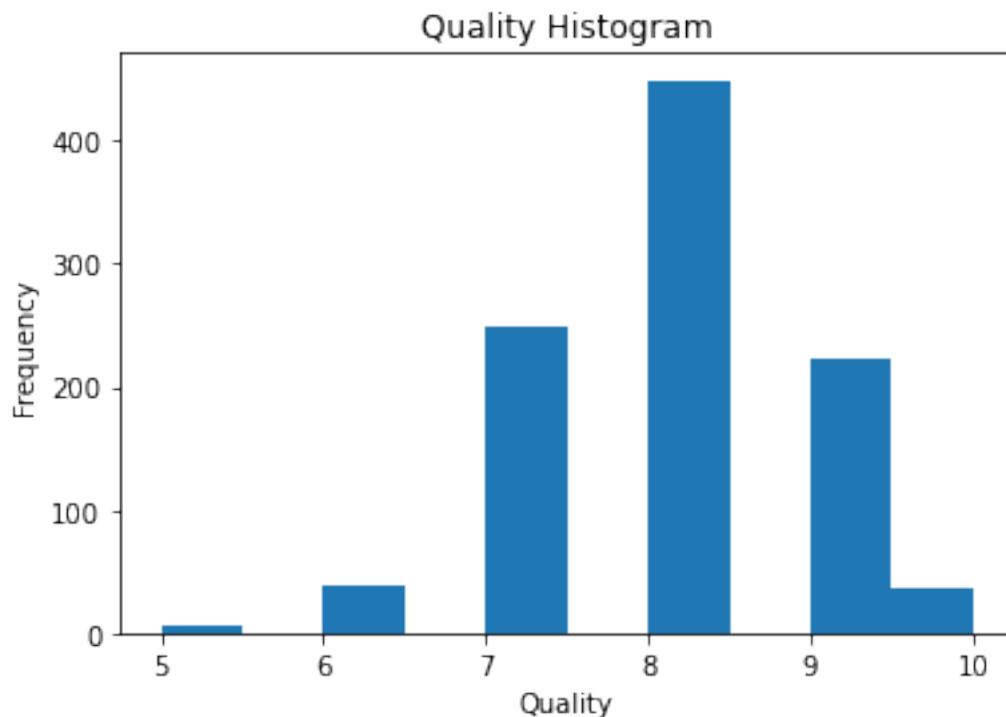
Berdasarkan histogram tersebut, terlihat bahwa distribusi Alcohol cenderung condong ke kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa beberapa data outlier berada dalam rentang sekitar 6-7 dan 14.5-15.1. Terlihat juga bahwa kuartil bawahnya berada di sekitar 9.6, kuartil tengahnya berada di sekitar 10.6, dan kuartil atasnya berada di sekitar 11.7. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 10.6 (sedikit di bawah median).

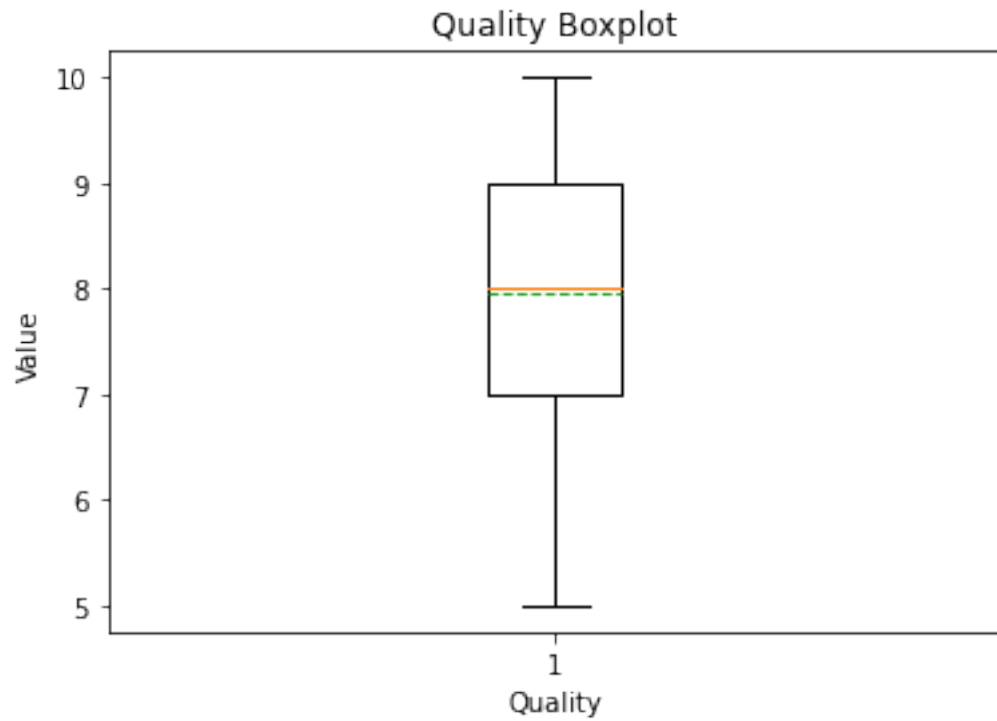
1.12 Kolom Quality

```
[13]: df_quality = df["quality"]

# Histogram
plt.hist(df_quality, bins = 10)
plt.title('Quality Histogram')
plt.xlabel('Quality')
plt.ylabel('Frequency')
plt.show()

# Boxplot
plt.boxplot(df_quality, showmeans = True, meanline = True)
plt.title('Quality Boxplot')
plt.xlabel('Quality')
plt.ylabel('Value')
plt.show()
```





Berdasarkan histogram tersebut, terlihat bahwa distribusi Quality cenderung condong ke kiri (negatively skewed). Selain itu, berdasarkan boxplot tersebut, terlihat bahwa tidak ada data outlier. Terlihat juga bahwa kuartil bawahnya berada di 7, kuartil tengahnya berada di 8, dan kuartil atasnya berada di 9. Meannya juga terlihat mendekati mediannya (kuartil tengah), yaitu sekitar 8 (sedikit di bawah median).