Nomor 1

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from scipy.stats import shapiro
```

In [ ]:
```
df = pd.read_csv('banana.csv')
df = df.drop(df.columns[0], axis=1)
df_num = df.select_dtypes(include=['number'])
df_str = df.select_dtypes(include=['object'])
```

In [ ]:
```
# df_num
IQR = {}
for column in df_num.columns:
  mode = stats.mode(df[column])
  print(column)
  print("Mean:", np.mean(df[column]))
  print("Median:", np.median(df[column]))
  print(f"Mode: {mode.mode} with {mode.count} occurences")
  print("Standard Deviation:", np.std(df[column]))
  print("Variance:", np.std(df[column]))
  print("Range:", np.ptp(df[column]))
  print("Min:", np.min(df[column]))
  print("Max:", np.max(df[column]))
  print("Q1:", np.percentile(df[column], 25))
  print("Q2:", np.percentile(df[column], 50))
  print("Q3:", np.percentile(df[column], 75))
  IQR[column] = np.percentile(df[column], 75) - np.percentile(df[column], 25)
  print("IQR:", IQR[column])
  print("Skewness:", stats.skew(df[column]))
  print("Kurtosis:", stats.kurtosis(df[column]))
  print()
```

Acidity
Mean: 8.014829628574718
Median: 8.005346574809552
Mode: 4.45611756155056 with 1 occurences
Standard Deviation: 1.1055048384965227
Variance: 1.1055048384965227
Range: 6.962518106354413
Min: 4.45611756155056
Max: 11.418635667904972
Q1: 7.259942415758844
Q2: 8.005346574809552
Q3: 8.758360762028273
IQR: 1.4984183462694292
Skewness: 0.05675052446787658
Kurtosis: -0.14976509292535845

Weight
Mean: 150.01154884061205
Median: 150.02286495781368
Mode: 146.06092220645047 with 1 occurences
Standard Deviation: 1.194681336922938
Variance: 1.194681336922938
Range: 8.009448077902334
Min: 146.06092220645047
Max: 154.0703702843528
Q1: 149.22711597797812
Q2: 150.02286495781368
Q3: 150.8276131389365
IQR: 1.6004971609583833
Skewness: -0.0847035441413555
Kurtosis: 0.021906412951779863

Length
Mean: 49.950434190968366
Median: 49.92368168187882
Mode: 46.418051592321525 with 1 occurences
Standard Deviation: 0.894375128686565
Variance: 0.894375128686565
Range: 6.6470993126980105
Min: 46.418051592321525
Max: 53.065150905019536
Q1: 49.34650755405077
Q2: 49.92368168187882
Q3: 50.57202687332001
IQR: 1.225519319269246
Skewness: 0.026858246315712334
Kurtosis: -0.05641464936749019

Appearance
Mean: 4.965594844869035
Median: 4.979533900424856
Mode: 1.7758640691873886 with 1 occurences
Standard Deviation: 1.0146090288990814
Variance: 1.0146090288990814
Range: 6.4581042192309805
Min: 1.7758640691873886

Max: 8.23396828841837
Q1: 4.2582095909622035
Q2: 4.979533900424856
Q3: 5.653875089976802
IQR: 1.3956654990145987
Skewness: -0.035362764473791115
Kurtosis: -0.005182153795968869

Tannin
Mean: 7.9654345706813405
Median: 8.02244846073121
Mode: 4.291273644087031 with 1 occurences
Standard Deviation: 1.2168833131537398
Variance: 1.2168833131537398
Range: 8.124903699290943
Min: 4.291273644087031
Max: 12.416177343377974
Q1: 7.167241274395545
Q2: 8.02244846073121
Q3: 8.792184148144969
IQR: 1.6249428737494238
Skewness: -0.06610241984906381
Kurtosis: 0.06318470232487527

Ripeness
Mean: 6.743433741509038
Median: 6.667617895134861
Mode: 4.862560328528312 with 1 occurences
Standard Deviation: 0.6801502035387855
Variance: 0.6801502035387855
Range: 4.619505546832993
Min: 4.862560328528312
Max: 9.482065875361306
Q1: 6.268257732243829
Q2: 6.667617895134861
Q3: 7.164812693580822
IQR: 0.8965549613369923
Skewness: 0.49522533656734163
Kurtosis: 0.27450916357317734

Sweetness
Mean: 6.226318700769242
Median: 6.312819065940845
Mode: 3.0331926464036347 with 1 occurences
Standard Deviation: 0.6628143613473609
Variance: 0.6628143613473609
Range: 4.6454959575079435
Min: 3.0331926464036347
Max: 7.678688603911578
Q1: 5.808027606800032
Q2: 6.312819065940845
Q3: 6.714660017511698
IQR: 0.9066324107116666
Skewness: -0.663193740375262
Kurtosis: 0.49087968766482026

```
Firmness
Mean: 0.5077895564331691
Median: 0.5154825330154527
Mode: 0.0002540323592254 with 1 occurences
Standard Deviation: 0.29215285905065547
Variance: 0.29215285905065547
Range: 1.9997459676407745
Min: 0.0002540323592254
Max: 2.0
Q1: 0.2543505163028975
Q2: 0.5154825330154527
Q3: 0.7587860576103286
IQR: 0.5044355413074311
Skewness: 0.0248544881761829
Kurtosis: -0.9056381260038568

Price
Mean: 19969.669241238404
Median: 19999.50831225203
Mode: 0.0 with 2 occurences
Standard Deviation: 777.1531031529252
Variance: 777.1531031529252
Range: 20282.431061619714
Min: -1.0
Max: 20281.431061619714
Q1: 19953.093529464502
Q2: 19999.50831225203
Q3: 20047.30194881691
IQR: 94.20841935240605
Skewness: -25.45013154938886
Kurtosis: 650.9997480753707
```

In [ ]:

```python
# df_str
for column in df_str.columns:
  print(f"{column}'s Unique Value:")
  print(df_str[column].unique())
  print("\nProportion:")
  print(df_str[column].value_counts(normalize=True))
  print()
```

```
Country_of_Origin's Unique Value:
['Costa Rica' 'Colombia' 'Ecuador' 'undefined']

Proportion:
Country_of_Origin
Ecuador       0.5605
Costa Rica    0.2850
Colombia      0.1530
undefined     0.0015
Name: proportion, dtype: float64

Grade's Unique Value:
['A' 'C' 'B']

Proportion:
Grade
A    0.3415
C    0.3390
B    0.3195
Name: proportion, dtype: float64
```