

April 18, 2023

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

dataAnggur = pd.read_csv('../data/anggur.csv')
```

Membuat Visualisasi plot distribusi, dalam bentuk histogram dan boxplot untuk setiap kolom numerik. Berikan uraian penjelasan kondisi setiap kolom berdasarkan kedua plot tersebut.

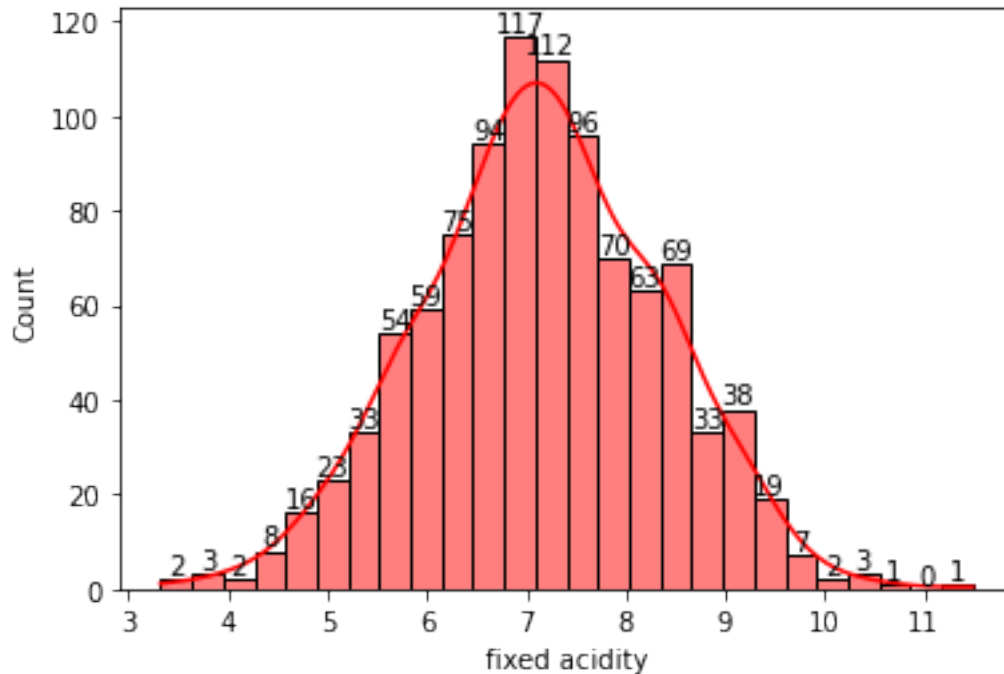
0.0.1 1. fixed acidity

```
[2]: # Data
dataFixedAcidity = dataAnggur['fixed acidity']

[3]: # ===== Histogram =====
ax = sns.histplot(dataFixedAcidity, color='red', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [3.32, 3.6342307692307685, 3.948461538461538,
4.2626923076923084, 4.576923076923077, 4.891153846153847, 5.205384615384615,
5.519615384615385, 5.833846153846153, 6.148076923076923, 6.462307692307691,
6.776538461538461, 7.090769230769231, 7.404999999999999, 7.719230769230769,
8.033461538461538, 8.347692307692308, 8.661923076923078, 8.976153846153846,
9.290384615384614, 9.604615384615386, 9.918846153846154, 10.233076923076922,
10.547307692307692, 10.861538461538462, 11.17576923076923]
```

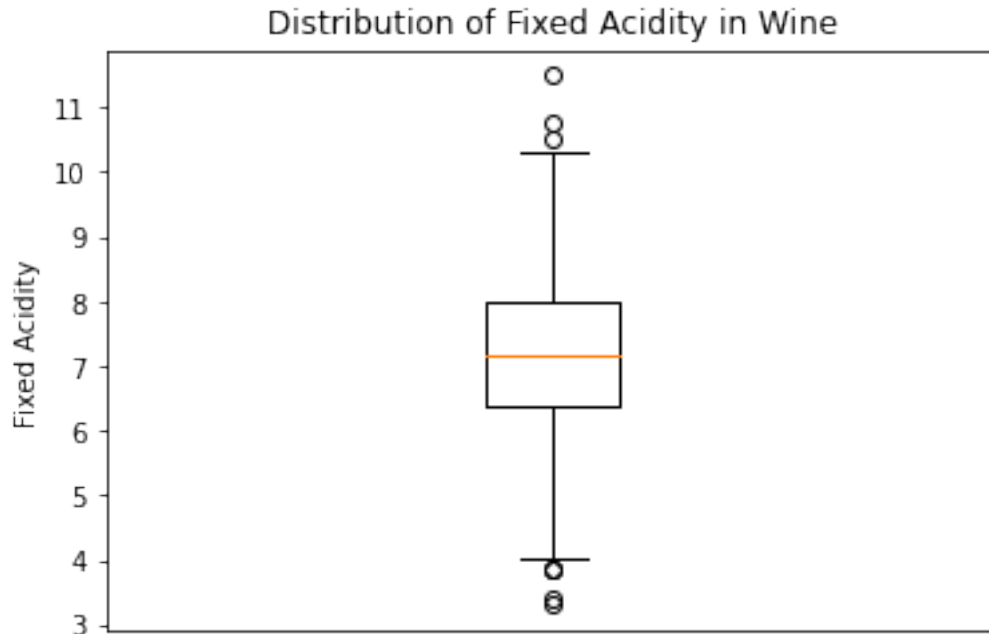


Histogram menunjukkan distribusi nilai *fixed acidity* atau nilai keasaman tetap dalam 1000 sampel anggur. Distribusi tersebut memiliki bentuk *bell-shaped* yang simetris dan memiliki nilai puncak pada 6.77-7.09 (dengan frekuensi 117). Rentang nilai berkisar antara 3,32 hingga 11,49. Ada beberapa yang memiliki nilai *fixed acidity* yang sangat rendah ataupun sangat tinggi tetapi tidak mempengaruhi bentuk keseluruhan distribusi.

```
[4]: # ===== Boxplot =====
plt.boxplot(dataFixedAcidity)

# Set attributes
plt.title('Distribution of Fixed Acidity in Wine')
plt.ylabel('Fixed Acidity')
plt.xticks([], [])

# Show graph
plt.show()
```



Boxplot menunjukkan nilai minimum dari *fixed acidity* adalah sekitar 4, sedangkan nilai maksimumnya sekitar 10,5. Walaupun begitu, terdapat lima *outlier*, tiga diantaranya berada dibawah nilai minimum. Nilai median terletak pada sekitar 7, dengan *interquartile range* sebesar 6,5 sampai 8.

0.0.2 2. volatile activity

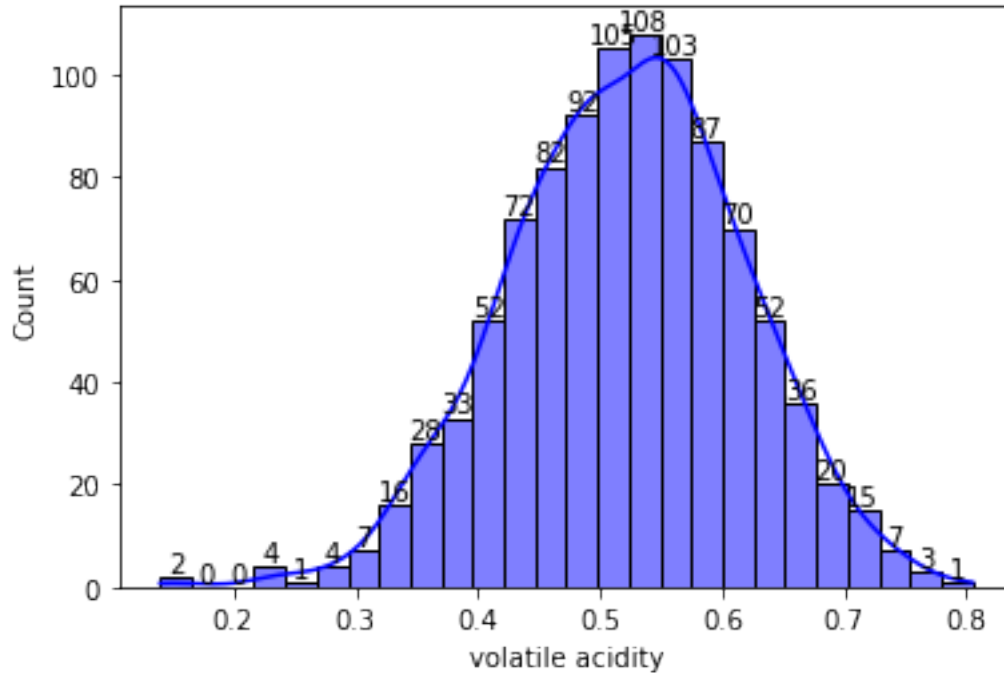
```
[5]: # Data
dataVolatileAcidity = dataAnggur['volatile acidity']

[6]: # ===== Histogram =====
ax = sns.histplot(dataVolatileAcidity, color='blue', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.13989999999999997, 0.1654846153846154,
0.19106923076923077, 0.21665384615384614, 0.2422384615384615,
0.26782307692307694, 0.29340769230769237, 0.3189923076923077,
0.3445769230769231, 0.37016153846153843, 0.39574615384615386,
0.4213307692307693, 0.4469153846153846, 0.47250000000000003,
0.49808461538461546, 0.5236692307692308, 0.5492538461538462, 0.5748384615384614,
0.600423076923077, 0.6260076923076924, 0.6515923076923076, 0.6771769230769231,
```

0.7027615384615385, 0.7283461538461538, 0.7539307692307693, 0.7795153846153847]

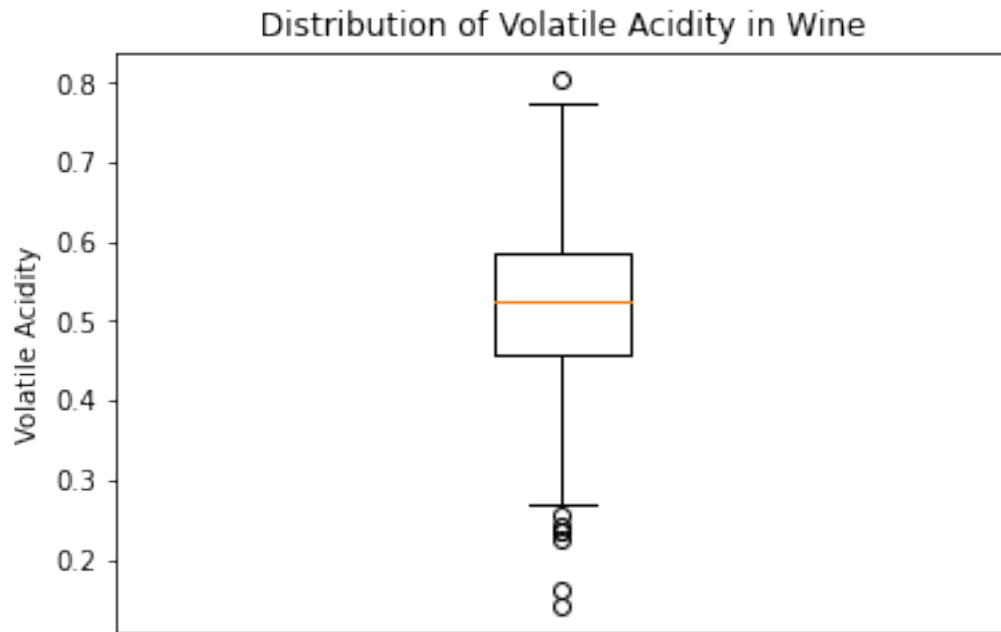


Histogram menunjukkan distribusi nilai *Volatile Acidity* dalam sampel 1000 anggur. Distribusi tersebut terlihat berbentuk *bell-shaped*, dengan distribusi normal. Walaupun begitu, jika dibandingkan dengan histogram kolom *fixed acidity*, bentuk ini sekilas terlihat lebih *negatively skewed*. Distribusi ini memiliki nilai puncak pada range keasaman 0.523 - 0.549, dengan frekuensi 108. Nilai *Volatile Acidity* berkisar antara 0.13 - 0.77.

```
[7]: # ===== Boxplot =====
plt.boxplot(dataVolatileAcidity)

# Set attributes
plt.title('Distribution of Volatile Acidity in Wine')
plt.ylabel('Volatile Acidity')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi adalah sekitar 0.27, sedangkan nilai maksimum terdapat pada 0.78. Nilai median *volatile acidity* berada pada 0.53, dengan *Interquartile Range* sebesar 0.46 - 0.58. Terdapat beberapa *outlier* pada distribusi, kebanyakan memiliki nilai dibawah minimum.

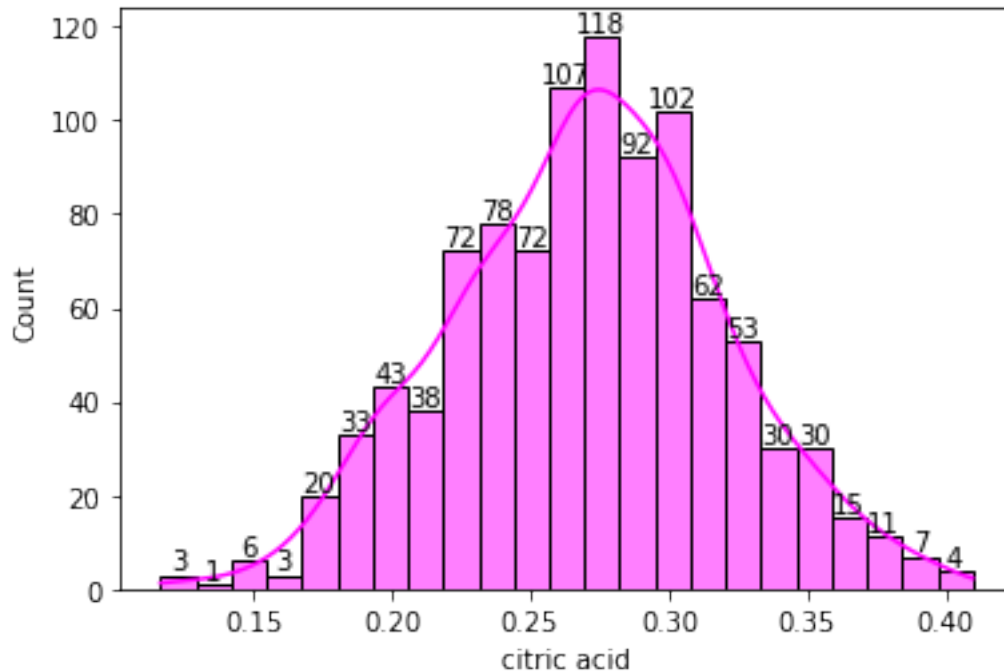
0.0.3 3. citric acid

```
[8]: # Data
dataCitricAcid = dataAnggur['citric acid']

[9]: # ===== Histogram =====
ax = sns.histplot(dataCitricAcid, color='magenta', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.1167, 0.12943478260869565, 0.14216956521739132,
0.15490434782608697, 0.16763913043478262, 0.18037391304347827,
0.19310869565217392, 0.20584347826086957, 0.21857826086956522,
0.23131304347826087, 0.24404782608695652, 0.25678260869565217,
0.2695173913043478, 0.28225217391304347, 0.2949869565217392, 0.3077217391304348,
0.3204565217391305, 0.3331913043478261, 0.3459260869565218, 0.3586608695652175,
0.3713956521739131, 0.3841304347826088, 0.3968652173913044]
```

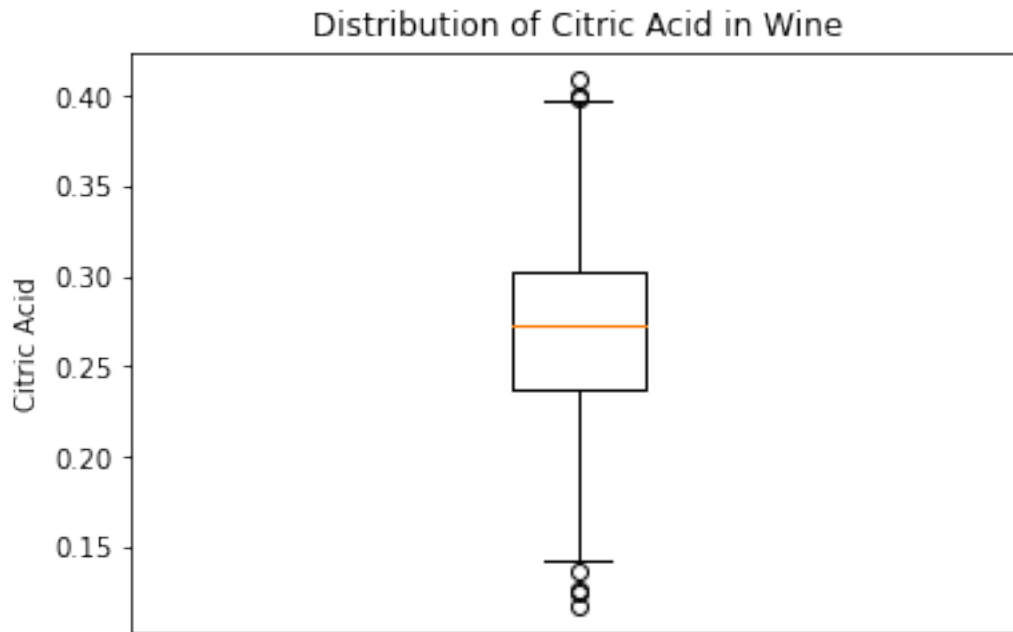


Histogram menunjukkan bahwa distribusi nilai *citric acid* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat keasaman 0.269 - 0.282 (dengan frekuensi sebanyak 118). Nilai *citric acid* memiliki range sekitar 0.11 - 0.39. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[10]: # ===== Boxplot =====
plt.boxplot(dataCitricAcid)

# Set attributes
plt.title('Distribution of Citric Acid in Wine')
plt.ylabel('Citric Acid')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum di sekitar 0.13 dan nilai maksimum di sekitar 0.40. Nilai median berada pada 0.27, dengan *interquartile range* sebesar 0.23 - 0.30. Terdapat beberapa outlier pada distribusi, sebagian besar memiliki nilai lebih kecil daripada nilai minimum.

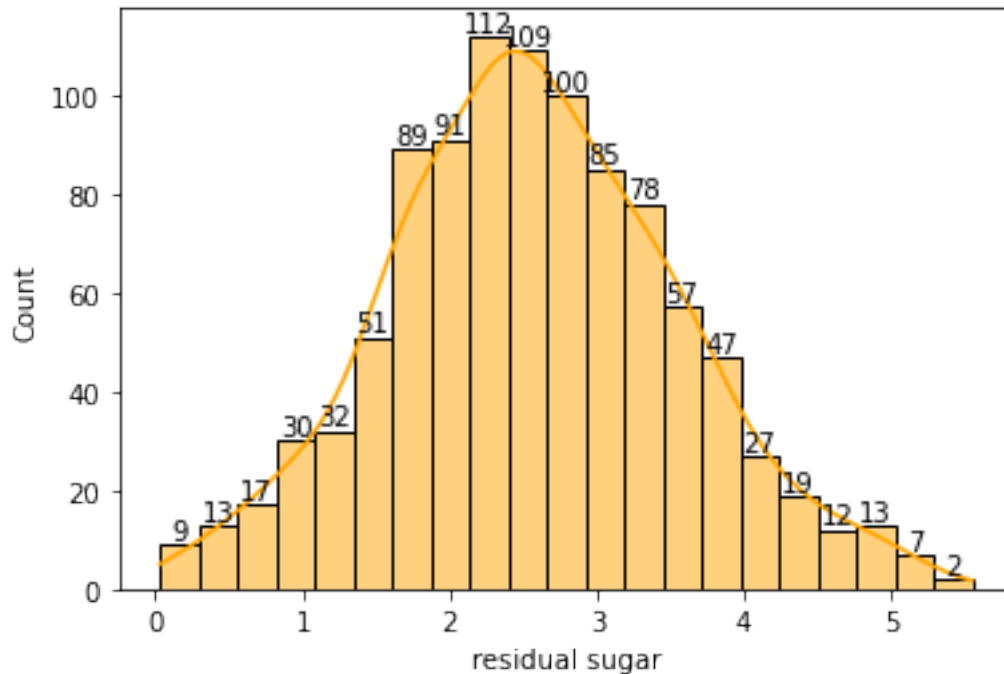
0.0.4 4. residual sugar

```
[11]: # Data
dataResidualSugar = dataAnggur['residual sugar']

[12]: # ===== Histogram =====
ax = sns.histplot(dataResidualSugar, color='orange', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.03255452501519501, 0.29532597309652175,
0.5580974211778486, 0.8208688692591752, 1.0836403173405023, 1.3464117654218293,
1.6091832135031559, 1.8719546615844829, 2.1347261096658094, 2.397497557747136,
2.6602690058284626, 2.92304045390979, 3.1858119019911166, 3.448583350072443,
3.71135479815377, 3.974126246235097, 4.236897694316424, 4.4996691423977495,
4.762440590479077, 5.025212038560404, 5.28798348664173]
```

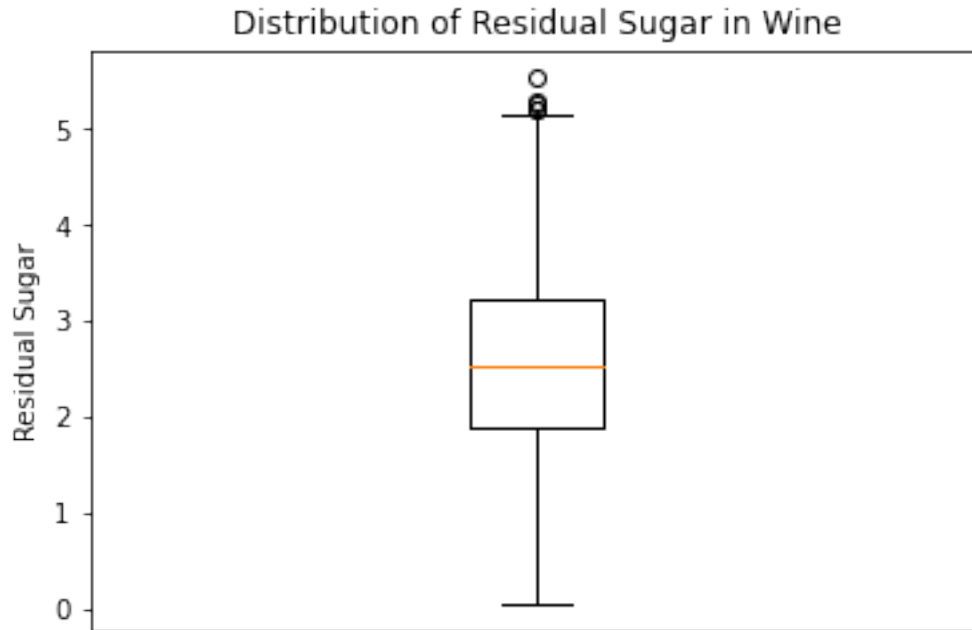


Histogram menunjukkan bahwa distribusi nilai *residual sugar* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat residu 2.134 - 2.397 (dengan frekuensi sebanyak 112). Nilai *residual sugar* memiliki range sekitar 0.03 - 5.287. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[13]: # ===== Boxplot =====
plt.boxplot(dataResidualSugar)

# Set attributes
plt.title('Distribution of Residual Sugar in Wine')
plt.ylabel('Residual Sugar')
plt.xticks([], [])

# Show graph
plt.show()
```

Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi adalah sekitar 0, sedangkan nilai maksimum terdapat pada 5.2. Nilai median *residual sugar* berada pada 2.5, dengan *Interquartile Range* diantara 1.8 - 3.2. Terdapat beberapa *outlier* pada distribusi, semua memiliki nilai di atas maksimum.

0.0.5 5. chlorides

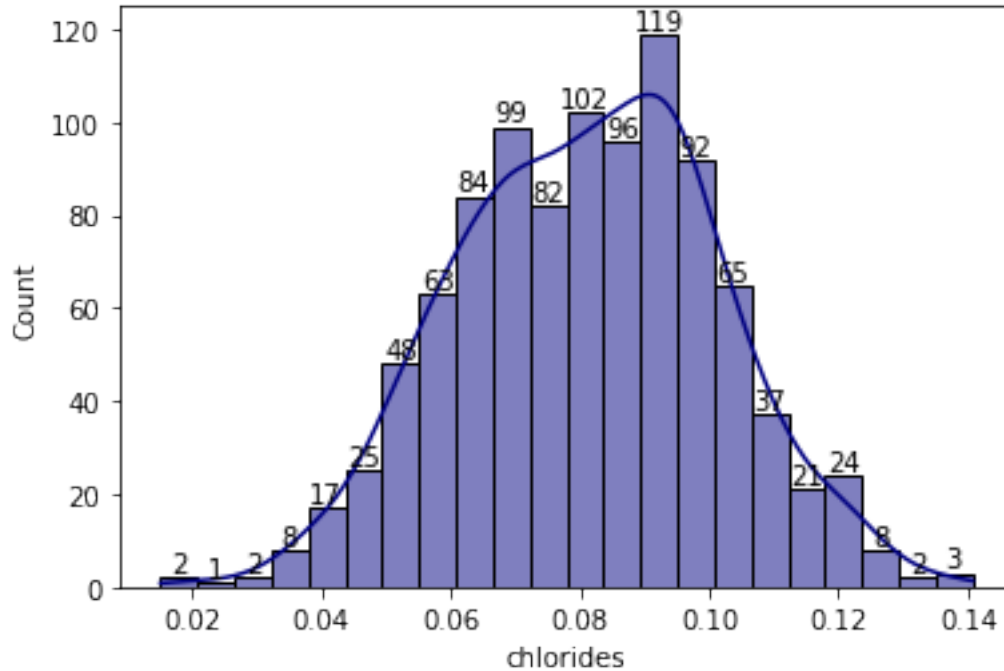
```
[14]: # Data
dataChlorides = dataAnggur['chlorides']

[15]: # ===== Histogram =====
ax = sns.histplot(dataChlorides, color='navy', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.0151224391657095, 0.02083312690504354,
0.02654381464437757, 0.03225450238371161, 0.03796519012304565,
0.04367587786237968, 0.049386565601713714, 0.05509725334104776,
0.06080794108038178, 0.06651862881971583, 0.07222931655904988,
0.07794000429838391, 0.08365069203771794, 0.08936137977705198,
0.09507206751638603, 0.10078275525572006, 0.10649344299505409,
0.11220413073438812, 0.11791481847372216, 0.12362550621305621,
```

0.12933619395239024, 0.13504688169172427]

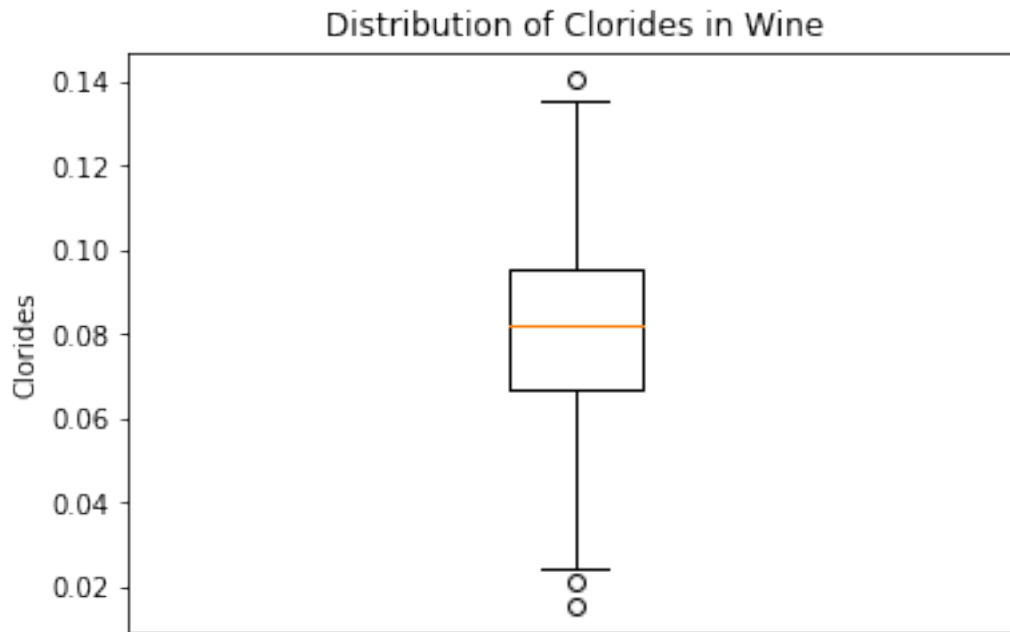


Histogram menunjukkan bahwa distribusi nilai *chlorides* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Walaupun begitu, jika dibandingkan dengan kolom lainnya, sekilas distribusi ini terlihat *negatively skewed*. Distribusi tersebut mencapai nilai puncak pada tingkat keasaman 0.089 - 0.095 (dengan frekuensi sebanyak 119). Nilai *chlorides* memiliki range sekitar 0.015 - 0.135. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[16]: # ===== Boxplot =====
plt.boxplot(dataChlorides)

# Set attributes
plt.title('Distribution of Chlorides in Wine')
plt.ylabel('Chlorides')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi adalah sekitar 0.02, sedangkan nilai maksimum terdapat pada 0.14. Nilai median *chlorides* berada pada 0.08, dengan *Interquartile Range* diantara 0.07 - 0.09. Terdapat tiga *outlier* pada distribusi, dua berada di bawah nilai minimum dan satu berada di atas nilai maksimum.

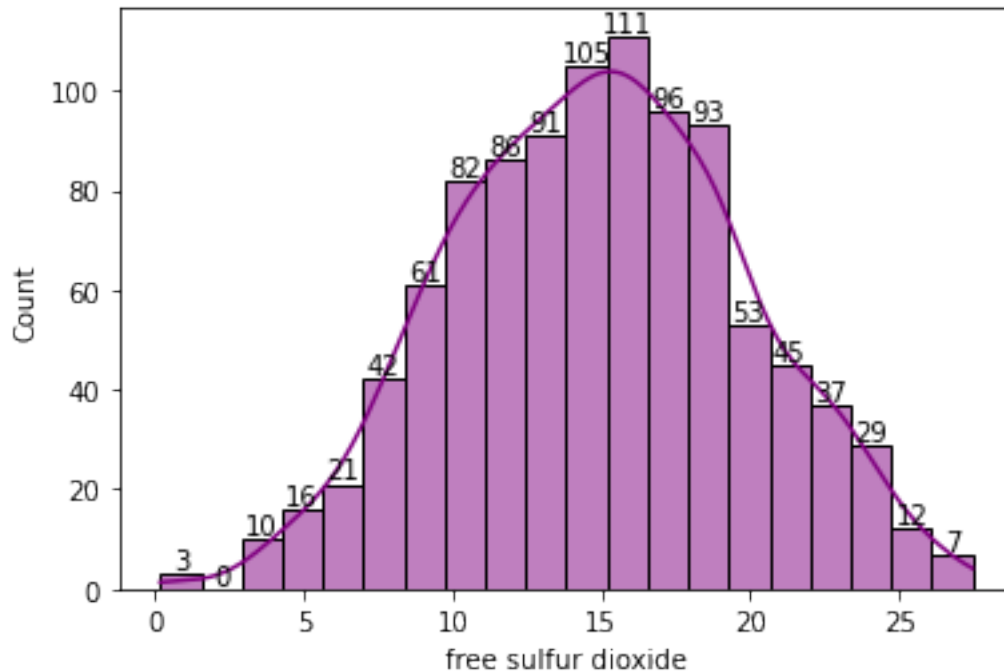
0.0.6 6. free sulfur dioxide

```
[17]: # Data
dataFreeSulfurDioxide = dataAnggur['free sulfur dioxide']

[18]: # ===== Histogram =====
ax = sns.histplot(dataFreeSulfurDioxide, color='purple', stat = 'count', kde = 
    ↪ True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.19467852332693703, 1.5580708683818822,
2.9214632134368275, 4.284855558491773, 5.648247903546719, 7.011640248601665,
8.37503259365661, 9.738424938711553, 11.1018172837665, 12.465209628821444,
13.828601973876392, 15.19199431893134, 16.55538666398628, 17.918779009041224,
19.28217135409617, 20.64556369915112, 22.00895604420606, 23.372348389261006,
24.735740734315954, 26.099133079370898]
```

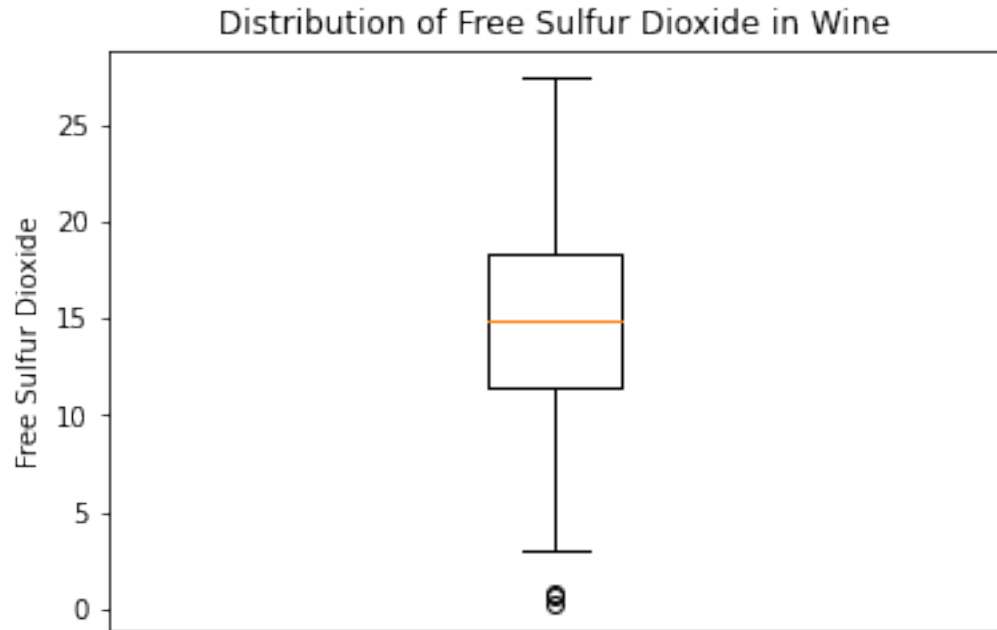


Histogram menunjukkan bahwa distribusi nilai *free sulfur dioxide* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat sulfur dioksida 15.19 - 16.55 (dengan frekuensi sebanyak 111). Nilai *free sulfur dioxide* memiliki range sekitar 0.194 - 26.099. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[19]: # ===== Boxplot =====
plt.boxplot(dataFreeSulfurDioxide)

# Set attributes
plt.title('Distribution of Free Sulfur Dioxide in Wine')
plt.ylabel('Free Sulfur Dioxide')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi adalah sekitar 2, sedangkan nilai maksimum terdapat pada 27. Nilai median *free sulfur dioxide* berada pada sekitar 15, dengan *Interquartile Range* diantara 10 - 17. Terdapat beberapa *outlier* pada distribusi, semua memiliki nilai di bawah minimum.

0.0.7 7. total sulfur dioxide

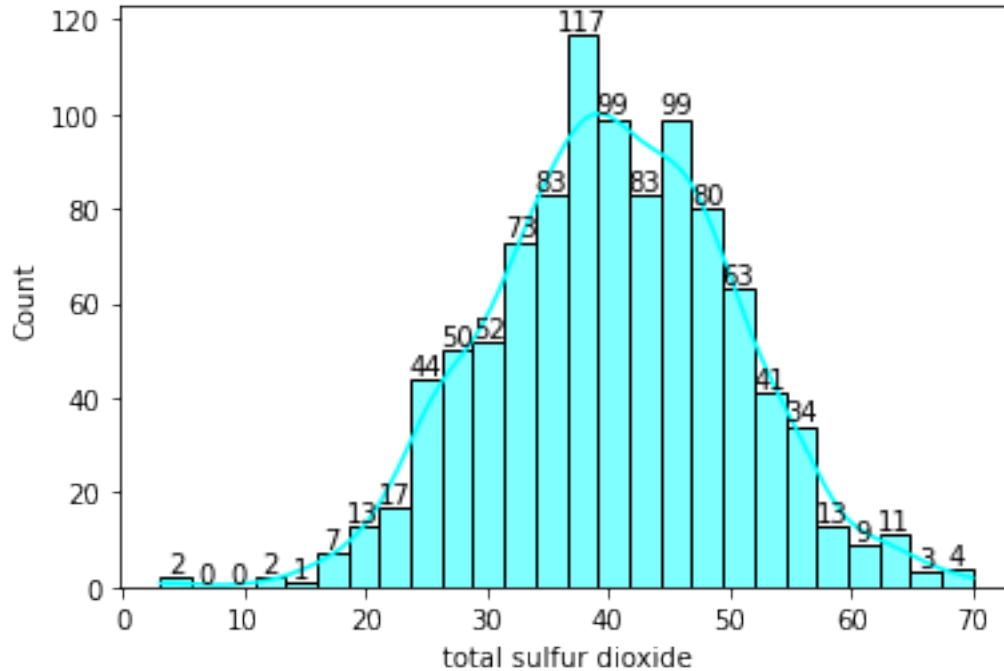
```
[20]: # Data
dataTotalSulfurDioxide = dataAnggur['total sulfur dioxide']

[21]: # ===== Histogram =====
ax = sns.histplot(dataTotalSulfurDioxide, color='cyan', stat = 'count', kde =
    ↪ True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [3.1500000000000004, 5.719615384615384,
8.289230769230768, 10.858846153846155, 13.428461538461537, 15.998076923076923,
18.567692307692305, 21.137307692307687, 23.70692307692307, 26.276538461538458,
28.84615384615384, 31.41576923076922, 33.9853846153846, 36.554999999999999,
39.124615384615375, 41.69423076923076, 44.263846153846146, 46.83346153846153,
49.40307692307691, 51.9726923076923, 54.54230769230768, 57.11192307692306,
```

59.68153846153845, 62.25115384615384, 64.82076923076923, 67.3903846153846]

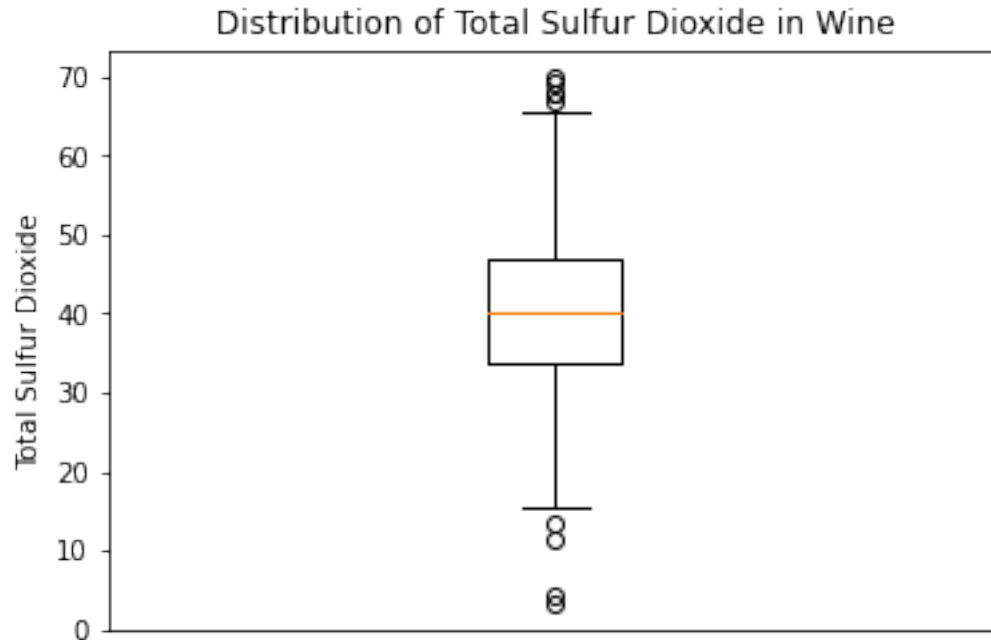


Histogram menunjukkan bahwa distribusi nilai *total sulfur dioxide* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat sulfur dioksida 36.55 - 39.12(dengan frekuensi sebanyak 117). Nilai *total sulfur dioxide* memiliki range sekitar 3.15 - 67.3. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[22]: # ===== Boxplot =====
plt.boxplot(dataTotalSulfurDioxide)

# Set attributes
plt.title('Distribution of Total Sulfur Dioxide in Wine')
plt.ylabel('Total Sulfur Dioxide')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi *total sulfur dioxide* adalah sekitar 15, sedangkan nilai maksimum terdapat pada 65. Nilai median *total sulfur dioxide* berada pada sekitar 40, dengan *Interquartile Range* diantara 35 - 45. Terdapat beberapa *outlier* pada distribusi, yang memiliki nilai di atas maksimum dan di bawah minimum.

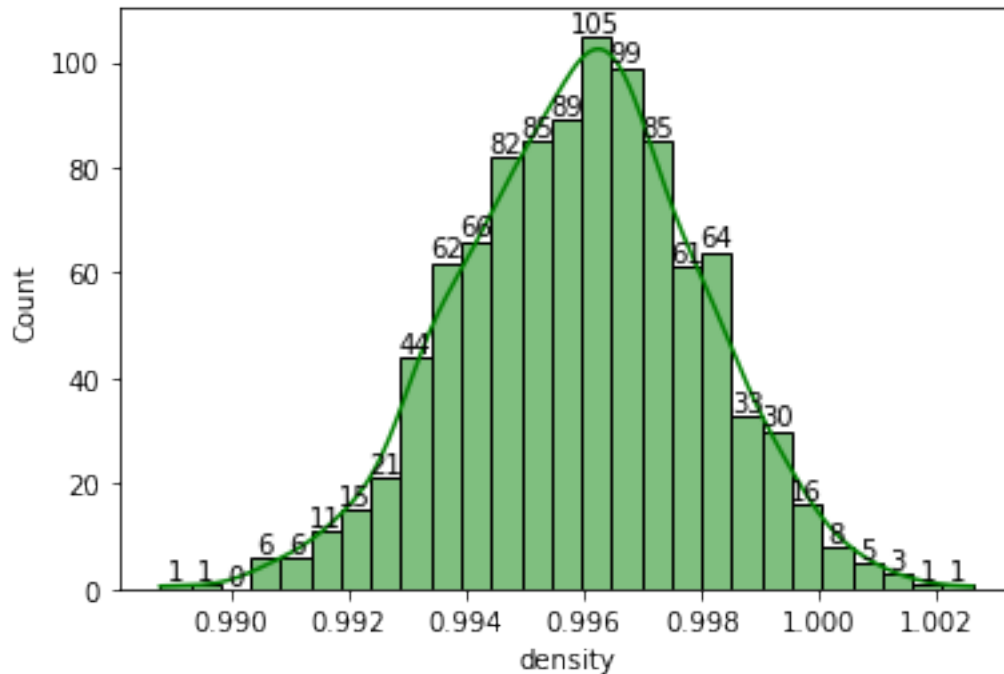
0.0.8 8. density

```
[23]: # Data
dataDensity = dataAnggur['density']
```

```
[24]: # ===== Histogram =====
ax = sns.histplot(dataDensity, color='green', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.9888, 0.9893111111111111, 0.9898222222222224,
0.9903333333333333, 0.9908444444444444, 0.9913555555555555, 0.9918666666666667,
0.9923777777777778, 0.9928888888888889, 0.9934, 0.9939111111111111,
0.9944222222222222, 0.9949333333333333, 0.9954444444444444, 0.9959555555555555,
0.9964666666666666, 0.9969777777777777, 0.9974888888888889, 0.998,
0.9985111111111111, 0.9990222222222221, 0.9995333333333334, 1.0000444444444443,
1.0005555555555556, 1.0010666666666665, 1.0015777777777778, 1.0020888888888888]
```

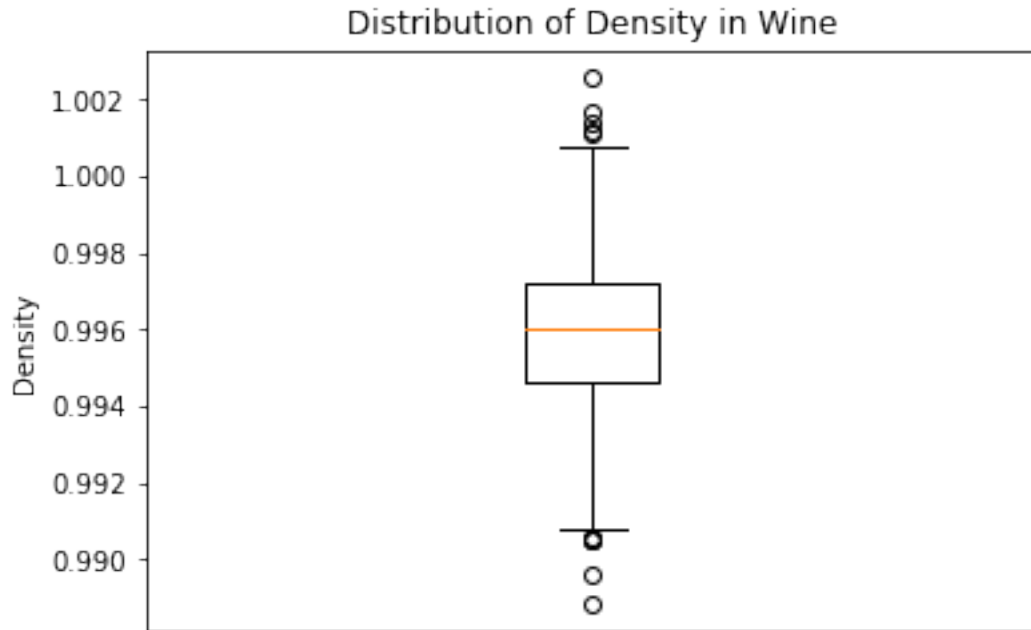


Histogram menunjukkan bahwa distribusi nilai *density* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat kepadatan 0.9959 - 0.9964 (dengan frekuensi sebanyak 105). Nilai *density* memiliki range sekitar 0.988 - 1.002. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[25]: # ===== Boxplot =====
plt.boxplot(dataDensity)

# Set attributes
plt.title('Distribution of Density in Wine')
plt.ylabel('Density')
plt.xticks([], [])

# Show graph
plt.show()
```

Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi *density* adalah sekitar 0.991, sedangkan nilai maksimum terdapat pada 1.001. Nilai median *density* berada pada sekitar 0.996, dengan *Interquartile Range* diantara 0.995 - 0.997. Terdapat beberapa *outlier* pada distribusi, yang memiliki nilai di atas maksimum dan di bawah minimum.

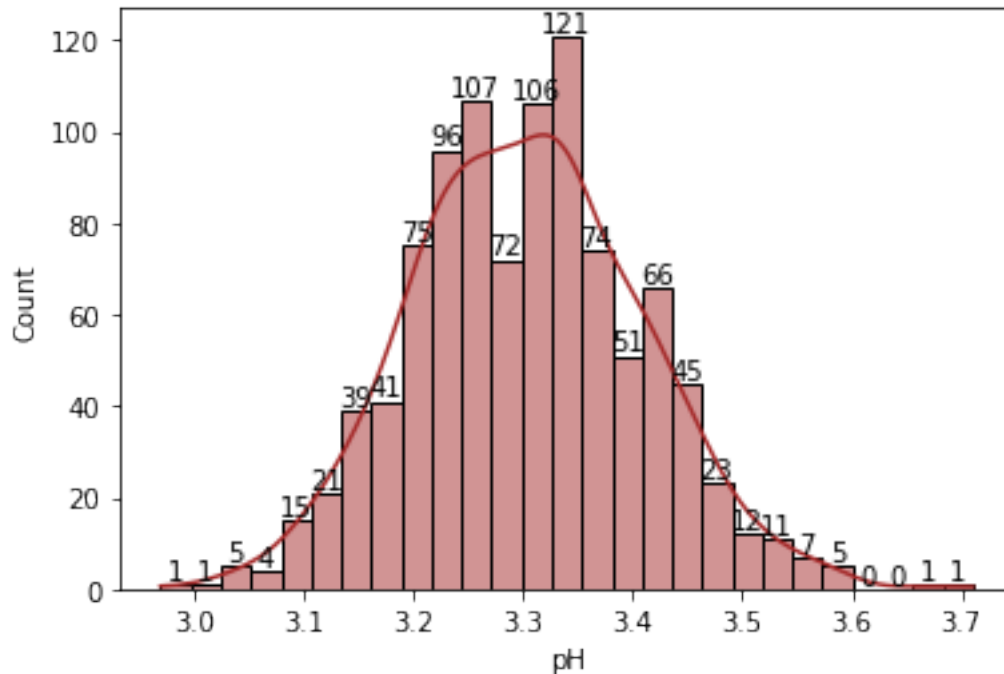
0.0.9 9. pH

```
[26]: # Data
      datapH = dataAnggur['pH']
```

```
[27]: # ===== Histogram =====
      ax = sns.histplot(datapH, color='brown', stat = 'count', kde = True)
      for i in ax.containers:
          ax.bar_label(i,)

      # Print the bin edges
      bin_edges = [patch.get_x() for patch in ax.patches]
      print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [2.97, 2.9974074074074073, 3.024814814814815,
3.0522222222222224, 3.07962962962963, 3.107037037037037, 3.1344444444444446,
3.1618518518518517, 3.1892592592592592, 3.2166666666666667, 3.2440740740740743,
3.2714814814814814, 3.2988888888888889, 3.326296296296296, 3.3537037037037036,
3.3811111111111111, 3.4085185185185187, 3.435925925925926, 3.4633333333333333,
3.4907407407407405, 3.518148148148148, 3.5455555555555556, 3.572962962962963,
3.60037037037037, 3.6277777777777773, 3.655185185185185, 3.6825925925925924]
```

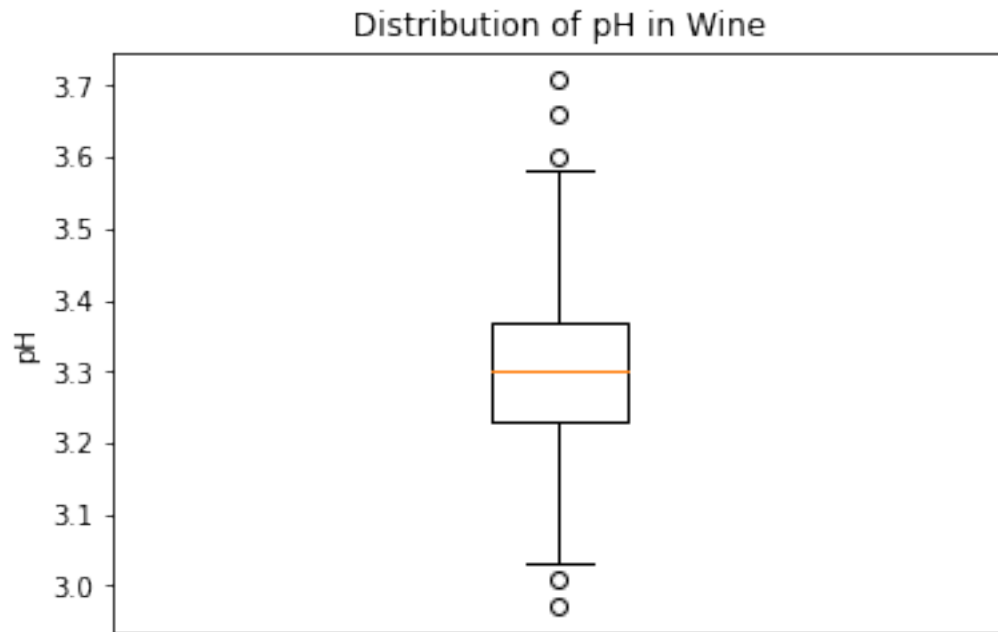


Histogram menunjukkan distribusi nilai pH dalam sampel 1000 anggur. Distribusi tersebut terlihat berbentuk *bell-shaped*, dengan distribusi normal. Walaupun begitu, jika dibandingkan dengan histogram kolom lainnya, bentuk ini sekilas terlihat lebih *positively skewed*. Distribusi ini memiliki nilai puncak pada range pH 3.32 - 3.35, dengan frekuensi 121. Nilai pH berkisar antara 2.97 - 3.68.

```
[28]: # ===== Boxplot =====
plt.boxplot(datapH)

# Set attributes
plt.title('Distribution of pH in Wine')
plt.ylabel('pH')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi pH adalah sekitar 3.05, sedangkan nilai maksimum terdapat pada 3.58. Nilai median pH berada pada sekitar 3.3, dengan *Interquartile Range* diantara 3.25 - 3.35 Terdapat beberapa *outlier* pada distribusi, yang memiliki nilai di atas maksimum dan di bawah minimum.

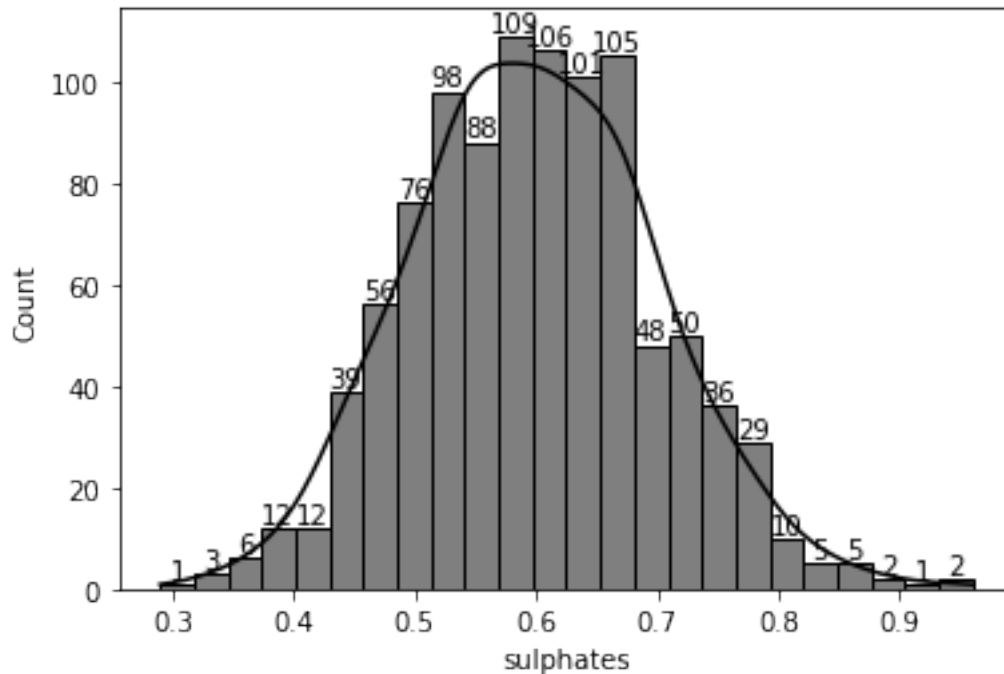
0.0.10 10. sulphates

```
[29]: # Data
dataSulphates = dataAnggur['sulphates']

[30]: # ===== Histogram =====
ax = sns.histplot(dataSulphates, color='black', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [0.29000000000000004, 0.31791666666666666,
0.34583333333333333, 0.37375, 0.40166666666666666, 0.42958333333333333, 0.4575,
0.48541666666666666, 0.51333333333333334, 0.54125, 0.56916666666666666,
0.59708333333333332, 0.625, 0.65291666666666666, 0.68083333333333332, 0.70875,
0.73666666666666666, 0.76458333333333332, 0.7925, 0.82041666666666666,
0.84833333333333332, 0.87625, 0.90416666666666666, 0.93208333333333332]
```

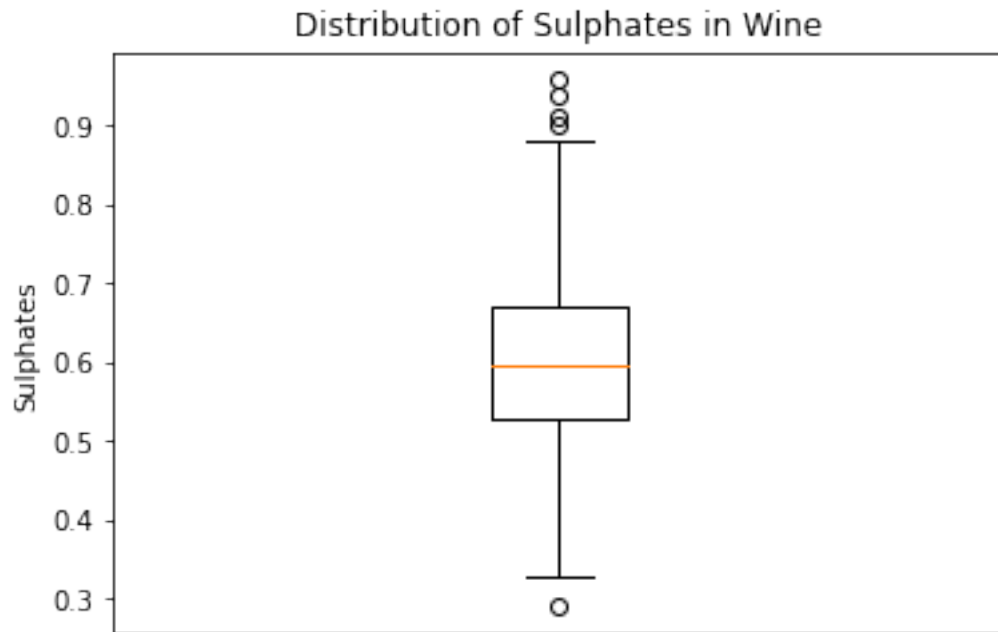


Histogram menunjukkan bahwa distribusi nilai *sulphates* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat kepadatan 0.569 - 0.597 (dengan frekuensi sebanyak 109). Nilai *sulphates* memiliki range sekitar 0.29 - 0.96. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[31]: # ===== Boxplot =====
plt.boxplot(dataSulphates)

# Set attributes
plt.title('Distribution of Sulphates in Wine')
plt.ylabel('Sulphates')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi *sulphates* adalah sekitar 0.33, sedangkan nilai maksimum terdapat pada 0.87. Nilai median *sulphates* berada pada sekitar 0.6, dengan *Interquartile Range* diantara 0.55 - 0.65. Terdapat beberapa *outlier* pada distribusi, sebagian besar memiliki nilai di atas nilai maksimum.

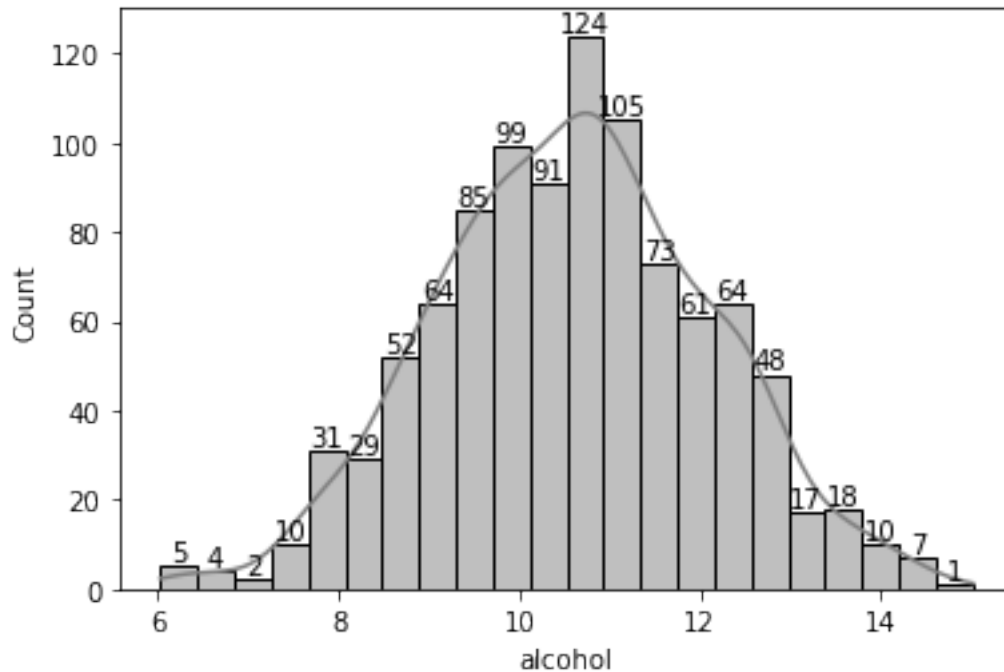
0.0.11 11. alcohol

```
[32]: # Data
dataAlcohol = dataAnggur['alcohol']
```

```
[33]: # ===== Histogram =====
ax = sns.histplot(dataAlcohol, color='grey', stat = 'count', kde = True)
for i in ax.containers:
    ax.bar_label(i,)

# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [6.03, 6.438636363636363, 6.847272727272728,
7.255909090909091, 7.664545454545454, 8.07318181818182, 8.48181818181818,
8.890454545454546, 9.29909090909091, 9.707727272727272, 10.116363636363637,
10.524999999999999, 10.933636363636364, 11.342272727272729, 11.75090909090909,
12.159545454545455, 12.568181818181817, 12.976818181818182, 13.385454545454547,
13.794090909090908, 14.202727272727273, 14.611363636363635]
```

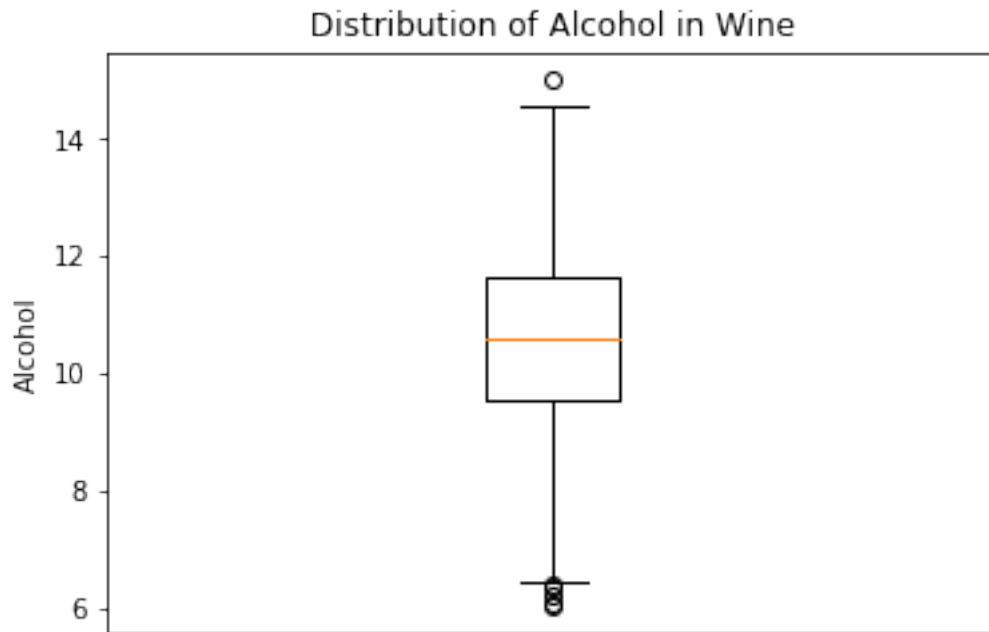


Histogram menunjukkan bahwa distribusi nilai *alcohol* dalam 1000 sampel anggur berbentuk *bell-shaped* atau memiliki distribusi normal. Distribusi tersebut mencapai nilai puncak pada tingkat kepadatan 10.52 - 10.93 (dengan frekuensi sebanyak 124). Nilai *alcohol* memiliki range sekitar 6.03 - 14.61. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[34]: # ===== Boxplot =====
plt.boxplot(dataAlcohol)

# Set attributes
plt.title('Distribution of Alcohol in Wine')
plt.ylabel('Alcohol')
plt.xticks([], [])

# Show graph
plt.show()
```



Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi *alcohol* adalah sekitar 6.5, sedangkan nilai maksimum terdapat pada 14.5. Nilai median *alcohol* berada pada sekitar 10.5, dengan *Interquartile Range* diantara 9.55 - 11.5. Terdapat beberapa *outlier* pada distribusi, sebagian besar memiliki nilai di bawah nilai minimum.

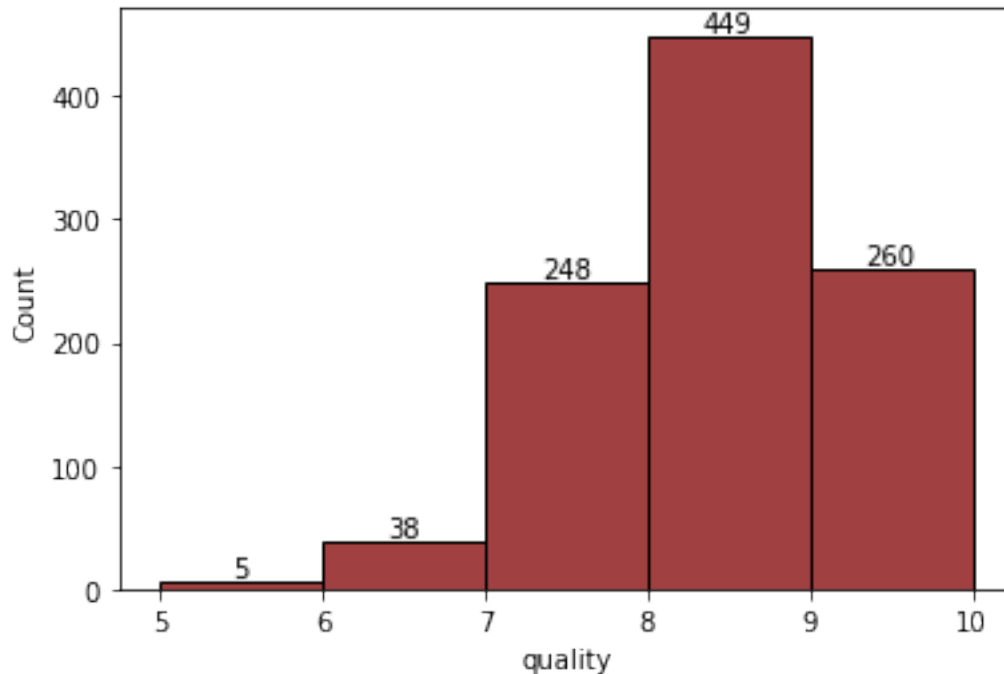
0.0.12 12. quality

```
[35]: # Data
dataQuality = dataAnggur['quality']

[36]: # ===== Histogram =====
ax = sns.histplot(dataQuality, color='maroon', stat = 'count', bins=5)
for i in ax.containers:
    ax.bar_label(i,)

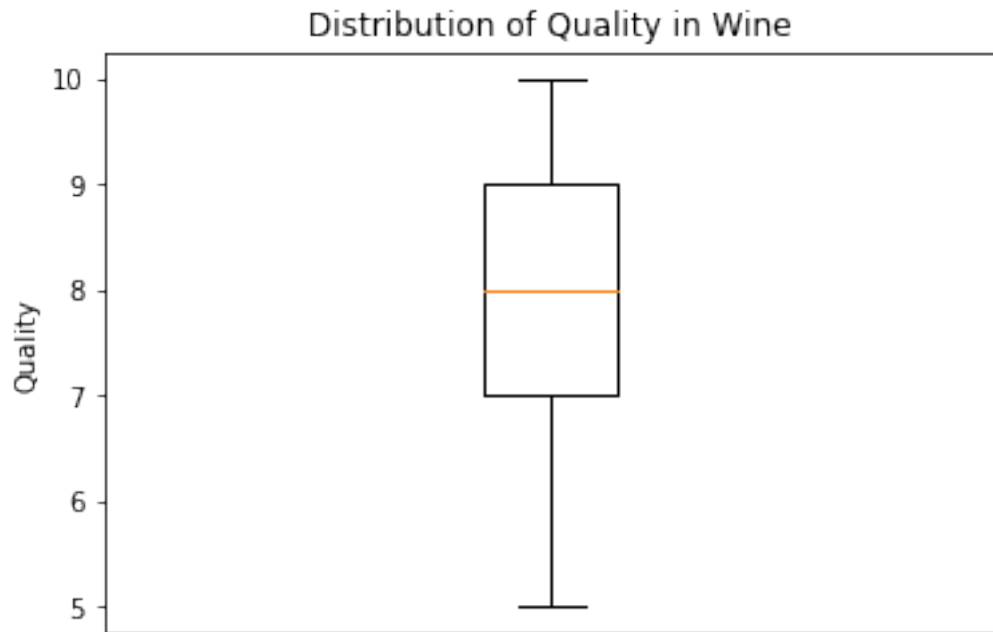
# Print the bin edges
bin_edges = [patch.get_x() for patch in ax.patches]
print("Bin Edges (from leftmost): ", bin_edges)
```

```
Bin Edges (from leftmost): [5.0, 6.0, 7.0, 8.0, 9.0]
```



Histogram menunjukkan bahwa distribusi nilai *quality* dalam 1000 sampel anggur. Data pada kolom ini bertipe diskrit. Distribusi tersebut mencapai nilai puncak pada tingkat kualitas 8 - 9 (dengan frekuensi sebanyak 449). Nilai *quality* memiliki range sekitar 5 - 10. Distribusi ini memiliki beberapa nilai dengan tingkat sangat tinggi ataupun sangat rendah, namun tidak mempengaruhi bentuk distribusi.

```
[37]: # ===== Boxplot =====  
plt.boxplot(dataQuality)  
  
# Set attributes  
plt.title('Distribution of Quality in Wine')  
plt.ylabel('Quality')  
plt.xticks([], [])  
  
# Show graph  
plt.show()
```

Berdasarkan visualisasi di atas, boxplot menunjukkan nilai minimum distribusi *quality* adalah 5, sedangkan nilai maksimum terdapat pada 10. Nilai median *quality* berada pada nilai 8, dengan *Interquartile Range* diantara 7 - 9. Tidak terdapat *outlier* pada distribusi.