PATTERN RECOGNITION ASSIGNMENT 1 2021-2022

At the following paragraph we will upload the packages that we will use during this assignment

In [84]:
```
import matplotlib.pyplot as plt
from sklearn import datasets
import pandas as pd
import numpy as np
import statistics as st
```

In [93]:
```
#load dataset

wine = pd.read_csv('wine.csv')
print(wine.shape)
wine.head(178)
```

```
(178, 14)
```

Out[93]:

	Category	Alcohol	Malic acid	Ash	Alcalinity of ash	Magnesium	Total phenols	Flavanoids	Nonflavanoid phenols	Proanthocyanins	Color intensity	Hue	OD280/OD315 of diluted wines
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93
...	...	...	...	...	...	...	...	...	...	...	...	...	...
173	3	13.71	5.65	2.45	20.5	95	1.68	0.61	0.52	1.06	7.70	0.64	1.74
174	3	13.40	3.91	2.48	23.0	102	1.80	0.75	0.43	1.41	7.30	0.70	1.56
175	3	13.27	4.28	2.26	20.0	120	1.59	0.69	0.43	1.35	10.20	0.59	1.56
176	3	13.17	2.59	2.37	20.0	120	1.65	0.68	0.53	1.46	9.30	0.60	1.62
177	3	14.13	4.10	2.74	24.5	96	2.05	0.76	0.56	1.35	9.20	0.61	1.60

178 rows × 14 columns

So in this case the datas et is an 178 Row's by 14 Columns Matrix

In [73]:
```
cat1_data = wine[0:59] # parounme ta dedomena apo tin katigoria 1
cat2_data = wine[60:130] # parounme ta dedomena apo tin katigoria 2
cat3_data = wine[131:178] # parounme ta dedomena apo tin katigoria 3

alcohol_1 = cat1_data['Alcohol']
malic_acid_1 = cat1_data['Malic acid']
ash_1 = cat1_data['Ash']
alcalinity_of_ash_1 = cat1_data['Alcalinity of ash']

alcohol_2 = cat2_data['Alcohol']
malic_acid_2 = cat2_data['Malic acid']
ash_2 = cat2_data['Ash']
alcalinity_of_ash_2 = cat2_data['Alcalinity of ash']

alcohol_3 = cat3_data['Alcohol']
malic_acid_3 = cat3_data['Malic acid']
ash_3 = cat3_data['Ash']
alcalinity_of_ash_3 = cat3_data['Alcalinity of ash']
```

Now lets go for Alcohol values

Category 1

In [106…
```
print("The minimum Value of Alchohol is",np.min(alcohol_1))
print("The maximum Value of Alchohol is",np.max(alcohol_1))
print("The range of min and max value(s) of Alchohol is",np.max(alcohol_1)- np.min(alcohol_1))
print("The Mean values of Alchohol is",st.mean(alcohol_1))
```

```
print("The Median values of Alchohol is",st.median(alcohol_1))
print("The Standard Deviation value of Alchohol is",np.std(alcohol_1))
print("The Dispersion or Variance value of Alchohol is",np.var(alcohol_1))
```

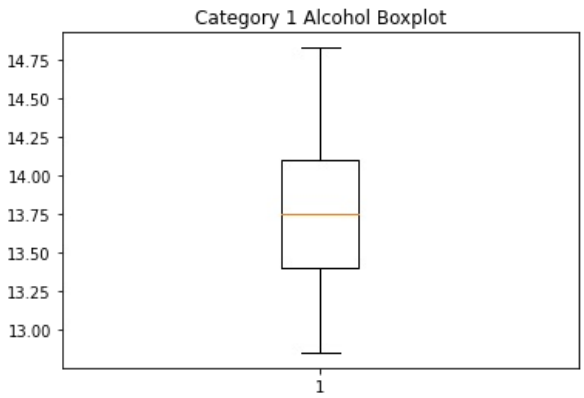
```
The minimum Value of Alchohol is 12.85
The maximum Value of Alchohol is 14.83
The range of min and max value(s) of Alchohol is 1.9800000000000004
The Mean values of Alchohol is 13.744745762711865
The Median values of Alchohol is 13.75
The Standard Deviation value of Alchohol is 0.4581923063525723
The Dispersion or Variance value of Alchohol is 0.20994018960068944
```

We could calculate the mean value by adding all the values and divide the sum of the values with the multitude of the values ex. (n1 + n2 + n3 + n4 + n5 + n6 + ... + n59 + n60)/ 60

For simplicity reasons we used mean function from the statistics library

In [94]:
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 1 Alcohol Boxplot')
ax1.boxplot(alcohol_1)
```

Out[94]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d28e2e0>,
  <matplotlib.lines.Line2D at 0x2290d28e670>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d28ea00>,
  <matplotlib.lines.Line2D at 0x2290d28ed90>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d281f10>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d29a160>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d29a4f0>],
 'means': []}
```



Category 1

In [89]:
```
print("The minimum Value of Alchohol is",np.min(alcohol_2))
print("The maximum Value of Alchohol is",np.max(alcohol_2))
print("The range of min and max value(s) of Alchohol is",np.max(alcohol_2)- np.min(alcohol_2))
print("The Mean values of Alchohol is",st.mean(alcohol_2))
print("The Median values of Alchohol is",st.median(alcohol_2))
print("The Standard Deviation value of Alchohol is",np.std(alcohol_2))
print("The Dispersion or Variance value of Alchohol is",np.var(alcohol_2))
```

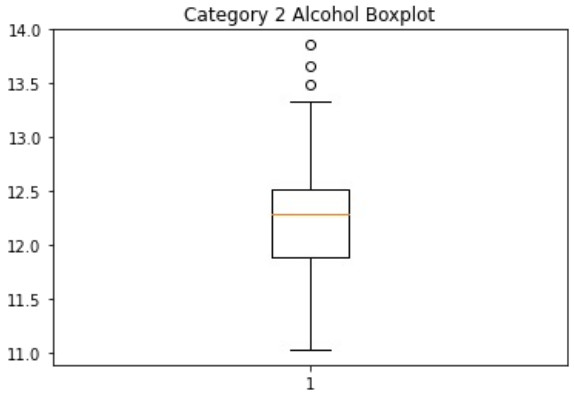
```
The minimum Value of Alchohol is 11.03
The maximum Value of Alchohol is 13.86
The range of min and max value(s) of Alchohol is 2.83
The Mean values of Alchohol is 12.27742857142857
The Median values of Alchohol is 12.29
The Standard Deviation value of Alchohol is 0.5378520394369911
The Dispersion or Variance value of Alchohol is 0.28928481632653064
```

The boxplot Alcohol of Category 2

In [95]:
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 2 Alcohol Boxplot')
ax1.boxplot(alcohol_2)
```

Out[95]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d2e5910>,
```

```
 <matplotlib.lines.Line2D at 0x2290d2e5ca0>],
'caps': [<matplotlib.lines.Line2D at 0x2290d2f3070>,
 <matplotlib.lines.Line2D at 0x2290d2f3400>],
'boxes': [<matplotlib.lines.Line2D at 0x2290d2e5580>],
'medians': [<matplotlib.lines.Line2D at 0x2290d2f3790>],
'fliers': [<matplotlib.lines.Line2D at 0x2290d2f3b20>],
'means': []}
```



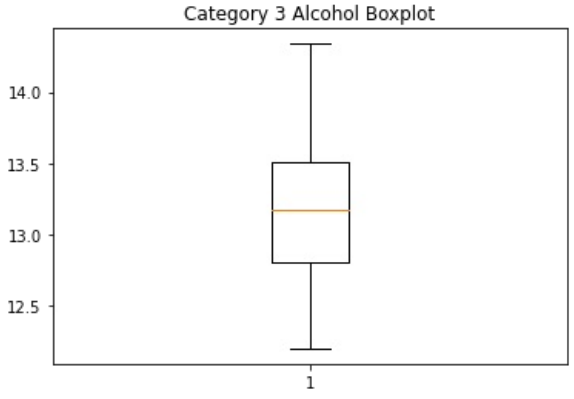
Category 3

In [91]:
```
print("The minimum Value of Alchohol is",np.min(alcohol_3))
print("The maximum Value of Alchohol is",np.max(alcohol_3))
print("The range of min and max value(s) of Alchohol is",np.max(alcohol_3)- np.min(alcohol_3))
print("The Mean values of Alchohol is",st.mean(alcohol_3))
print("The Median values of Alchohol is",st.median(alcohol_3))
print("The Standard Deviation value of Alchohol is",np.std(alcohol_3))
print("The Dispersion or Variance value of Alchohol is",np.var(alcohol_3))
```

```
The minimum Value of Alchohol is 12.2
The maximum Value of Alchohol is 14.34
The range of min and max value(s) of Alchohol is 2.1400000000000006
The Mean values of Alchohol is 13.16
The Median values of Alchohol is 13.17
The Standard Deviation value of Alchohol is 0.5284702934544476
The Dispersion or Variance value of Alchohol is 0.27928085106382994
```

In [98]:
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 3 Alcohol Boxplot')
ax1.boxplot(alcohol_3)
```

Out[98]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d3af7f0>,
  <matplotlib.lines.Line2D at 0x2290d3afb80>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d3aff10>,
  <matplotlib.lines.Line2D at 0x2290d3bb2e0>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d3af4c0>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d3bb670>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d3bba00>],
 'means': []}
```



Now lets go to do the same for Malic Acid

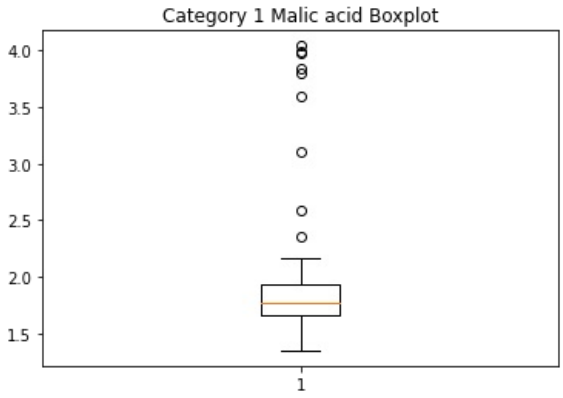
Category 1

In [100…
```
print("The minimum Value of Malic acid is",np.min(malic_acid_1))
print("The maximum Value of Malic acid is",np.max(malic_acid_1))
print("The range of min and max value(s) of Malic acid is",np.max(malic_acid_1)- np.min(malic_acid_1))
print("The Mean values of Malic acid is",st.mean(malic_acid_1))
print("The Median values of Malic acid is",st.median(malic_acid_1))
print("The Standard Deviation value of Malic acid is",np.std(malic_acid_1))
print("The Dispersion or Variance value of Malic acid is",np.var(malic_acid_1))
```

```
The minimum Value of Malic acid is 1.35
The maximum Value of Malic acid is 4.04
The range of min and max value(s) of Malic acid is 2.69
The Mean values of Malic acid is 2.010677966101695
The Median values of Malic acid is 1.77
The Standard Deviation value of Malic acid is 0.6826887630111
The Dispersion or Variance value of Malic acid is 0.4660639471416259
```

In [101…
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 1 Malic acid Boxplot')
ax1.boxplot(malic_acid_1)
```

Out[101…
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d415430>,
  <matplotlib.lines.Line2D at 0x2290d4157c0>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d415b50>,
  <matplotlib.lines.Line2D at 0x2290d415ee0>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d4150a0>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d41c2b0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d41c5e0>],
 'means': []}
```



Category 2

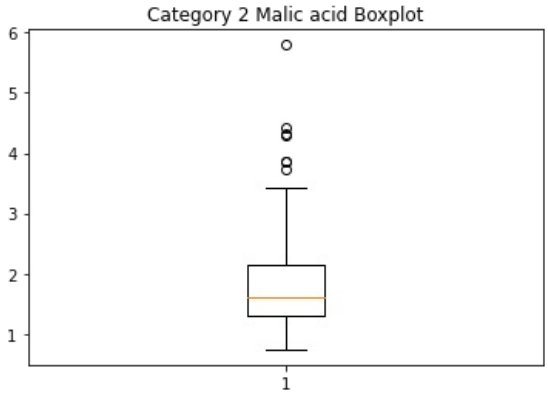
In [102…
```
print("The minimum Value of Malic acid is",np.min(malic_acid_2))
print("The maximum Value of Malic acid is",np.max(malic_acid_2))
print("The range of min and max value(s) of Malic acid is",np.max(malic_acid_2)- np.min(malic_acid_2))
print("The Mean values of Malic acid is",st.mean(malic_acid_2))
print("The Median values of Malic acid is",st.median(malic_acid_2))
print("The Standard Deviation value of Malic acid is",np.std(malic_acid_2))
print("The Dispersion or Variance value of Malic acid is",np.var(malic_acid_2))
```

```
The minimum Value of Malic acid is 0.74
The maximum Value of Malic acid is 5.8
The range of min and max value(s) of Malic acid is 5.06
The Mean values of Malic acid is 1.9468571428571428
The Median values of Malic acid is 1.62
The Standard Deviation value of Malic acid is 1.0085145269258187
The Dispersion or Variance value of Malic acid is 1.017101551020408
```

In [103…
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 2 Malic acid Boxplot')
ax1.boxplot(malic_acid_2)
```

Out[103…
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d471dc0>,
```

```
 <matplotlib.lines.Line2D at 0x2290d483190>],
'caps': [<matplotlib.lines.Line2D at 0x2290d483520>,
 <matplotlib.lines.Line2D at 0x2290d4838b0>],
'boxes': [<matplotlib.lines.Line2D at 0x2290d471a30>],
'medians': [<matplotlib.lines.Line2D at 0x2290d483c40>],
'fliers': [<matplotlib.lines.Line2D at 0x2290d483fd0>],
'means': []}
```



Category 3

In [104…

```
print("The minimum Value of Malic acid is",np.min(malic_acid_3))
print("The maximum Value of Malic acid is",np.max(malic_acid_3))
print("The range of min and max value(s) of Malic acid is",np.max(malic_acid_3)- np.min(malic_acid_3))
print("The Mean values of Malic acid is",st.mean(malic_acid_3))
print("The Median values of Malic acid is",st.median(malic_acid_3))
print("The Standard Deviation value of Malic acid is",np.std(malic_acid_3))
print("The Dispersion or Variance value of Malic acid is",np.var(malic_acid_3))
```

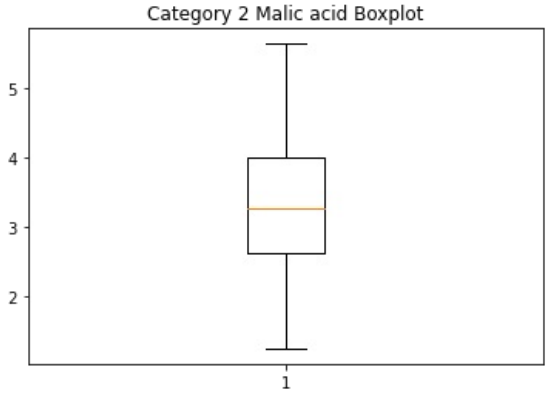
```
The minimum Value of Malic acid is 1.24
The maximum Value of Malic acid is 5.65
The range of min and max value(s) of Malic acid is 4.41
The Mean values of Malic acid is 3.3759574468085107
The Median values of Malic acid is 3.27
The Standard Deviation value of Malic acid is 1.047868473910136
The Dispersion or Variance value of Malic acid is 1.0980283386147573
```

In [105…

```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 2 Malic acid Boxplot')
ax1.boxplot(malic_acid_3)
```

Out[105…

```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d4e2700>,
  <matplotlib.lines.Line2D at 0x2290d4e2a90>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d4e2e20>,
  <matplotlib.lines.Line2D at 0x2290d4ed1f0>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d4e2370>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d4ed580>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d4ed910>],
 'means': []}
```



Now lets go to do the same for Ash

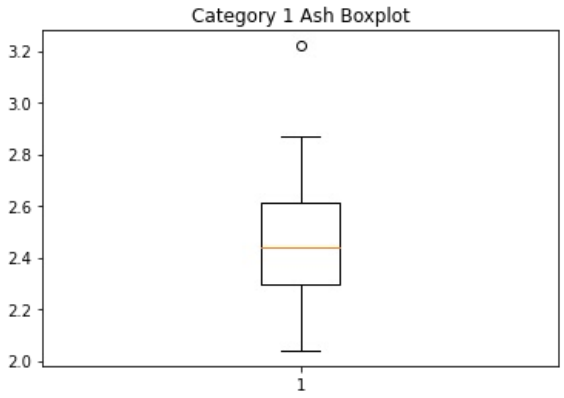
Category 1

In [109…
```
print("The minimum Value of Ash is",np.min(ash_1))
print("The maximum Value of Ash is",np.max(ash_1))
print("The range of min and max value(s) of Ash is",np.max(ash_1)- np.min(ash_1))
print("The Mean values of Ash is",st.mean(ash_1))
print("The Median values of Ash is",st.median(ash_1))
print("The Standard Deviation value of Ash is",np.std(ash_1))
print("The Dispersion or Variance value of Ash is",np.var(ash_1))
```

```
The minimum Value of Ash is 2.04
The maximum Value of Ash is 3.22
The range of min and max value(s) of Ash is 1.1800000000000002
The Mean values of Ash is 2.455593220338983
The Median values of Ash is 2.44
The Standard Deviation value of Ash is 0.22523261938580777
The Dispersion or Variance value of Ash is 0.05072973283539215
```

In [110…
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 1 Ash Boxplot')
ax1.boxplot(ash_1)
```

Out[110…
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d549580>,
  <matplotlib.lines.Line2D at 0x2290d549910>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d549ca0>,
  <matplotlib.lines.Line2D at 0x2290d556070>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d5491f0>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d556400>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d556790>],
 'means': []}
```



Category 2

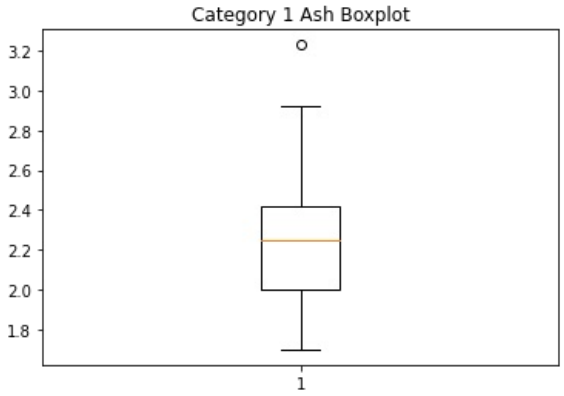
In [112…
```
print("The minimum Value of Ash is",np.min(ash_2))
print("The maximum Value of Ash is",np.max(ash_2))
print("The range of min and max value(s) of Ash is",np.max(ash_2)- np.min(ash_2))
print("The Mean values of Ash is",st.mean(ash_2))
print("The Median values of Ash is",st.median(ash_2))
print("The Standard Deviation value of Ash is",np.std(ash_2))
print("The Dispersion or Variance value of Ash is",np.var(ash_2))
```

```
The minimum Value of Ash is 1.7
The maximum Value of Ash is 3.23
The range of min and max value(s) of Ash is 1.53
The Mean values of Ash is 2.2574285714285716
The Median values of Ash is 2.25
The Standard Deviation value of Ash is 0.29694485161046175
The Dispersion or Variance value of Ash is 0.08817624489795915
```

In [113…
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 2 Ash Boxplot')
ax1.boxplot(ash_2)
```

Out[113…
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d516550>,
```

```
 <matplotlib.lines.Line2D at 0x2290d5168e0>],
'caps': [<matplotlib.lines.Line2D at 0x2290d516c70>,
 <matplotlib.lines.Line2D at 0x2290d5c0040>],
'boxes': [<matplotlib.lines.Line2D at 0x2290d5161c0>],
'medians': [<matplotlib.lines.Line2D at 0x2290d5c03d0>],
'fliers': [<matplotlib.lines.Line2D at 0x2290d5c0760>],
'means': []}
```



Category 3

In [114…

```
print("The minimum Value of Ash is",np.min(ash_3))
print("The maximum Value of Ash is",np.max(ash_3))
print("The range of min and max value(s) of Ash is",np.max(ash_3)- np.min(ash_3))
print("The Mean values of Ash is",st.mean(ash_3))
print("The Median values of Ash is",st.median(ash_3))
print("The Standard Deviation value of Ash is",np.std(ash_3))
print("The Dispersion or Variance value of Ash is",np.var(ash_3))
```

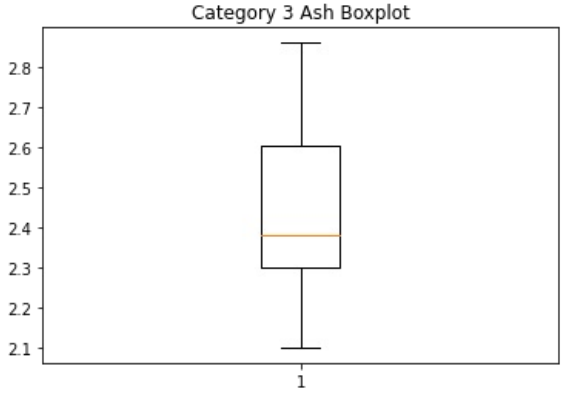
```
The minimum Value of Ash is 2.1
The maximum Value of Ash is 2.86
The range of min and max value(s) of Ash is 0.7599999999999998
The Mean values of Ash is 2.4395744680851066
The Median values of Ash is 2.38
The Standard Deviation value of Ash is 0.18388198628478059
The Dispersion or Variance value of Ash is 0.03381258488003623
```

In [117…

```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 3 Ash Boxplot')
ax1.boxplot(ash_3)
```

Out[117…

```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d6940a0>,
  <matplotlib.lines.Line2D at 0x2290d694430>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d6947c0>,
  <matplotlib.lines.Line2D at 0x2290d694b50>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d685cd0>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d694ee0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d69d2b0>],
 'means': []}
```



Now lets go to do the same for Alcalinity of ash

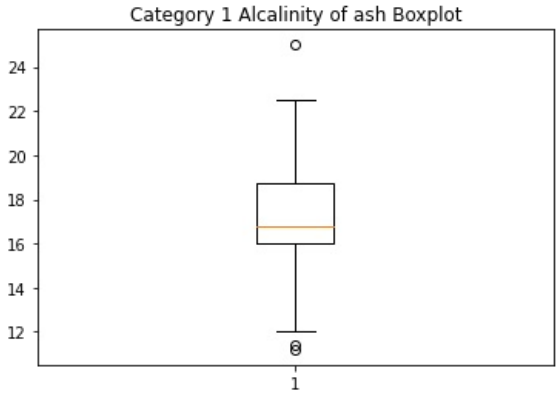
Category 1

In [116…
```
print("The minimum Value of Alcalinity of ash is",np.min(alcalinity_of_ash_1))
print("The maximum Value of Alcalinity of ash is",np.max(alcalinity_of_ash_1))
print("The range of min and max value(s) of Alcalinity of ash is",np.max(alcalinity_of_ash_1)- np.min(alcalinity_
print("The Mean values of Alcalinity of ash is",st.mean(alcalinity_of_ash_1))
print("The Median values of Alcalinity of ash is",st.median(alcalinity_of_ash_1))
print("The Standard Deviation value of Alcalinity of ash is",np.std(alcalinity_of_ash_1))
print("The Dispersion or Variance value of Alcalinity of ash is",np.var(alcalinity_of_ash_1))
```

```
The minimum Value of Alcalinity of ash is 11.2
The maximum Value of Alcalinity of ash is 25.0
The range of min and max value(s) of Alcalinity of ash is 13.8
The Mean values of Alcalinity of ash is 17.03728813559322
The Median values of Alcalinity of ash is 16.8
The Standard Deviation value of Alcalinity of ash is 2.524651229820095
The Dispersion or Variance value of Alcalinity of ash is 6.373863832232119
```

In [118…
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 1 Alcalinity of ash Boxplot')
ax1.boxplot(alcalinity_of_ash_1)
```

Out[118…
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290ca87940>,
  <matplotlib.lines.Line2D at 0x2290ca875e0>],
 'caps': [<matplotlib.lines.Line2D at 0x2290ca870a0>,
  <matplotlib.lines.Line2D at 0x2290d0a9e20>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290c384820>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d0a9ee0>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d0a9ac0>],
 'means': []}
```



Category 2

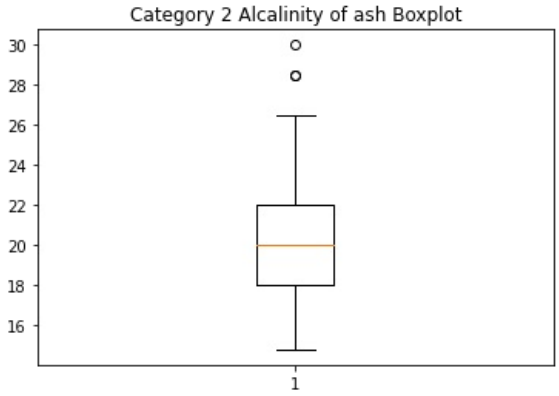
In [119…
```
print("The minimum Value of Alcalinity of ash is",np.min(alcalinity_of_ash_2))
print("The maximum Value of Alcalinity of ash is",np.max(alcalinity_of_ash_2))
print("The range of min and max value(s) of Alcalinity of ash is",np.max(alcalinity_of_ash_2)- np.min(alcalinity_
print("The Mean values of Alcalinity of ash is",st.mean(alcalinity_of_ash_2))
print("The Median values of Alcalinity of ash is",st.median(alcalinity_of_ash_2))
print("The Standard Deviation value of Alcalinity of ash is",np.std(alcalinity_of_ash_2))
print("The Dispersion or Variance value of Alcalinity of ash is",np.var(alcalinity_of_ash_2))
```

```
The minimum Value of Alcalinity of ash is 14.8
The maximum Value of Alcalinity of ash is 30.0
The range of min and max value(s) of Alcalinity of ash is 15.2
The Mean values of Alcalinity of ash is 20.375714285714285
The Median values of Alcalinity of ash is 20.0
The Standard Deviation value of Alcalinity of ash is 3.1424483500374443
The Dispersion or Variance value of Alcalinity of ash is 9.874981632653057
```

In [120…
```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 2 Alcalinity of ash Boxplot')
ax1.boxplot(alcalinity_of_ash_2)
```

Out[120…
```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d6d17c0>,
```

```
 <matplotlib.lines.Line2D at 0x2290d6d1af0>],
'caps': [<matplotlib.lines.Line2D at 0x2290d6d1d90>,
 <matplotlib.lines.Line2D at 0x2290cf6a130>],
'boxes': [<matplotlib.lines.Line2D at 0x229711e66a0>],
'medians': [<matplotlib.lines.Line2D at 0x2290cf6a4c0>],
'fliers': [<matplotlib.lines.Line2D at 0x2290cf6a850>],
'means': []}
```



Category 3

In [121…

```
print("The minimum Value of Alcalinity of ash is",np.min(alcalinity_of_ash_3))
print("The maximum Value of Alcalinity of ash is",np.max(alcalinity_of_ash_3))
print("The range of min and max value(s) of Alcalinity of ash is",np.max(alcalinity_of_ash_3)- np.min(alcalinity_
print("The Mean values of Alcalinity of ash is",st.mean(alcalinity_of_ash_3))
print("The Median values of Alcalinity of ash is",st.median(alcalinity_of_ash_3))
print("The Standard Deviation value of Alcalinity of ash is",np.std(alcalinity_of_ash_3))
print("The Dispersion or Variance value of Alcalinity of ash is",np.var(alcalinity_of_ash_3))
```

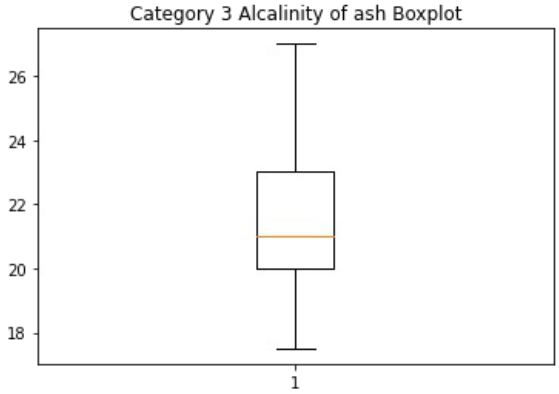
```
The minimum Value of Alcalinity of ash is 17.5
The maximum Value of Alcalinity of ash is 27.0
The range of min and max value(s) of Alcalinity of ash is 9.5
The Mean values of Alcalinity of ash is 21.48936170212766
The Median values of Alcalinity of ash is 21.0
The Standard Deviation value of Alcalinity of ash is 2.2012795159791594
The Dispersion or Variance value of Alcalinity of ash is 4.8456315074694425
```

In [122…

```
fig1,ax1 = plt.subplots()
ax1.set_title('Category 3 Alcalinity of ash Boxplot')
ax1.boxplot(alcalinity_of_ash_3)
```

Out[122…

```
{'whiskers': [<matplotlib.lines.Line2D at 0x2290d710c40>,
  <matplotlib.lines.Line2D at 0x2290d710fa0>],
 'caps': [<matplotlib.lines.Line2D at 0x2290d720370>,
  <matplotlib.lines.Line2D at 0x2290d720700>],
 'boxes': [<matplotlib.lines.Line2D at 0x2290d7108b0>],
 'medians': [<matplotlib.lines.Line2D at 0x2290d720a90>],
 'fliers': [<matplotlib.lines.Line2D at 0x2290d720e20>],
 'means': []}
```



5. Scale data into [0, 1] range

5. Scale data into [0, 1] range

In [132…

```
from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler()
A_minmax_1 = min_max_scaler.fit_transform(cat1_data[['Alcohol','Malic acid']]) # it could be all the data...
```

Out[132…

```
array([[0.6969697 , 0.133829  ],
       [0.17676768, 0.1598513 ],
       [0.15656566, 0.37546468],
       [0.76767677, 0.22304833],
       [0.1969697 , 0.46096654],
       [0.68181818, 0.15241636],
       [0.77777778, 0.19330855],
       [0.61111111, 0.29739777],
       [1.        , 0.10780669],
       [0.51010101, 0.        ],
       [0.63131313, 0.30111524],
       [0.64141414, 0.04832714],
       [0.45454545, 0.14126394],
       [0.95959596, 0.14126394],
       [0.77272727, 0.19330855],
       [0.39393939, 0.17100372],
       [0.73232323, 0.21189591],
       [0.49494949, 0.08178439],
       [0.67676768, 0.08921933],
       [0.3989899 , 0.65055762],
       [0.61111111, 0.10408922],
       [0.04040404, 0.91078067],
       [0.43434343, 0.18959108],
       [0.        , 0.0929368 ],
       [0.32828283, 0.17100372],
       [0.1010101 , 0.26022305],
       [0.27272727, 0.15613383],
       [0.22727273, 0.13754647],
       [0.51515152, 0.20446097],
       [0.59090909, 0.12267658],
       [0.44444444, 0.05576208],
       [0.36868687, 0.11524164],
       [0.41919192, 0.17843866],
       [0.45959596, 0.0669145 ],
       [0.33333333, 0.16728625],
       [0.31818182, 0.17100372],
       [0.21717172, 0.10780669],
       [0.1010101 , 0.11152416],
       [0.11111111, 0.05576208],
       [0.69191919, 0.98141264],
       [0.35858586, 0.133829  ],
       [0.28282828, 0.92565056],
       [0.52020202, 0.20074349],
       [0.1969697 , 0.97769517],
       [0.1010101 , 0.15613383],
       [0.68686869, 1.        ],
       [0.77272727, 0.83271375],
       [0.53030303, 0.12267658],
       [0.63131313, 0.24907063],
       [0.55050505, 0.14126394],
       [0.1010101 , 0.14126394],
       [0.49494949, 0.11152416],
       [0.48989899, 0.14869888],
       [0.46464646, 0.20446097],
       [0.44949495, 0.11895911],
       [0.35858586, 0.14126394],
       [0.69191919, 0.13011152],
       [0.22222222, 0.23048327],
       [0.43939394, 0.02973978]])
```

Split data into a training (70%) and a test set (30%).

In [151…

```
X = cat1_data[['Alcohol','Malic acid']]
y = cat1_data[['Ash','Alcalinity of ash']]

from sklearn.model_selection import train_test_split
X_train ,X_test , y_train , y_test = train_test_split(X,y,test_size=0.2,random_state=5)#random state = value gia
```

and little training to our model will not be bad

In [152…

```
from sklearn.linear_model import LinearRegression

clc = LinearRegression()
clc.fit(X_train,y_train)

clc.predict(X_test)
clc.score(X_test,y_test)
```

Out[152… -0.2636626058781737