In [ ]:
```
#Omkar Annasaheb Shinde
#practical 3
```

In [2]:
```
import pandas as pd
import numpy as np
```

In [3]:
```
df1 = pd.read_csv("C:/Users/avcoe/OneDrive/Desktop/Prac3.csv")
df1
```

Out[3]:

	Age_Group	Income
0	18-30	742
1	51-70	688
2	51-70	663
3	31-50	169
4	18-30	683
5	31-50	104
6	18-30	747
7	18-30	503
8	31-50	95
9	51-70	503
10	51-70	652
11	31-50	44
12	51-70	361
13	51-70	661
14	51-70	729
15	31-50	192
16	18-30	143
17	51-70	241
18	51-70	193
19	18-30	503
20	51-70	52
21	31-50	154
22	51-70	199
23	18-30	387
24	51-70	205
25	31-50	708
26	51-70	749
27	51-70	636
28	51-70	361
29	51-70	368

In [4]:
```
df1.Age_Group.unique()
```

Out[4]:
```
array(['18-30', '51-70', '31-50'], dtype=object)
```

In [ ]:
```
#Using groupby()
```

In [5]:
```
#count number of non null income value in each age group

df1.groupby(df1.Age_Group).count()
```

Out[5]:

	Income
Age_Group	
18-30	7
31-50	7
51-70	16

In [6]:
```
#minimum value of income in each age group

df1.groupby(df1.Age_Group).min()
```

Out[6]:

Age_Group	Income
18-30	143
31-50	44
51-70	52

In [7]:
```
#maximum value of income in each age group

df1.groupby(df1.Age_Group).max()
```

Out[7]:

Age_Group	Income
18-30	747
31-50	708
51-70	749

In [8]:
```
#mean of income in each age group

df1.groupby(df1.Age_Group).mean()
```

Out[8]:

Age_Group	Income
18-30	529.714286
31-50	209.428571
51-70	453.812500

In [9]:
```
#standard deviation of income in each age group

df1.groupby(df1.Age_Group).std()
```

Out[9]:

Age_Group	Income
18-30	218.753068
31-50	225.521512
51-70	231.729503

In [10]:
```
#.describe() method

df1.groupby(df1.Age_Group).describe()
```

Out[10]:

	Income							
	count	mean	std	min	25%	50%	75%	max
Age_Group								
18-30	7.0	529.714286	218.753068	143.0	445.0	503.0	712.5	747.0
31-50	7.0	209.428571	225.521512	44.0	99.5	154.0	180.5	708.0
51-70	16.0	453.812500	231.729503	52.0	232.0	435.5	661.5	749.0

In [11]:
```
#load iris
from sklearn import datasets
data = datasets.load_iris()
df = pd.DataFrame(data.data,columns=data.feature_names)
df['species'] = pd.Series(data.target)
df.head(
```

```
  Cell In[11], line 6
    df.head(
           ^
SyntaxError: incomplete input
```

In [12]:
```
from sklearn import datasets
data = datasets.load_iris()
df = pd.DataFrame(data.data,columns=data.feature_names)
df['species'] = pd.Series(data.target)
df.head()
```

Out[12]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

In [13]:
```
df.species.unique()
```

Out[13]:
```
array([0, 1, 2])
```

In [14]:
```
df.groupby(df.species)
```

Out[14]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x00000201EEFCF7D0>
```

In [15]:
```
#count()() for species

df.groupby(df.species).count()
```

Out[15]:

species	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	50	50	50	50
1	50	50	50	50
2	50	50	50	50

In [16]:
```
#Max() for species

df.groupby(df.species).max()
```

Out[16]:

species	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.8	4.4	1.9	0.6
1	7.0	3.4	5.1	1.8
2	7.9	3.8	6.9	2.5

In [17]:
```
#Min() for species

df.groupby(df.species).min()
```

Out[17]:

species	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	4.3	2.3	1.0	0.1
1	4.9	2.0	3.0	1.0
2	4.9	2.2	4.5	1.4

In [18]:
```
#Mean() for Species

df.groupby(df.species).mean()
```

Out[18]:

species	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.006	3.428	1.462	0.246
1	5.936	2.770	4.260	1.326
2	6.588	2.974	5.552	2.026

In [19]:
```
#standard deviation  for species


df.groupby(df.species).std()
```

Out[19]:

species	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	0.352490	0.379064	0.173664	0.105386
1	0.516171	0.313798	0.469911	0.197753
2	0.635880	0.322497	0.551895	0.274650

In [20]:
```
##describe() Method for sepal length

df.groupby(df.species)["sepal length (cm)"].describe()
```

Out[20]:

species	count	mean	std	min	25%	50%	75%	max
0	50.0	5.006	0.352490	4.3	4.800	5.0	5.2	5.8
1	50.0	5.936	0.516171	4.9	5.600	5.9	6.3	7.0
2	50.0	6.588	0.635880	4.9	6.225	6.5	6.9	7.9

In [21]:
```
##describe() Method for petal length

df.groupby(df.species)["petal length (cm)"].describe()
```

Out[21]:

species	count	mean	std	min	25%	50%	75%	max
0	50.0	1.462	0.173664	1.0	1.4	1.50	1.575	1.9
1	50.0	4.260	0.469911	3.0	4.0	4.35	4.600	5.1
2	50.0	5.552	0.551895	4.5	5.1	5.55	5.875	6.9

In [22]:
```
#describe() Method for petal width

df.groupby(df.species)["petal width (cm)"].describe()
```

Out[22]:

species	count	mean	std	min	25%	50%	75%	max
0	50.0	0.246	0.105386	0.1	0.2	0.2	0.3	0.6
1	50.0	1.326	0.197753	1.0	1.2	1.3	1.5	1.8
2	50.0	2.026	0.274650	1.4	1.8	2.0	2.3	2.5

In [ ]:
```

```