Pandas

In [1]:

```
#packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
np.random.seed(123)
```

Series

In [2]:

```
#create a series from
#list
#an array
#dictionary
```

In [3]:

```
ml = [78,85,79,90]
myser1 =pd.Series(ml)
```

In [4]:

```
myser1
```

Out[4]:

```
0    78
1    85
2    79
3    90
dtype: int64
```

In [5]:

```
myind = ['Stud_A', 'Stud_B', 'Stud_C', 'Stud_D']
```

In [6]:

```
myser1 =pd.Series(ml, myind)
myser1
```

Out[6]:

```
Stud_A    78
Stud_B    85
Stud_C    79
Stud_D    90
dtype: int64
```

In [7]:

```
#array

myarr =np.random.randint(0,100, size=(4,))
```

In [8]:

```
myarr
```

Out[8]:

```
array([66, 92, 98, 17])
```

In [9]:

```
myser2 =pd.Series(myarr)
```

In [10]:

```
myser2
```

```
Out[10]: 0    66
         1    92
         2    98
         3    17
         dtype: int32
```

```
In [11]: myser2 = pd.Series(myarr, myind)
         myser2
```

```
Out[11]: Stud_A    66
         Stud_B    92
         Stud_C    98
         Stud_D    17
         dtype: int32
```

```
In [12]: #dictionary
         mydict = {'Stud_A' :77, 'Stud_B': 98, 'Stud_C': 67, 'Stud_D': 88}
         myser3 = pd.Series(mydict)
         myser3
```

```
Out[12]: Stud_A    77
         Stud_B    98
         Stud_C    67
         Stud_D    88
         dtype: int64
```

Basic Operations or methods on series

```
In [13]: myser3.keys()
```

```
Out[13]: Index(['Stud_A', 'Stud_B', 'Stud_C', 'Stud_D'], dtype='object')
```

```
In [15]: myser2.keys()
```

```
Out[15]: Index(['Stud_A', 'Stud_B', 'Stud_C', 'Stud_D'], dtype='object')
```

```
In [16]: myser1.index
```

```
Out[16]: Index(['Stud_A', 'Stud_B', 'Stud_C', 'Stud_D'], dtype='object')
```

```
In [17]: myser1 +myser2
```

```
Out[17]: Stud_A    144
         Stud_B    177
         Stud_C    177
         Stud_D    107
         dtype: int64
```

```
In [18]: len(myser1)
```

```
Out[18]: 4
```

```
In [19]: myser2.agg(['min', 'max'])
```

```
Out[19]: min    17
         max    98
         dtype: int64
```

In [22]:

```
#to know more about series
#run - help(pd.Series)
```

DataFrame

In [28]:

```
##create a dataframe from an array
myarr1 = np.random.randint(0,100, size=(5,5))
col = ['River', 'Oyo', 'Niger', 'Lagos', 'Delta']

df = pd.DataFrame(myarr1)
```

In [26]:

```
df
```

Out[26]:

	0	1	2	3	4
0	47	73	32	46	96
1	25	83	78	36	96
2	80	68	49	55	67
3	2	84	39	66	84
4	47	61	48	7	99

In [29]:

```
ind = ['Jan', 'Feb', 'Mar', 'Apr', 'May']

df = pd.DataFrame(myarr1, index=ind, columns=col)
df
```

Out[29]:

	River	Oyo	Niger	Lagos	Delta
Jan	92	52	97	85	94
Feb	27	34	97	76	40
Mar	3	69	64	75	34
Apr	58	10	22	77	18
May	15	27	30	52	70

In [33]:

```
myarrt = np.random.random((5,5))
```

In [34]:

```
myarrt
```

Out[34]:

```
array([[0.91669867, 0.10892895, 0.49549179, 0.23283593, 0.43686066],
       [0.75154299, 0.48089213, 0.79772841, 0.28270293, 0.43341824],
       [0.00975735, 0.34079598, 0.68927201, 0.86936929, 0.26780382],
       [0.45674792, 0.26828131, 0.8370528 , 0.27051466, 0.53006201],
       [0.17537266, 0.31496645, 0.8911091 , 0.18033628, 0.4943162 ]])
```

In [39]:

```
mydata = pd.read_csv("autos_mpg.csv")
mydata.head()
```

Out[39]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino

In [40]:

```
mydata.tail()
```

Out[40]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
393	27.0	4	140.0	86	2790	15.6	82	1	ford mustang gl
394	44.0	4	97.0	52	2130	24.6	82	2	vw pickup
395	32.0	4	135.0	84	2295	11.6	82	1	dodge rampage
396	28.0	4	120.0	79	2625	18.6	82	1	ford ranger
397	31.0	4	119.0	82	2720	19.4	82	1	chevy s-10

In [41]:

```
mydata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 398 entries, 0 to 397
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   mpg           398 non-null    float64
 1   cylinders     398 non-null    int64  
 2   displacement  398 non-null    float64
 3   horsepower    398 non-null    object 
 4   weight        398 non-null    int64  
 5   acceleration  398 non-null    float64
 6   model_year    398 non-null    int64  
 7   origin        398 non-null    int64  
 8   car_name      398 non-null    object 
dtypes: float64(3), int64(4), object(2)
memory usage: 28.1+ KB
```

In [42]:
```
mydata.describe()
```

Out[42]:

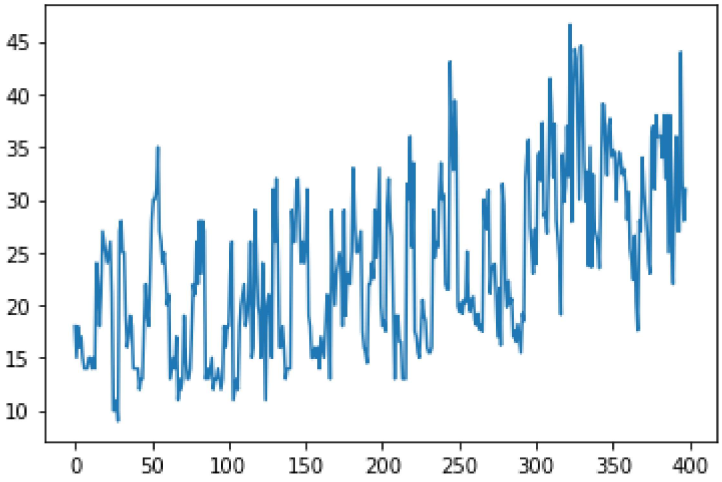
	mpg	cylinders	displacement	weight	acceleration	model_year	origin
count	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000
mean	23.514573	5.454774	193.425879	2970.424623	15.568090	76.010050	1.572864
std	7.815984	1.701004	104.269838	846.841774	2.757689	3.697627	0.802055
min	9.000000	3.000000	68.000000	1613.000000	8.000000	70.000000	1.000000
25%	17.500000	4.000000	104.250000	2223.750000	13.825000	73.000000	1.000000
50%	23.000000	4.000000	148.500000	2803.500000	15.500000	76.000000	1.000000
75%	29.000000	8.000000	262.000000	3608.000000	17.175000	79.000000	2.000000
max	46.600000	8.000000	455.000000	5140.000000	24.800000	82.000000	3.000000

In [43]:
```
#Referencing a column
mydata['mpg']
```

Out[43]:
```
0      18.0
1      15.0
2      18.0
3      16.0
4      17.0
       ...
393    27.0
394    44.0
395    32.0
396    28.0
397    31.0
Name: mpg, Length: 398, dtype: float64
```

In [46]:
```
plt.plot(mydata['mpg'])
```

Out[46]:
```
[<matplotlib.lines.Line2D at 0x289fb8c33a0>]
```



In [44]:
```
mydata['cylinders']
```

Out[44]:

```
0      8
1      8
2      8
3      8
4      8
      ..
393    4
394    4
395    4
396    4
397    4
Name: cylinders, Length: 398, dtype: int64
```

In [48]:

```
twodata=mydata[['mpg', 'cylinders']]
```

In [49]:

```
twodata
```

Out[49]:

	mpg	cylinders
0	18.0	8
1	15.0	8
2	18.0	8
3	16.0	8
4	17.0	8
...	...	...
393	27.0	4
394	44.0	4
395	32.0	4
396	28.0	4
397	31.0	4

398 rows × 2 columns

In [51]:

```
#adding a new column
mydata['is_overspeed'] = False
```

In [52]:

```
mydata
```

Out[52]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino
...	...	...	...	...	...	...	...	...	...
393	27.0	4	140.0	86	2790	15.6	82	1	ford mustang gl
394	44.0	4	97.0	52	2130	24.6	82	2	vw pickup
395	32.0	4	135.0	84	2295	11.6	82	1	dodge rampage
396	28.0	4	120.0	79	2625	18.6	82	1	ford ranger
397	31.0	4	119.0	82	2720	19.4	82	1	chevy s-10

398 rows × 10 columns

In [53]:

```
mydata.drop(columns='is_overspeed')
```

Out[53]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino
...	...	...	...	...	...	...	...	...	...
393	27.0	4	140.0	86	2790	15.6	82	1	ford mustang gl
394	44.0	4	97.0	52	2130	24.6	82	2	vw pickup
395	32.0	4	135.0	84	2295	11.6	82	1	dodge rampage
396	28.0	4	120.0	79	2625	18.6	82	1	ford ranger
397	31.0	4	119.0	82	2720	19.4	82	1	chevy s-10

398 rows × 9 columns

In [54]:

```
mydata
```

Out[54]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino
...	...	...	...	...	...	...	...	...	...
393	27.0	4	140.0	86	2790	15.6	82	1	ford mustang gl
394	44.0	4	97.0	52	2130	24.6	82	2	vw pickup
395	32.0	4	135.0	84	2295	11.6	82	1	dodge rampage
396	28.0	4	120.0	79	2625	18.6	82	1	ford ranger
397	31.0	4	119.0	82	2720	19.4	82	1	chevy s-10

398 rows × 10 columns

```
In [59]: mydata.drop(columns='is_overspeed', inplace=True, axis=1)
```

```
In [60]: mydata
```

Out[60]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino
...	...	...	...	...	...	...	...	...	...
393	27.0	4	140.0	86	2790	15.6	82	1	ford mustang gl
394	44.0	4	97.0	52	2130	24.6	82	2	vw pickup
395	32.0	4	135.0	84	2295	11.6	82	1	dodge rampage
396	28.0	4	120.0	79	2625	18.6	82	1	ford ranger
397	31.0	4	119.0	82	2720	19.4	82	1	chevy s-10

398 rows × 9 columns

In [61]:

```
mydata
```

Out[61]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino
...	...	...	...	...	...	...	...	...	...
393	27.0	4	140.0	86	2790	15.6	82	1	ford mustang gl
394	44.0	4	97.0	52	2130	24.6	82	2	vw pickup
395	32.0	4	135.0	84	2295	11.6	82	1	dodge rampage
396	28.0	4	120.0	79	2625	18.6	82	1	ford ranger
397	31.0	4	119.0	82	2720	19.4	82	1	chevy s-10

398 rows × 9 columns

In [63]:

```
#Saving the data
mydata.to_csv('mpg_updated.csv')
```

In [64]:

```
#groupby method
mydata.groupby('cylinders').count()
```

Out[64]:

	mpg	displacement	horsepower	weight	acceleration	model_year	origin	car_name
cylinders								
3	4	4	4	4	4	4	4	4
4	204	204	204	204	204	204	204	204
5	3	3	3	3	3	3	3	3
6	84	84	84	84	84	84	84	84
8	103	103	103	103	103	103	103	103

In [ ]: