

Data Mining

Lab - 1

Name : Vyas Bhagyesh Y.

Div : 5-B

Batch : 2

Roll-No : 423

Enrollment No : 23010101662

Introduction to Pandas Library Function:

Step-1 Import the pandas Libraries

In [1]:

```
import pandas as pd
```

Step-2 Import the dataset from this:....

In [2]:

```
data=('titanic.csv')
```

Step-3 Read csv or excel File

In [3]:

```
df=pd.read_csv('titanic.csv')
```

Step-4 Print Data from csv or excel File

In [4]:

```
df
```

Out [4]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

Step-5 See the First 10 Rows

In [5]:

```
df.head(10)
```

Out [5]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

Step-6 See the Last 10 Rows

In [6]:

```
df.tail(10)
```

Out[6]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	NaN	S
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

Step-7 Data type of each columns

In [7]:

```
df.dtypes
```

Out[7]:

```
PassengerId      int64
Survived          int64
Pclass            int64
Name              object
Sex               object
Age              float64
SibSp             int64
Parch             int64
Ticket            object
Fare              float64
Cabin             object
Embarked          object
dtype: object
```

Step-8 Display Summary Information

In [8]:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null    int64
1   Survived        891 non-null    int64
2   Pclass         891 non-null    int64
3   Name            891 non-null    object
4   Sex             891 non-null    object
5   Age            714 non-null    float64
6   SibSp          891 non-null    int64
7   Parch          891 non-null    int64
8   Ticket         891 non-null    object
9   Fare           891 non-null    float64
10  Cabin          204 non-null    object
11  Embarked       889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [9]:

```
df.describe()
```

Out[9]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

Step-9 Access a specific column

In [10]:

```
df.Name
```

Out[10]:

```
0      Braund, Mr. Owen Harris
1  Cumings, Mrs. John Bradley (Florence Briggs Th...
2      Heikkinen, Miss. Laina
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)
4      Allen, Mr. William Henry
...
886      Montvila, Rev. Juozas
887      Graham, Miss. Margaret Edith
888  Johnston, Miss. Catherine Helen "Carrie"
889      Behr, Mr. Karl Howell
890      Dooley, Mr. Patrick
Name: Name, Length: 891, dtype: object
```

Step-10 Access rows by their integer location

In [11]:

1	PassengerId	2	Survived	Pclass	Name	Sex	Age	SibSp	Parch	0	Ticket	Fare	Cabin	Embarked	Fare2
					Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	0	PC 17599	71.2833	C85		
2		3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	0	STON/O2. 3101282	7.9250	NaN	S	1
3		4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	0	113803	53.1000	C123	S	1
4		5	0	3	Allen, Mr. William Henry	male	35.0	0	0	0	373450	8.0500	NaN	S	1
...	
886		887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	0	211536	13.0000	NaN	S	1
887		888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	0	112053	30.0000	B42	S	1
888		889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	0	W./C. 6607	23.4500	NaN	S	1
889		890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	0	111369	30.0000	C148	C	1
890		891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	0	370376	7.7500	NaN	Q	1

891 rows × 13 columns

Step-13 Perform Condition Selection on DataFrame

In [14]:

```
df['Age']>10
```

Out[14]:

```
0      True
1      True
2      True
3      True
4      True
...
886    True
887    True
888   False
889    True
890    True
Name: Age, Length: 891, dtype: bool
```

In [15]:

```
df['Sex']=='female'
```

Out[15]:

```
0      False
1       True
2       True
3       True
4      False
...
886   False
```

887 True
888 True
889 False
890 False
Name: Sex, Length: 891, dtype: bool

In [18]:

```
df[(df['Sex']=='female') & (df['Age']>10)]
```

Out[18]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Fare2
1	2	1	1Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	1
2	3	1	3Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	1
3	4	1	1Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S	1
8	9	1	3Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S	1
9	10	1	2Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C	1
...
879	880	1	1Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C	1
880	881	1	2Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S	1
882	883	0	3Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S	1
885	886	0	3Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q	1
887	888	1	1Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S	1

230 rows x 13 columns

In [22]:

```
df[(df['Fare']>300) & (df['Age']>35)]
```

Out[22]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Fare2
679	680	1	1Cardeza, Mr. Thomas Drake	male	36.0	0	1	PC 17755	512.3292	B51 B53	C	1

PassengerId	Survived	Pclass	Martinez Name	Sex	Age	SibSp	Parch	Ticket	Fare	B55 Cabin	Embarked	Fare2
-------------	----------	--------	------------------	-----	-----	-------	-------	--------	------	--------------	----------	-------

Step-14 Compute the sum of value

In [23]:

```
df.Age.sum()
```

Out[23]:

21205.17

Step-15 Compute the mean of value

In [24]:

```
df.Age.mean()
```

Out[24]:

29.69911764705882

Step-16 Count non-null value (column)

In [28]:

```
df.Age.count()
```

Out[28]:

714

In [27]:

```
df.count()
```

Out[27]:

```
PassengerId      891
Survived          891
Pclass            891
Name              891
Sex               891
Age              714
SibSp             891
Parch             891
Ticket            891
Fare              891
Cabin            204
Embarked          889
Fare2             891
dtype: int64
```

Step-17 Find Minimun or Maximum values

In [29]:

```
df.Age.min()
```

Out[29]:

0.42

In [30]:

```
df.Age.max()
```



```
Out[30]:
```

```
80.0
```