Darshan
UNIVERSITY
योगः कर्मसु कौशलम्

`Data Mining`

`Lab - 1`

Name : Vora Yagnik

Enrollment no : 23010101661

Introduction to Pandas Library Function:

⌄ Step-1 Import the pandas Libraries

```
import pandas as pd
```

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

```
file = "titanic.csv"
```

⌄ Step-3 Read csv or excel File

```
df = pd.read_csv(file)
```

⌄ Step-4 Print Data from csv or excel File

```
df
```

	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

⌄ Step-5 See the First 10 Rows

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

	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

⌄ Step-6 See the Last 10 Rows

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

	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

Step-7 Data type of each columns

```
df.dtypes
```

```
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

```
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
```

Step-9 Access a specific column

```
df["Ticket"]
```

```
0              A/5 21171
1               PC 17599
2       STON/O2. 3101282
3                 113803
4                 373450
             ...
886               211536
887               112053
888           W./C. 6607
889               111369
890               370376
Name: Ticket, Length: 891, dtype: object
```

Step-10 Access rows by their integer location

```
df.iloc[100]
```

```
PassengerId                           102
Survived                                0
Pclass                                  3
Name           Petroff, Mr. Pastcho ("Pentcho")
Sex                                  male
Age                                   NaN
SibSp                                   0
Parch                                   0
Ticket                             349215
Cabin                                 NaN
Embarked                                S
Name: 101, dtype: object
```

Step-11 Delete a specific Column

```
df = df.drop(columns= ["Fare"])
df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	NaN	S
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	NaN	Q
...	...	...	...		...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	NaN	Q

Step-12 Create a new Column

```
df["Fare"] = "1"
```

```
df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Cabin	Embarked	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	NaN	S	1
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	NaN	S	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	C123	S	1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	NaN	S	1
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	NaN	Q	1
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	NaN	S	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	B42	S	1
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	NaN	S	1
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	C148	C	1

Step-13 Perform Condition Selection on DataFrame

```
data = df[df['Age'] > 65]
print(data)
```

```
     PassengerId  Survived  Pclass                                    Name  \
33            34         0       2                   Wheadon, Mr. Edward H
96            97         0       1                   Goldschmidt, Mr. George B
116          117         0       3                   Connors, Mr. Patrick
493          494         0       1                   Artagaveytia, Mr. Ramon
630          631         1       1   Barkworth, Mr. Algernon Henry Wilson
672          673         0       2                   Mitchell, Mr. Henry Michael
745          746         0       1                   Crosby, Capt. Edward Gifford
851          852         0       3                   Svensson, Mr. Johan

       Sex   Age  SibSp  Parch     Ticket Cabin Embarked Fare
33    male  66.0      0      0  C.A. 24579   NaN        S    1
```

```
96   male  71.0    0    0    PC 17754    A5     C    1
116  male  70.5    0    0      370369    NaN    Q    1
493  male  71.0    0    0    PC 17609    NaN    C    1
630  male  80.0    0    0       27042    A23    S    1
672  male  70.0    0    0  C.A. 24580    NaN    S    1
745  male  70.0    1    1    WE/P 5735   B22    S    1
851  male  74.0    0    0      347060    NaN    S    1
```

⌄ Step-14 Compute the sum of value

```
data = df['Survived'].sum()
print(data)
```

    341

⌄ Step-15 Compute the mean of value

```
data = df['Age'].mean()
print(data)
```

    29.687475455820476

⌄ Step-16 Count non-null value (column)

```
data =  df['Age'].count()
print(data)
```

    713

⌄ Step-17 Find Minimun or Maximum values

```
minAge = df['Age'].min()
maxAge = df['Age'].max()
print(minAge)
print(maxAge)
```

    0.42
    80.0