Darshan
UNIVERSITY
योग: कर्मसु कौशलम्

Data Mining

Lab - 1

Name: Smit Maru

Enrollment No: 23010101161

Introduction to Pandas Library Function:

Step-1 Import the pandas Libraries

```
In [4]: import pandas as pd
```

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

```
In [5]: df = pd.read_csv('titanic.csv')
```

Step-3 Read csv or excel File

```
In [6]: df = pd.read_csv('titanic.csv')
```

Step-4 Print Data from csv or excel File

```
In [7]: df
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75

891 rows × 12 columns

Step-5 See the First 10 Rows

In [8]:

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

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.458
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.862
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.075
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.133
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.070

Step-6 See the Last 10 Rows

In [9]:

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

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.

Step-7 Data type of each columns

```
In [10]: df.dtypes
```

```
Out[10]: 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 [11]: df.info
```

Out[11]:

```
<bound method DataFrame.info of      PassengerId  Survived  Pclass  \
0              1         0       3   
1              2         1       1   
2              3         1       3   
3              4         1       1   
4              5         0       3   
..           ...       ...     ...   
886          887         0       2   
887          888         1       1   
888          889         0       3   
889          890         1       1   
890          891         0       3   

                                                  Name     Sex   Age  SibSp  \
0                              Braund, Mr. Owen Harris    male  22.0      1   
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1   
2                               Heikkinen, Miss. Laina  female  26.0      0   
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1   
4                             Allen, Mr. William Henry    male  35.0      0   
..                                                 ...     ...   ...    ...   
886                              Montvila, Rev. Juozas    male  27.0      0   
887                       Graham, Miss. Margaret Edith  female  19.0      0   
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1   
889                              Behr, Mr. Karl Howell    male  26.0      0   
890                                Dooley, Mr. Patrick    male  32.0      0   

     Parch            Ticket     Fare Cabin Embarked  
0        0         A/5 21171   7.2500   NaN        S  
1        0          PC 17599  71.2833   C85        C  
2        0  STON/O2. 3101282   7.9250   NaN        S  
3        0            113803  53.1000  C123        S  
4        0            373450   8.0500   NaN        S  
..     ...               ...      ...   ...      ...  
886      0            211536  13.0000   NaN        S  
887      0            112053  30.0000   B42        S  
888      2        W./C. 6607  23.4500   NaN        S  
889      0            111369  30.0000  C148        C  
890      0            370376   7.7500   NaN        Q  

[891 rows x 12 columns]>
```

Step-9 Access a specific column

In [12]:

```python
df['Name']
```

Out[12]:

```
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 [13]:

```
print(df.iloc[1])
```

```
PassengerId                                                    2
Survived                                                       1
Pclass                                                         1
Name           Cumings, Mrs. John Bradley (Florence Briggs Th...
Sex                                                       female
Age                                                         38.0
SibSp                                                          1
Parch                                                          0
Ticket                                                  PC 17599
Fare                                                     71.2833
Cabin                                                        C85
Embarked                                                       C
Name: 1, dtype: object
```

Step-11 Delete a specific Column

In [14]:

```
df = df.drop('Cabin', axis=1)
```

Step-12 Create a new Column

In [15]:

```
df['Status'] = 'Active'
df
```

Out[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

891 rows × 12 columns

Step-13 Perform Condition Selection on DataFrame

In [16]:
```
male = df[df['Sex'] == 'male']
print(male)
```

```
     PassengerId  Survived  Pclass                            Name   Sex  \
0              1         0       3         Braund, Mr. Owen Harris  male
4              5         0       3        Allen, Mr. William Henry  male
5              6         0       3                Moran, Mr. James  male
6              7         0       1         McCarthy, Mr. Timothy J  male
7              8         0       3  Palsson, Master. Gosta Leonard  male
..           ...       ...     ...                             ...   ...
883          884         0       2   Banfield, Mr. Frederick James  male
884          885         0       3          Sutehall, Mr. Henry Jr  male
886          887         0       2           Montvila, Rev. Juozas  male
889          890         1       1           Behr, Mr. Karl Howell  male
890          891         0       3             Dooley, Mr. Patrick  male

      Age  SibSp  Parch            Ticket     Fare Embarked  Status
0    22.0      1      0         A/5 21171   7.2500        S  Active
4    35.0      0      0            373450   8.0500        S  Active
5     NaN      0      0            330877   8.4583        Q  Active
6    54.0      0      0             17463  51.8625        S  Active
7     2.0      3      1            349909  21.0750        S  Active
..    ...    ...    ...               ...      ...      ...     ...
883  28.0      0      0  C.A./SOTON 34068  10.5000        S  Active
884  25.0      0      0   SOTON/OQ 392076   7.0500        S  Active
886  27.0      0      0            211536  13.0000        S  Active
889  26.0      0      0            111369  30.0000        C  Active
890  32.0      0      0            370376   7.7500        Q  Active

[577 rows x 12 columns]
```

Step-14 Compute the sum of value

In [17]:
```
total_fare = df['Fare'].sum()
print("Total Fare:", total_fare)
```

```
Total Fare: 28693.9493
```

Step-15 Compute the mean of value

In [18]:
```
mean_age = df['Age'].mean()
print("Mean Age:", mean_age)
```

```
Mean Age: 29.69911764705882
```

Step-16 Count non-null value (column)

In [19]:

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

```
714
```

Step-17 Find Minimun or Maximum values

In [20]:

```
print("Min Age:", df['Age'].min())
print("Max Age:", df['Age'].max())
```

```
Min Age: 0.42
Max Age: 80.0
```