Data Mining Lab - 1

Name: Sumit Chuahan

Enrollment No: 23010101610

Batch: 5B-3

Date: 20-6-24

# Introduction to Pandas Library Function:

## Step-1 Import the pandas Libraries

```
import pandas as pd

pd.__version__
```

```
'2.2.2'
```

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

## Step-3 Read csv or excel File

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

## Step-4 Print Data from csv or excel File

```
df
```

```
     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-5 See the First 10 Rows

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

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3
5            6         0       3
6            7         0       1
7            8         0       3
8            9         1       3
9           10         1       2

                                                 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
5                                    Moran, Mr. James    male   NaN
0
6                             McCarthy, Mr. Timothy J    male  54.0
0
7                      Palsson, Master. Gosta Leonard    male   2.0
3
8  Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0
0
9                 Nasser, Mrs. Nicholas (Adele Achem)  female  14.0
1

   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
5      0            330877   8.4583   NaN        Q
6      0             17463  51.8625   E46        S
7      1            349909  21.0750   NaN        S
8      2            347742  11.1333   NaN        S
9      0            237736  30.0708   NaN        C
```

# Step-6 See the Last 10 Rows

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

```
     PassengerId  Survived  Pclass
Name  \
881          882         0       3                      Markun, Mr.
Johann
882          883         0       3             Dahlberg, Miss. Gerda
Ulrika
883          884         0       2            Banfield, Mr. Frederick
James
884          885         0       3                   Sutehall, Mr.
Henry Jr
885          886         0       3      Rice, Mrs. William (Margaret
Norton)
886          887         0       2                    Montvila, Rev.
Juozas
887          888         1       1             Graham, Miss. Margaret
Edith
888          889         0       3  Johnston, Miss. Catherine Helen
"Carrie"
889          890         1       1                    Behr, Mr. Karl
Howell
890          891         0       3                      Dooley, Mr.
Patrick

        Sex   Age  SibSp  Parch            Ticket     Fare Cabin
Embarked
881    male  33.0      0      0            349257   7.8958   NaN
S
882  female  22.0      0      0              7552  10.5167   NaN
S
883    male  28.0      0      0  C.A./SOTON 34068  10.5000   NaN
S
884    male  25.0      0      0   SOTON/OQ 392076   7.0500   NaN
S
885  female  39.0      0      5            382652  29.1250   NaN
Q
886    male  27.0      0      0            211536  13.0000   NaN
S
887  female  19.0      0      0            112053  30.0000   B42
S
888  female   NaN      1      2        W./C. 6607  23.4500   NaN
S
889    male  26.0      0      0            111369  30.0000  C148
C
890    male  32.0      0      0            370376   7.7500   NaN
Q
```

# 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
```

```
df.describe()
```

```
       PassengerId    Survived      Pclass         Age       SibSp
Fare
count   891.000000  891.000000  891.000000  714.000000  891.000000
891.000000
```

```
mean    446.000000    0.383838    2.308642   29.699118    0.523008
32.204208
std     257.353842    0.486592    0.836071   14.526497    1.102743
49.693429
min       1.000000    0.000000    1.000000    0.420000    0.000000
0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
7.910400
50%     446.000000    0.000000    3.000000   28.000000    0.000000
14.454200
75%     668.500000    1.000000    3.000000   38.000000    1.000000
31.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000
512.329200
```

# Step-9 Access a specific column

```
df.Age
```

```
0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888     NaN
889    26.0
890    32.0
Name: Age, Length: 891, dtype: float64
```

# Step-10 Access rows by their integer location

```
df.iloc[10]
```

```
PassengerId                                 11
Survived                                     1
Pclass                                       3
Name           Sandstrom, Miss. Marguerite Rut
Sex                                       female
Age                                          4.0
SibSp                                        1
Parch                                        1
Ticket                                    PP 9549
Fare                                        16.7
```

```
Cabin                                           G6
Embarked                                         S
Name: 10, dtype: object
```

# Step-11 Delete a specific Column

```
df.drop('Parch',axis=1,inplace=False)
```

# Step-12 Create a new Column

```
df["Demo"] = 0
df
```

```
     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

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

[891 rows x 12 columns]
```

# Step-13 Perform Condition Selection on DataFrame

```
df['Fare'] > 10
```

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

# Step-14 Compute the sum of value

```
df["Fare"].sum()
```

```
28693.9493
```

## Step-15 Compute the mean of value

```
df['Fare'].mean()
```

```
32.204207968574636
```

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

```
df["Age"].count()
```

```
714
```

## Step-17 Find Minimun or Maximum values

```
df["Age"].min()
```

```
0.42
```

```
df["Age"].max()
```

```
80.0
```