Darshan UNIVERSITY
योग: कर्मसु कौशलम्

(https://www.darshan.ac.in/)

Data Mining

Lab - 1

Name : Gohel Krish V.

Roll No. : 216

Enrollment NO. : 22010101060

Date : 13/06/2024

Introduction to Pandas Library Function:

Step-1 Import the pandas Libraries

In [2]:
```
import pandas as pd
```

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

In [ ]:
```

```

Step-3 Read csv or excel File

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

Step-4 Print Data from csv or excel File

In [6]:
```
df
```

Out[6]:

	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	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

891 rows × 12 columns

Step-5 See the First 10 Rows

In [7]:
```
df.head(10)
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	N
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	N
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	N
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	N
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	N

Step-6 See the Last 10 Rows

In [8]:
```
df.tail(10)
```

Out[8]:

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

Step-7 Data type of each columns

In [12]:
```
df.dtypes
```

Out[12]:
```
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 [15]:
```
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

In [17]:
```
df.Ticket
```

Out[17]:
```
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

In [23]:
```
df[4:10] # use for multiple columns
# or
df.iloc[173] # use for specified column
```

Out[23]:
```
PassengerId                          174
Survived                               0
Pclass                                 3
Name           Sivola, Mr. Antti Wilhelm
Sex                                 male
Age                                 21.0
SibSp                                  0
Parch                                  0
Ticket                  STON/O 2. 3101280
Fare                                7.925
Cabin                                 NaN
Embarked                                S
Name: 173, dtype: object
```

Step-11 Delete a specific Column

In [25]:
```
# df.drop('Cabin',axis=1,inplace = True)
df.pop('Age')
```

Out[25]:
```
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-12 Create a new Column

In [26]:
```
import random
alist = []
for i in range(0,891):
    alist.append(random.choice(['yes','no']))

df['Alive'] = alist
```

Step-13 Perform Condition Selection on DataFrame

In [31]:
```
df[df['Survived'] > 0]
```

Out[31]:

	PassengerId	Survived	Pclass	Name	Sex	SibSp	Parch	Ticket	Fare	Embar
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	1	0	113803	53.1000	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	0	2	347742	11.1333	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	1	0	237736	30.0708	
...	...	...	...	...	...	...	...	...	...	...
875	876	1	3	Najib, Miss. Adele Kiamie "Jane"	female	0	0	2667	7.2250	
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	0	1	11767	83.1583	
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	0	1	230433	26.0000	
887	888	1	1	Graham, Miss. Margaret Edith	female	0	0	112053	30.0000	
889	890	1	1	Behr, Mr. Karl Howell	male	0	0	111369	30.0000	

342 rows × 11 columns

Step-14 Compute the sum of value

In [35]:
```
print(df['Fare'].sum())
print(df['Pclass'].sum())
```

```
28693.9493
2057
```

In [36]:
```
df
```

Out[36]:

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

891 rows × 11 columns

Step-15 Compute the mean of value

In [42]:
```
df['Fare'].mean()
```

Out[42]:
```
32.204207968574636
```

Step-16 Count non-null value (column)

In [43]:
```
df['PassengerId'].count()
```

Out[43]:
```
891
```

Step-17 Find Minimun or Maximum values

In [46]:
```
print("Maximum Survived : ",df['Survived'].max())
print("Minimum Survived : ",df['Survived'].min())
```

```
Maximum Survived :  1
Minimum Survived :  0
```