Darshan
UNIVERSITY
योग: कर्मसु कौशलम्

Data Mining

Lab - 1

Mohil Parmar

23010101192

Introduction to Pandas Library Function:

Step-1 Import the pandas Libraries

```
In [3]: import pandas as pd
```

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

```
In [ ]:
```

Step-3 Read csv or excel File

```
In [13]: df = pd.read_csv('titanic.csv')
```

Step-4 Print Data from csv or excel File

```
In [24]: df
```

Out[24]:

	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 [26]:

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

Out[26]:

	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
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 [28]:

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

Out[28]:

	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 [36]:

```
df.dtypes
```

Out[36]:

```
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 [32]:

```
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 [56]:

```
# df.shape // return tuple of (row , column)
# df.shape[0] // return total number of rows
```

In [54]:

```
df.describe()
```

Out[54]:

	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 [65]:

```
# list pass
df[['Age','SibSp']]
```

Out[65]:

	Age	SibSp
0	22.0	1
1	38.0	1
2	26.0	0
3	35.0	1
4	35.0	0
...	...	...
886	27.0	0
887	19.0	0
888	NaN	1
889	26.0	0
890	32.0	0

891 rows × 2 columns

Step-10 Access rows by their integer location

In [77]:

```
df.iloc[10:20]
```

Out[77]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750	NaN	S
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0000	NaN	S
16	17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29.1250	NaN	Q
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	NaN	S
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18.0000	NaN	S
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250	NaN	C

Step-11 Delete a specific Column

In [129…

```
# df.drop('Age',axis=1,inplace=True) permananet delete
df.drop(columns='SibSp')
```

Out[129…

	PassengerId	Survived	Pclass	Name	Parch	Ticket	Fare	Cabin	Embarked	NewFare	Gender
0	1	0	3	Braund, Mr. Owen Harris	0	A/5 21171	7.2500	NaN	S	7.97500	male
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	PC 17599	71.2833	C85	C	78.41163	male
2	3	1	3	Heikkinen, Miss. Laina	0	STON/O2. 3101282	7.9250	NaN	S	8.71750	male
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	113803	53.1000	C123	S	58.41000	male
4	5	0	3	Allen, Mr. William Henry	0	373450	8.0500	NaN	S	8.85500	male
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	0	211536	13.0000	NaN	S	14.30000	male
887	888	1	1	Graham, Miss. Margaret Edith	0	112053	30.0000	B42	S	33.00000	male
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	2	W./C. 6607	23.4500	NaN	S	25.79500	male
889	890	1	1	Behr, Mr. Karl Howell	0	111369	30.0000	C148	C	33.00000	male
890	891	0	3	Dooley, Mr. Patrick	0	370376	7.7500	NaN	Q	8.52500	male

891 rows × 11 columns

Step-12 Create a new Column

In [137…

```
df['NewFare']=df['Fare']*1.1

df
```

Out[137…

	PassengerId	Survived	Pclass	Name	SibSp	Parch	Ticket	Fare	Cabin	Embarked	NewFare
0	1	0	3	Braund, Mr. Owen Harris	1	0	A/5 21171	7.2500	NaN	S	7.97500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	0	PC 17599	71.2833	C85	C	78.41163
2	3	1	3	Heikkinen, Miss. Laina	0	0	STON/O2. 3101282	7.9250	NaN	S	8.71750
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	0	113803	53.1000	C123	S	58.41000
4	5	0	3	Allen, Mr. William Henry	0	0	373450	8.0500	NaN	S	8.85500
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	0	0	211536	13.0000	NaN	S	14.30000
887	888	1	1	Graham, Miss. Margaret Edith	0	0	112053	30.0000	B42	S	33.00000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	1	2	W./C. 6607	23.4500	NaN	S	25.79500
889	890	1	1	Behr, Mr. Karl Howell	0	0	111369	30.0000	C148	C	33.00000
890	891	0	3	Dooley, Mr. Patrick	0	0	370376	7.7500	NaN	Q	8.52500

891 rows × 11 columns

Step-13 Perform Condition Selection on DataFrame

In [141…

```
df[(df['Fare']>8) & (df['Fare'] < 10)]
```

Out[141…

	PassengerId	Survived	Pclass	Name	SibSp	Parch	Ticket	Fare	Cabin	Embarked	NewFare
4	5	0	3	Allen, Mr. William Henry	0	0	373450	8.0500	NaN	S	8.85500
5	6	0	3	Moran, Mr. James	0	0	330877	8.4583	NaN	Q	9.30413
12	13	0	3	Saundercock, Mr. William Henry	0	0	A/5. 2151	8.0500	NaN	S	8.85500
22	23	1	3	McGowan, Miss. Anna "Annie"	0	0	330923	8.0292	NaN	Q	8.83212
37	38	0	3	Cann, Mr. Ernest Charles	0	0	A./5. 2152	8.0500	NaN	S	8.85500
...	...	...	...	...	...	...	...	...	...	...	...
844	845	0	3	Culumovic, Mr. Jeso	0	0	315090	8.6625	NaN	S	9.52875
855	856	1	3	Aks, Mrs. Sam (Leah Rosen)	0	1	392091	9.3500	NaN	S	10.28500
868	869	0	3	van Melkebeke, Mr. Philemon	0	0	345777	9.5000	NaN	S	10.45000
873	874	0	3	Vander Cruyssen, Mr. Victor	0	0	345765	9.0000	NaN	S	9.90000
876	877	0	3	Gustafsson, Mr. Alfred Ossian	0	0	7534	9.8458	NaN	S	10.83038

95 rows × 11 columns

Step-14 Compute the sum of value

In [149…

```
df.Fare.sum()
# df['Fare'].sum()
```

Out[149…

```
28693.9493
```

Step-15 Compute the mean of value

In [145…

```
df.Fare.mean()
```

Out[145…

```
32.204207968574636
```

Step-16 Count non-null value (column)

In [147…

```
df.isnull().sum()
```

Out[147…

```
PassengerId      0
Survived         0
Pclass           0
Name             0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
NewFare          0
dtype: int64
```

Step-17 Find Minimun or Maximum values

In [151…

```
df.Fare.max()
```

Out[151…

```
512.3292
```

In [153…

```
df.Fare.min()
```

Out[153…

```
0.0
```

In [ ]: