Data Mining

Lab - 1

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

Step-3 Read csv or excel File

In [64]:
```
pd.read_csv("titanic.csv")
```

Out[64]:

	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-4 Print Data from csv or excel File

In [25]:
```
df
```

Out[25]:

	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	C
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	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	
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	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	

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 [9]:
```
df.dtypes
```

Out[9]:
```
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 [10]:
```
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 [18]:
```
df.shape[1]
```

Out[18]:
```
12
```

Step-9 Access a specific column

In [16]:
```
df[['Age','PassengerId','Name']]
# df.Age
```

Out[16]:

	Age	PassengerId	Name
0	22.0	1	Braund, Mr. Owen Harris
1	38.0	2	Cumings, Mrs. John Bradley (Florence Briggs Th...
2	26.0	3	Heikkinen, Miss. Laina
3	35.0	4	Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	35.0	5	Allen, Mr. William Henry
...	...	...	...
886	27.0	887	Montvila, Rev. Juozas
887	19.0	888	Graham, Miss. Margaret Edith
888	NaN	889	Johnston, Miss. Catherine Helen "Carrie"
889	26.0	890	Behr, Mr. Karl Howell
890	32.0	891	Dooley, Mr. Patrick

891 rows × 3 columns

Step-10 Access rows by their integer location

In [22]:
```
df.iloc[10]
```

Out[22]:
```
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
```

In [26]:
```
#10 to 20 Rows.
df.iloc[10:20]
```

Out[26]:

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

In [27]:
```
# 10 to 20 Rows and 0 to 5
df.iloc[10:20,0:5]
```

Out[27]:

	PassengerId	Survived	Pclass	Name	Sex
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female
11	12	1	1	Bonnell, Miss. Elizabeth	female
12	13	0	3	Saundercock, Mr. William Henry	male
13	14	0	3	Andersson, Mr. Anders Johan	male
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female
16	17	0	3	Rice, Master. Eugene	male
17	18	1	2	Williams, Mr. Charles Eugene	male
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female
19	20	1	3	Masselmani, Mrs. Fatima	female

Step-11 Delete a specific Column

In [36]:
```
df.drop('Age',axis=1)
```

Out[36]:

	PassengerId	Survived	Pclass	Name	Sex	SibSp	Parch	Ticket	Fare	Cal
0	1	0	3	Braund, Mr. Owen Harris	male	1	0	A/5 21171	7.2500	N
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	0	0	STON/O2. 3101282	7.9250	N
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May	female	1	0	113803	53.1000	C1

In [41]:
```
#permanent Delete
df.drop('Pclass',axis=1,inplace=True)
```

In [42]:
```
df
```

Out[42]:

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

891 rows × 10 columns

In [43]:
```
#delete Row
df.drop(4)
```

Out[43]:

	PassengerId	Survived	Name	Sex	SibSp	Parch	Ticket	Fare	Cabin	Embar
0	1	0	Braund, Mr. Owen Harris	male	1	0	A/5 21171	7.2500	NaN	
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	1	0	PC 17599	71.2833	C85	
2	3	1	Heikkinen, Miss. Laina	female	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	1	0	113803	53.1000	C123	
5	6	0	Moran, Mr. James	male	0	0	330877	8.4583	NaN	
...	...	...	...	...	...	...	...	...	...	...
886	887	0	Montvila, Rev. Juozas	male	0	0	211536	13.0000	NaN	
887	888	1	Graham, Miss. Margaret Edith	female	0	0	112053	30.0000	B42	
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	1	2	W./C. 6607	23.4500	NaN	
889	890	1	Behr, Mr. Karl Howell	male	0	0	111369	30.0000	C148	
890	891	0	Dooley, Mr. Patrick	male	0	0	370376	7.7500	NaN	

890 rows × 10 columns

Step-12 Create a new Column

In [45]:
```
df['Amount']=df['Fare']*100
df
```

Out[45]:

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

891 rows × 11 columns

Step-13 Perform Condition Selection on DataFrame

In [63]:
```
df.query('Fare>10 & Fare<50')
```

Out[63]:

	PassengerId	Survived	Name	Sex	SibSp	Parch	Ticket	Fare	Cabin	Embarke
7	8	0	Palsson, Master. Gosta Leonard	male	3	1	349909	21.0750	NaN	
8	9	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	0	2	347742	11.1333	NaN	
9	10	1	Nasser, Mrs. Nicholas (Adele Achem)	female	1	0	237736	30.0708	NaN	
10	11	1	Sandstrom, Miss. Marguerite Rut	female	1	1	PP 9549	16.7000	G6	
11	12	1	Bonnell, Miss. Elizabeth	female	0	0	113783	26.5500	C103	
...	...	...	...	...	...	...	...	...	...	...
885	886	0	Rice, Mrs. William (Margaret Norton)	female	0	5	382652	29.1250	NaN	
886	887	0	Montvila, Rev. Juozas	male	0	0	211536	13.0000	NaN	
887	888	1	Graham, Miss. Margaret Edith	female	0	0	112053	30.0000	B42	
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	1	2	W./C. 6607	23.4500	NaN	
889	890	1	Behr, Mr. Karl Howell	male	0	0	111369	30.0000	C148	

394 rows × 11 columns

Step-14 Compute the sum of value

In [46]:
```
df['Amount'].sum()
```

Out[46]: 2869394.93

Step-15 Compute the mean of value

In [47]:
```
df['Fare'].mean()
```

Out[47]:
```
32.204207968574636
```

Step-16 Count non-null value (column)

In [57]:
```
df.count()
```

Out[57]:
```
PassengerId    891
Survived       891
Name           891
Sex            891
SibSp          891
Parch          891
Ticket         891
Fare           891
Cabin          204
Embarked       889
Amount         891
dtype: int64
```

In [61]:
```
#null value
(~df.isnull()).sum()
```

Out[61]:
```
PassengerId    891
Survived       891
Name           891
Sex            891
SibSp          891
Parch          891
Ticket         891
Fare           891
Cabin          204
Embarked       889
Amount         891
dtype: int64
```

Step-17 Find Minimun or Maximum values

In [54]:
```
df['Fare'].min()
```

Out[54]:
```
0.0
```

In [55]:
```
df['Fare'].max()
```

Out[55]:
```
512.3292
```