Data Mining

Lab - 1

23010101407

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 [10]:
```
df = pd.read_csv("titanic.csv")
df
```

Out[10]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 12 columns

Step-3 Read csv or excel File

In [14]:
```
df = pd.read_csv("titanic.csv")
df
```

Out[14]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 12 columns

Step-4 Print Data from csv or excel File

In [16]:
```
df = pd.read_csv("titanic.csv")
df
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 12 columns

Step-5 See the First 10 Rows

In [24]:
```
tp=df.head(10)
tp
```

Out[24]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0

Step-6 See the Last 10 Rows

In [26]:
```
tp=df.tail(10)
tp
```

Out[26]:

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

Step-7 Data type of each columns

In [48]:
```
df.dtypes
```

Out[48]:
```
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]:
```
tp=df.info()
tp
```

```
<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 [53]:
```
cl=df[["Name","Parch"]]
cl
```

Out[53]:

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

891 rows × 2 columns

Step-10 Access rows by their integer location

In [44]:
```
ri=df.iloc[1]
ri
```

Out[44]:
```
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 [55]:
```
ri=df.drop("Parch",axis=1,inplace=True)
ri
```

Step-12 Create a new Column

In [61]:
```
df["isCabin"]=~df['Cabin'].isnull()
df
```

Out[61]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	370376	7.7500

891 rows × 12 columns

Step-13 Perform Condition Selection on DataFrame

In [65]:
```
df[(df['Pclass']==1)]
df[(df['Age']>25)]
```

Out[65]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Ticket	Fare
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	373450	8.0500
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	17463	51.8625
...	...	...	...	...	...	...	...	...	...
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	C.A./SOTON 34068	10.5000
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	382652	29.1250
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	211536	13.0000
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	370376	7.7500

413 rows × 12 columns

In [73]:
```
df[(df['Pclass']==1) & (df['Sex']=='male')]
```

Out[73]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Ticket	Fare	C
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	17463	51.8625	
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	113788	35.5000	
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	19950	263.0000	
30	31	0	1	Uruchurtu, Don. Manuel E	male	40.0	0	PC 17601	27.7208	
34	35	0	1	Meyer, Mr. Edgar Joseph	male	28.0	1	PC 17604	82.1708	
...	...	...	...	...	...	...	...	...	...	
839	840	1	1	Marechal, Mr. Pierre	male	NaN	0	11774	29.7000	
857	858	1	1	Daly, Mr. Peter Denis	male	51.0	0	113055	26.5500	
867	868	0	1	Roebling, Mr. Washington Augustus II	male	31.0	0	PC 17590	50.4958	
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	695	5.0000	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	111369	30.0000	

122 rows × 12 columns

Step-14 Compute the sum of value

In [75]:
```
df["Age"].sum()
```

Out[75]:
```
21205.17
```

Step-15 Compute the mean of value

In [77]:
```
df["Age"].mean()
```

Out[77]:
```
29.69911764705882
```

Step-16 Count non-null value (column)

In [79]:
```
df["Age"].count()
```

Out[79]: 714

Step-17 Find Minimun or Maximum values

In [83]:
```
df["Age"].max()
df["Age"].min()
```

Out[83]: 0.42

In [ ]: