Data Mining

Karan Sonagara

Enrollment -- 24010101694

Rollno -- 138

Lab - 1

In [ ]:

Introduction to Pandas Library Function:

In [ ]:
```
The Pandas library is a powerful and popular Python library used for data manipu
```

Step-1 Import the pandas Libraries

In [4]:
```
import pandas as pd
```

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

In [30]:
```
import pandas as pd
```

Step-3 Read csv or excel File

In [27]:
```
##read CSV File

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

Out[27]:

	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 [32]:
```
##read CSV File

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

Out[32]:

	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 [36]:
```
df = pd.read_csv("titanic.csv")
df.head(10)
```

Out[36]:

	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 [43]:
```
df = pd.read_csv("titanic.csv")
df.tail(10)
```

Out[43]:

	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 [194…
```
df.dtypes
```

Out[194…
```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
isCabin           bool
dtype: object
```

Step-8 Display Summary Information

In [45]:
```
df = pd.read_csv("titanic.csv")
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 [83]:
```
df["Age"]
```

Out[83]:
```
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

In [99]:
```
df = pd.read_csv("titanic.csv")
df.iloc[2]
```

Out[99]:
```
PassengerId                     3
Survived                        1
Pclass                          3
Name           Heikkinen, Miss. Laina
Sex                           female
Age                             26.0
SibSp                              0
Parch                              0
Ticket               STON/O2. 3101282
Fare                           7.925
Cabin                            NaN
Embarked                           S
Name: 2, dtype: object
```

Step-11 Delete a specific Column

In [121…
```
df.drop("Sex",axis=1,inplace=True)
```

Out[121...

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

891 rows × 10 columns

Step-12 Create a new Column

In [126…
```
df["isCabin"] = ~df['Cabin'].isnull()
df
```

Out[126…

	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 × 13 columns

Step-13 Perform Condition Selection on DataFrame

In [150…

```
df[df['Pclass']==1]
```

Out[150...

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.
...	...	...	...	...	...	...	...	...	...	
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.

216 rows × 13 columns

Step-14 Compute the sum of value

In [136…
```
df["Age"].sum()
```

Out[136…
```
21205.17
```

Step-15 Compute the mean of value

In [138…
```
df["Age"].mean()
```

Out[138…
```
29.69911764705882
```

Step-16 Count non-null value (column)

In [152…
```
df["Age"].count()
```

Out[152…
```
714
```

Step-17 Find Minimun or Maximum values

In [142…
```
df["Age"].min()
```

Out[142…
```
0.42
```

In [144…
```
df["Age"].max()
```

Out[144…
```
80.0
```

In [214…
```
df[(df['Pclass']==1) & (df['Age']>25)]
```

Out[214...

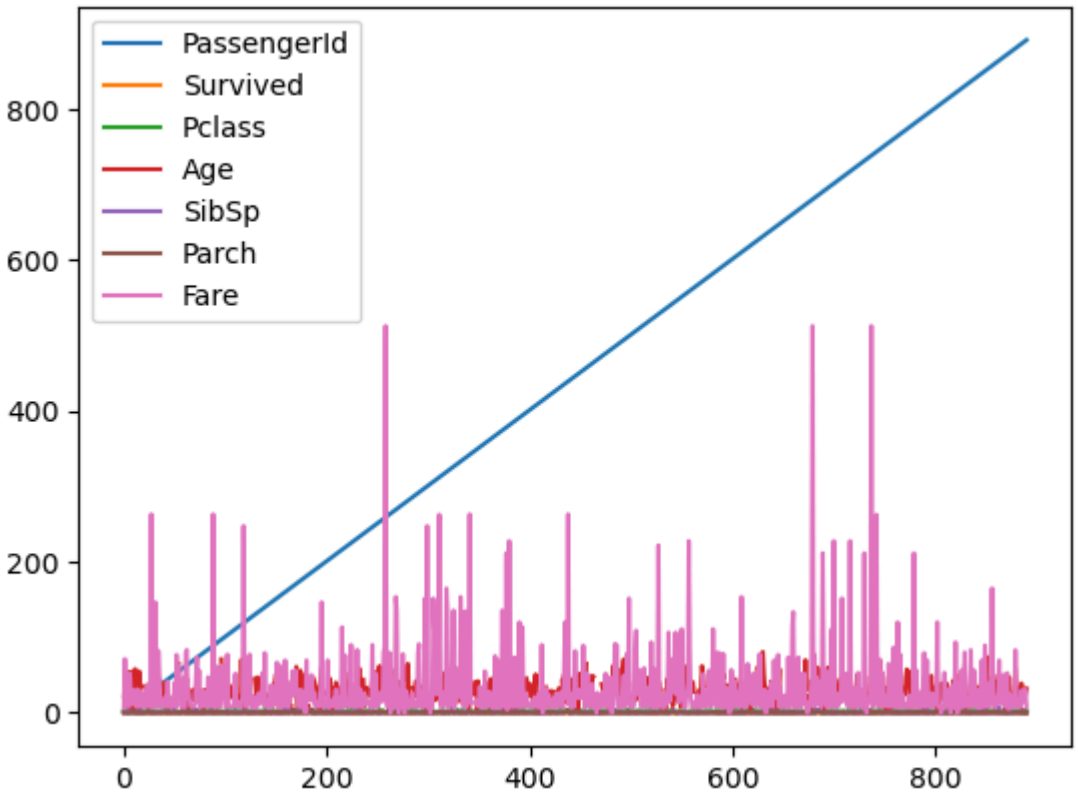
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35
...	...	...	...	...	...	...	...	...	...	
867	868	0	1	Roebling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30

144 rows × 13 columns

In [222...

```
import matplotlib.pyplot as plt
import pandas as df
df = pd.read_csv('titanic.csv')
```

```
df.plot()
plt.show()
```



In [ ]:

In [ ]: