



Data Mining

Lab - 1

Mahek Bhatti

23010101024

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

Step-3 Read csv or excel File

```
In [4]: df
```

Out[4]:

	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 [3]: df

Out[3]:

	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 [4]: df.head(10)

Out[4]:

	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	7.0
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.0
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.0
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.0
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	8.0
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.0
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	11.0
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 [6]: `df.tail(10)`

Out[6]:

	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 [8]: `df.dtypes`

```
Out[8]: 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 [9]: df.describe()
```

```
Out[9]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
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	32.200000
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693000
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	7.910000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.450000
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	512.320000

Step-9 Access a specific column

```
In [10]: df['Pclass']
```

```
Out[10]: 0      3
1      1
2      3
3      1
4      3
..
886    2
887    1
888    3
889    1
890    3
Name: Pclass, Length: 891, dtype: int64
```

how many row and column

```
In [13]: df.shape
```

```
Out[13]: (891, 12)
```

Step-10 Access rows by their integer location

```
In [14]: df.iloc[0]
```

```
Out[14]: PassengerId      1
Survived      0
Pclass       3
Name      Braund, Mr. Owen Harris
Sex      male
Age      22.0
SibSp      1
Parch      0
Ticket      A/5 21171
Fare       7.25
Cabin      NaN
Embarked     S
Name: 0, dtype: object
```

Step-11 Delete a specific Column

```
In [5]: df.drop("Name",axis='columns',inplace=True)
```

Step-12 Create a new Column

```
In [21]: df['isCabin'] = ~ df['Cabin'].isnull()
```

Step-13 Perform Condition Selection on DataFrame

```
In [23]: df["Pclass"]
```

```
Out[23]: 0      3
1      1
2      3
3      1
4      3
..
886    2
887    1
888    3
889    1
890    3
Name: Pclass, Length: 891, dtype: int64
```

Step-14 Compute the sum of value

```
In [24]: df['Fare'].sum()
```

```
Out[24]: np.float64(28693.9493)
```

Step-15 Compute the mean of value

```
In [25]: df['Fare'].mean()
```

```
Out[25]: np.float64(32.204207968574636)
```

Step-16 Count non-null value (column)

```
In [28]: (~df.isnull()).sum()
```

```
Out[28]: PassengerId      891  
Survived      891  
Pclass      891  
Name      891  
Sex      891  
Age      714  
SibSp      891  
Parch      891  
Ticket      891  
Fare      891  
Cabin      204  
isCabin      891  
dtype: int64
```

Step-17 Find Minimum or Maximum values

```
In [26]: df['Fare'].max()
```

```
Out[26]: np.float64(512.3292)
```

```
In [27]: df['Fare'].min()
```

```
Out[27]: np.float64(0.0)
```