

Data Mining

Lab - 1

abbas vakhariya

23010101288

#### **Introduction to Pandas Library Function:**

# Step-1 Import the pandas Libraries

In [5]: import pandas as pd

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

In [ ]:

#### Step-3 Read csv or excel File

In [7]: data = pd.read\_csv("titanic.csv")

#### Step-4 Print Data from csv or excel File

In [9]: data

Out

•		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	9
	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	(1)
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
ł	391 rc	ows × 12 colur	mns								
	4 6			_							

# Step-5 See the First 10 Rows

In [11]: data.head(10)

Out

	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.
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.
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0
4										•

# Step-6 See the Last 10 Rows

In [13]: data.tail(10)

Out[13]:		Passengerld	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
	4			_		_				•

# Step-7 Data type of each columns

In [15]: data.dtypes

```
Out[15]: 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**

[21]:	data.describe()									
]:		PassengerId	Survived	Pclass	Age	SibSp	Parch			
	count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000		
	mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204		
	std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693		
	min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000		
	25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.91(		
	50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454		
	75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000		
	max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329		
	4							•		

In [25]: data.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
44	£1+C4/2	\ : -+	+/F\

dtypes: float64(2), int64(5), object(5)

memory usage: 83.7+ KB

# Step-9 Access a specific column

```
In [37]: data['Age']
Out[37]: 0
                 22.0
                 38.0
          2
                 26.0
          3
                 35.0
                 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 [41]: data.iloc[100:230]
```

•	Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
100	101	0	3	Petranec, Miss. Matilda	female	28.0	0	0	349245	
101	102	0	3	Petroff, Mr. Pastcho ("Pentcho")	male	NaN	0	0	349215	
102	103	0	1	White, Mr. Richard Frasar	male	21.0	0	1	35281	7
103	104	0	3	Johansson, Mr. Gustaf Joel	male	33.0	0	0	7540	
104	105	0	3	Gustafsson, Mr. Anders Vilhelm	male	37.0	2	0	3101276	
•••										
225	226	0	3	Berglund, Mr. Karl Ivar Sven	male	22.0	0	0	PP 4348	
226	227	1	2	Mellors, Mr. William John	male	19.0	0	0	SW/PP 751	1
227	228	0	3	Lovell, Mr. John Hall ("Henry")	male	20.5	0	0	A/5 21173	
228	229	0	2	Fahlstrom, Mr. Arne Jonas	male	18.0	0	0	236171	1
229	230	0	3	Lefebre, Miss. Mathilde	female	NaN	3	1	4133	2
130 r	ows × 12 colur	mns								
4 6						_				

# Step-11 Delete a specific Column

In [43]: data.drop('Age',axis=1)

,		. assengena	Survived	r Class	Name	Sex	JibSh	Parch	Ticket	Fare
	0	1	0	3	Braund, Mr. Owen Harris	male	1	0	A/5 21171	7.2500
	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	1	0	PC 17599	71.2833
	2	3	1	3	Heikkinen, Miss. Laina	female	0	0	STON/O2. 3101282	7.9250
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	1	0	113803	53.1000
	4	5	0	3	Allen, Mr. William Henry	male	0	0	373450	8.0500
	•••									
	886	887	0	2	Montvila, Rev. Juozas	male	0	0	211536	13.0000
	887	888	1	1	Graham, Miss. Margaret Edith	female	0	0	112053	30.0000
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	1	2	W./C. 6607	23.4500
	889	890	1	1	Behr, Mr. Karl Howell	male	0	0	111369	30.0000
	890	891	0	3	Dooley, Mr. Patrick	male	0	0	370376	7.7500
	891 rd	ows × 11 colui	mns							
	4 =									•

# Step-12 Create a new Column

```
In [61]: data['amount']=data['Fare']
  data['amount']=data['Fare'].apply(lambda x:"high"if x>10 else "low")
  data
```

-17.					_	_				
1]:	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
C	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	2 3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	<b>3</b> 4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
		0	3	Allen, Mr. William Henry	male 	35.0	0	0	373450	
886	<b>5</b> 887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	7 888	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 [77]: data[data['Fare']>10]

Out[77]

	Daccongorld	Survived	Delace	Name	Sav	Λαο	SihSn	Darch	Ticket	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0			PC 17599	71.2
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.{
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.(
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.
•••									•••	
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.(
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
	3 6 7 885 886 887	1 2  3 4  6 7  7 8  8 9  885 886  886 887  888 888	1 2 1  3 4 1  6 7 0  7 8 0  88 9 1   885 886 0  886 887 0  887 888 1	3       4       1       1         6       7       0       1         7       8       0       3         8       9       1       3               885       886       0       3         886       887       0       2         887       888       1       1         888       889       0       3	1       2       1       1       Cumings, Mrs. John Bradley (Florence Briggs Th         3       4       1       1       Futrelle, Mrs. Jacques Heath (Lily May Peel)         6       7       0       1       McCarthy, Mr. Timothy J Mr. Timothy J Mrs. Gosta Leonard         8       9       1       3       Palsson, Master. Gosta Leonard         8       9       1       3       Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)                 885       886       0       3       Rice, Mrs. William (Margaret Norton)       Montvila, Rev. Juozas         887       888       1       1       Montvila, Miss. Margaret Edith         888       889       0       3       Catherine Helen "Carrie"         889       890       1       1       Behr, Mr. Karl	1       2       1       1       Cumings, Mrs. John Bradley (Florence Briggs Th       female Riggs Th         3       4       1       1       Futrelle, Mrs. Jacques Th       female Leath (Lily May Peel)         6       7       0       1       McCarthy, Mrs. Jacques Heath (Lily May Peel)       male         7       8       0       3       Palsson, Master. Gosta Leonard       male         8       9       1       3       Palsson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)       female         885       886       0       3       Rice, Mrs. William (Margaret Norton)       male         886       887       0       2       Rev. Juozas Montvila, Miss. Margaret Edith       female         887       888       1       1       Montvila, Miss. Margaret Edith       female Edith         888       889       0       3       Johnston, Miss. Catherine Helen "Carrie"       female Helen "Carrie"	1       2       1       1       Cumings, Mrs. John Bradley (Florence Briggs Th       female       38.0         3       4       1       1       Futrelle, Mrs. Jacques Heath (Lily May Peel)       female       35.0         6       7       0       1       McCarthy, Mr. Timothy J       male       54.0         7       8       0       3       Palsson, Master. Gosta Leonard       male       2.0         8       9       1       3       Oscar W (Elisabeth Vilhelmina Berg)       female       27.0         885       886       0       3       Rice, Mrs. William (Margaret Norton)       female       39.0         886       887       0       2       Rev. Juozas Margaret Edith       female       27.0         887       888       1       1       Graham, Miss. Margaret Edith       female       19.0         888       889       0       3       Catherine Helen "Carrie"       female       NaN         889       890       1       Behr, Mr. Karl       male       26.0	1       2       1       1       Cumings, Mrs. John Bradley (Florence Briggs Th       female Briggs Th       38.0       1         3       4       1       1       Futrelle, Mrs. Jacques Hemale (Lily May Peel)       female St.0       1         6       7       0       1       McCarthy, Mr. Timothy J       male St.0       0         7       8       0       3       Palsson, Mrs. Gosta Leonard       male Leonard       2.0       3         8       9       1       3       Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)       female St.0       0       27.0       0         885       886       0       3       Rice, Mrs. William (Margaret Norton)       female St.0       0       0         886       887       0       2       Montvila, Rev. Juozas       male St.0       0       0         887       888       1       1       Graham, Miss. Margaret Female Edith       19.0       0         888       889       0       3       Johnston, Miss. Genale Helen "Carrie"       Female Pemale St.0       NaN       1         889       890       1       Behr, Mr. Karl male St.0       0       0	1       2       1       1       Cumings, Mrs. John Bradley (Florence Briggs Th       female 38.0       1       0         3       4       1       1       Futrelle, Mrs. Jacques Heath (Lily May Peel)       female 35.0       1       0         6       7       0       1       McCarthy, Peel)       male 54.0       0       0       0         7       8       0       3       Msstr. Gosta Leonard       male 2.0       3       1         8       9       1       3       Oscar W (Elisabeth Vilhelmina Berg)       female 27.0       0       2         885       886       0       3       Rice, Mrs. William (Margaret Norton)       female 27.0       0       5         886       887       0       2       Montvila, Margaret Edith       female 27.0       0       0       0         887       888       1       1       Miss. Margaret Edith       female NaN       1       2         888       889       0       3       Catherine Helen "Carrie"       female NaN       1       2         889       890       1       1       Behr, Mr. Mr. Margaret Margaret Edith       Rev. Gosta Warm       Rev. Gosta Warm       Rev. Gosta Warm       Rev. Gosta	The second color of the

555 rows × 13 columns

# Step-14 Compute the sum of value

```
In [67]: data['Fare'].sum()
Out[67]: 28693.9493
```

#### Step-15 Compute the mean of value

```
In [69]: data['Fare'].mean()
Out[69]: 32.204207968574636
```

## Step-16 Count non-null value (column)

```
In [71]: data['Fare'].count()
Out[71]: 891
```

# Step-17 Find Minimun or Maximum values

```
In [73]: data['Fare'].max()
Out[73]: 512.3292
In []:
```