

Data Mining

Lab - 1

Jeet Bhalodi_23031701006

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 [11]: df = pd.read_csv('titanic.csv')

Step-4 Print Data from csv or excel File

In [13]: df

]:		PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	E
	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	53.10
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0!
	•••										
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00
	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	30.00
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7!
8	391 rc	ows × 12 colur	mns								
											•

Step-5 See the First 10 Rows

In [33]: df.head(10)

#df.head() #only 5 record print

Out[33]:		Passengerld	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	7.925(
	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	8.050(
	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	51.862!
	7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.075(
	8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.133:
	9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708
	4										•

Step-6 See the Last 10 Rows

In [39]: df.tail(10)

9]:		Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
	881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7
	882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10
	883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10
	884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7
	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
8	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7
	4										•

Step-7 Data type of each columns

```
df.dtypes
In [52]:
Out[52]:
         PassengerId
                           int64
          Survived
                           int64
         Pclass
                           int64
         Name
                          object
         Sex
                          object
                         float64
         Age
         SibSp
                           int64
         Parch
                           int64
         Ticket
                          object
         Fare
                         float64
         Cabin
                          object
          Embarked
                          object
          dtype: object
In [56]: 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
             Survived
                          891 non-null
                                          int64
         1
             Pclass
                          891 non-null
                                          int64
         3
             Name
                          891 non-null
                                          object
         4
             Sex
                          891 non-null
                                          object
         5
                                          float64
             Age
                          714 non-null
                                          int64
             SibSp
                          891 non-null
             Parch
                          891 non-null
                                          int64
             Ticket
                          891 non-null
                                          object
         9
                          891 non-null
                                          float64
             Fare
                          204 non-null
         10
             Cabin
                                          object
             Embarked
                          889 non-null
                                          object
```

Step-8 Display Summary Information

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

memory usage: 83.7+ KB

In [44]: 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 [66]: df.shape

Out[66]: (891, 12)

In [68]: df.shape[0] #row

Out[68]: 891

In [70]: df.shape[1] #column

Out[70]: **12**

In [72]: df.describe()

Out[72]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
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	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
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	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
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	512.329200
4							•

Step-9 Access a specific column

```
In [85]:
          df['Name']
          df.Name
Out[85]:
                                                Braund, Mr. Owen Harris
                   Cumings, Mrs. John Bradley (Florence Briggs Th...
           2
                                                 Heikkinen, Miss. Laina
           3
                        Futrelle, Mrs. Jacques Heath (Lily May Peel)
                                              Allen, Mr. William Henry
           886
                                                  Montvila, Rev. Juozas
           887
                                          Graham, Miss. Margaret Edith
           888
                             Johnston, Miss. Catherine Helen "Carrie"
           889
                                                  Behr, Mr. Karl Howell
           890
                                                    Dooley, Mr. Patrick
          Name: Name, Length: 891, dtype: object
In [87]:
          df[['Name','Age']]
Out[87]:
                                                      Name
                                                              Age
             0
                                      Braund, Mr. Owen Harris
                                                              22.0
                Cumings, Mrs. John Bradley (Florence Briggs Th...
                                                              38.0
             2
                                        Heikkinen, Miss. Laina
                                                              26.0
                      Futrelle, Mrs. Jacques Heath (Lily May Peel)
                                                              35.0
             4
                                      Allen, Mr. William Henry
                                                              35.0
          886
                                                              27.0
                                         Montvila, Rev. Juozas
           887
                                 Graham, Miss. Margaret Edith
                                                              19.0
          888
                        Johnston, Miss. Catherine Helen "Carrie"
                                                              NaN
          889
                                         Behr, Mr. Karl Howell
                                                              26.0
          890
                                           Dooley, Mr. Patrick
                                                              32.0
```

891 rows × 2 columns

Step-10 Access rows by their integer location

```
In [109... df.iloc[15]
```

Out[109... PassengerId 16 Survived 1 Pclass 2 Hewlett, Mrs. (Mary D Kingcome) Name Sex female 55.0 Age 0 SibSp Parch 0 Ticket 248706 Fare 16.0 Cabin NaN Embarked S Name: 15, dtype: object

In [113... df.iloc[10:20]

Out[113		PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
	10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.70
	11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.550
	12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.050
	13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.27
	14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.85 ₄
	15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.00
	16	17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29.12
	17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.00
	18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande	female	31.0	1	0	345763	18.000
	19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.22
	4										•
In [115	df.	iloc[10:20,3	:6]								

Out[115...

	Name	Sex	Age
10	Sandstrom, Miss. Marguerite Rut	female	4.0
11	Bonnell, Miss. Elizabeth	female	58.0
12	Saundercock, Mr. William Henry	male	20.0
13	Andersson, Mr. Anders Johan	male	39.0
14	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0
15	Hewlett, Mrs. (Mary D Kingcome)	female	55.0
16	Rice, Master. Eugene	male	2.0
17	Williams, Mr. Charles Eugene	male	NaN
18	Vander Planke, Mrs. Julius (Emelia Maria Vande	female	31.0
19	Masselmani, Mrs. Fatima	female	NaN

In [119...

df.iloc[:,3:6]

Out[119...

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

891 rows × 3 columns

Step-11 Delete a specific Column

In [125... df.drop('Ticket',axis=1)

125		PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Fare	Cabin
	0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	7.2500	NaN
	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	71.2833	C85
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	7.9250	NaN
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	C123
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	8.0500	NaN
	•••										
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	13.0000	NaN
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	30.0000	B42
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	23.4500	NaN
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	30.0000	C148
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	7.7500	NaN
	891 rd	ows × 11 colur	mns								
	4										•

In [131... df.drop('Pclass',axis=1,inplace=True)
In [133... df

Out[133...

	PassengerId	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cal
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	N
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C 1
4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N
•••		•••			•••				•••	
886	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	N
887	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	E
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	N
889	890	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C 1
890	891	0	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	N
891 rc	ows × 11 colur	nns								
4										•

In [135... df.drop(columns='Fare')

> Out[135... PassengerId Survived Name Sex Age SibSp Parch Ticket Cabin Emb Braund, A/5 0 1 Mr. Owen male 22.0 1 NaN 21171 Harris Cumings, Mrs. John Bradley 2 1 female 38.0 1 PC 17599 C85 (Florence Briggs Th... Heikkinen, STON/O2. 2 3 0 1 Miss. female 26.0 NaN 3101282 Laina Futrelle, Mrs. Jacques 3 4 female 35.0 0 113803 C123 Heath (Lily May Peel) Allen, Mr. 5 4 0 William male 35.0 0 0 373450 NaN Henry Montvila, 886 887 0 Rev. male 27.0 0 0 211536 NaN Juozas Graham, Miss. 0 0 887 888 female 19.0 112053 B42 Margaret Edith Johnston, Miss. W./C. 888 889 Catherine female NaN 1 2 NaN 6607 Helen "Carrie" Behr, Mr. 889 890 1 Karl male 26.0 0 0 111369 C148 Howell Dooley, 890 891 0 Mr. male 32.0 0 0 370376 NaN **Patrick** 891 rows × 10 columns

Step-12 Create a new Column

```
In [139... df['NewFare'] = df['Fare']*1.1

In [141... df
```

L41		PassengerId	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cal
	0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
	1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.2833	C
	2	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	N
	3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C 1
	4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N
	•••										
	886	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	N
	887	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	E
	888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	N
	889	890	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C 1
	890	891	0	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	N
	891 rc	ows × 12 colur	nns								
	4										•

Step-13 Perform Condition Selection on DataFrame

In [162...

df[df['Fare']>10]

Out[162		Passengerld	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabir
	1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.2833	C85
	3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
	6	7	0	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46
	7	8	0	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN
	8	9	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN
	•••										
	885	886	0	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN
	886	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN
	887	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42
	888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN
	889	890	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148

555 rows × 12 columns

In [168... df[(df['Fare']>10) & (df['Fare']<12)]

Out[168...

		Passengerld	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare
	8	9	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333
	33	34	0	Wheadon, Mr. Edward H	male	66.0	0	0	C.A. 24579	10.5000
	39	40	1	Nicola- Yarred, Miss. Jamila	female	14.0	1	0	2651	11.2417
	56	57	1	Rugg, Miss. Emily	female	21.0	0	0	C.A. 31026	10.5000
	66	67	1	Nye, Mrs. (Elizabeth Ramell)	female	29.0	0	0	C.A. 29395	10.5000
	70	71	0	Jenkin, Mr. Stephen Curnow	male	32.0	0	0	C.A. 33111	10.5000
	84	85	1	llett, Miss. Bertha	female	17.0	0	0	SO/C 14885	10.5000
	125	126	1	Nicola- Yarred, Master. Elias	male	12.0	1	0	2651	11.2417
	144	145	0	Andrew, Mr. Edgardo Samuel	male	18.0	0	0	231945	11.5000
	172	173	1	Johnson, Miss. Eleanor Ileen	female	1.0	1	1	347742	11.1333
	205	206	0	Strom, Miss. Telma Matilda	female	2.0	0	1	347054	10.4625
	219	220	0	Harris, Mr. Walter	male	30.0	0	0	W/C 14208	10.5000
	226	227	1	Mellors, Mr. William John	male	19.0	0	0	SW/PP 751	10.5000
	234	235	0	Leyson, Mr. Robert William Norman	male	24.0	0	0	C.A. 29566	10.5000
	238	239	0	Pengelly, Mr. Frederick	male	19.0	0	0	28665	10.5000

	PassengerId	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare
			William						
242	243	0	Coleridge, Mr. Reginald Charles	male	29.0	0	0	W./C. 14263	10.5000
251	252	0	Strom, Mrs. Wilhelm (Elna Matilda Persson)	female	29.0	1	1	347054	10.4625
265	266	0	Reeves, Mr. David	male	36.0	0	0	C.A. 17248	10.5000
398	399	0	Pain, Dr. Alfred	male	23.0	0	0	244278	10.5000
439	440	0	Kvillner, Mr. Johan Henrik Johannesson	male	31.0	0	0	C.A. 18723	10.5000
458	459	1	Toomey, Miss. Ellen	female	50.0	0	0	F.C.C. 13531	10.5000
516	517	1	Lemore, Mrs. (Amelia Milley)	female	34.0	0	0	C.A. 34260	10.5000
526	527	1	Ridsdale, Miss. Lucy	female	50.0	0	0	W./C. 14258	10.5000
529	530	0	Hocking, Mr. Richard George	male	23.0	2	1	29104	11.5000
570	571	1	Harris, Mr. George	male	62.0	0	0	S.W./PP 752	10.5000
619	620	0	Gavey, Mr. Lawrence	male	26.0	0	0	31028	10.5000
672	673	0	Mitchell, Mr. Henry Michael	male	70.0	0	0	C.A. 24580	10.5000
687	688	0	Dakic, Mr. Branko	male	19.0	0	0	349228	10.1708
717	718	1	Troutt, Miss. Edwina Celia "Winnie"	female	27.0	0	0	34218	10.5000
757	758	0	Bailey, Mr. Percy Andrew	male	18.0	0	0	29108	11.5000

	Passengerld	Survived	Name	Sex	Age	SibSp	Parch	Ticket	Fare
772	773	0	Mack, Mrs. (Mary)	female	57.0	0	0	S.O./P.P. 3	10.5000
812	813	0	Slemen, Mr. Richard James	male	35.0	0	0	28206	10.5000
841	842	0	Mudd, Mr. Thomas Charles	male	16.0	0	0	S.O./P.P. 3	10.5000
861	862	0	Giles, Mr. Frederick Edward	male	21.0	1	0	28134	11.5000
869	870	1	Johnson, Master. Harold Theodor	male	4.0	1	1	347742	11.1333
882	883	0	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167
883	884	0	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000

Step-14 Compute the sum of value

In [180... df.Survived.sum()

Out[180... 342

Step-15 Compute the mean of value

In [172... df.Age.mean()

Out[172... 29.69911764705882

Step-16 Count non-null value (column)

In [190... df.count()

```
Out[190...
           PassengerId
                            891
           Survived
                            891
           Name
                            891
                            891
           Sex
                            714
           Age
                            891
           SibSp
           Parch
                            891
                            891
           Ticket
           Fare
                            891
           Cabin
                            204
            Embarked
                            889
           NewFare
                            891
           dtype: int64
In [196...
           df.isnull().sum()
                              0
Out[196...
           PassengerId
           Survived
                              0
           Name
                              0
           Sex
                              0
           Age
                            177
           SibSp
                              0
           Parch
                              0
           Ticket
                              0
           Fare
                              0
           Cabin
                            687
            Embarked
                              2
           NewFare
           dtype: int64
In [198...
           (~df.isnull()).sum()
Out[198...
           PassengerId
                            891
           Survived
                            891
           Name
                            891
           Sex
                            891
           Age
                            714
           SibSp
                            891
           Parch
                            891
           Ticket
                            891
           Fare
                            891
           Cabin
                            204
            Embarked
                            889
           NewFare
                            891
           dtype: int64
```

Step-17 Find Minimun or Maximum values

```
In [174... df.Age.max()
Out[174... 80.0
In [176... df.Age.min()
```

Out[176... 0.42

In []: