

# pandas

December 23, 2022

```
[3]: import pandas as pd
```

Two major data types 1.Series 2.DataFrame

```
[6]: lst=[1,2,3,4]
     print(type(lst))
```

```
<class 'list'>
```

```
[9]: ser=pd.Series(lst)
     print(type(ser))
```

```
<class 'pandas.core.series.Series'>
```

```
[10]: ser1=pd.Series([1,2,3,4,5,6])
```

```
[11]: print(ser1)
```

```
0    1
1    2
2    3
3    4
4    5
5    6
dtype: int64
```

Series is a one dimensional data type

```
[13]: lst1=[[1,2,3,4],[5,6,7,8]]
```

```
[14]: ser2=pd.Series(lst1)
```

```
[15]: print(ser2)
```

```
0    [1, 2, 3, 4]
1    [5, 6, 7, 8]
dtype: object
```

```
[16]: df1=pd.DataFrame(lst1)
```

```
[17]: df1
```

```
[17]:    0  1  2  3
0    1  2  3  4
1    5  6  7  8
```

DataFrame is a data type having more than 1 dimensions What is dimension? Dimensions are nothing but the no of columns. If thr Dataset is having one column then we can consider that dataset as Series else it will be dataframe

```
[4]: #reading data in pandas
import pandas as pd
df = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/
master/titanic.csv")
df
```

```
[4]:    PassengerId  Survived  Pclass  \
0             1         0        3
1             2         1        1
2             3         1        3
3             4         1        1
4             5         0        3
..          ...         ...      ...
886          887         0        2
887          888         1        1
888          889         0        3
889          890         1        1
890          891         0        3
```

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

```
    Parch    Ticket   Fare Cabin Embarked
0      0  A/5 21171   7.2500   NaN        S
1      0    PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0    113803  53.1000  C123        S
```

4	0	373450	8.0500	NaN	S
...	...	...	...	...	...
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
888	2	W./C. 6607	23.4500	NaN	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

[891 rows x 12 columns]

[20]: `#top 5 rows`

[22]: `df.head()`

[22]:

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

[4]: `#top 10 rows`  
`df.head(10)`

[4]:

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
5	6	0	3	
6	7	0	1	
7	8	0	3	

8	9	1	3
9	10	1	2

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	
5	Moran, Mr. James	male	NaN	0	
6	McCarthy, Mr. Timothy J	male	54.0	0	
7	Palsson, Master. Gosta Leonard	male	2.0	3	
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
5	0	330877	8.4583	NaN	Q
6	0	17463	51.8625	E46	S
7	1	349909	21.0750	NaN	S
8	2	347742	11.1333	NaN	S
9	0	237736	30.0708	NaN	C

```
[5]: #bottom 5 rows
df.tail()
```

	PassengerId	Survived	Pclass	Name	\
886	887	0	2	Montvila, Rev. Juozas	
887	888	1	1	Graham, Miss. Margaret Edith	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	
889	890	1	1	Behr, Mr. Karl Howell	
890	891	0	3	Dooley, Mr. Patrick	

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	male	27.0	0	0	211536	13.00	NaN	S
887	female	19.0	0	0	112053	30.00	B42	S
888	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	male	26.0	0	0	111369	30.00	C148	C
890	male	32.0	0	0	370376	7.75	NaN	Q

```
[25]: #bottom 10 rows
df.tail(10)
```

```
[25]:
```

	PassengerId	Survived	Pclass	Name \
881	882	0	3	Markun, Mr. Johann
882	883	0	3	Dahlberg, Miss. Gerda Ulrika
883	884	0	2	Banfield, Mr. Frederick James
884	885	0	3	Sutehall, Mr. Henry Jr
885	886	0	3	Rice, Mrs. William (Margaret Norton)
886	887	0	2	Montvila, Rev. Juozas
887	888	1	1	Graham, Miss. Margaret Edith
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"
889	890	1	1	Behr, Mr. Karl Howell
890	891	0	3	Dooley, Mr. Patrick

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
881	male	33.0	0	0	349257	7.8958	NaN	S
882	female	22.0	0	0	7552	10.5167	NaN	S
883	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S
884	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN	S
885	female	39.0	0	5	382652	29.1250	NaN	Q
886	male	27.0	0	0	211536	13.0000	NaN	S
887	female	19.0	0	0	112053	30.0000	B42	S
888	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	male	26.0	0	0	111369	30.0000	C148	C
890	male	32.0	0	0	370376	7.7500	NaN	Q

```
[31]: df[50:100:5]
```

```
[31]:
```

	PassengerId	Survived	Pclass	\
50	51	0	3	
55	56	1	1	
60	61	0	3	
65	66	1	3	
70	71	0	2	
75	76	0	3	
80	81	0	3	
85	86	1	3	
90	91	0	3	
95	96	0	3	

  

	Name	Sex	Age	SibSp	\
50	Panula, Master. Juha Niilo	male	7.0	4	
55	Woolner, Mr. Hugh	male	NaN	0	
60	Sirayanian, Mr. Orsen	male	22.0	0	
65	Moubarek, Master. Gerios	male	NaN	1	
70	Jenkin, Mr. Stephen Curnow	male	32.0	0	
75	Moen, Mr. Sigurd Hansen	male	25.0	0	
80	Waelens, Mr. Achille	male	22.0	0	
85	Backstrom, Mrs. Karl Alfred (Maria Mathilda Gu...	female	33.0	3	

90	Christmann, Mr. Emil	male	29.0	0
95	Shorney, Mr. Charles Joseph	male	NaN	0

	Parch	Ticket	Fare	Cabin	Embarked
50	1	3101295	39.6875	NaN	S
55	0	19947	35.5000	C52	S
60	0	2669	7.2292	NaN	C
65	1	2661	15.2458	NaN	C
70	0	C.A. 33111	10.5000	NaN	S
75	0	348123	7.6500	F G73	S
80	0	345767	9.0000	NaN	S
85	0	3101278	15.8500	NaN	S
90	0	343276	8.0500	NaN	S
95	0	374910	8.0500	NaN	S

```
[6]: # checking the datatypes of the columns
df.dtypes
```

```
[6]: 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
```

```
[28]: # getting insights of the dataset
df.describe()
```

```
[28]:
```

	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

  

	Parch	Fare
count	891.000000	891.000000

mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

```
[30]: # get the shape of dataframe
df.shape
```

```
[30]: (891, 12)
```

```
[7]: # Selecting a column
df.PassengerId
```

```
[7]: 0      1
     1      2
     2      3
     3      4
     4      5
     ...
    886    887
    887    888
    888    889
    889    890
    890    891
     Name: PassengerId, Length: 891, dtype: int64
```

```
[8]: df['Pclass'].unique()
```

```
[8]: array([3, 1, 2])
```

```
[9]: df['Embarked'].unique()
```

```
[9]: array(['S', 'C', 'Q', nan], dtype=object)
```

```
[10]: df['PassengerId']
```

```
[10]: 0      1
     1      2
     2      3
     3      4
     4      5
     ...
    886    887
    887    888
```

```

888      889
889      890
890      891
Name: PassengerId, Length: 891, dtype: int64

```

```

[14]: #accessing the columns
df[['PassengerId', 'Name', 'Sex']]

```

```

[14]:      PassengerId      Name      Sex
0              1      Braund, Mr. Owen Harris    male
1              2  Cumings, Mrs. John Bradley (Florence Briggs Th...  female
2              3      Heikkinen, Miss. Laina    female
3              4  Futrelle, Mrs. Jacques Heath (Lily May Peel)  female
4              5      Allen, Mr. William Henry    male
..          ...
886          887      Montvila, Rev. Juozas    male
887          888      Graham, Miss. Margaret Edith  female
888          889  Johnston, Miss. Catherine Helen "Carrie"  female
889          890      Behr, Mr. Karl Howell    male
890          891      Dooley, Mr. Patrick    male

[891 rows x 3 columns]

```

```

[11]: #accessing via rows
df.loc[100]

```

```

[11]: PassengerId      101
Survived              0
Pclass                3
Name      Petranec, Miss. Matilda
Sex              female
Age              28.0
SibSp              0
Parch              0
Ticket      349245
Fare        7.8958
Cabin      NaN
Embarked      S
Name: 100, dtype: object

```

```

[12]: df.loc[30:40]

```

```

[12]:      PassengerId  Survived  Pclass  \
30          31          0          1
31          32          1          1
32          33          1          3
33          34          0          2

```



34	35	0	1
35	36	0	1
36	37	1	3
37	38	0	3
38	39	0	3
39	40	1	3
40	41	0	3

	Name	Sex	Age	SibSp	\
30	Uruchurtu, Don. Manuel E	male	40.0	0	
31	Spencer, Mrs. William Augustus (Marie Eugenie)	female	NaN	1	
32	Glynn, Miss. Mary Agatha	female	NaN	0	
33	Wheadon, Mr. Edward H	male	66.0	0	
34	Meyer, Mr. Edgar Joseph	male	28.0	1	
35	Holverson, Mr. Alexander Oskar	male	42.0	1	
36	Mamee, Mr. Hanna	male	NaN	0	
37	Cann, Mr. Ernest Charles	male	21.0	0	
38	Vander Planke, Miss. Augusta Maria	female	18.0	2	
39	Nicola-Yarred, Miss. Jamila	female	14.0	1	
40	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40.0	1	

	Parch	Ticket	Fare	Cabin	Embarked
30	0	PC 17601	27.7208	NaN	C
31	0	PC 17569	146.5208	B78	C
32	0	335677	7.7500	NaN	Q
33	0	C.A. 24579	10.5000	NaN	S
34	0	PC 17604	82.1708	NaN	C
35	0	113789	52.0000	NaN	S
36	0	2677	7.2292	NaN	C
37	0	A./5. 2152	8.0500	NaN	S
38	0	345764	18.0000	NaN	S
39	0	2651	11.2417	NaN	C
40	0	7546	9.4750	NaN	S

```
[44]: #Selecting a row using default index
df.iloc[100]
```

```
[44]: PassengerId      101
Survived              0
Pclass               3
Name      Petranec, Miss. Matilda
Sex                female
Age                28.0
SibSp              0
Parch              0
Ticket            349245
Fare              7.8958
```

```

Cabin          NaN
Embarked       S
Name: 100, dtype: object

```

```

[15]: #changing 'passengerId' column to index
df.set_index('PassengerId')

```

```

[15]:      Survived  Pclass \
PassengerId
1             0         3
2             1         1
3             1         3
4             1         1
5             0         3
...
887          ...      ...
888             1         1
889             0         3
890             1         1
891             0         3

```

```

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

```

```

      SibSp  Parch      Ticket    Fare Cabin Embarked
PassengerId
1           1     0      A/5 21171    7.2500   NaN      S
2           1     0      PC 17599   71.2833   C85      C
3           0     0  STON/O2. 3101282    7.9250   NaN      S
4           1     0      113803   53.1000  C123      S
5           0     0      373450    8.0500   NaN      S
...
887          0     0      211536   13.0000   NaN      S
888          0     0      112053   30.0000   B42      S
889          1     2      W./C. 6607   23.4500   NaN      S
890          0     0      111369   30.0000  C148      C

```

```
891          0      0          370376    7.7500    NaN      Q
```

```
[891 rows x 11 columns]
```

```
[47]: df
```

```
[47]:      PassengerId  Survived  Pclass  \
0              1         0         3
1              2         1         1
2              3         1         3
3              4         1         1
4              5         0         3
..          ...     ...     ...
886          887         0         2
887          888         1         1
888          889         0         3
889          890         1         1
890          891         0         3
```

```

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

```

      Parch      Ticket    Fare Cabin Embarked
0         0    A/5 21171    7.2500   NaN      S
1         0      PC 17599   71.2833   C85      C
2         0  STON/O2. 3101282    7.9250   NaN      S
3         0     113803   53.1000  C123      S
4         0     373450    8.0500   NaN      S
..      ...     ...     ...     ...
886         0     211536   13.0000   NaN      S
887         0     112053   30.0000   B42      S
888         2  W./C. 6607   23.4500   NaN      S
889         0     111369   30.0000  C148      C
890         0     370376    7.7500   NaN      Q
```

```
[891 rows x 12 columns]
```

```
[6]: df.set_index('PassengerId',inplace=True)
```

```
[7]: df
```

```
[7]:      Survived  Pclass  \
```

```
PassengerId
```

```
1          0         3
2          1         1
3          1         3
4          1         1
5          0         3
...
887         0         2
888         1         1
889         0         3
890         1         1
891         0         3
```

```
      Name      Sex  Age  \
```

```
PassengerId
```

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

```
      SibSp  Parch      Ticket    Fare Cabin Embarked
```

```
PassengerId
```

```
1          1         0      A/5 21171    7.2500   NaN      S
2          1         0      PC 17599   71.2833   C85      C
3          0         0  STON/O2. 3101282    7.9250   NaN      S
4          1         0      113803   53.1000  C123      S
5          0         0      373450    8.0500   NaN      S
...
887         0         0      211536   13.0000   NaN      S
888         0         0      112053   30.0000   B42      S
889         1         2  W./C. 6607   23.4500   NaN      S
890         0         0      111369   30.0000  C148      C
891         0         0      370376    7.7500   NaN      Q
```

```
[891 rows x 11 columns]
```

```
[17]: #consider system index
df.iloc[100]
```

```
[17]: Survived          0
Pclass          3
Name      Petranec, Miss. Matilda
Sex          female
Age          28.0
SibSp          0
Parch          0
Ticket      349245
Fare          7.8958
Cabin          NaN
Embarked      S
Name: 101, dtype: object
```

```
[15]: #consider user index --user index is one more than the system index
df.loc[100]
```

```
[15]: Survived          0
Pclass          2
Name      Kantor, Mr. Sinai
Sex          male
Age          34.0
SibSp          1
Parch          0
Ticket      244367
Fare          26.0
Cabin          NaN
Embarked      S
Name: 100, dtype: object
```

```
[12]: df
```

```
[12]:      Survived  Pclass \
PassengerId
1           0         3
2           1         1
3           1         3
4           1         1
5           0         3
...         ...     ...
887          0         2
888          1         1
889          0         3
890          1         1
891          0         3
```

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

PassengerId	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	0	A/5 21171	7.2500	NaN	S
2	1	0	PC 17599	71.2833	C85	C
3	0	0	STON/O2. 3101282	7.9250	NaN	S
4	1	0	113803	53.1000	C123	S
5	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...
887	0	0	211536	13.0000	NaN	S
888	0	0	112053	30.0000	B42	S
889	1	2	W./C. 6607	23.4500	NaN	S
890	0	0	111369	30.0000	C148	C
891	0	0	370376	7.7500	NaN	Q

[891 rows x 11 columns]

```
[19]: #checking the null values
df.isnull()
```

```
[19]:
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	\
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	
5	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	
887	False	False	False	False	False	False	False	False	
888	False	False	False	False	False	False	False	False	
889	False	False	False	False	True	False	False	False	
890	False	False	False	False	False	False	False	False	
891	False	False	False	False	False	False	False	False	

	Fare	Cabin	Embarked
PassengerId			
1	False	True	False
2	False	False	False
3	False	True	False
4	False	False	False
5	False	True	False
...	...	...	...
887	False	True	False
888	False	False	False
889	False	True	False
890	False	False	False
891	False	True	False

[891 rows x 11 columns]

```
[20]: #getting the sum of null values
df.isnull().sum()
```

```
[20]: Survived      0
Pclass           0
Name             0
Sex              0
Age            177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin          687
Embarked        2
dtype: int64
```

```
[22]: #creating a new column
df['new_col']=df['Survived']+df['Pclass']
df.head()
```

```
[22]:
```

	Survived	Pclass	\
PassengerId			
1	0	3	
2	1	1	
3	1	3	
4	1	1	
5	0	3	

	Name	Sex	Age	\
PassengerId				

1		Braund, Mr. Owen Harris	male	22.0
2	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	
3		Heikkinen, Miss. Laina	female	26.0
4	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	
5		Allen, Mr. William Henry	male	35.0

	SibSp	Parch	Ticket	Fare	Cabin	Embarked	new_col
PassengerId							
1	1	0	A/5 21171	7.2500	NaN	S	3
2	1	0	PC 17599	71.2833	C85	C	2
3	0	0	STON/O2. 3101282	7.9250	NaN	S	4
4	1	0	113803	53.1000	C123	S	2
5	0	0	373450	8.0500	NaN	S	3

```
[26]: #deleting the column
df.drop('new_col',axis=1,inplace=True)
```

```
[27]: df
```

```
[27]:
```

	Survived	Pclass	\
PassengerId			
1	0	3	
2	1	1	
3	1	3	
4	1	1	
5	0	3	
...	...	...	
887	0	2	
888	1	1	
889	0	3	
890	1	1	
891	0	3	

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



PassengerId	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	0	A/5 21171	7.2500	NaN	S
2	1	0	PC 17599	71.2833	C85	C
3	0	0	STON/O2. 3101282	7.9250	NaN	S
4	1	0	113803	53.1000	C123	S
5	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...
887	0	0	211536	13.0000	NaN	S
888	0	0	112053	30.0000	B42	S
889	1	2	W./C. 6607	23.4500	NaN	S
890	0	0	111369	30.0000	C148	C
891	0	0	370376	7.7500	NaN	Q

[891 rows x 11 columns]

```
[28]: #groupby operation
data1 = {'Name':['Jai', 'Anuj', 'Jai', 'Princi', 'Gaurav', 'Anuj', 'Princi', 'Abhi'],
        'Age':[27, 24, 22, 32, 33, 36, 27, 32],
        'Address':['Nagpur', 'Kanpur', 'Allahabad', 'Kannuaj', 'Jaunpur', 'Kanpur', 'Allahabad', 'Aligarh'],
        'Qualification':['Msc', 'MA', 'MCA', 'Phd', 'B.Tech', 'B.com', 'Msc', 'MA']}
}
```

```
[29]: type(data1)
```

[29]: dict

```
[30]: df_new=pd.DataFrame(data1)
df_new
```

```
[30]:
```

	Name	Age	Address	Qualification
0	Jai	27	Nagpur	Msc
1	Anuj	24	Kanpur	MA
2	Jai	22	Allahabad	MCA
3	Princi	32	Kannuaj	Phd
4	Gaurav	33	Jaunpur	B.Tech
5	Anuj	36	Kanpur	B.com
6	Princi	27	Allahabad	Msc
7	Abhi	32	Aligarh	MA

```
[31]: #groupby the data by column name
df_new.groupby('Name')
```

[31]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f6b403c4a60>

```
[32]: print(df_new.groupby('Name').groups)
```

```
{'Abhi': [7], 'Anuj': [1, 5], 'Gaurav': [4], 'Jai': [0, 2], 'Princi': [3, 6]}
```

```
[34]: gk=df_new.groupby('Name')
gk
```

```
[34]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f68f329e920>
```

```
[35]: for i in gk:
        print(i)
```

```
('Abhi',      Name  Age  Address Qualification
7  Abhi    32  Aligarh             MA)
('Anuj',      Name  Age  Address Qualification
1  Anuj    24  Kanpur              MA
5  Anuj    36  Kanpur              B.com)
('Gaurav',    Name  Age  Address Qualification
4  Gaurav   33  Jaunpur            B.Tech)
('Jai',       Name  Age  Address Qualification
0  Jai     27   Nagpur             Msc
2  Jai     22  Allahabad           MCA)
('Princi',    Name  Age  Address Qualification
3  Princi   32   Kannuaj           Phd
6  Princi   27  Allahabad           Msc)
```

```
[36]: #groupby using multiple keys
data1 = {'Name':['Jai', 'Anuj', 'Jai', 'Princi', 'Gaurav', 'Anuj', 'Princi',
↪ 'Abhi'],
        'Age':[27, 24, 22, 32, 33, 36, 27, 32],
        'Address':['Nagpur', 'Kanpur', 'Allahabad', 'Kannuaj', 'Jaunpur',
↪ 'Kanpur', 'Allahabad', 'Aligarh'],
        'Qualification':['Msc', 'MA', 'MCA', 'Phd', 'B.Tech', 'B.com', 'Msc',
↪ 'MA']}
}
```

```
[37]: df_new=pd.DataFrame(data1)
```

```
[38]: df_new
```

```
[38]:      Name  Age  Address Qualification
0     Jai    27   Nagpur             Msc
1    Anuj    24   Kanpur              MA
2     Jai    22  Allahabad           MCA
3  Princi    32   Kannuaj           Phd
4  Gaurav    33   Jaunpur          B.Tech
5    Anuj    36   Kanpur              B.com
```

6	Princi	27	Allahabad	Msc
7	Abhi	32	Aligarh	MA

```
[40]: df_new.groupby(['Name', 'Qualification'])
```

```
[40]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f68f33053c0>
```

```
[43]: print(df_new.groupby(['Name', 'Qualification']).groups)
```

```
{('Abhi', 'MA'): [7], ('Anuj', 'B.com'): [5], ('Anuj', 'MA'): [1], ('Gaurav', 'B.Tech'): [4], ('Jai', 'MCA'): [2], ('Jai', 'Msc'): [0], ('Princi', 'Msc'): [6], ('Princi', 'Phd'): [3]}
```

```
[44]: df_new.groupby('Address').sum()
```

```
/tmp/ipykernel_128/1899709775.py:1: FutureWarning: The default value of
numeric_only in DataFrameGroupBy.sum is deprecated. In a future version,
numeric_only will default to False. Either specify numeric_only or select only
columns which should be valid for the function.
```

```
df_new.groupby('Address').sum()
```

```
[44]:
```

	Age
Address	
Aligarh	32
Allahabad	49
Jaunpur	33
Kannauj	32
Kanpur	60
Nagpur	27

```
[ ]: #merging two dataframes
```

```
[45]: data1 = {'Name': ['Jai', 'Anuj', 'Jai', 'Princi'],
              'Age': [27, 24, 22, 32],
              'Address': ['Nagpur', 'Kanpur', 'Allahabad', 'Kannauj'],
              'Qualification': ['Msc', 'MA', 'MCA', 'Phd']}

data2 = {'Name': ['Gaurav', 'Anuj', 'Princi', 'Abhi'],
         'Age': [33, 36, 27, 32],
         'Address': ['Jaunpur', 'Kanpur', 'Allahabad', 'Aligarh'],
         'Qualification': ['B.Tech', 'B.com', 'Msc', 'MA']}
```

```
[50]: df1=pd.DataFrame(data1,index=[0,1,2,3])
df2=pd.DataFrame(data2,index=[4,5,6,7])
```

```
[51]: df1
```

```
[51]:
```

	Name	Age	Address	Qualification
0	Jai	27	Nagpur	Msc
1	Anuj	24	Kanpur	MA
2	Jai	22	Allahabad	MCA
3	Princi	32	Kannuaj	Phd

```
[52]: df2
```

```
[52]:
```

	Name	Age	Address	Qualification
4	Gaurav	33	Jaunpur	B.Tech
5	Anuj	36	Kanpur	B.com
6	Princi	27	Allahabad	Msc
7	Abhi	32	Aligarh	MA

```
[53]: df_concat=[df1,df2]  
res=pd.concat(df_concat)
```

```
[54]: res
```

```
[54]:
```

	Name	Age	Address	Qualification
0	Jai	27	Nagpur	Msc
1	Anuj	24	Kanpur	MA
2	Jai	22	Allahabad	MCA
3	Princi	32	Kannuaj	Phd
4	Gaurav	33	Jaunpur	B.Tech
5	Anuj	36	Kanpur	B.com
6	Princi	27	Allahabad	Msc
7	Abhi	32	Aligarh	MA

```
[ ]:
```