

# Pandas

1. Pandas is a library used for working with datasets
2. It has different functions for analyzing, cleaning, exploring, and manipulating data
3. Pandas is a popular Python-based data analysis library
4. Use with `[import pandas as pd]`
5. Pandas presents a diverse range of utilities ranging from parsing multiple file formats

## Why Pandas?

1. Pandas allows us to analyze big data and make conclusion-based statistical analysis
2. Pandas can clean messy data sets, and make them readable and relevant
3. Relevant is very important in data science

## What can Pandas do

1. Is there correlation between two or more columns
2. Average value
3. Max value
4. Min value
5. Remove Noise
6. Cleaning data

## There are different operations perform with the help of Pandas:

1. Basic:
  1. Pandas Series
  2. DataFrame
  3. Read CSV
  4. Read JSON
  5. Analyze Data
2. Cleaning Data:
  1. Clean Data
  2. Clean Empty Cell
  3. Clean Wrong Data
  4. Remove Duplicates
3. Advance:
  1. Correaltion
  2. Plotting

```
In [1]: import pandas as pd
```

```
In [2]: my_dataset = {  
        'name': ['Rohan', 'Rahul', 'Ram', 'Shital'],  
        'gender': ['male', 'male', 'male', 'female']  
    }  
  
    my_df = pd.DataFrame(my_dataset)  
    my_df
```

Out[2]:

	name	gender
0	Rohan	male
1	Rahul	male
2	Ram	male
3	Shital	female

```
In [16]: df = my_df.to_csv('Data.csv')  
df = my_df.to_json('Data.json')
```

```
In [9]: df = pd.read_csv('Data.csv')
```

```
In [10]: df
```

Out[10]:

	Unnamed: 0	name	gender
0	0	Rohan	male
1	1	Rahul	male
2	2	Ram	male
3	3	Shital	female

```
In [11]: df = pd.read_json('Data.json')
```

```
In [13]: print(df)
```

	name	gender
0	Rohan	male
1	Rahul	male
2	Ram	male
3	Shital	female

```
In [19]: df = pd.read_csv('titanic_train.csv')
```

In [20]: df

Out[20]:

	PassengerId	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	7.9250
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	8.0500
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

891 rows × 12 columns



```
In [21]: df.isnull().sum()
```

```
Out[21]: PassengerId      0
         Survived        0
         Pclass         0
         Name           0
         Sex            0
         Age           177
         SibSp          0
         Parch          0
         Ticket         0
         Fare           0
         Cabin         687
         Embarked       2
         dtype: int64
```

```
In [22]: pd.__version__
```

```
Out[22]: '1.2.4'
```

```
In [33]: import pandas as pd
         a = [10,20,30,40]
         df = pd.Series(a)
```

```
In [24]: pd
```

```
Out[24]: 0    10
         1    20
         2    30
         3    40
         dtype: int64
```

```
In [31]: import pandas as pd
         my_num = [10,20,30,40,50]
         my_var = pd.Series(my_num)
```

```
In [32]: my_var
```

```
Out[32]: 0    10
         1    20
         2    30
         3    40
         4    50
         dtype: int64
```

```
In [ ]:
```