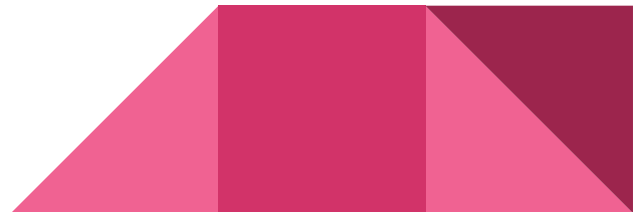



PYTHON PANDAS

© Pratik Joshi



Pandas Module

- Pandas is a Python library used for working with data sets.
 - It is used for analyzing, cleaning, exploring, and manipulating data.
 - The word Pandas has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.
 - Pandas module is widely used in Data Science to analyze big data set and clean it to get relevant data.
 - Relevant data is very crucial in case of Data Science.
 - Data Science: It is a branch of computer science where we study how to store, use and analyze data for deriving information from it.
 - Pandas Source Code: <https://github.com/pandas-dev/pandas>
- 

Getting Started with Pandas

Installation:

```
pip install pandas
```

Import:

```
import pandas
```

Import with Alias:

```
import pandas as pd
```

pd is the general alias for pandas, which is used widely.



Pandas Dataframe

A Pandas DataFrame is a 2 dimensional data structure like a table with rows and column.

Creation of Dataframe:

```
import pandas as pd                #Importing pandas module

data={"id":[101,102,103]
      ,"name" : ["King", "Jack", "Joker"] }
df=pd.DataFrame(data)              #Creation of Dataframe
print(df)
```

OP:

	id	name
0	101	King
1	102	Jack
2	103	Joker

Read CSV File

CSV files (comma separated files) is a simple way to store big data sets. Using pandas the csv file can be converted to dataframe using **read_csv()**.

Let's assume simple csv file containing emp data(id,name,location,salary).

```
import pandas as pd
emp_df=pd.read_csv("emp.csv")
print(emp_df)
print(emp_df['ename'][0]) #King
```

OP:

	empid	ename	eloc	esal
0	101	King	BLR	20000
1	102	Jack	HYD	50000
2	103	Queen	BBSR	10000
3	104	Ace	CTC	30000
4	105	Knight	KOL	90000

Read Excel File

To read .xls file in windows OS we need to install **openpyxl** module.


```
pip install openpyxl
```

Lets try to read one excel file emp.xls with two sheets, emp_details and dept_details containing employee data and department data respectively.

```
import pandas as pd
```

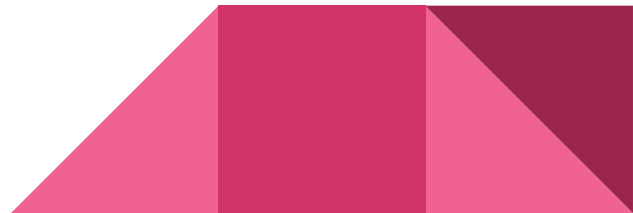
```
sheet1=pd.read_excel('emp.xls',sheet_name='emp_details')    #Read data from sheet1  
print(sheet1)
```

```
sheet2=pd.read_excel('emp.xls',sheet_name='dept_details')    #Read data from sheet2  
print(sheet2)
```



PYTHON NUMPY

© Pratik Joshi



Numpy Module

- NumPy stands for Numerical Python.
- It is used for working in Arrays.
- It was created in 2005 by Travis Oliphant.
- Numpy Arrays are 50 times faster than Python List. Numpy Arrays are widely used in Data Science.
- Numpy is written in C/C++.
- Numpy Code Base: <https://github.com/numpy/numpy>

© Pratik Joshi



Getting Started with Numpy

Installation: (No Need to Install Numpy if pandas is already installed)
`pip install numpy`

Import:
`import numpy`

Import with Alias:
`import numpy as np`
`np` is the general alias for numpy, which is used widely.



Numpy Array

- The array object in NumPy is called **ndarray**.
- NumPy ndarray object can be created by using the `array()` function.

Creating Numpy Array:

```
import numpy as np
```

```
arr=np.array([1,2,3,7])
```

```
print(arr)
```

```
print(type(arr))
```

```
#<class 'numpy.ndarray'>
```



Python List VS Numpy Array

```
import numpy as np
import time
```

```
py_list=list(range(30000000))      #Creation of Python List
arr=np.arange(30000000)            #Arrange function is used to Create Numpy Array in range
```

```
begin=time.time()
new_lst=[i*2 for i in py_list]      #Squaring items of Python List
end=time.time()
print("Time of execution for Python List:",end-begin)  # 2.19 Sec.
```

```
begin=time.time()
new_arr=arr*2                       #Squaring items of Numpy Array
end=time.time()
print("Time of execution for Numpy Array:",end-begin)  #0.05 Sec.
```

****Numpy Array is Taking very less time and occupied less memory then Python List**