

Python pandas

- ✓ Pandas is an open-source, BSD-licensed Python library providing
 - high-performance,
 - easy-to-use data structures and data analysis toolsfor the Python programming language.
- ✓ Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Library Pandas of Python

Dahmani Djaouida

```
import pandas
```

File sudents.csv

Name,LastName,Sexe,Mark

Ilie,Jean,M,12

Akli,Ali,M,8

Sarco,Michele,F,18

Messa,Francine,F,2

Messa,Cathy,F,7

.....

```
df = pandas.read_csv("C:/Users/dze/Desktop/persons.csv")  
    print(df)  
    print(df.shape)
```

```
df.head(3)  
df.tail(2)
```

```
df.columns  
df.dtypes
```

```
import pandas as pd
df1=pd.DataFrame([["Ilié","Jean",10],["Abed","Ali",12],["Boutier","Maxime",18],["Boutier","Vanessa",2]],columns=['firstName','lastName','mark' ])
print(df1)
```

```
df2=pd.DataFrame([["Ilié","Jean",10],["Abed","Ali",12],["Boutier","Maxime",18],["Boutier","Vanessa",2]],index=['s1','s2','s3','s4'],columns=['firstName','lastName','mark' ])
print(df2)
```

`df.firstName`

`df.firstName[1:3]`

`df['firstName'].describe()`

```
df1.columns
```

```
df1.index
```

```
df1.mark.sort_values()
```

```
df1.sort_values(mark)
```

```
df1.loc['s1']
```

```
df1.loc[['s1','s3']]
```

```
df1.sum()
```

```
df1.mark.sum()
```

```
df1.matk+2
```

```
df1.firstName.apply(lambda x:x+'###')
```

```
df1.loc[df1['firstName']=="Boutier",:]
```

```
df1.loc[df1['mark']<10,:]
```

```
df1.loc[df1['mark']>10,:]
```

```
df1.loc[df1['mark']>10,:].count()
```

```
df1.loc[(df1['firstName']=='Boutier') & (df1['note']  
>10),:]
```

```
pd.crosstab(df.name,df.mark)
```

```
pd.crosstab(df.mark,df1.Name)
```

```
df.mark.sort_values()
```

```
import matplotlib.pyplot as plt
```

```
plt.hist(df.name)
```

```
plt.hist(df.mark)
```

```
s=pd.Series(np.linspace(0,15,5),name='series',index=['s1'  
            , 's2', 's3', 's4', 's5'])  
            print(s)  
            type(s)
```

```
s1=df1['mark']  
            type(s1)  
            print(s1)
```

```
df4=pd.DataFrame([[1,2,3,4],[5,6,7,8]],index=['s1','s2'],
                  columns=['A','B','C','D'])
df5=pd.DataFrame([[-1,-2,-3,-4],[-5,-6,-7,-8]],index=['d1','d2'],columns=['A','B','C','D'])
df6=pd.concat((df4,df5))
df6
```

```
df4=pd.DataFrame([[1,2,3,4],[5,6,7,8]],index=['s1','s2'],
                  columns=['A','B','C','D'])
df5=pd.DataFrame([[-1,-2,-3,-4],[-5,-6,-7,-8]],index=['d1','d2'],columns=['E','F','G','H'])
df6=pd.concat((df4,df5),sort=True)
df6
```

```
df4=pd.DataFrame([[1,2,3,4],[5,6,7,8]],index=['s1','s2'],columns=['A','B','C','D'])
df5=pd.DataFrame([[-1,-2,-3,-4],[-5,-6,-7,-8]],index=['d1','d2'],columns=['A','F','G','C'])
df6=pd.concat((df4,df5),join='inner')
print(df6)
```

```
df4=pd.DataFrame([[1,2,3,4],[5,6,7,8]],index=['s1','s2'],columns=['A','B','C','D'])
df5=pd.DataFrame([[-1,-2,-3,-4],[-5,-6,-7,-8]],index=['d1','d2'],columns=['A','F','G','C'])
df6=pd.concat((df4,df5),join='inner',keys=['a','b'])
Print(df6)
```

```
df4=pd.DataFrame([[1,2,3,4],[5,6,7,8]],index=['s1','s2'],columns=['A','B','C','D'])
df5=pd.DataFrame([[-1,-2,-3,-4],[-5,-6,-7,-8]],index=['d1','d2'],columns=['A','F','G','C'])
df6=pd.concat((df4,df5),join='inner')
%time df6.mean()
```