Programação com Python

Prof. Erneson A. Oliveira

MBA em Ciência de Dados
Universidade de Fortaleza

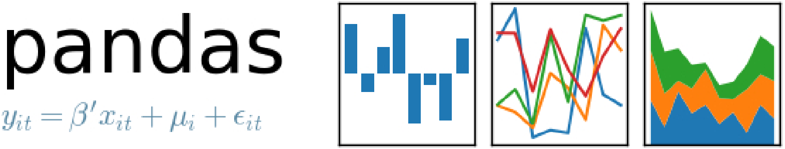
18 de Janeiro de 2020

FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA
ENSINANDO E APRENDENDO

DPDI - DIRETORIA DE PESQUISA,
DESENVOLVIMENTO E INOVAÇÃO

Aula 4 - Introdução à Python

http://www.python.org



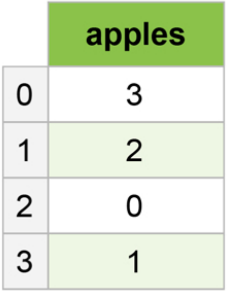
https://pandas.pydata.org

Pandas *é um módulo para manipulação de tabelas (planilhas).*

X
Excel

Séries

Series



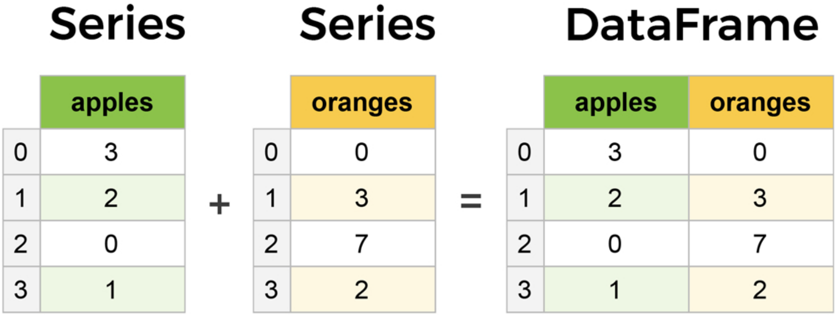
	apples
0	3
1	2
2	0
3	1

```python
import numpy as np
import pandas as pd

apples=[3,2,0,1]
s=pd.Series(apples,name='apples',dtype=np.int64)

print(s)

# print(type(s))
# print(dir(s))
```

```python

print(s.head(2))
print(s.tail(2))

print(s.ndim)
print(s.shape)
print(s.mean())
print(s.max())
print(s.min())

```

https://pandas.pydata.org/pandas-docs/stable

DataFrame



Series

	apples
0	3
1	2
2	0
3	1

+

Series

	oranges
0	0
1	3
2	7
3	2

=

DataFrame

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

```python
import numpy as np
import pandas as pd

d={'apples': [3,2,0,1],
   'oranges': [0,3,7,2]}

df=pd.DataFrame(d)
print(df)

# print(type(df))
```

```python
# print(dir(df))

print(df.ndim)
print(df.shape)
print(df.mean())
print(df.max())
print(df.min())
print(df.T)

```

```
import numpy as np
import pandas as pd

d={'apples': [3,2,0,1],
   'oranges': [0,3,7,2]}

df=pd.DataFrame(d)
print(df)

print([x for x in df.index])
print([x for x in df.columns])

d={'apples': {'A': 3,
              'B': 2,
              'C': 0,
              'D': 1},
   'oranges': {'A': 0,
               'B': 3,
               'C': 7,
               'D': 2}}

df=pd.DataFrame(d)
print(df)

print([x for x in df.index])
print([x for x in df.columns])

```

Entrada e Saída

```python
import numpy as np
import pandas as pd

d={'apples': [3,2,0,1],
   'oranges': [0,3,7,2],
   'bananas': [1,3,5,4],
   'avocados': [9,0,0,1]}

df0=pd.DataFrame(d)
print(df0)

```

```python
df0.to_csv('output.csv',sep=';',encoding='utf-8',index=False) # Escrevendo um CSV

df1=pd.read_csv('output.csv',sep=';',usecols=['apples','bananas'],encoding='utf-8') # Lendo um CSV
print(df1)

```

Ordenando e filtrando DataFrames

```python
import numpy as np
import pandas as pd

d={'apples': {'A': 3,'B': 2,'C': 0,'D': 1},
   'oranges': {'A': 0,'B': 3,'C': 7,'D': 2},
   'bananas': {'A': 1,'B': 3,'C': 5,'D': 4},
   'avocados': {'A': 9,'B': 0,'C': 0,'D': 1}}

df=pd.DataFrame(d)
print(df)

print(df.sort_index(axis=0, ascending=False))
print(df.sort_index(axis=1, ascending=False))
# continua

# continuacao

print(df.sort_values(by='oranges'))
print(df.sort_values(by=['avocados','apples']))
print(df.sort_values(by='oranges',axis=0))

print(df)
print(df.sort_values(by='B',axis=1))
print(df)

df.sort_values(by='B',axis=1,inplace=True)
print(df)

```

```python
import numpy as np
import pandas as pd

d={'apples': {'A': 3,'B': 2,'C': 0,'D': 1},
   'oranges': {'A': 0,'B': 3,'C': 7,'D': 2},
   'bananas': {'A': 1,'B': 3,'C': 5,'D': 4},
   'avocados': {'A': 9,'B': 0,'C': 0,'D': 1}}

df=pd.DataFrame(d)

print(df)

print(df['oranges'])
print(df['oranges'][0])
print(df.oranges)
print(df.oranges[0:2])
print(df[0:3])
# continua

# continuacao

df['oranges'][0]=99
print(df)

linha=df.loc['A']
print(type(linha))
print(linha)

linhas=df.loc[:,['oranges','bananas']]
print(type(linhas))
print(linhas)

print(df.loc['A','oranges']) # df.at['A','oranges']

print(df.iloc[3])
print(df.iloc[0:2, 0:2])
print(df.iloc[2,2]) # df.iat[2,2]
```

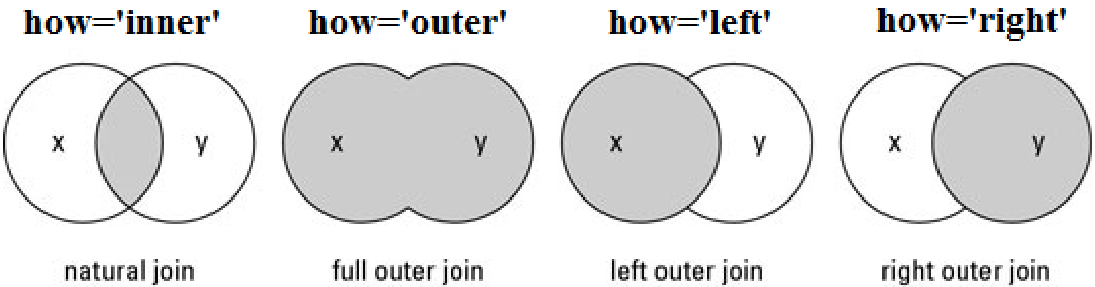
Iterando sobre DataFrames

```python
import numpy as np
import pandas as pd

d={'apples': {'A': 3,'B': 2,'C': 0,'D': 1},
   'oranges': {'A': 0,'B': 3,'C': 7,'D': 2},
   'bananas': {'A': 1,'B': 3,'C': 5,'D': 4},
   'avocados': {'A': 9,'B': 0,'C': 0,'D': 1}}

df=pd.DataFrame(d)
print(df)

for indice in df.index:
    print(df.loc[indice,'apples'])

for indice in df.index:
    print(df['apples'][indice])
# continua
# continuacao

for i in range(df.shape[0]):
    print(df.iloc[i,0])

for indice,linha in df.iterrows():
    print(indice,linha["apples"])

```

Higienização de DataFrames

```python
import numpy as np
import pandas as pd

d={'apples': [2,np.nan,1,5,1],
   'oranges': [3,2,1,3,1],
   'grapes': [3,2,1,3,1]}

df=pd.DataFrame(d)
print(df)

print(df.dropna(axis=0,how='any')) # Removing rows with missing values
print(df.dropna(axis=1,how='any')) # Removing columns with missing values
# continua

# continuacao

print(df.drop_duplicates(keep='first')) # Removing duplicated rows (but keeping the first)
print(df.T.drop_duplicates(keep='first').T) # Removing duplicated columns (but keeping the first)

print(df[df.apples > 1]) # Removing rows with zero values on column 'apples'

print(df[df.apples > 1].reset_index(drop=True)) # ...and reseting the indexes

```

Relacionando DataFrames


how='inner'
how='outer'
how='left'
how='right'
x
y
x
y
x
y
x
y
natural join
full outer join
left outer join
right outer join

```python
import pandas as pd
import numpy as np

d={'Customer_id':pd.Series([1,2,3,4,5,6]),
    'Product':pd.Series(['Oven','Oven','Oven','Television','Television','Television'])}
df0=pd.DataFrame(d)
print(df0)

d={'Customer_id':pd.Series([2,4,6]),
    'State':pd.Series(['California','California','Texas'])}
df1 = pd.DataFrame(d)
print(df1)

print(pd.merge(df0, df1, on='Customer_id', how='inner')) #inner join
print(pd.merge(df0, df1, on='Customer_id', how='outer')) #outter join
print(pd.merge(df0, df1, on='Customer_id', how='left')) # left join
print(pd.merge(df0, df1, on='Customer_id', how='right')) # right join

```

https://docs.python.org/3/tutorial/index.html
https://pandas.pydata.org
https://youtu.be/eykoKxsYtow?list=PLeo1K3hjS3usILfyvQlvUBokXkHPSve6S