



Școala
informală
de IT

Python Development

Pandas



Școala
informală
de IT

Pandas



Pandas

- Pandas este o librerie.
- Scopul este de analiza si procesare de seturi de date (data sets).
- Aceasta librerie are in componenta sa functii cu scop de analiza, curatare, explorare si manipulare de date.
- Numele are referinte atat in Panel Data, cat si in Python Data Analysis si a fost create in 2008.

Cum instalam Pandas?

```
pip install pandas
```



Ce este o serie (Pandas Series)?

O serie in Pandas reprezinta o coloana a unui table.

Este un array unidimensional care contine date.

```
1 import pandas as pd
2 a = [1, 7, 2]
3 variabila = pd.Series(a)
4 print(variabila)
```

```
0    1
1    7
2    2
dtype: int64
```

Etichetele (labels) reprezinta modalitatea prin care putem sa identificam o valoare. Daca nu este specificat nicio eticheta, valoarea este identificata prin indexul propriu. Primul index este 0, al doilea este 1, s.am.d.

```
print(variabila[0])
```

```
1
```



Ce este o serie (Pandas Series)?

Utilizand argumentul index putem sa adaugam propriile etichete.

```
import pandas as pd
a = [1, 7, 2]
variabila = pd.Series(a, index=["x", "y", "z"])
print(variabila)
```

```
x    1
y    7
z    2
dtype: int64
```

Dupa adaugarea propriului label, putem accesa valoarea prin referinta la acesta, dar si prin intermediul indexului.

```
print(variabila[1])
print(variabila["y"])
```

```
7
7
```



Ce este o serie (Pandas Series)?

În Python pentru adăugarea datelor într-o serie utilizăm dicționarele.

```
import pandas as pd
taskuri = {"ziua1": 2, "ziua2": 4, "ziua3": 1}
variabila = pd.Series(taskuri)
print(variabila)
```

```
ziua1    2
ziua2    4
ziua3    1
dtype: int64
```

Dacă dorim afișarea doar primelor două zile, adăugăm în argumentul index lista de zile.

```
import pandas as pd
taskuri = {"ziua1": 2, "ziua2": 4, "ziua3": 1}
variabila = pd.Series(taskuri, index=["ziua2", "ziua3"])
print(variabila)
```

```
ziua2    4
ziua3    1
dtype: int64
```



Ce este un dataframe?

Un Dataframe reprezinta un tabel multi-dimensional.

Seria reprezenta o coloana, dataframeul este intregul tabel.

```
1 import pandas as pd
2 taskuri = {
3     "zile": [2, 4, 5],
4     "durata": [50, 40, 45]
5 }
6 variabila = pd.DataFrame(taskuri)
7 print(variabila)
```

	zile	durata
0	2	50
1	4	40
2	5	45



Ce este un dataframe?

Utilizand atributul loc putem accesa una sau mai multe randuri dintr-un dataframe.

```
import pandas as pd
taskuri = {
    "zile": [2, 4, 5],
    "durata": [50, 40, 45]
}
variabila = pd.DataFrame(taskuri)
print(variabila.loc[0])
```

```
zile      2
durata    50
Name: 0, dtype: int64
```

Daca dorim accesarea mai multor randuri:

```
import pandas as pd
taskuri = {
    "zile": [2, 4, 5],
    "durata": [50, 40, 45]
}
variabila = pd.DataFrame(taskuri)
print(variabila.loc[[0, 1]])
```

```
   zile  durata
0     2     50
1     4     40
```



Ce este un dataframe?

Daca dorim sa numim indecsii, folosim argumentul index.

```
import pandas as pd
taskuri = {
    "zile": [2, 4, 5],
    "durata": [50, 40, 45]
}
variabila = pd.DataFrame(taskuri, index=["ziua1", "ziua2", "ziua3"])
print(variabila)
```

	zile	durata
ziua1	2	50
ziua2	4	40
ziua3	5	45

Localizarea randurilor.

```
import pandas as pd
taskuri = {
    "zile": [2, 4, 5],
    "durata": [50, 40, 45]
}
variabila = pd.DataFrame(taskuri, index=["ziua1", "ziua2", "ziua3"])
print(variabila.loc["ziua2"])
```

zile	4
durata	40
Name: ziua2, dtype: int64	



Importuri de date intr-un dataframe

Daca datele sunt stocate intr-un fisier, putem incarca datele cu ajutorul functiei read_csv.

```
import pandas as pd
df = pd.read_csv('data.csv')
print(df)
```

	Prop	producator1	producator2	producator3
0	masina	Dacia	Volvo	Renault
1	culoare	rosu	alb	verde

- max_rows: afiseaza numarul maxim de randuri

