Data manipulation

Using Python Libraries

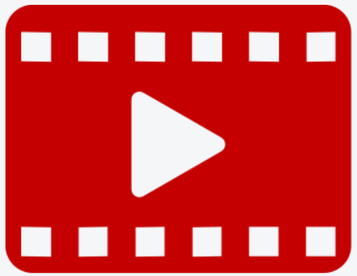
Start Now


DATA

Numbers

huge collections

R u serious ?



Qualitative / Quantitative

Quantitative (Numerical data)

- is represented by numbers.
- takes an infinite number of values that can be discrete ir continuous.

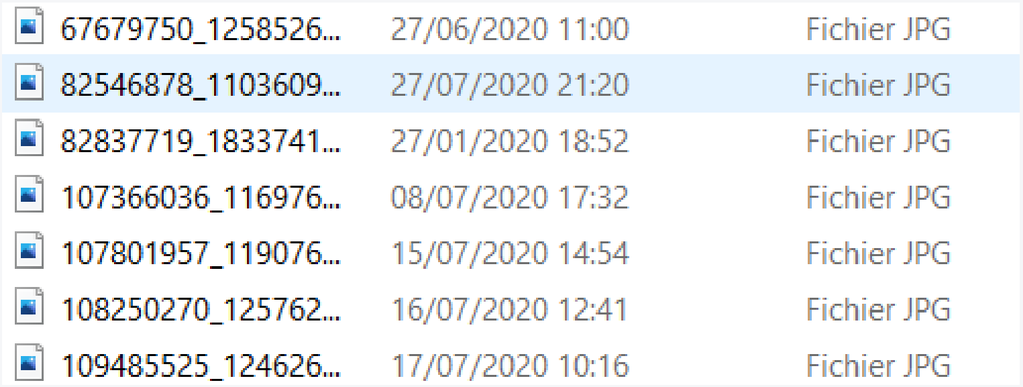
	Age
0	23
1	20
2	21
3	22
4	29
5	24
0	40

	price
0	221900.0
1	538000.0
2	180000.0
3	604000.0
4	510000.0
5	1225000.0

Quantitative (Categorical data)

- Is represented by numbers, words or text.
- Takes a limited number of values

	Name	Last Name	Age	Role	Home Adress	Full Name
0	John	Stott	23	AI	Los Angeles	John Stott
1	Jack	Mcbride	20	AI	California	Jack Mcbride
2	Angelica	Newman	21	WEB1	New York	Angelica Newman
3	Andrew	Hines	22	WEB2	Florida	Andrew Hines
4	Isac	Justice	23	Manager	Los Angeles	Isac Justice



What do we mean by

Data manipulation

Python Libraries

NumPy

- NumPy Array is an alternative to Python List
- Unlike Python Arrays, NumPy arrays can perform calculation on entire arrays
- It serves as ground base essential for other packages example: scikit-learn
- Many functionalities are available in NumPy with more convenient usage
- useful linear algebra, Fourier transform, and random number capabilities

Pandas

- Pandas is an open source python library which provides high performance, easy to use, flexible and expressive data structures designed to make working with structured (tabular, multi-dimensional, potentially heterogeneous) and time series data easily and intuitively.
- It aims to be the fundamental high-level building block for doing practical, real world data analysis in python.
- Pandas is built on Numpy, which is also an open source python library.

Series

- Is a one-dimensional labelled object.
- Has homogeneous data: Same type of data.
- His size is immutable : Unchangeable size.
- Values are mutable: You can change the data itself.

```
import pandas as pd

S=pd.Series(["AI","DataScience","ML"])
print(S)
```

```
0              AI
1    DataScience
2              ML
dtype: object
```

```
import pandas as pd

S=pd.Series(["25","60","70"])
print(S)
```

```
0    25
1    60
2    70
dtype: object
```

Series

Exemple

```
import pandas as pd

fruits=["bananas","apple","orange"]
quantities=[20,14,8]
S=pd.Series(quantities,index=fruits)
print(S)
```

```
bananas    20
apple      14
orange      8
dtype: int64
```

DataFrame

- DataFrame is a two-dimensional labeled data structure with columns of potentially different types.
- Dataframe columns are made up of pandas Series.
- You can think of it like a spreadsheet or an SQL table,or a dictionary of Series.

	Dates	ID	Names	Designation
0	2020-01-01	1	A	Manager
1	2020-01-02	2	B	Employee
2	2020-01-03	3	C	CEO
3	2020-01-04	4	D	CTO
4	2020-01-05	5	E	Manager
5	2020-01-06	6	F	Employee

Matplotlib

- Is built on NumPy arrays.
- Designed to create a wide variety of 2D graphs and plots.

Seaborn

- Is built on matplotlib.
- Introduces additional plot types.
- Makes your traditional Matplotlib plots look prettier.

Let's practice!

Youupi

Let's practice!

1. Manipulate data with NumPy
2. Manipulate data with Pandas
3. Visualize data with Matplotlib and Seaborn

Youupi