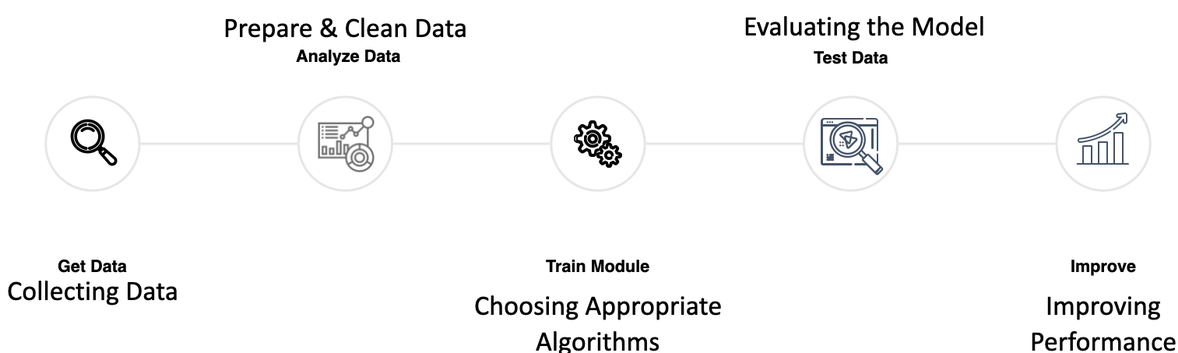


# pandas

أهم الخطوات في مشاريع علم البيانات



## مقدمة عن مكتبة Pandas

مكتبة Pandas مبنية فوق مكتبة NumPy وتوفر أدوات سهلة الاستخدام لتحليل وهيكلة البيانات (data structures and data analysis tools) بلغة python.

## هيكلة البيانات (Pandas Data Structures)

### النوع الأول: Series

مصفوفة أحادية الأبعاد نستخدمها في حفظ أي نوع من البيانات مثل: (integer, string, float) هذا النوع يحتوي index و data كما في الشكل التالي.

Series	
index	value
0	12
1	-4
2	7
3	9

../images/336498\_2\_En\_4\_Chapter/336498\_2\_En\_4\_Fig1\_HTML.jpg

#### النوع الثاني: DataFrame

هيكل ثنائي الأبعاد يحتوي على أنواع مختلفة من البيانات ويمكن تشبيهه بملف الأكسل أو جدول SQL أو dictionary مكون من Series objects.

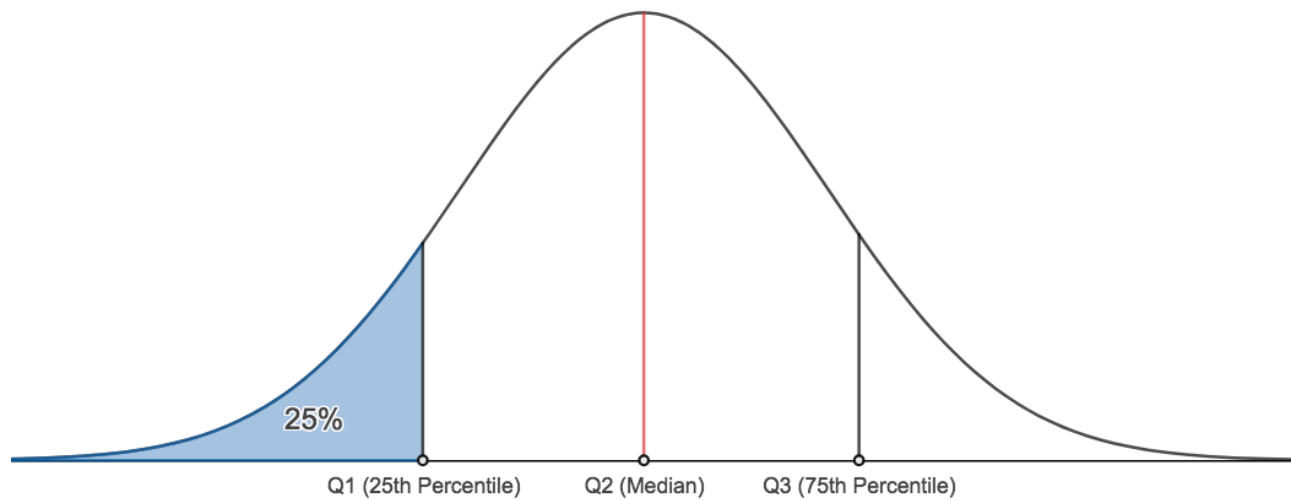
Columns →

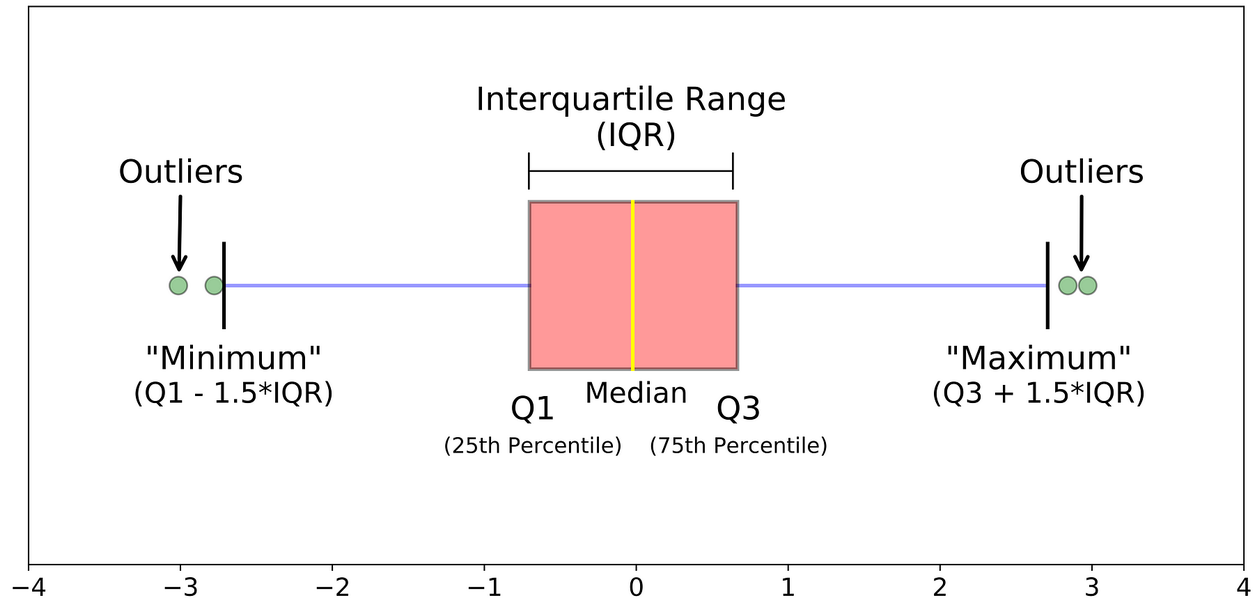
Country	Capital	Population
---------	---------	------------

Index →

0	Belgium	Brussels	11190846
1	India	New Delhi	1303171035
2	Brazil	Brasilia	207847528

## Quantiles and Percentiles مفهوم





*Understanding Boxplots. The image above is a boxplot. A boxplot... / by Michael Galarnyk / Towards Data Science*

مصادر أخرى للاطلاع:

- Introduction to Data Science by Laura Igual and Santi Seguí
  - <https://github.com/DataScienceUB/introduction-datascience-python-book>
- pandas Documentation
  - <https://pandas.pydata.org/>
- [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/dsintro.html#dataframe](https://pandas.pydata.org/pandas-docs/stable/user_guide/dsintro.html#dataframe)
- <https://www.datacamp.com/community/tutorials/pandas-tutorial-dataframe-python>