The most important Python libraries and packages for Data Science

These are the five most essential Data Science libraries on Python

- Numpy

- Pandas

- Matplotlib

- Scikit-Learn

- Scipy

Numpy:

Numpy will help you to manage multi-dimensional arrays very efficiently. Maybe you won't do that directly, but since the concept is a crucial part of data science

Pandas

To analyze data, we like to use two-dimensional tables – like in SQL and in Excel. Originally, Python didn't have this feature. Weird, isn't it? But that's why Pandas is so important! I like to say, Pandas is the "*SQL of Python*

Matplotlib

Hope I don't have to detail why data visualization is important. Data visualization helps you to better understand your data, discover things that you wouldn't discover in raw format and communicate your findings more efficiently to others.

Scikit-Learn

Without any doubt the fanciest things in Python are Machine Learning and Predictive Analytics. And the best library for that is Scikit-Learn, which simply defines itself as "Machine Learning in Python

Scipy

This is kind of confusing, but there is a Scipy *library* and there is a Scipy *stack*. Most of the libraries and packages I wrote about in this article are part of the Scipy *stack* (that is for scientific computing in Python)