الجمعيّة السوريّة للبحث العلميّ
Syrian Society for Scientific Research

Intro to Data Analysis

October 2022

AGENDA

- Intro & Meet the lecturers
- Challenges to be a DS
- Why Cloud is essential?
- Why Data Analysis?
- Why Pandas and NumPy?
- Common IDEs and tools
- Installation
- Q&A

Motivation

Reslan Tinawi

- BSc degree in information engineering from Damascus University.
- Software Engineer from 2018 to 2020.
- Data Scientist from 2020 to present.
- Data Science Mentor at Syrian Youth Assembly
- Personal blog: reslan-tinawi.github.io

Soubhi Hadri

- Data Scientist at Microsoft, USA.
- Former CV/AI Eng. with Shiseido.
- MSc in ECE, University of Oklahoma.
- BSc degree in Information Engineering from the University of Aleppo.
- Soubhihadri.com

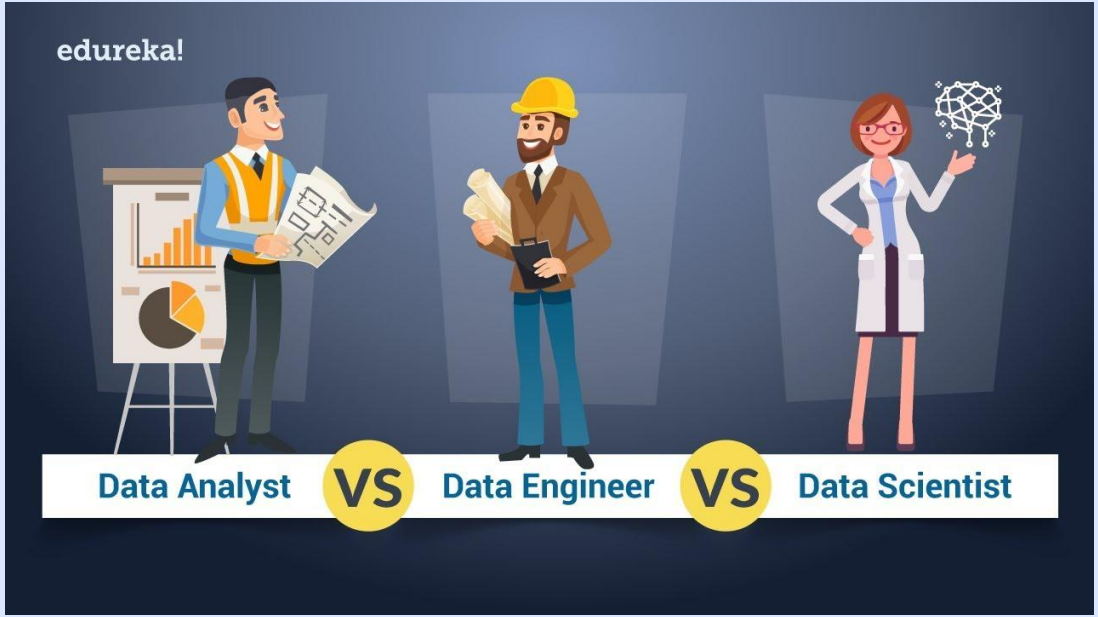
Before DS?

Challenges to be a DS

- **MENA**:
  - Few opportunities.
  - Data is not the focus.
- **AMERICA**:
  - Many opportunities.
  - VERY competitive.
  - PhD, MSc.

What are all these titles, Data Eng, Data SC, Data and Applied SC, ML Eng?

- **Data** Engineer**.**
- **Data** Analyst.
- Data and **Applied** SC.
- ML Eng, ML Researcher, DS.
- ML**Ops** Eng: DevOps



Data Analyst vs Data Engineer vs Data Scientist | Edureka

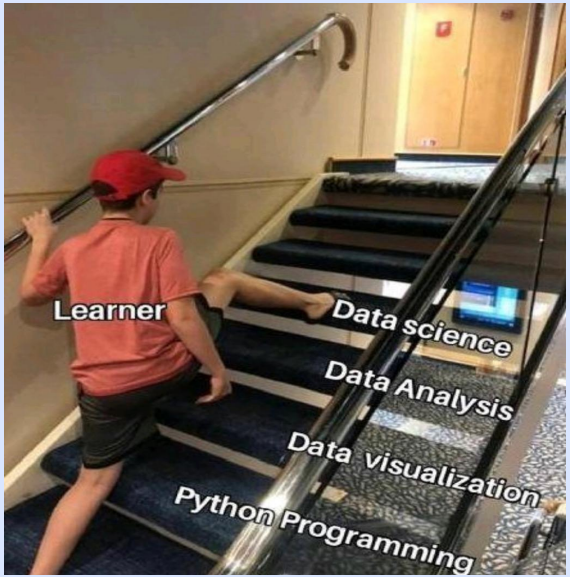
Why Cloud is essential?

Why Python?

- Easy to learn language with simple syntax (compared to other languages).
- Wide community of researchers and academics use it.
- Large number of open-source libraries for data science tasks (don’t reinvent the wheel).
  - Deep learning and ML: TensorFlow, scikit-learn.
  - Data manipulation: pandas.
  - Data visualization: Matplotlib, seaborn.

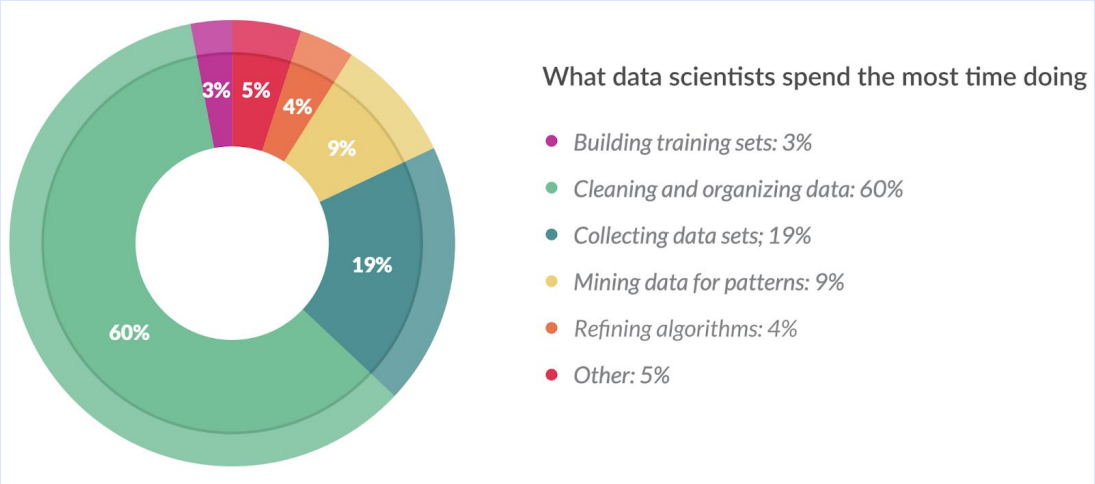
Why Data Analysis?

Learners often jump immediately to machine learning before understanding the basics, such as Python programming, data analysis, and visualization.



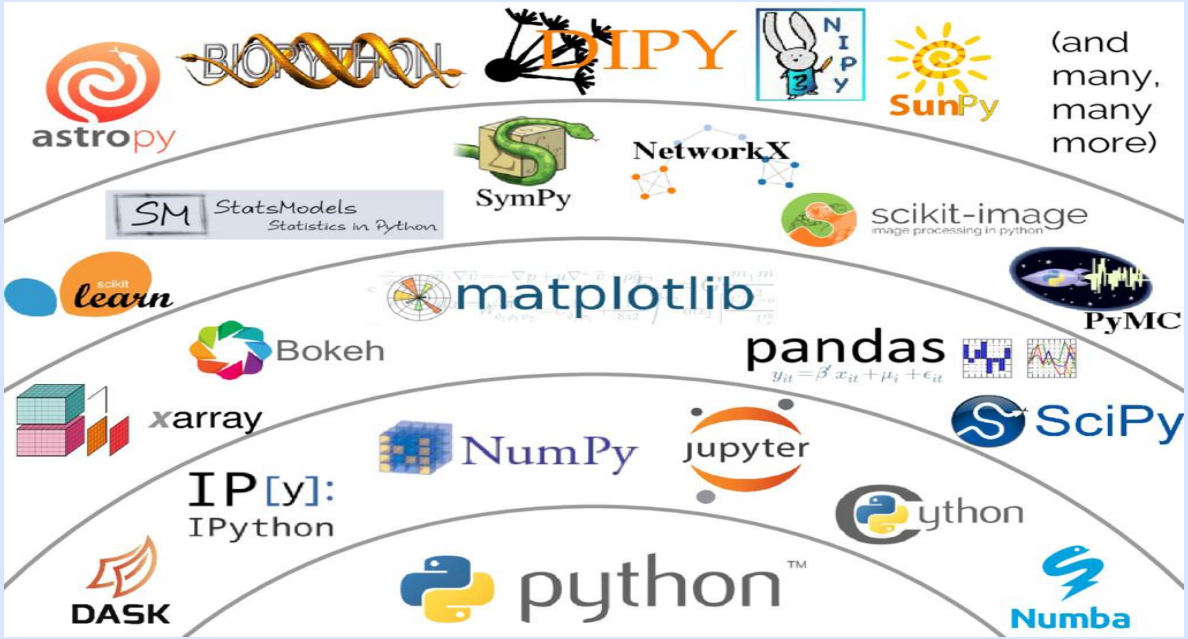
Why Data Analysis?

In one study, it was found that data scientists spend almost 60% of their time cleaning and manipulating data. Thus, it's very important to become knowledgeable in the tools to analyze data.



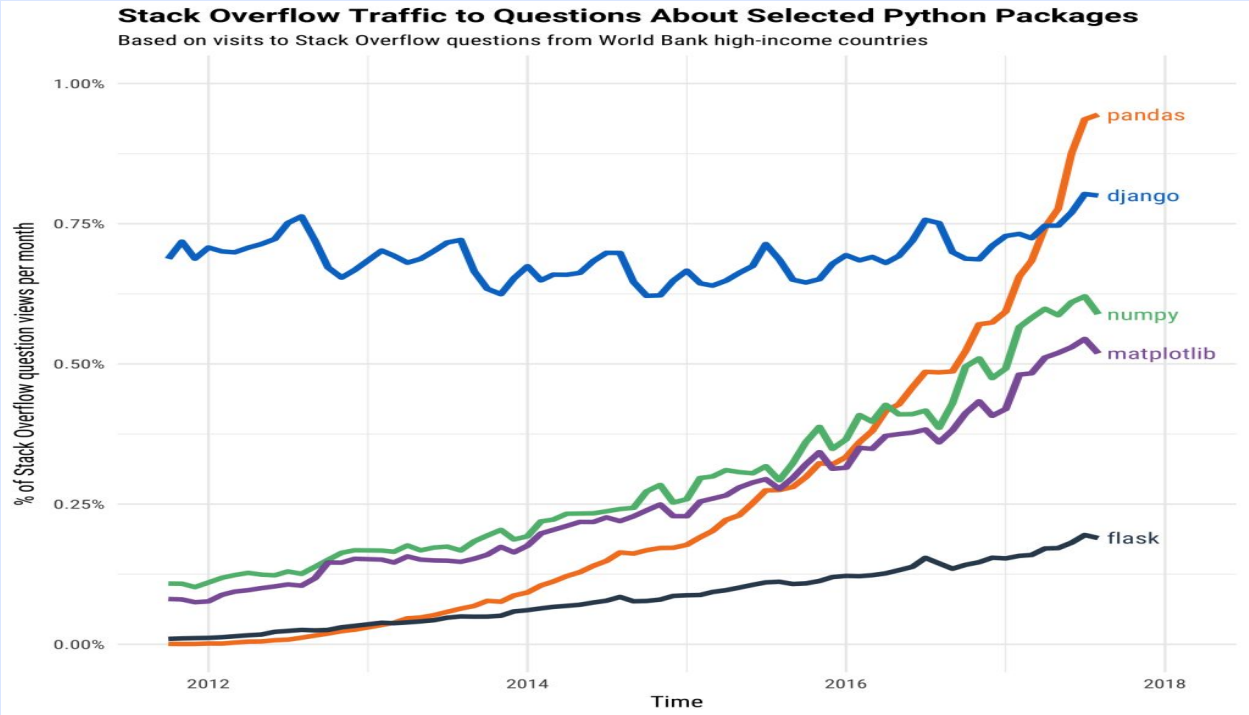
Why Pandas and NumPy?

They form the basis of scientific computing in in Python



Why Pandas and NumPy?

Both libraries have been growing in popularity in the last 10 years.



What is Jupyter Notebook?

- Data Analysis tasks often involves interacting with a dataset.
- Different from traditional programming context.
- A platform (or tool) is needed to facilitate this interactivity.
- Jupyter notebook is a web-based interactive computing tool for creating language-agnostic notebooks.
- One can create notebook in Python, R, or Julia.
- A notebook is a combination of code cells and Markdown cells.

Tour of jupyter notebook

- What is a notebook?
- What is a kernel?
- Working on multiple notebooks simultaneously.
- Writing code and markdown.

Common IDEs and tools

VS code

Google Colab

PyCharm

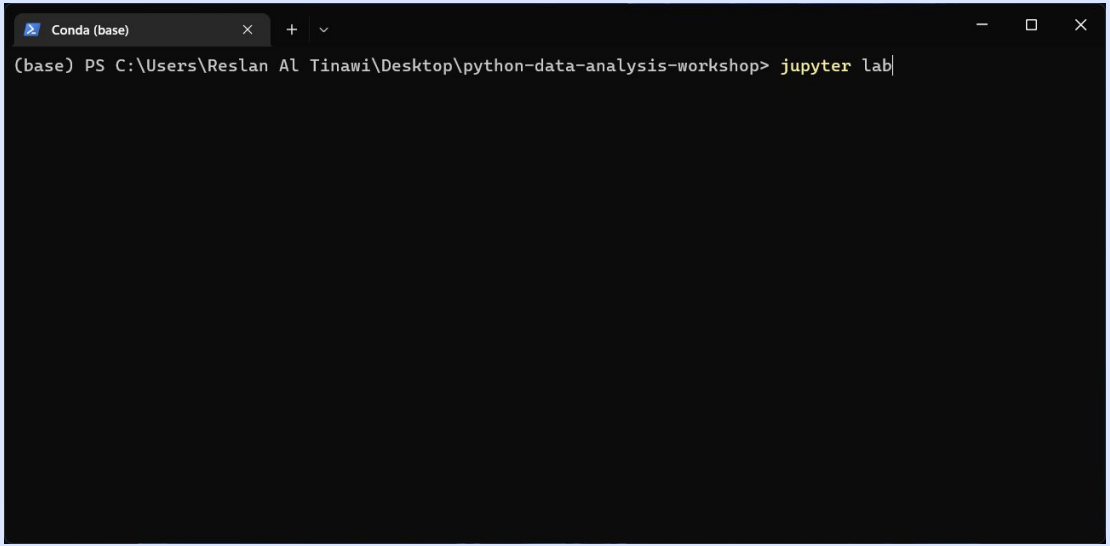
Anaconda

Installation:

- Go to: https://www.anaconda.com/products/distribution
- Depending on your OS, choose the right installation.

Installation (launching Jupyter):

- In your terminal type: jupyter lab



Installation:

- VS Code supports working with notebooks.
- First, you should install the jupyter package (if not already installed) by running:

  ```
  pip install jupyter
  ```

- Then, you need to install the jupyter extension for VS Code from here.

Google Colab

- If you don’t want to install anything at the moment, you can just use google colab by opening: https://colab.research.google.com/
- All the notebooks for this workshop will be available on colab.

https://github.com/Reslan-Tinawi/python-data-analysis-workshop

reslan.tinawi@syssr.org

https://www.linkedin.com/in/reslan-al-tinawi/

soubhi.hadri@syssr.org

https://www.linkedin.com/in/soubhihadri/

الجمعيّة السوريّة للبحث العلميّ
Syrian Society for Scientific Research