

# Python ecosystem

Python can be used to make programs which perform certain tasks. An economist will often need to do the following:

- Import data
- Clean data
- Fit models to the data (like regression models)
- Generate outputs (like tables and graphs)
- Make a report based on the economists' findings (a presentation or white paper)

All of this can be done in Python as long as you know how to program and which modules to use.

## Import data

- There are several ways to get hold of data.
- Data base. Here, you often have to use a programming language called SQL to extract data directly from a data base.
- Web scraping. In Python, you would use the library “Beautiful Soup” for this. Beware! Not all web pages allow web scraping.
- Web-API
  - o "pandas\_datareader" from the Pandas module

- "ticker" function from the yfinance module
  - "get" function from the quandl module. (check out [quandl.com](https://www.quandl.com) which is now owned by Nasdaq)
  - There are many, many more web-APIs out there where you can download data.
- Files on shared sources like the web. Often, you will find data in different formats. Sometimes the files are in .csv format (comma separated values), sometimes in Excel format or text format.
  - Sometimes you want to keep the data in Excel and perform the analysis in this application. Often, there are two reasons for this, either that the rest of the organization uses Excel (facilitates communication and collaboration), or that you need to make a small implementation fast (in Excel you get a solution up and running in no time). The downsides of Excel are size limitations, limited analytical methods, lack of automatization, and execution speed.

## **Clean data**

- You will both use the library Pandas and programming logic to clean your data in Python.

## **Fit models to data**

- There are several modules which are important when you fit models to data or want to generate data sets.
- The modules NumPy and SciPy for mathematical operations and simple data handling.
- The module statsmodels for fitting statistical models, like regression models.

- The module scikit-learn, keras, and tensorflow helps you fit statistical models, also machine learning models.

### **Generate outputs (good enough for academic papers)**

- Matplotlib is a great starter tool for graphs in Python. It is also more used than Plotly. Faster to make simple plots.
- Plotly is more advanced than Matplotlib and makes it faster to create elaborate plots.
- However, if you use Holoviews or Seaborn, then making elaborate graphs in Matplotlib goes very fast.

### **Make a report or dashboards (business intelligence)**

- Jupyter notebooks. The benefit of using Jupyter notebooks is that the reader can examine the code behind the numbers. The notebook can contain text, mathematical equations, graphs and tables.
- Flask in Python. This is the simplest framework to make web pages in Python. It can also be used to create dashboards. The web page itself will not be Python code but it will be HTML, CSS, and Javascript.
- Django in Python. Django is a more advanced framework to make web pages in Python. It is also used to make dashboards. The web page itself will not be Python code but it will be HTML, CSS, and Javascript.
- Power BI. One of the two most popular programs to make dashboards (has a free version).
- Tableau. The other of the two most popular programs to make dashboards.

- Excel. Yes, often results are presented using figures from Excel...

## **Other programming tasks**

### Building web pages.

- For this you would use HTML, CSS and Javascript to build the front-end (the web page).
- Backend could often be Python. Where the program has to react to the user input from the web page and interact with a data base or a statistical model (for recommendation of new movies for example).

### Making games.

- For this you can use the PyGames module in Python. If you are really serious about game dev then Unity or Unreal Engine would suit you.
  - o Unity uses C#, whereas Unreal Engine uses C++

### Other programming languages for economists.

- Statistical modelling is often done using R, mostly in academia but also in business. This is the second most popular language for data analysts/data scientists, after Python.
- Stata is very popular in academic circles, as it is quite streamlined and powerful. You do not have to import several modules as in Python or R.
- Learn Excel, period. It is very good for modelling Results, Balance, and Cashflow statements. All companies use it at some level.