42578 Advanced Business Analytics

Refreshing Python and Machine Learning

Today: Practical Exercises Only

- **Part 1 - Python basics**
- **Part 2 - Numpy and Matplotlib**
- **Part 3 - Pandas**
- **Part 4 - Machine Learning with scikit-learn**

** Parts 1-3 are taken from the 42577 Introduction to Business Analytics E19 / 42184 Data Science and GIS for Mobility E19*

Installation instructions

- Anaconda (easy to use but resource-demanding)
  - https://www.anaconda.com/distribution/ (Python 3)

- **OR, ALTERNATIVELY,**

- Python 3 and pip
  - https://www.python.org/downloads/
  - https://docs.python.org/3/installing/index.html
  - https://packaging.python.org/tutorials/installing-packages/

- venv/virtualenv (optionally, a cleaner way to manage installed packages)
  - https://packaging.python.org/guides/installing-using-pip-and-virtual-environments/
  - https://docs.python.org/3/library/venv.html

- Jupyter notebook / JupyterLab
  - https://jupyter.org/install

Packages

For the first notebooks some python packages are necessary to be installed:

- numpy
- matplotlib
- scipy
- pandas
- scikit-learn (sklearn)

What's next for today

- Help with installing Python and running Jupyter notebooks
- Quickly going through Parts 1-3 (from 42577 Intro to BA E19)
- Slowly going through Part 4

Appendix: Extra stuff

- **Tutorials**
  https://www.datacamp.com/community/tutorials
  (many nice Python, Jupyter notebooks, Pandas, and sci-kit learn tutorials)

- **Cheat sheets**
  https://www.datacamp.com/community/data-science-cheatsheets
  (I downloaded some relevant for the course)

- **Some useful materials from 42577 Introduction to Business Analytics**
  (Lectures with exercises)
  - Linear regression
  - Classification
  - Nuts and bolts of Machine Learning
  - Dimensionality reduction
  - Clustering

Appendix: Anaconda Documentation

The Jupyter Notebook App (or newer JupyterLab) can be launched by clicking on the Jupyter Notebook icon installed by Anaconda in the start menu.

To execute our notebooks follow the steps below:

- Launch the Jupyter Notebook (or JupyterLab) App
- In the Notebook Dashboard navigate to find the notebook: clicking on its name will open in a new browser tab.
- You can run the notebook document step-by-step (one cell a time) by pressing **shift + enter**.
- You can run the whole notebook in a single step by clicking on the menu **Cell → Run All**.

Modifications to the notebooks are automatically saved every few minutes.

Appendix: Installing Packages in Anaconda

The easiest way to install them is:

- Open **Anaconda Navigator**
- Environments → root
- Search for the package you want to install (e.g. Numpy), select it, and press “Apply”