## Software requirements

All of the software requirements for this course are open-source and/or free. All of the lecture notebooks will be accessible in Kaggle.

You can also download the class notebooks and run them locally on your PC. My suggestion is to install Anaconda with Python 3.* accessible here.

## Syllabus

The suggested syllabus:

1. Intro to python: review of basics
2. Data science toolkit (Pandas, Numpy, Matplotly, etc.)

→ Case Study 1

3. Statistics and hypothesis testing
4. Data acquisition:
   (a) APIs
   (b) Website scraping
5. Managing big data: SQL and Python

→ Case Study 2

6. Textual analysis:
   (a) Sentiment Analysis
   (b) Topic Modeling

→ Case Study 3

7. Machine Learning:
   (a) Concepts: bias/variance tradeoff, how good is good?
   (b) Regressions: Lasso, Ridge, Trees

## Textbook and other readings

we try to make lecture notebooks as detailed as possible so that you won't need additional books. Fortunately, Python has a great programming community that regardless of your coding issue, you can find a solution with some searches in Stack overflow!

Still, the lectures will closely follow the following resources:

- "**Data Analysis for Business, Economics, and Policy**"(Gábor Békés , Gábor Kézdi ) Cambridge University Press

- "**Python for Finance: Mastering Data-Driven Finance [2nd ed.]**" (Yves Hilpisch) - O'Reilly Media
- "**Web Scraping with Python: Collecting Data from the Modern Web**" (Ryan Mitchell) - O'Reilly Media
- "**Applied Text Analysis with Python**" (Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda) - O'Reilly Media
- "**Deep Learning with Python**" (Francois Chollet) - O'Reilly Media
- Etc.

## Academic integrity

Plagiarizing would be checked manually and using the software. Students caught cheating or plagiarizing will automatically be assigned a failed grade.