

# An Intro to Data Science in Finance

October 4, 2020

If this works we can do the same MOOC with R.

## 1 Week 1 (Setup)

### 1.1 Session 1

- Course Overview: Note most other courses use stock data and create trading algos. Here we are going to focus on core financial analyst tasks, e.g. what is the average PPP loan size in Texas?
- Python Relative to other tools.
  - Excel
  - Other Programming Languages
  - Comparison with R
  - Using Javascript for Visualization

### 1.2 Session 2

- Ways to run python
  - Google's Colab and IPython
  - Locally
    - \* The REPL
    - \* As a script file
- Python Packages
  - Extending the core language
  - Our most used packages.
  - Installing and Loading.

### 1.3 Session 3

- An Introduction to Useful Packages:
  - Pandas
  - Numpy
  - Scipy and Scikit Learn
  - Pytorch
  - Tensorflow and Keras

## 2 Week 2 (Pandas and Data Description)

Get spreadsheet type data and perform a corporate finance related task.

### 2.1 Session 1

- An intro to Pandas
- Reading csv and excel files.
  - Locally and over the web.

### 2.2 Session 2

- Example: PPP Loan Statistics by State

Alternative:

- What is the average price at which a power plant sells electricity?
- What is the NPV of a Power Plant over a Year?

### 2.3 Session 3

- Subsetting data
- Pivot Tables
- Summary Statistics

## 3 Week 3 (Numpy/Scipy)

### 3.1 Session 1

- An introduction to Application Programming Interfaces (APIs) and online data
  - Example: Coinbase API
  - Example: Interactive Brokers API
  - EIA API.
- Connecting to databases.

### 3.2 Session 2

- Example: Simple Regression, the CAPM.
- Make it better with pytorch/pyro and the Kalman filter.

### 3.3 Session 3

- Visualization:
  - Matplotlib
  - Seaborn
  - Plotly
  - ggplot (using plotnine library)

## 4 Week 4

Where to go from here. How to share code, get help, and create reproducible environments.

### 4.1 Session 1

Virtual Environments

- Ensuring compatible python and package versions.

## 4.2 Session 2

- Git/Github and IDE Integration
  - Why use version control?
    - \* Collaborate with others more easily.
    - \* Show others your work.

## 4.3 Session 3

- The Integrated Development Environment (IDE)
  - What does an IDE provide?
    - \* Code Completion
  - Git/Github and IDE Integration