**Overview**

This is a very interactive hands on course, there are 2 sets of slides that can be skimmed over.

For extra material, go to a few of your fav python/pandas/numpy web sites,

e.g.

YouTube channels

Kaggle

Stack Overflow

Investopedia

**Labs**

All Labs are in the Labs folder. This includes the datasets and instructions for each lab and a set of partially completed notebooks for the user to complete.

Instructions for the Labs are in pdf format the exercises folders and these should be viewable inside the Jupyter Lab environment – just double click on a pdf and it will open up inside the Jupyter environment.

A suggested approach is to spend around 30 to 45 mins going through each of the notebooks, explaining concepts as you go along and allow about 5 to 10 mins / Lab.

The instructor pack will have solutions, the student pack does not.

**VM Set up.**

The VMs can download timeseries data using the yfinance package.

Sample datasets are in the Data directory.

If a user opens a terminal in the VM they will be logged in at ec2-user with sudo permissions therefore users can also install any extra packages they want – though for this course its probably not needed.

**Module 1**

Python for Data Analysis

**Demo of a virtual environment**

**Emphasize that a virtual environment is simply a directory that includes a python interpreter and a let of all the non-standard python packages that are to be included in a deployment**

Contents, Jupyter notebooks, cell types, accessing online help

Variables, if/then/else, collections (lists, tuple, dictionaries and sets), looping, functions

Suggested Labs – Lab 01 – 5 mins max

**Module 2**

Data Science Using Python

Pandas – to shape data – DataFrames, import, rows, columns, indexes, selecting Data

Numpy – to model data – vector arithmetic, numpy where

Every column in a DataFrame is a numpy array

Matplotlib – to visualize Data – using the pandas wrapper module for matplotlib

Suggested Labs – Lab 02 – 10 mins max

**Module 3**

TimeSeries

Datetime indexes and slicing into a timeseries DataFrame

Date ranges and frequencies, shifting time, rolling functions, moving windows, grouping by time.

Suggested Labs – Lab 03 – 10 mins max

**Module 4**

Merging and grouping data

Concatenation

Joins

Merges

More on Grouping and Aggregation

Suggested Labs – Lab 05 – 10 mins max

**Module 5**

Introduction to Plotly

Line Plots

Time Series Plots

Range sliders

Suggested Labs – None.

**Module 6**

**Introduction to Machine Learning**

This is just a simple introduction to machine learning. Given the time constraints, I would just flick thru the slides and get as quickly as possible to the Money and Happiness section.

For this notebook emphasise the following

The sklearn web site

Naming convention for dependent and independent variables (UPPERCASE for X and lowercase for y

There are many algorithms to perform the same function, in the example here, we are using OLS and k-means to perform a regression

**Module 7 – C++ vs Python**

Just spend 5 mins on this, no need for demos.

**Appendices**

These 3 appendices are the first 2 things I teach when teaching this to Quants, so depending on the individual, they will either be extremely difficult or extremely easy.

However, if some students find the content troo easy, then tell them to work their way these 2 notebooks.

**Appendix 1**

Financial Modelling Essentials

Built in statistics in pandas

Pandas apply – with user defined functions, with lambdas

Calculating returns

Normalized pricing

Daily Returns

Log of Daily Returns

Correlation and Covariance and Heatmaps

Moving Covariances

**Appendix 2**

Plotly for Financial Analysis

The yfinance package

Institutional holders

Using calibrated rangesliders

OHLC plots

Candlesticks

Technical Indicators

Simple Moving Averages

Exponentially weighed moving average

Relative Strength Index

The MACD Signal

**Appendix 3**

A quick tour of the python type hints package.

Fairly self-explanatory.

**Extra Material**

Some cheat sheets