**Code 501 Course Desc - Python**

Overview

It's crucial for modern companies to have clean, easy-to-understand data to inform business direction and measure outcomes. Data literacy and organization allow for better decision-making, faster interpretation, and more widespread comprehension throughout an organization. Being able to collect, manage, understand, contemplate, and communicate with data will separate those who experience change to those who drive it.

In this course, you will learn how to harness the power of Python to gain highly coveted skill in data analysis and visualization. We’ll cover how to use standard packages for the organization, analysis, and visualization of data, such as Numpy, Scipy, Matplotlib, and Scikit-Learn. You’ll get to apply these skills on a daily basis and, at the end, produce a substantial project showcasing your new abilities as a data analyst.

Outcomes

At the end of this course students will:

* Be familiar with the standard data analysis tools of Python.
* Know how to visualize their data, whether processed or not, so that they can communicate its relevance to those with and without the appropriate domain knowledge.
* Understand how to clean data and prune for quality without losing depth of meaning (or at least be able to adequately explain information loss) such that any analysis using that data isn't hampered by regular or irregular artifacts.
* Be equipped to attack small to mid-size data sets with one of the most popular modern programming languages, empowering them with the agency to handle mass amount of data that the world generates daily.
* Be able to back up claims based in data with solid reporting and data-driven analyses, adding to the legitimacy of their work and the credibility of their decisions.
* Complete a deep dive and thorough analysis of a publicly-available data set. This will be done in class as a group project. It should include at least three meaningful figures describing their data and relevant to their analysis, along with the hypotheses (if there were any) that drove them to that analysis, and any insightful conclusions they may have gained from their analysis (even if it was a null result).

Prerequisites

Students enrolling in this coures should:

* Know arithmetic and basic algebra
* Have seen a mathematical matrix (even if it's not yet understood)
* Understand Git and GitHub
* Have a basic understanding of Python:
  + Built-in Python data structures and functions (list, tuple, dict, int, float, string, len, sum, min, max)
  + Performing basic mathematical operations in Python (with and without math from the standard library)
  + Writing custom functions and classes
  + Importing packages
  + Running scripts
  + Reading/writing files
  + Installing new packages

Topics

Day 1: Aggregate Analysis

* Learn to convert an unruly dataset into meaningful numbers with Descriptive Statistics
* Leverage the math superpowers of the NumPy package
* See the when, why, and how of:
  + Linear graphs
  + Scatter plots
  + Plots with multiple data sets
* Package and present the code ⇔ data story graphically with Jupyter Notebook

Day 2: Making Reliable Hypothesis

* Utilize your data within the Scientific Method: Hypothesis, model, experimentation, and conclusion
* Build a foundation for assessment with basic probability
* Construct testable data-driven hypotheses
* Engage with the data using Pandas DataFrames
* Visualize distributions with histograms, bar charts, and box plots with Matplotlib (including: what they all mean, and when to use them)
* Build better hypotheses with data patterns common in the industry

Day 3: Build Meaningful Models

* Implement the Gradient Descent Algorithm to hone in on parameters for your model
* Learn just enough matrix math to quickly perform Linear Regression
* Harness the power of the SciPy package to model data
* See how the Scikit-Learn package can do the heavy lifting for you

Day 4: Data Munging & Data Display

* Know the limits:
  + Rescale data for better decision making
  + Understand when, why, and how
  + Execute with and without Scikit-Learn
* Identify and manage problematic data
* Visualize complex multi-dimensional datasets via heatmaps in Matplotlib
* Revisit reporting with Jupyter Notebook
* Pitch and evaluate portfolio project ideas

Day 5: Build Out Your Own Data Report

* Put it all together with a portfolio-worthy data project
* Find and clean a data set of interest to you
* Create graphics to tell your data story
* Write a robust analysis and interpretation of the data and your findings
* Share your analysis with a final public Jupyter Notebook

Learn with Stacked Modules

Concepts in each of our courses are taught using stacked modules, where a new concept is introduced in each class session, building upon what came before it. This is a challenging style that requires persistence, practice, and collaboration, but allows more concepts to be introduced over the length of the course. This method helps students learn and retain more information in a short period of time. [Learn more about stacked modules »](https://www.codefellows.org/blog/how-to-accelerate-your-learning-with-stacked-modules)