**Python Data Science Course Outline**

**Course Overview**

**This course will introduce you to the field of data science and will prepare you for the next three courses in the MicroMasters: Statistics, Machine Learning, and Spark.**

**First, and foremost, you'll learn how to conduct data science by learning how to analyze data. That includes knowing how to import data, explore it, analyze it, learn from it, visualize it, and ultimately generate easily shareable reports.**

**We'll also introduce you to two powerful areas of data analysis: machine learning and natural language processing.**

**To conduct data analysis, you'll learn a collection of powerful, open-source, tools including:**

**● python**

**● jupyter notebooks**

**● pandas**

**● numpy**

**● matplotlib**

**● scikit learn**

**● nltk**

**● And many other tools**

**And you won't be learning these tools in isolation. You will learn these tools all within the context of solving compelling data science problems.**

**Learning Objectives**

**● Basic process of data science**

**● Python and Jupyter notebooks**

**● An applied understanding of how to manipulate and analyze uncurated datasets**

**● Basic statistical analysis and machine learning methods**

**● How to effectively visualize results**

**By the end of the course, you should be able to find a dataset, formulate a research question, use the tools and techniques of this course to explore the answer to that question, and share your findings.**

**Course Outline**

**The course is broken into 10 Modules. The beginning of the course is heavily focused on learning the basic tools of data science, but we firmly believe that you learn the most about data science by doing data science. So the latter half of the course is a combination of working on large projects and introductions to advanced data analysis techniques.**

**● Module 1 - Introduction: Welcome and overview of the course. Introduction to the data science process and the value of learning data science.**

**● Module 2 - Background: In this optional Module, we provide a brief background in python or unix to get you up and running. If you are already familiar with python and/or unix, feel free to skip this content.**

**● Module 3 - Jupyter and Numpy: Jupyter notebooks are one of the most commonly used tools in data science as they allow you to combine your research notes with the code for the analysis. After getting started in Jupyter, we'll learn how to use numpy for data analysis. numpy offers many useful functions for processing data as well as data structures which are time and space efficient.**

**● Module 4 - Pandas: Pandas, built on top of numpy, adds data frames which offer critical data analysis functionality and features.**

**● Module 5 - Visualization: When working with large datasets, you often need to visualize your data to gain a better understanding of it. Also, when you reach conclusions about the data, you'll often wish to use visualizations to present your results.**

**● Module 6 - Mini Project: With the tools of Jupyter notebooks, numpy, pandas, and Visualization, you're ready to do sophisticated analysis on your own. You'll pick a dataset we've worked with already and perform an analysis for this first project.**

**● Module 7 - Machine Learning: To take your data analysis skills one step further, we'll introduce you to the basics of machine learning and how to use sci-kit learn - a powerful library for machine learning.**

**● Module 8 - Working with Text and Databases: You'll find yourself often working with text data or data from databases. This Module will give you the skills to access that data. For text data, we'll also give you a preview of how to analyze text data using ideas from the field of Natural Language Processing and how to apply those ideas using the Natural Language Processing Toolkit (NLTK) library.**

**● Module 9 and 10 - Final Project: These Modules let you showcase all your new skills in an end-to-end data analysis project. You'll pick the dataset, do the data munging, ask the research questions, visualize the data, draw conclusions, and present your results.**