Syllabus
Python for Data Science

Welcome!
We are delighted to welcome you into the first course of the EdX / UC San Diego MicroMasters in Data Science: Python for Data Science. In this course, you will learn both the basics of conducting data science and how to perform data analysis in python.

Course Staff
Instructors
Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center (SDSC)
Leo Porter, Assistant Teaching Professor, Computer Science and Engineering Department

Teaching Assistants
Alok Singh, Computational Data Science Research Specialist, SDSC
Andrea Zonca, HPC Applications Specialist, SDSC

Prerequisites
This course is intended for learners who have a basic knowledge of programming in any language (Java, C, C++, Pascal, Fortran, Javascript, PHP, python, etc.). You could have learned these basic programming skills on your own or taken a course in programming in high school or college.

Your knowledge need not be extensive, but we'll assume you already know how to:

- Create an assign variables.
- Write programs with loops
- Write programs with conditions
- Author and use functions (methods)

If you are unfamiliar with python, we have an entire week (Week 2) dedicated to getting you up to speed with basic programming in python. If you find that Week 2 progresses too quickly and you need more help with basic programming, you may wish to try an introductory programming course in python before starting this course on Python for Data Science.

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 weeks. 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.

- **Week 1 - Introduction**: Welcome and overview of the course. Introduction to the data science process and the value of learning data science.
- **Week 2 - Background**: In this optional week, 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.

- **Week 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.
- **Week 4 - Pandas**: Pandas, built on top of numpy, adds data frames which offer critical data analysis functionality and features.
- **Week 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.
- **Week 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.
- **Week 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.
- **Week 8 - Working with Text and Databases**: You'll find yourself often working with text data or data from databases. This week 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.
- **Week 9 and 10 - Final Project**: These weeks 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.

Assessing and Assisting Your Learning

We know you're interested in Data Science because, well, you signed up for this course. But we also know many MOOC learners already have full lives – and that sometimes that can make it hard to "stick with" a course.

We want you to know that we have specifically designed this course to give you the excuse and the incentive to stick with us to the end! Let us tell you about that:

- **You will get points for "spending time and effort learning".** If you've taken traditional university courses, you know that "going to class" is a big motivation to stay caught up. We will try to emulate that by giving you points for "Engagement" – for units where we'll simply give you the opportunity to click a button and say "Mark as complete". We hope this will help with the satisfaction of recognizing work you put in watching videos, participating in polls, etc. as you learn. 20% of your grade comes from engaging in course material (marking learning items as complete and participating in discussions).
- **You will get points for making sure you learned the content**. We know that MOOC learners like quizzes as a manner of checking "Did I get what the professor wanted me

to?" We also know from research on learning that to really learn something you have to test yourself on it (called retrieval learning). So we've made two kinds of quizzes. You'll find "practice quizzes" (presented as polls) sprinkled throughout the course and you'll have end-of-week, Check Your Knowledge, quizzes to test your knowledge. 30% of your grade comes from six Check Your Knowledge quizzes (you can drop one).

- **You will get points for doing data science projects**. You'll hear this from us a lot - the best way to learn data science is to do data science. We have a small mini-project designed for you to engage in the data science process using the tools of this course (python, Juypyter, numpy, pandas, matplotlib). Then you'll have a large end of course final project where you'll do your own data analysis from end-to-end, possibly using knowledge from machine learning and natural language processing. These projects will give you the practice you need to be confident in your knowledge of the course material and will be outcomes you can show off to friends or colleagues. 10% of your grade comes from the Mini-Project and 20% of your grade comes from Final Project.
- **You'll get points for completing the final exam.** Exams are great for learning. If you've done the work throughout the course, you'll be in great shape to succeed on the final. So why have it? For two reasons. First, the act of studying for the final exam is good for your learning - you organize the material, you do more retrieval practice, and you solidify your understanding of concepts. Second, succeeded on the final exam gives you confidence in your own knowledge. Realize that if you've done the work up to the final, you could pass with a zero on the final. But if you're aiming for that "A", you'll want to do fairly well on the exam. 20% of your grade comes from the final exam.

The projects and final exam are necessary for those of you interested in the Verified Course Certificate.

Finally, there are a number of practice activities for you to do throughout the course. We already mentioned the practice quizzes (aka polls) sprinkled throughout the weeks, but there will also be exercise notebooks for you to work through on your own and, in the notebooks we work through in the videos, suggestions on ways to explore the dataset(s) in more depth.

Verified Learner- Earning a Certificate

To earn a verified certificate for this course, you need to be enrolled as part of the verified track, complete identity verification before you take the proctored final exam, and earn a passing grade. If you are auditing the course, you will not receive a certificate.

Grading

This course is offered for a letter grade as follows:

Grade	Percent of Available Points
A	90%-100%

B	80%-89.9%
C	70%-79.9%
F	<= 69.9%

Passing Grade: You must score 70% or above to pass the course.

Effort
We expect this course will take you 8-10 hours per week to complete.

Deadlines
This course is self-paced, so there are no deadlines for any specific assignment. This course will be open until December, 2017. Shortly after that we expect to re-open the course, self-paced, with updates.

Communication
General Discussion Forum: Please reach out to us through the discussion forum with any questions about the course content. Please allow 24 hours to receive a response from the Teaching Team. Also, all communication on the discussion forum must follow the edX Honor Code. Never post code or solutions to assignments on the discussion forum.

Honor Code - Excelling with Integrity
You are in this course because you want to learn and we want to do everything we can to help you learn. In order for that to happen, you need to do your own work and not help other people do work they should be doing.

As we tell our in-person students, focusing on course grades and course credit is short-sighted. We take courses to learn new skills and to learn new ways of thinking about the world. Yes, you get grades for your work in a course. But grades and course credit just helps open doors for you - with these grades comes the expectation you know the course material. When it comes time for you to use your knowledge and skills, you'll want the pride and personal confidence of knowing you did the work yourself.

We care a lot of about this and so should you. Please be sure you read the edX terms of service agreement. If you're unclear about what is permitted and what is not, check out the agreement and if still in doubt, just ask.

We care about your learning!
This online course is new for us and for UC San Diego. We believe we are providing you with world-class materials and resources to learn data science or we wouldn't have created the course. But, being experienced educators, we know things often go wrong in new courses. We

value your patience as we resolve issues that come up and we promise we'll do our best to resolve them both fairly and quickly.

If you encounter problems, please post on the forums and fill in the surveys which occur periodically in the course. We take your feedback seriously and will use it to improve this, and future iterations, of this course.

This is all to say: Thank you for your patience and feedback!

Self-Paced == Under Construction

Unlike in-person or synchronous classes, not everyone will be at the same place at the same time. With our in-person classes, we just take notes every term on how to make things better for students next time around. Here, we get the benefit of being able to make changes as soon as we get feedback or identify the need. This means you might go back to a prior week and find a whole new practice exercise or video there, and that's okay.

We'll try our best to not let these changes interfere with learners who have already advanced past that point in the course while improving the experience for learners who haven't gotten there yet. Again, thank you for your patience.