**Introduction to Data Science in Python**

University of Michigan

**Week 1**

In this week you'll get an introduction to the field of data science, review common Python functionality and features which data scientists use, and be introduced to the Coursera Jupyter Notebook for the lectures. All of the course information on grading, prerequisites, and expectations are on the course syllabus, and you can find more information about the Jupyter Notebooks on our Course Resources page.

**Key Concepts**

* Recognize the meaning of the term "Data Science"
* Develop basic Python programs using strings, functions, lists, dictionaries, date/time features, and files
* Use advanced Python features, including lambdas, list comprehensions and the numpy library

**Week 2**

In this week of the course you'll learn the fundamentals of one of the most important toolkits Python has for data cleaning and processing -- pandas. You'll learn how to read in data into DataFrame structures, how to query these structures, and the details about such structures are indexed. The module ends with a programming assignment and a discussion question.

**Key Concepts**

* Create Series and DataFrame Data Structures
* Use pandas math functions, as well as broadcasting features
* Employ the pandas library to import and manipulate data
* Apply indexing and querying to DataFrames, and deal with missing values

**Week 3**

In this week you'll deepen your understanding of the python pandas library by learning how to merge DataFrames, generate summary tables, group data into logical pieces, and manipulate dates. We'll also refresh your understanding of scales of data, and discuss issues with creating metrics for analysis. The week ends with a more significant programming assignment.

**Key Concepts**

* Apply merge and join on DataFrames
* Employ slicing and indexing on DataFrames
* Analyze data with groupby and understand categorical variables
* Produce the entire process of data source to elucidation
* Examine the data by manipulating, cutting, and applying aggregate functions to DataFrames

**Week 4**

In this week of the course you'll be introduced to a variety of statistical techniques such a distribution, sampling and t-tests. The majority of the week will be dedicated to your course project, where you'll engage in a real-world data cleaning activity and provide evidence for (or against!) a given hypothesis. This project is suitable for a data science portfolio, and will test your knowledge of cleaning, merging, manipulating, and test for significance in data. The week ends with two discussions of science and the rise of the fourth paradigm -- data driven discovery.

**Key Concepts**

* Identify your understanding of basic statistics
* Recognize different distributions such as binomial, uniform, normal, and chi-squared
* Interpret data to evaluate hypothesis tests

**Summary**

* Understand techniques such as lambdas and manipulating csv files
* Describe common Python functionality and features used for data science
* Query Data Frame structures for cleaning and processing
* Explain distributions, sampling, and t-tests