**Data Scientist** **specialization**:

### WHAT YOU WILL LEARN

* Describe common Python functionality and features used for data science.
* Explain distributions, sampling, and t-tests
* Query DataFrame structures for cleaning and processing
* Understand techniques such as lambdas and manipulating csv files
* Use GitHub to manage data science projects.

There are 10 courses in this program

* **Course 1: The Data Scientist’s Toolbox**
  + In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, GitHub, Python, and [Anaconda](https://www.anaconda.com/distribution/#windows).

### WHAT YOU WILL LEARN

* + - Create a Github repository
    - Explain essential study design concepts
    - Set up Python, Anaconda, Github
    - Understand the data, problems, and tools that data analysts work with
  + **Data Science Fundamentals** "3h"
    - In this module, we'll introduce and define data science and data itself. We'll also go over some of the resources that data scientists use to get help when they're stuck.
      * What is Data Science?
      * What is Data?
      * Data Science Process

## Python and Anaconda "2h"

* + - In this module, we'll help you get up and running with both Python and Anconda. Along the way, you'll learn some basics about both and why data scientists use them.
      * [Installing Anaconda](https://www.coursera.org/lecture/data-scientists-tools/installing-r-y6mU2)
      * Python Packages

## Version Control and GitHub "2h"

* + - During this module, you'll learn about version control and why it's so important to data scientists. You'll also learn how to use Git and GitHub to manage version control in data science projects.
      * [Version Control](https://www.coursera.org/lecture/data-scientists-tools/version-control-PjlHw)
      * Github and Git
      * Linking Github and Python IDE
      * Projects under Version Control

## Python Markdown, Scientific Thinking, and Big Data "5h"

* + - During this final module, you'll learn to use Python Markdown and get an introduction to three concepts that are incredibly important to every successful data scientist: asking good questions, experimental design, and big data.
      * [Python Markdown](https://www.coursera.org/lecture/data-scientists-tools/r-markdown-134kE)
      * Types of Data Science Questions
      * Experimental Design
      * Big Data
* **Course 2: Python Programming**
  + This course will introduce the learner to the basics of the python programming environment, including fundamental python programming techniques such as lambdas, reading and manipulating csv files, and the numpy library. The course will introduce data manipulation and cleaning techniques using the popular python pandas data science library and introduce the abstraction of the Series and DataFrame as the central data structures for data analysis, along with tutorials on how to use functions such as groupby, merge, and pivot tables effectively. By the end of this course, students will be able to take tabular data, clean it, manipulate it, and run basic inferential statistical analyses. This course should be taken before any of the other Applied Data Science with Python courses: Applied Plotting, Charting & Data Representation in Python, Applied Machine Learning in Python, Applied Text Mining in Python, Applied Social Network Analysis in Python.

### WHAT YOU WILL LEARN

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

## Week 1: "3h"

## In this week you'll get an introduction to the field of data science, review common Python functionality and features which data scientists use. "1h"

* + - * Python Functions
      * Python Types and Sequences
      * Python More on Strings
      * Python Demonstration: Reading and Writing CSV files
      * Python Dates and Times
      * Advanced Python Objects, map()
      * Advanced Python Lambda and List Comprehensions
      * Advanced Python Demonstration: The Numerical Python Library (NumPy)

## Week 2: "3h"

* + - 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.
      * [Introduction](https://www.coursera.org/lecture/python-data-analysis/introduction-JvRF7)
      * The Series Data Structure
      * Querying a Series
      * The DataFrame Data Structure
      * DataFrame Indexing and Loading
      * Querying a DataFrame
      * Indexing Dataframes
      * Missing Values

## Week 3: "3h"

* + - 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.
      * [Merging Dataframes](https://www.coursera.org/lecture/python-data-analysis/merging-dataframes-08sf6)
      * Pandas Idioms
      * Group by
      * Scales
      * Pivot Tables
      * Date Functionality

## Week 4: "6h"

* + - In this week of the course you'll be introduced to a variety of statistical techniques such a distributions, 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 weekends with two discussions of science and the rise of the fourth paradigm -- data driven discovery.
      * [Introduction](https://www.coursera.org/lecture/python-data-analysis/introduction-CrBP0)
      * Distributions
      * More Distributions
      * Hypothesis Testing in Python