Course Objectives:

1. Exploration – Use modern data science software (Python) to input, manipulate, summarize, and visualize data
2. Inference – Draw conclusions about populations based on sampled data
3. Prediction – Make sound predictions based on collected data

Lesson 1

Title: Intro to Data Science, Python

Objectives:

1. Introduce the course
2. Introduce Python & Jupyter notebooks
3. Set classroom expectations

Reading: None

Videos: None

Augmented Material:

Practical Application: None

Lesson 2

Title: Causality & Experiments

Objectives:

1. Define the terms *observational study*, *observation*, and *treatment*
2. Understand the difference between *association* and *causality*
3. Understand *confounding* and how it is mitigated by *randomization*

Reading: AD Chapters 1-2

Videos: edX Data8.1x; Foundations of Data Science: Computational Thinking with Python; Lectures 1.1-1.6, 2.1-2.4

Augmented Material: This lesson might require some upfront Jupyter/Python help

Practical Application: None

Lesson 3

Title: Python Programming

Objectives:

1. Understand the basics of Python, including expressions, operators and variable assignment
2. Apply basics of Python to solve basic math operations
3. Understand the basics of Tables in Python

Reading: AD Chapter 3

Videos: edX Data8.1x; Lectures 3.1-3.7, 4.1-4.5

Augmented Material: Python/Jupyter help?

Practical Application:

1. Use basic Python operators to find the number of seconds in the month of June. Find the number of seconds that have lapsed between 14 Aug, 0930 and 1 June, 1200.
2. Import a table (‘world\_population.csv’) and find:
   1. The population in 2016.
   2. The population growth between 1950 (beginning of dataset) and 2016.
   3. The annual growth rates over the course of the dataset

Lesson 4

Title: Data Types

Objectives:

1. Understand the difference between ‘number’ and ‘string’ data types
2. Apply basic string methods
3. Define and apply *comparisons*

Reading: AD Chapter 4

Videos: edX Data8.1x; Lectures 5.1-5.4

Augmented Material:

Practical Application:

1. Recreate the following sentence using variable assignment and the ‘print’ command in Python:
   1. “I am <Your name here>. I am <age> years old and I am from <hometown>”
2. Using the ‘replace’ method, replace all “I am” with “he is” or “she is”.
3. Return a boolean response indicating whether the length of your original sentence is greater than your age plus 40.

Lesson 5

Title: Arrays

Objectives:

1. Create your own array from scratch using the make\_array command.
2. Apply the various functions from the numpy package to work with arrays.
3. Understand the difference between arrays and lists

Reading: AD Chapter 5

Videos: edX Data8.1x; Lectures 5.5-5.6 (will lead into more Tables)

Augmented Material:

Practical Application:

1. Create an array containing 4 string elements: your first name, your last name, your birth state and your favorite color.
   1. Use numpy to make all elements uppercase.
   2. Use numpy to count the number of times the letter “a” appears in each element of your array
2. Create an array containing EVERY OTHER odd number starting with 1 and ending with (and including) 45.
   1. Use numpy to find the sum and the product of the elements of this array
   2. Use numpy to take the natural log of each element

Lesson 6

Title: Tables

Objectives:

1. Import tables from a file (location on computer or on web)
2. Apply the various table methods to add/remove columns, filter/sort rows based on criteria, and create new columns from existing columns

Reading: AD Chapter 6

Videos: edX Data8.1x; Lectures 5.7-5.9, 6.1-6.4

Augmented Material:

Practical Application:

1. Read in the `nba\_salaries.csv’ file and answer the following questions, using the appropriate Table methods from the datascience package:
   1. What percent of players make more than $7 million?
   2. What was the salary of Aaron Gordon (PF for the Orlando Magic)?
   3. Who was the third highest paid player on the Philadelphia 76ers?
2. Create and answer two other questions you might have about this data set.

Lesson 7

Title: Visualization

Objectives:

1. Build and interpret various visualizations of data (including line graphs and scatterplots).
2. Build, interpret and modify histograms

Reading: AD Chapter 7

Videos: edX Data8.1x; Lectures 7.1-7.8, 8.1-8.7

Augmented Material:

Practical Application:

1. Import the 2014 census data (use the code given below). Create two plots, each with two lines; one depicting the 2010 and 2014 populations by age for males and one depicting the same for females.

full\_census\_table = Table.read\_table("NC-EST2014-AGESEX-RES.csv")

partial = full\_census\_table.select(['SEX', 'AGE', 2, 8])

us\_pop = partial.relabeled(2, '2010').relabeled(3, '2014')

ratio = (us\_pop.column(3) / us\_pop.column(2))

census = us\_pop.with\_columns(

'Change', us\_pop.column(3) - us\_pop.column(2),

'Total Growth', ratio - 1,

'Annual Growth', ratio \*\* (1/5) - 1)

census.set\_format([2, 3, 4], NumberFormatter)

census.set\_format([5, 6], PercentFormatter)

1. Import the top\_movies dataset and build two histograms. One histogram should show the distribution of adjusted gross income for movies released by Fox studios and the other should show the same for Paramount studios. Interpret the two plots and compare them.

Lesson 8

Title: Functions and ‘Apply’

Objectives:

1. Understand functions in Python; write your own functions
2. Understand how to apply functions to arrays or table columns in Python

Reading: AD Chapter 8.1-8.2

Videos: edX Data8.1x; Lectures 9.1-9.6

Augmented Material:

Practical Application:

1. Q
2. Q
3. Q

Lesson 9

Title: Advanced Tables

Objectives:

1. O
2. O

Reading: AD Chapter 8.3-8.5

Videos: edX Data8.1x; Lectures 10.1-10.6, 11.1-11.4, 12.1-12.4

Augmented Material:

Practical Application:

1. Q
2. Q

Lesson 10

Title: Introduction to iteration

Objectives:

1. Understand and apply conditional statements in Python
2. Understand and apply basic iteration (for loops) in Python

Reading: AD Chapter 9

Videos: edX Data8.1x; Lectures 13.1-13.7

Augmented Material:

Practical Application:

1. Q
2. Q