University of Illinois - Deloitte Foundation Center for Business Analytics

Instructor's Guide for Data Analytics Foundations for Accountancy I

Course Description

Data Analytics Foundations for Accountancy I will introduce students to the basic concepts needed to complete common data analytic tasks in accountancy and business in general. To complete these tasks, students will learn to develop data analytic scripts by using the Python programming language and the standard data analytic Python modules, including Pandas, NumPy, SciPy, Matplotlib, and Seaborn. Specific concepts that are covered include data exploration with Pandas and NumPy, visual exploration with Matplotlib and Seaborn, and statistical analysis with NumPy, SciPy, and Seaborn. Throughout this course, the focus is on basic concepts that will have direct, practical benefit to problems in accountancy.

Course Goals and Objectives

Upon successful completion of this course, you will be able to:

- Articulate the importance of data analytics in the accountancy profession,
- Write effective data analytic scripts by using the Python programming language,
- Visually explore one- and two-dimensional data by using histograms, box plots, scatter plots, and density estimation, and
- Statistically quantify one- and two-dimensional data by using descriptive statistics, probability density functions, and correlations measures.

Course Organization

This course is organized into eight modules. Each module consists of multiple lessons, which each contain an associated video explaining the lesson content, one or more external readings, and included course Jupyter notebooks. Each module also includes a quiz (or assessment) that tests basic mastery of the lesson contents, and a programming assignment that tests synthesis of the lesson contents, especially into the broader module and course content.

Typically, a module is covered in one week, although this pace can be either faster or slower based on student needs. In addition, content can be selectively used from lessons and modules to provide supplemental learning content for an existing course. The module quiz is traditionally given through the course LMS in a timed fashion (e.g., thirty minutes). The assignments are completed on the course JupyterHub server, and are collected at the assignment deadline and processed by an autograder.

Original Content

The original content for this course consists of Jupyter notebooks, which are rendered to standard HTML webpages by github automatically.

HTML Webpage Content

The Jupyter notebook system can convert the original notebooks into a variety of different formats, including HTML, which is the standard format used to create web pages.

Special Notes

Software Versions

This course employs multiple Python modules to demonstrate important data analytics concepts. These open source software modules are undergoing continual development, and occasionally new versions of the software will cause the code in the course notebooks to throw warnings and potentially cause error conditions that might prevent a notebook from working. The notebooks in this course release have been tested with the following versions of Python and associated Python library (note, there is a special notebook called test-version.ipynb in the modified directory that will display your installed software versions for comparison).

<pre># Display Python version import sys print(sys.version)</pre>	
3.7.6 (default, Jan [Clang 4.0.1 (tags/I	8 2020, 13:42:34) RELEASE_401/final)]
# Display basic Pyt	non library versions
import numpy as np	
import pandas as pd	
import seaborn as s	ıs
import matplotlib as	s mpl
<pre>print(f'Numpy version)</pre>	on = {npversion}')
<pre>print(f'Pandas version = {pdversion}')</pre>	
print(f'Matplotlib	version = {snsversion}')
<pre>print(f'Seaborn vers</pre>	sion = {mplversion}')
Numpy version	= 1.18.1
Pandas version	
Matplotlib version	= 0.10.0
Seaborn version	

Course Contents

Module 1: Foundations

This module serves as the introduction to the course content and the course Jupyter server, where you will run your analytics scripts. First, you will read about specific examples of how analytics is being employed by accounting firms. Next, you will learn about the capabilities of the course Jupyter server, and how to create, edit, and run notebooks on the course server. After this, you will learn how to write Markdownformatted documents, which is an easy way to quickly write formatted text, including descriptive text inside a course notebook.

Learning Objectives

By the end of this module, you should be able to:

- Appreciate the importance of data analytics in accountancy,
- Work effectively with a Jupyter notebook on the course server, and
- Write basic Markdown-encoded documents.

Module 2: Introduction to Python

This module focuses on the basic features in the Python programming language that underlie most data analytics programs (or scripts). First, you will read about why accounting students should learn to write computer programs. In the first lesson, you will also learn the basic concepts of the Python programming language, including how to create variables, basic data types and mathematical operators, and how to document your programs with comments. Next, you will learn about functions and how they can simplify developing and maintaining computer programs. You will also learn how to create and call functions in Python. Finally, you will learn about Boolean and logical operators in Python and how they can be used to control the flow of a Python program by using conditional statements.

Learning Objectives

By the end of this module, you should be able to:

- Articulate how accounting students can benefit from learning to write computer programs,
- Write simple Python scripts,
- Effectively create and use variables of different data types in a Python program, and
- Apply Boolean and logical operators within conditional statements in a Python program.

Module 3: Introduction to Python Programming

This module introduces two important programming concepts: data structures and iteration. Mastering data structures is an important task because a challenging problem can often be made much simpler by using data structures. Python provides several built-in data structures including the list, tuple, string, and dictionary. With such a rich set of data structures, iteration becomes an important concept as we can rapidly process data by iterating through the items in a data structure. This process is generally done by using the "for" statement, but this module also introduces the "while" statement and comprehensions. Finally, we will review a number of functions, which are built into these data structures, that increase the efficacy and general applicability of these data structures.

Module 4: Python Programming

This module introduces fundamental concepts in Python programming. First, you will learn about several advanced concepts relating to Python functions. These concepts, such as treating functions as objects, lambda functions, and variable-length function arguments, are frequently used in Python to analyze data. Second, you will learn about writing Python programs, which includes concepts drawn from the field of software engineering, as well as Python-specific concepts like modules and exception handling. Finally, you will learn about the NumPy module, which provides support for fast numerical operations within Python. This module will focus on using NumPy with one-dimensional data (i.e., vectors or 1-D arrays), but a later module will explore using NumPy for higher-dimensional data.

Learning Objectives

By the end of this module, you should be able to:

- Create and use the Python list, tuple, string, and dictionary data structures,
- Explain the difference between mutable and immutable data structures and why both types are important,
- Effectively use iteration to process data in these data structures, and
- Apply built-in functions to manipulate data in these data structures.

Learning Objectives

By the end of this module, you should be able to:

- Apply advanced function concepts in Python, like lambda functions and variable-length function arguments,
- · Use different Python modules effectively,
- Use Python's exception-handling mechanism, and
- Work with one-dimensional numerical data by using the NumPy module.

Module 5: Introduction to Data Persistence

This module introduces how to persist data in different ways by using Python. First, you will learn about the Unix file system, which is the operating system used for most big data processing (as well as Linux and Mac OSX desktops and many mobile phones). Second, you will learn how to read and write data to a file from within a Python program. Finally, you will learn about persisting more complex data by using the pickle module and how diverse text-based data formats such as CSV, JSON, and XML can be used to persist data.

Module 6: Introduction to Data Analysis

This module introduces fundamental concepts in data analysis. First, you will learn about the Pandas Python module that can simplify many challenging data analysis tasks. The Pandas module includes the DataFrame, which programmatically mimics many of the features of a traditional spreadsheet. Second, you will learn about advanced functionality within the Pandas module including reading, writing, masking, grouping, stacking, and sorting. Finally, you will learn about descriptive statistics, which can be used to characterize a data set by using a few specific measurements.

Learning Objectives

By the end of this module, you should be able to:

- Work with the Unix file system and understand the concept of directories and files,
- Read and write data to files from within a Python script,
- · Pickle python data to and from a file, and
- Read and write data to standard text-based file formats like CSV, JSON, and XML.

Learning Objectives

By the end of this module, you should be able to:

- Explain the benefits of using the Pandas module for data analysis,
- Apply advanced features to more effectively analyze data by using a Pandas DataFrame, and
- · Compute and interpret descriptive statistics.

Module 7: Introduction to Visualization

This module introduces visualization as an important tool for exploring and understanding data. First, the basic components of visualizations are introduced with an emphasis on how they can be used to convey information. Also, you will learn how to identify and avoid ways that a visualization can mislead or confuse a viewer. Next, you will learn how to actually create a simple visualization (basic line plot) in Python, which will introduce creating and displaying a visualization within a notebook, how to annotate a plot, and how to improve the visual aesthetics of a plot by using the seaborn module. Finally, you will learn how to explore a one-dimensional data set by using rug plots, box plots, and histograms.

Module 8: Exploring Two-Dimensional Data

This module extends what you have learned in previous modules to the visual exploration of two-dimensional data. First, you will learn more about conveying information to a user visually, specifically the importance of form, color, and location. Finally, you will learn how to make twodimensional scatter plots in Python and how they can be used to graphically identify a correlation and outlier points.

Learning Objectives

By the end of this module, you should be able to:

- Understand the basic concepts in conveying information visually,
- Identify and avoid visualization techniques that can mislead viewers,
- Create basic visualizations by using Python and the matplotlib module, and
- Visually explore one-dimensional data by using Python and the seaborn module.

Learning Objectives

By the end of this module, you should be able to:

- Explain how different plotting components can be used to convey information visually, and
- Display and interpret two-dimensional data by using Python.

© 2018-2019: Robert J. Brunner at the University of Illinois. This notebook is released under the Creative Commons license CC BY-NC-SA 4.0. Any reproduction, adaptation, distribution, dissemination or making available of this notebook for commercial use is not allowed unless authorized in writing by the copyright holder.

ILLINOIS Gies College of Business