Week Eight Reflection Journal

Michael Surdek

Southern New Hampshire University

IT 697: Python Experiential Learning Activity

Dr. Bhanu Kapoor

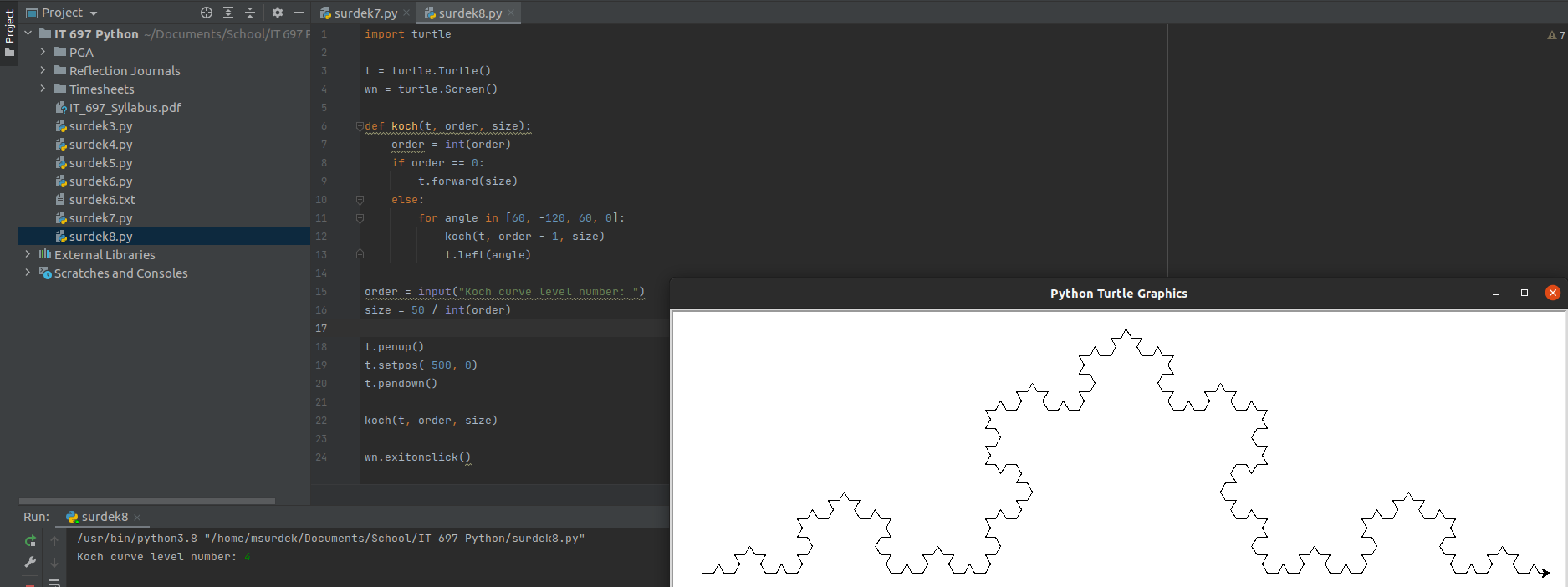
July 4, 2021

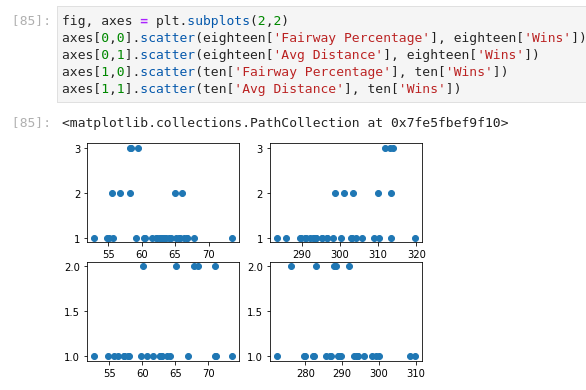
This week was mostly spent getting myself into a solid position to wrap up this experiential learning activity. I have learned a lot about python throughout this course, but with the executive brief due next Sunday, I did not yet have a piece of work that could demonstrate my experience in relation to my program competencies. Along with my usual weekly activities consisting of our textbook and assignment script, the discussion board, and my Python for Data Analysis book, I also started working on a data visualization project that I have been planning to include in the executive brief. With my current position, I should have no problem completing my project, the executive brief, and any other weekly activities in module nine.

**Learning Experiences This Week**

Although I knew I had a lot of activities to complete this week, I was unable to dedicate any time on Monday or Tuesday to this experience, so I was in a challenging spot heading into Wednesday when I began reading section 9.4 of our Introduction to Scripting textbook and chapter 9 of my Python for Data Analysis, which took parts of three days to get through. Chapter 9 of Python for Data Analysis covered data aggregation and group operations. I was familiar with many of the concepts of grouping from my experience in R and SQL, but I found out that it works a little differently in python by creating a group by object and then pulling information from the object with various arguments and methods. I was also introduced to other data aggregation strategies such as transformations and apply methods.

Section 9.4 of the Introduction to Scripting textbook continued with recursive functions and demonstrated concepts such as the Koch Curve and L-Systems. All of these concepts are brand new to me and although I am starting to understand them a little more. Our assignment script was to create a program that draws the Koch Curve based on a given level number from the user. I needed some assistance to complete the script, and I found an online walkthrough of this exact problem. I did not copy the code from the walkthrough, but I very much based my own code on their solution, since the problem was identical. I came across many errors in my attempts, but I was finally able to get the program to draw a visible Koch Curve in turtle for any given level, although levels 0-4 worked the best and anything larger became a little too big for my turtle graphics window. Here is my program in the PyCharm IDE and its output for a level 4 Koch Curve:

 My last and most important activity of this week was the beginning stages of my data visualization project. Prior to this week, I had read chapter 8 of Python for Data Analysis and watched two courses on Pluralsight, all of which covered data visualization in Python with matplotlib. To get the actual project going, I knew that I needed to find a dataset that interested me, import it into python, explore it, create one or more plots, and format them into a visualization that communicates my findings. Yesterday, I looked through places where I might find datasets, including github and Twitter, and I settled on a dataset of player statistics in PGA Tour seasons from 2010-2018. Once I committed to this dataset, I downloaded it and imported it into a data frame using a Jupyer notebook. Today I began creating scatterplots from the data, analyzing variables such as wins, top 10s, driving accuracy and distance, greens in regulation, and others. From my outside knowledge, I knew that a longer driving distance leads to better results in professional golf today, but I wanted to see if that was the case 10 years ago as well. I created scatterplots that show the relationships between wins and both driving distance and driving accuracy in 2018 and 2010. The data from 2018 shows that a longer driving distance leads to more wins, whereas a higher driving accuracy does not. The data from 2010 interestingly shows the opposite. The process of visualizing this data provides evidence to the common hypothesis that driving distance is becoming more important on the PGA Tour, regardless of driving accuracy. In my current plots, driving accuracy percentage vs wins is on the left, driving distance in yards vs wins is on the right, 2018 is on top, and 2010 is on bottom.



**Record of Project/Work Ideas and Their Current Status**

* Guessing game program
  + Starting next module
* Learn Python The Hard Way
  + 18/53 Python exercises
  + 15/15 Command Line Crash Course exercises
* Visualization with real data set in python
  + Completed 2/2 Pluralsight courses on data visualization
  + Read & followed code of Python for Data Visualization Ch 8
  + Chose a topic/dataset
  + Created 4 subplots that show comparison I plan to demonstrate
  + Still need to format subplots and overall plot into final visualization

References

McKinney, William-Wes. 2012. Python for Data Analysis. *O’Reilly*.

Miller, Bradley and Ranum, David. 2017. Introduction to Scripting. *Jones & Bartlett Learning.*

Shaw, Zed A. 2014. Learn Python the Hard Way. *Addison-Wesley*.