Joe Donato

May 6th, 2020

MCIS 540 – Programming for Data Science

Final Project

**Introduction/Purpose**:

As this course has progressed throughout the weeks, we have worked toward becoming more comfortable with using Python and the analysis behind this powerful tool. The goal of this final project was to use the skills that we have been practicing since the beginning of this course to preform data analysis to solve actual real-world problems. By using the US census data for the state of Pennsylvania, I was able to dive in and get to the bottom of the various questions asked for this project along with some other analysis.

**Program/Function descriptions**:

This program asked to analyze this data based on three components, find the county with the highest percentage of children in poverty, the lowest percentage of children in poverty, and to find a county, poverty rate, population of children in poverty and the median household income. The first step to this project was to import the data from the US census website. I did this by downloading the data to Excel and importing to Python. Once imported I ran the descriptive statistics to determine what I was looking at on a whole. The descriptive statistics provided me with the mean, standard deviation, min and max of each county based of three different criteria, Poverty Estimate, Poverty Percentage and Median Household Income. I also set the index of this Datafame to the County Name column to represent the unique identifier.

From there, I worked on finding the counties that had the minimum and maximum poverty percentage. I was able to do this using the Index Min and Index Max calls in the Pandas library. Based on these calls, I was able to determine that Chester County had the lowest percent of children in poverty while Forest County had the highest percent of children in poverty. I also made functions for each Index Min/Max call if I needed to recall the data at a later point.

Once I had this information, I created a loop to continually prompt for a county and print its name, percentage, count of children in poverty and the median household income regardless of if the entered value was case sensitive and only the county name is used. To do this I first created a loop with case sensitivity and where a full county name needed to be provided in order to have my loop in place. Once that was established, I created a second loop to look up a county without case sensitivity. This loop only provided me with the county name so I knew that my task was not completed. Since I had two loops I combined the both of them to create what we were looking for. At first glance I knew my continuous loop worked with the county lookup and case insensitivity loops but realized that I had a redundant if statement in the loop from my original While loop. The final product that I came to was a continuous While loop exiting only if “quit” was entered and a For loop to call the county when entered. I also created a separate function for this loop to be used if need be. Lastly, I was able to create a function from there that will give the Min/Max poverty rate and all of the county information so that these functions can be called all at once.

I was still curious about the information and was wondering what this data set looked like visually. I imported the MatPlotLib package so that I can plot a scatter chart. With the X-axis being the median household income and the Y-axis being the child poverty rate I was able to determine that the lower the median household income the higher the poverty rate for counties in Pennsylvania.

**Program Output/Data Analysis**:











