**Directions**

* Submit a word document that contains your codes and outputs. Include comments or brief sentences explaining what your codes are addressing where necessary. Also submit the Python file. Your file names should follow this format: LastName\_FirstName\_Final\_Project
* Be efficient with your code as much as possible. For example, do not use five lines of code with a loop if you could efficiently solve the problem with a single line of code. However, if you can use multiple lines of code and that is how you think you can best solve the problem, that is fine.

**Rubrics**

* You are required to extract data from an online source and use it for this project. Do not download the data into your computer. Use the url or weblink to read the data into Python.
* Your data should have at least two columns with categorical variables and two columns with quantitative or numerical variables. Check for missing data by finding out how many missing values are in each column. Prepare and clean your data for analysis as you see necessary.
* Do descriptive statistics on your data. Implement. agg() or apply() function on your data with statistics of your choice (at least, 5 statistics not found in the output of the .describe() function).
* Compute correlation and covariance. You may select only columns that make sense for these computations. What do these values tell you about your data?
* Create an appropriate function of your choice and use .apply () to apply the function to a suitable portion of your data. You can store the results in another dataframe.
* Call the pandas crosstab() function on your data in a way you find suitable. You can store the results in another dataframe. Also use a pivot table to reshape your data in a meaningful way. Store this reshaped data in another dataframe so you can continue to use your original data for subsequent analysis. Choose a value on your pivot table or crosstab dataframe and interpret what that value means.
* Select a certain portion of your data (at least two rows) using a loop and store these rows into a different dataframe. You may include a user input function in your code for this part.
* Create a line plot, scatter plot, bar charts, histogram, and boxplot for your data with suitable variables of your choice. Your plots should be well labelled and formatted. You should also interpret and tell a brief story from each plot.
* Chose a dependent and an independent variable from your data and run a regression analysis with these variables. Make sure you evaluate your regression model with suitable residual plots, etc.
* Find another suitable dataset online and use a hypothesis testing to compare two levels of a categorical variable in relation to a continuous variable in that data. Report your results.
* Find a table embedded in a webpage and use BeautifulSoup to extract that table into Python
* You can include more analysis to this project as you desire

**May God bless your future endeavors, and I hope you apply the skills learned in this course in your future career or projects.**