

# Assignment 1: Python basics

Econ 294A

April 17, 2023

## Important

- **Deadline of submission:** April 30, 2023
- Use any IDE for the coding. Upload Jupyter code file or .py file on Canvas. Use comment option or markdown inside Python code for the answers. Embed your plots inside Jupyter notebook.
- Grading will be done based on your attempt. Show your attempts in details

**Question 1:** Use "Restated\_Bilateral\_External\_Portfolios.xlsx" for this question.

1. Load the data as pandas data frame in Python. Create a new data frame where Investor name is United States. Provide the summary statistics of (i) Position\_Residency (ii) Restatement\_TH\_Only (iii) Restatement\_Full. Report the summary in the Jupyter file
2. Tabulate the frequency distribution of Asset class from the filtered data. Embed the output in file
3. Plot "Position Residency" and "Restatement Full" using matplotlib. Embed the plot inside the Jupyter file
4. Create a scatter plot of "Position Residency" and "Restatement Full" using Seaborn package in python
5. Estimate the regression equation between "Restatement Full" and "Position Residency". Report the regression summary. Use any package for fitting regression line.
6. Take a sample from the filtered data with replacement and then run the same regression. Report the regression fit and coefficient estimates. Insert a markdown to comment on the regression fit from full sample and sub-sample.

**Question 2:** Use the Players.csv data for this question

1. Load the data in Python as Pandas data frame. Now create a frequency table of players, the frequency represents the number of years that player played.
2. Create a function in Python so that you can input any player name and the function returns players who played during same years. For example, if player 'A' was active during 2000-2008, then the function should identify those players who were active during the same time. The output should be pandas data frame with player name.

**Question 3:** Use TFP\_Temperature.xlsx for this question

1. Load the data as Pandas data frame. Provide summary statistics of TFP Index, Land Index, Capital Index and Temp Anomalies
2. Create a correlation plot between these variables using Seaborn package
3. Run a quantile regression between TFP Index and Temp Anomalies. Create a train-test split on the data, then fit the regression on training data and check the forecast accuracy on test data.

**End**