

# ECON 0150 | Fall 2025 | Homework 2.1

---

*Due: Friday, September 26, at 5PM*

Homework is designed to both test your knowledge and challenge you to apply familiar concepts in new applications. Answer clearly and completely. You are welcomed and encouraged to work in groups so long as your work is your own. Submit your figures and answers to Gradescope.

## Q1. Handling Missing Values

In the dataset 'world\_in\_data.csv' the Employment in agriculture (% of total employment) variable has many missing values. We want a dataset with these missing values removed. After removing the countries with no records in the Employment in agriculture (% of total employment) column, how many countries remain in your dataset? (Hint: Use the `.dropna()` function on the dataset to remove *Nan* values.)

## Q2. Handling Dates and Times

In the dataset `Monthly_Coffee_Prices.csv` the date variable isn't in the right format for our purposes. Construct a year variable to create a boxplot (code in Notebook). (Hint: Use `pd.to_datetime()` to convert to a date and `.dt.year` to extract the year.)

## Q3. Handling Strings and Numerical Values

In the dataset `Starbucks_Location_Hours_raw.csv`, use the `olsonTimeZoneId` variable to construct a new variable GMT, which contains the GMT time (time of day when it is midnight in Greenwich). Plot this new variable GMT as a histogram (code in Notebook). (Hint: Use `.str.replace('GMT', '')` to remove text from the start of the strings and `.to_numeric()` to turn the text into numbers.)