# **Final Project Assignment (30 points)**

This final project draws upon the material covered throughout the course. I recommend working backward, to some extent. In other words, I recommend working through your complete analysis in Python and then responding to assignment questions.

These directions are stylized to be formatted into a Word document. You should generate a Jupyter notebook that includes the answers the following questions. When answering questions that require operations in Python, include the Python code.

## **Part I: Background Information**

1. (1 points) State the research question: What is the relationship between \_\_\_\_\_ and \_\_\_\_\_\_\_?
2. (1 points) Write a short paragraph that provides some context for the research question. Why is this an interesting relationship to study? Why do you expect there to be a relationship?
3. (1 points) Cite the dataset you’ll be using:
4. (1 points) Complete Table 1. Only include those variables you’ll be using in your analysis. Adjust the number of rows as necessary.

### **Table 1: Variable Definitions**

| **Variable** | **Definition** |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## **Part 2: Descriptive Information**

1. (2 points) Complete the table of summary statistics. Add or delete rows as necessary.

### **Table 2: Summary Statistics**

| **Variable** | **Minimum** | **Maximum** | **Mean** | **Standard Deviation** | ***N*** |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

1. (2 points) Create one visualization of the dependent variable (e.g. bar graph, line graph, boxplot, histogram). Be sure the visualization includes all the critical components, including a title, meaningful axes labels and any other useful labels.
2. (1 points) Write a one-sentence summary of the visualization above. What is something interesting about the variable that the visualization highlights?
3. (2 points) Create one visualization of the independent variable of interest (e.g. bar graph, line graph, histogram). Be sure the visualization includes all the critical components, including a title, meaningful axes labels and any other useful labels. If you have more than one independent variable of interest, select one for this question.
4. (1 points) Write a one-sentence summary of the visualization above. What is something interesting about the variable that the visualization highlights?
5. (2 points) Create one bivariate relationship that shows the relationship between the two variables used above (e.g. scatter plot, side-by-side bar plot, line graph).
6. (1 points) Write a one-sentence summary of the visualization above. What is something interesting about the relationship that the visualization highlights?

**Part 3: Regression Analysis**

1. (5 points) Run a bivariate regression model that includes your dependent variable and one independent variable of interest.

Run a multivariate regression model that includes your dependent variable, one or more independent variables of interest and at least one control variable.

Complete Table 3 with your regression results. List the independent variables (using meaningful variable names) in the first column. Include your bivariate model results in the second column and the multivariate model results in the third column. *For each variable, list the estimated coefficient and the estimated standard error beneath in parentheses. Include only two digits beyond the decimal point.* Add or delete rows as necessary.

**Table 3: Regression Results**

**Dependent variable:**

| **Independent Variable** | **Bivariate Model** | **Multivariate Model** |
| --- | --- | --- |
| *Example row: variable name* | *(standard error)* | *(standard error)* |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| *N* |  |  |

1. (3 points) Explain your justification for one of the control variables that you chose to include in the multivariate model.
2. (2 points) Interpret the coefficient on the independent variable of interest in the bivariate model.
3. (2 points) Interpret the coefficient on the independent variable of interest in the multivariate model.
4. (2 points) Conduct a hypothesis test on the independent variable of interest in the multivariate model. State the null hypothesis, alternative hypothesis, *t*-score and result of the test.

* *H0*:
* *HA*:
* *t*-score:
* Result of the test:

1. (1 points) In a few sentences, describe your overall findings from your analysis.