**Scenario**

Smart businesses in all industries use data to provide an intuitive analysis of how they can get a competitive advantage. The real estate industry heavily uses linear regression to estimate home prices, as cost of housing is currently the largest expense for most families. Additionally, in order to help new homeowners and home sellers with important decisions, real estate professionals need to go beyond showing property inventory. They need to be well versed in the relationship between price, square footage, build year, location, and so many other factors that can help predict the business environment and provide the best advice to their clients.

**Prompt**

You have been recently hired as a junior analyst by D.M. Pan Real Estate Company. The sales team has tasked you with preparing a report that examines the relationship between the selling price of properties and their size in square feet. You have been provided with a [Real Estate Data](https://snhu.brightspace.com/content/enforced/955893-MAT-240-H3911-OL-TRAD-UG.22EW3/course_documents/MAT%20240%20Real%20Estate%20Data.xlsx?_&d2lSessionVal=3aTp3JApTcvuhjcWMBNnGJSVC&ou=955893) spreadsheet that includes properties sold nationwide in recent years. The team has asked you to select a region, complete an initial analysis, and provide the report to the team.

**Note**: In the report you prepare for the sales team, the response variable (y) should be the listing price and the predictor variable (x) should be the square feet.

Specifically you must address the following rubric criteria, using the [Module Two Assignment Template](https://snhu.brightspace.com/content/enforced/955893-MAT-240-H3911-OL-TRAD-UG.22EW3/course_documents/MAT%20240%20Module%20Two%20Assignment%20Template.docx?_&d2lSessionVal=3aTp3JApTcvuhjcWMBNnGJSVC&ou=955893):

* **Generate a Representative Sample of the Data**
  + Select a region and generate a simple random sample of 30 from the data.
  + Report the mean, median, and standard deviation of the listing price and the square foot variables.
* **Analyze Your Sample**
  + Discuss how the regional sample created is or is not reflective of the national market.
    - Compare and contrast your sample with the population using the [National Statistics and Graphs](https://snhu.brightspace.com/content/enforced/955893-MAT-240-H3911-OL-TRAD-UG.22EW3/course_documents/MAT%20240%20National%20Statistics%20and%20Graphs.pdf?_&d2lSessionVal=3aTp3JApTcvuhjcWMBNnGJSVC&ou=955893) document.
  + Explain how you have made sure that the sample is random.
    - Explain your methods to get a truly random sample.
* **Generate Scatterplot**
  + Create a scatterplot of the *x* and *y* variables noted above and include a trend line and the regression equation
* **Observe patterns**
  + Answer the following questions based on the scatterplot:
    - Define *x* and *y*. Which variable is useful for making predictions?
    - Is there an association between *x* and *y*? Describe the association you see in the scatter plot.
    - What do you see as the shape (linear or nonlinear)?
    - If you had a 1,800 square foot house, based on the regression equation in the graph, what price would you choose to list at?
    - Do you see any potential outliers in the scatterplot?
      * Why do you think the outliers appeared in the scatterplot you generated?
      * What do they represent?