**Tips for Week 7 Assignment**

Use the exact structure that you find in the Week 7 tutorial “simple\_linear\_regression\_for\_salary\_codefile”

Read in the correct file for each of the two “problems” presented in the instructions (you will create two separate notebooks.

Decide on whether you want to use sns or matplotlib for the plotting and stick with that

Import hint: because the kc housing data includes prices over 1 million, you may find that your plot only shows integers for the y axis (with a small 1e6 indicator in the upper left corner) – meaning that the numbers are in exponential notation format. Include this line just in front of EVERY plt.show() line in your code:

plt.ticklabel\_format(axis="y", style="plain"). This should make your y-axis labels look much like the data.

Be sure to include comments that explain your code. Much of what you can use to explain linear regression, train\_test\_split, etc. can be found in the readings for the week. You can also Google to find explanations that make sense to you and would be easy for a non-specialist to understand. I will be looking for these comments in your notebooks.

Perhaps your most important comment will relate to the r-squared value and what it means for your results.

Answer all the questions posed in each of the two exercises. Follow the instructions for submission of your assignment.

What is expected in your Week 7 assignment. Getting the plots correctly created (step 4) was a matter of just following the structure of the tutorial. But explaining the steps and analyzing the results was the more challenging part.   
  
For each of the 2 problems you needed to **comment input and output of each code cell or cells that are associated with the 4 numbered elements**

1. **model building/linear regression process**
2. **Model Results:** Print the intercept and coefficient of the line.
3. **Model Accuracy:** Print accuracy metrics from scikit learn, including r-squared\
4. **Visualizations:** Create a least 3 visualizations

You are creating a model that will predict the outcome, given a target dependent variable and one or more independent variables. To build the model, you divide the dataset into two parts (train/test) and run a simple linear regression to enable you to predict an outcome. For LR-1, 80% of the data was used to create a model, 20% for testing it. What you were predicting in LR-1 was the cost of a house based solely on square footage. In LR-2 you were predicting a final score based on the number of hours spent studying.