Here you are!

I hoped you enjoyed playing Data Scientist at a Real Estate Investment Trust. Well done!

This rubric will provide you with a grade breakdown for the evaluation of the final project of your peers.

This project is worth 13% of your final grade.

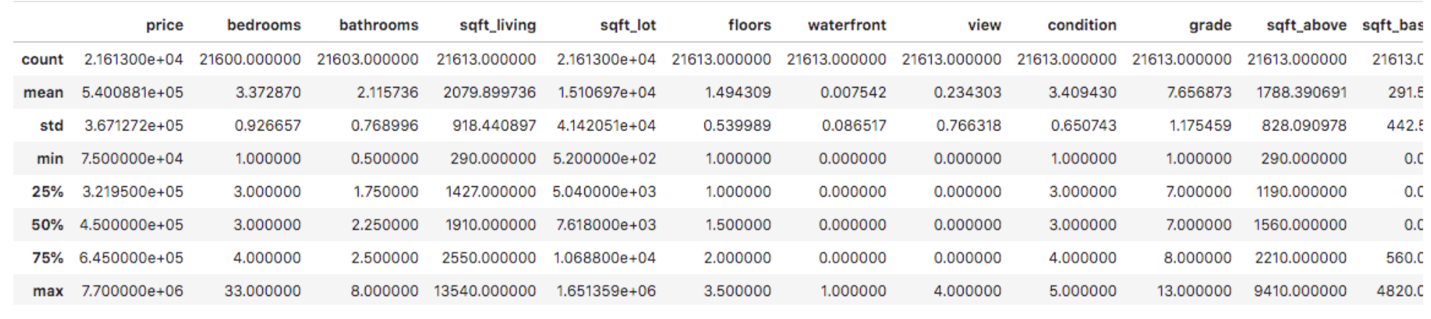
**Review criteria**

There are ten questions. Simple questions are worth one mark and harder questions are worth two marks. For each item, you should take a screen shot of the output of each cell with the code that generated it. The provided code must be run. Most of the questions are independent i.e. if you miss one you can still do the rest of the problems.

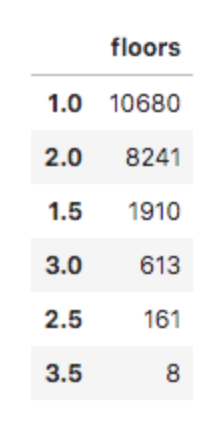
In the last prompt, you will also be asked to upload your Jupyter Notebook that you have created. You will receive marks for displaying your notebook. You can paste the individual screenshots of the solutions from your Jupyter Notebook as Your Response to the Questions. Your peers will review your Jupyter Notebook and evaluate Your Responses to each Question in the Assignment.

### Question 1) Display the data types of each column using the attribute dtypes, then take a screenshot and submit it, include your code in the image.

Question 2) Drop the columns "id" and "Unnamed: 0" from axis 1 using the method drop(), then use the method describe() to obtain a statistical summary of the data. Take a screenshot and submit it, make sure the inplace parameter is set to True. Your output should look like this:

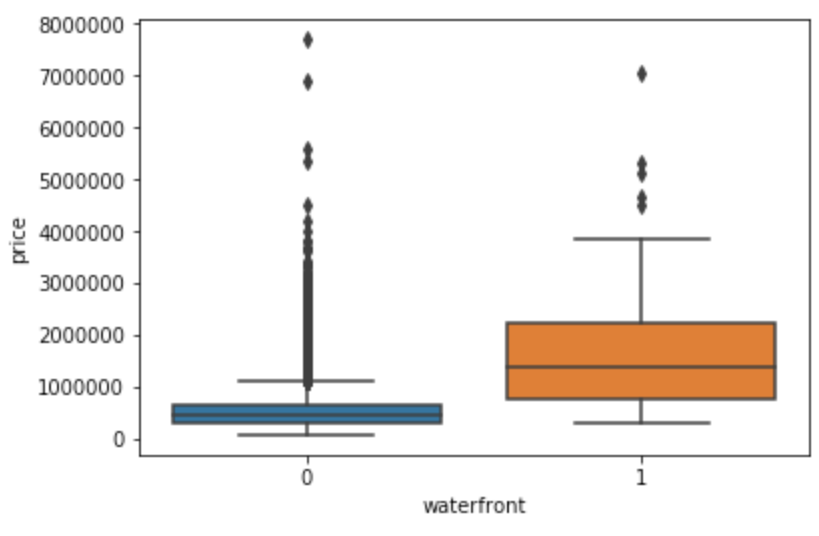


### Question 3) use the method **value\_counts** to count the number of houses with unique floor values, use the method .to\_frame() to convert it to a dataframe. Your output should look like this :



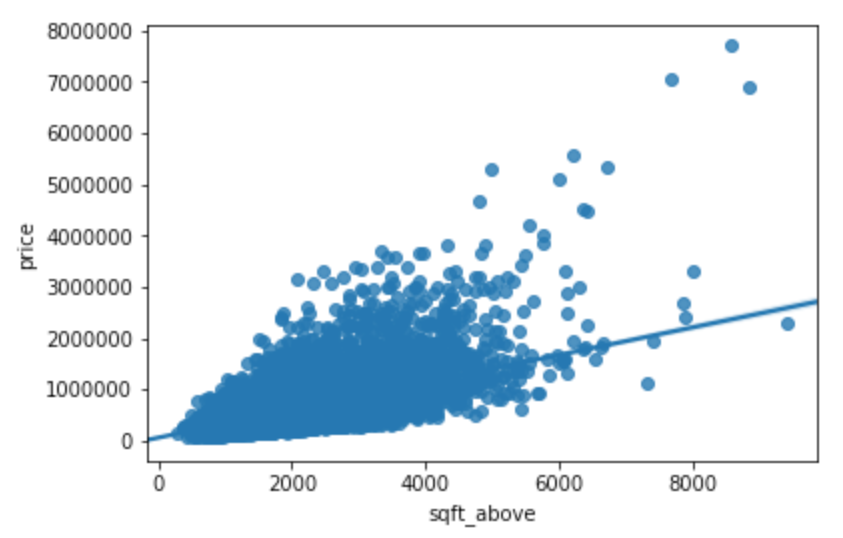
Y

### Question 4) use the function boxplot in the seaborn library to produce a plot that can be used to determine whether houses with a waterfront view or without a waterfront view have more price outliers. Your output should look like this with the code that produced it (the colors may be different ) :



### Question 5) Use the function regplot in the seaborn library to determine if the feature sqft\_above is negatively or positively correlated with price. Take a screenshot of the plot and the code used to generate it.

### Your output should look like this with the code that produced it :



Question 6) Fit a linear regression model to predict the price using the feature 'sqft\_living' then calculate the R^2. Take a screenshot of your code and the value of the R^2.

Question 7) Fit a linear regression model to predict the 'price' using the list of features:

* "floors"
* "waterfront"
* "lat"
* "bedrooms"
* "sqft\_basement"
* "view"
* "bathrooms"
* "sqft\_living15"
* "sqft\_above"
* "grade"
* "sqft\_living"

The calculate the R^2. Take a screenshot of your code and the value of the R^2.

Question 8) Create a pipeline object that scales the data performs a polynomial transform and fits a linear regression model. Fit the object using the features in the question above, then fit the model and calculate the R^2. Take a screenshot of your code and the R^2.

There are some hints in the notebook

Question 9) Create and fit a Ridge regression object using the training data, setting the regularization parameter to 0.1 and calculate the R^2 using the test data. Take a screenshot for your code and the R^2

Question 10) Perform a second order polynomial transform on both the training data and testing data. Create and fit a Ridge regression object using the training data, setting the regularisation parameter to 0.1. Calculate the R^2 utilising the test data provided. Take a screenshot of your code and the R^2.

Upload your notebook.