CS 7310

Final Exam Assignment

Create a Jupyter notebook named: <last_name>.final.ipynb

Create a cell with:
Name:
Student ID:

Organize the notebook to create a Linear Regression Model to predict the price per square foot for the following housing data set: **housing_data.csv** found in 7310.final.zip

The data appears as follows:

:

	No	X1 date	X2 home age	X3 dist to station	X4 num nearby stores	X5 latitude	X6 longitude	Y price per sqft	condition
0	1	2012.917	32.0	84.87882	10	24.98298	121.54024	37.9	Average
1	2	2012.917	19.5	306.59470	9	24.98034	121.53951	42.2	Average
2	3	2013.583	13.3	561.98450	5	24.98746	121.54391	47.3	Average
3	4	2013.500	13.3	561.98450	5	24.98746	121.54391	54.8	Poor
4	5	2012.833	5.0	390.56840	5	24.97937	121.54245	43.1	Excellent

Your task is to:

A. Prepare the data for machine learning
    a. Remove the column 'No' with transaction sequence numbers since they do not contribute to home prices
    b. Convert the column 'condition' to appropriate numeric values since ML cannot work with text values, only numbers
B. Divide the data into training and testing data sets
    a. Display the first five lines of the training set
    b. Display the first five lines of the testing set
C. Create a regression model based on the training data to predict' Y price per sqft'
D. Test the model with the testing data set and compute the accuracy of the predictions
    a. Display the accuracy value for both the test and training data sets

Submit: A zip file that contains your notebook. Name the zip file: <last_name>.<smu_id>.zip

Note: do not put the angle brackets in the zip file name