Machine Learning and Fintech Hw 6. Due 2019/10/26

**Download the “House Prices Dataset.csv” in new E3. Go over MLfinTech\_Regression.ipynb to get more information on the data.**

Descriptions on each variable:

(1) OverallCond: Rates the overall condition of the house

10 Very Excellent

9 Excellent

8 Very Good

7 Good

6 Above Average

5 Average

4 Below Average

3 Fair

2 Poor

1 Very Poor

(2) TotalBsmtSF: Total square feet of basement area

(3) 1stFlrSF: First Floor square feet

(4) GrLivArea: Above grade (ground) living area square feet

(5) FullBath: Full bathrooms above grade

(6) TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

(7) GarageCars: Size of garage in car capacity

(8)GarageArea: Size of garage in square feet

Analyze the data using Multiple Linear Regression.

1. Perform and summarize EDA for the data.
2. Preprocess the data, e.g. missing data and categorical data, etc. Explain how you deal if missing data or categorical data exist.
3. Would you like to consider any transformation on the data? Why or why not? Explain for each variable.
4. Use the full data to fit the full model (called model A), write down your fitted model.
5. With the following values:

OverallQual: 6

TotalBsmtSF: 1262

1stFlrSF: 1262

GrLivArea: 1262

FullBath: 2

TotRmsAbvGrd: 6

GarageCars: 2

GarageArea: 460

What is your predicted house price using model A?