



Neighborhood Battle

*Predicting Housing Price In a Well
Defined Neighborhood*

Judy Dae

May, 5th, 2020



Predicting housing price is important

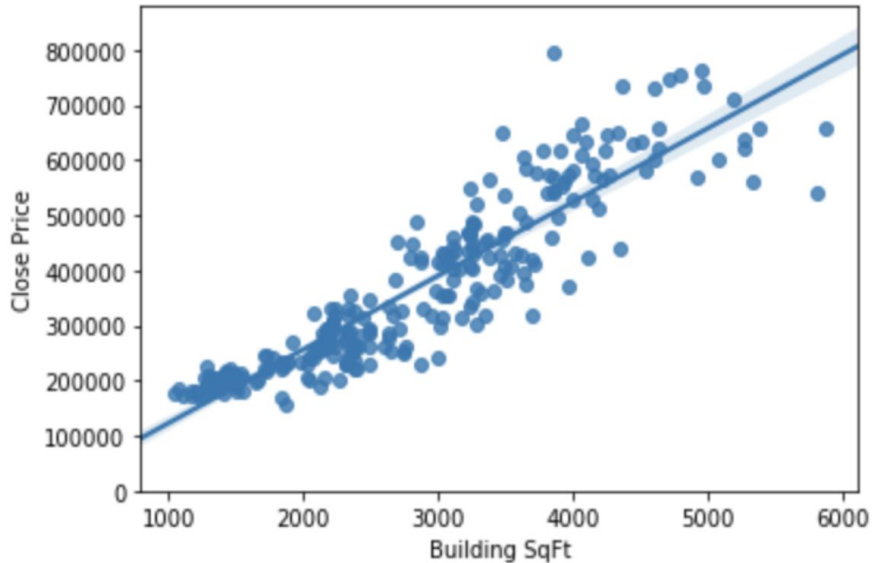
- In the Real Estate market, one of the most interesting questions is “How much is that house?”. That question can be asked in various situations and the answer to that question is critical to many perspectives.
- The current tool many Real Estate Professionals use is called CMA. There are many manual processes involved so it often causes inaccurate data. It can also cause disagreement between realtors and their clients.



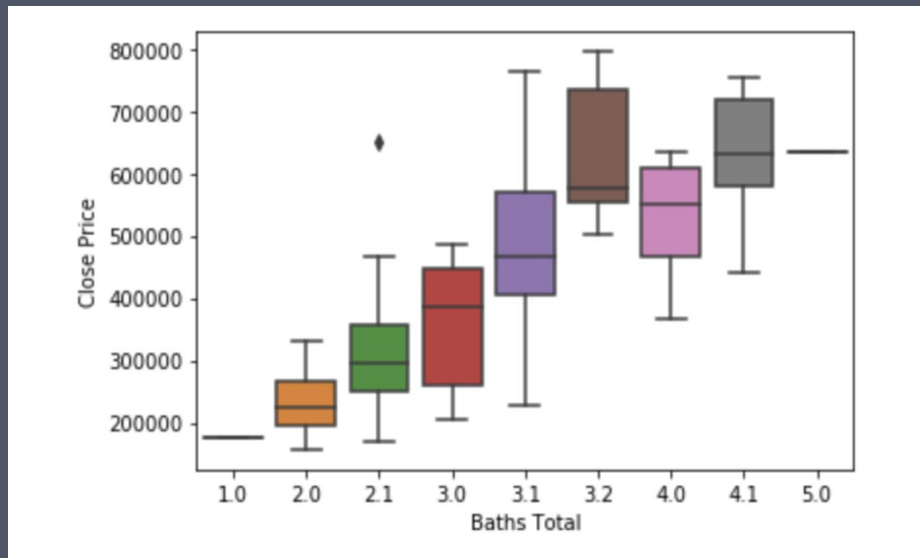
Data acquisition and cleaning

- From har.com login as real estate professionals, I am able to get all the raw data I need.
- There were 289 rows and 54 features in the raw data;
- After cleaning up the irrelevant data and wrong data, the dataset has total 289 rows with 9 most correlated data

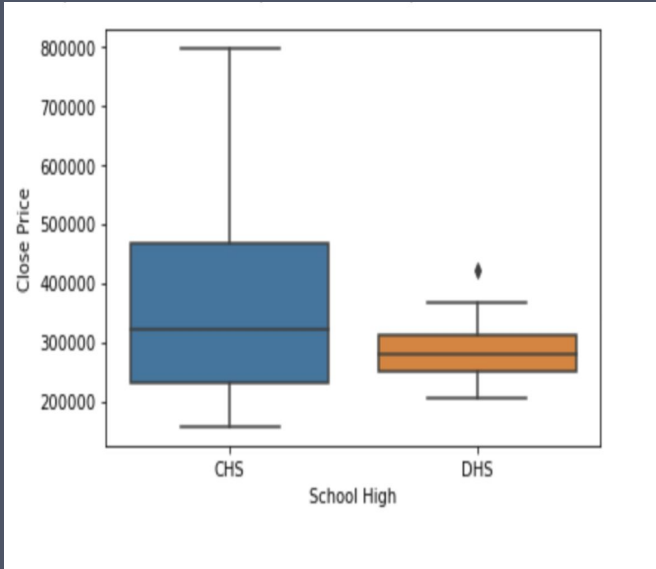
Exploratory Data Analysis



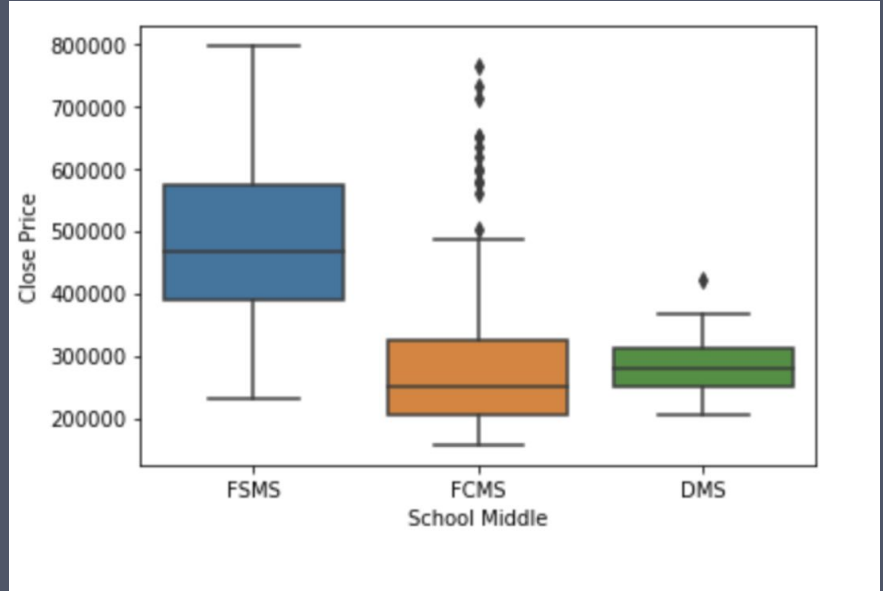
The correlation
between *Sale
Price* and
Building SqFt



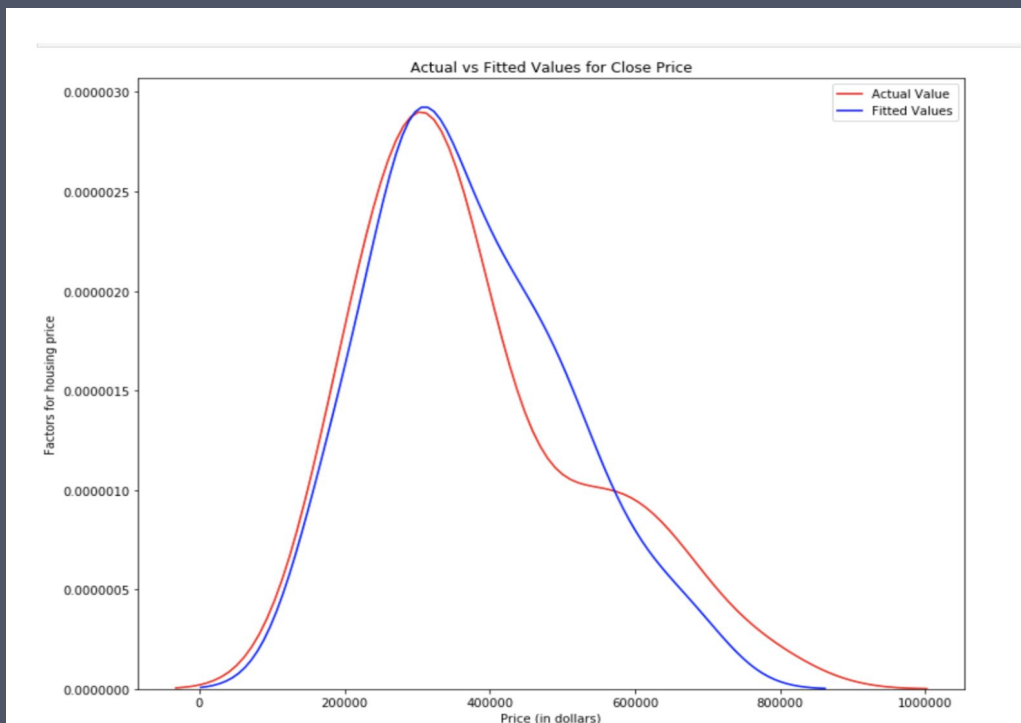
Correlation between *number of bathrooms* and the *sale price*.



Correlation between *High Schools*,
Middle schools and *Sale Price*



Linear Regression Model



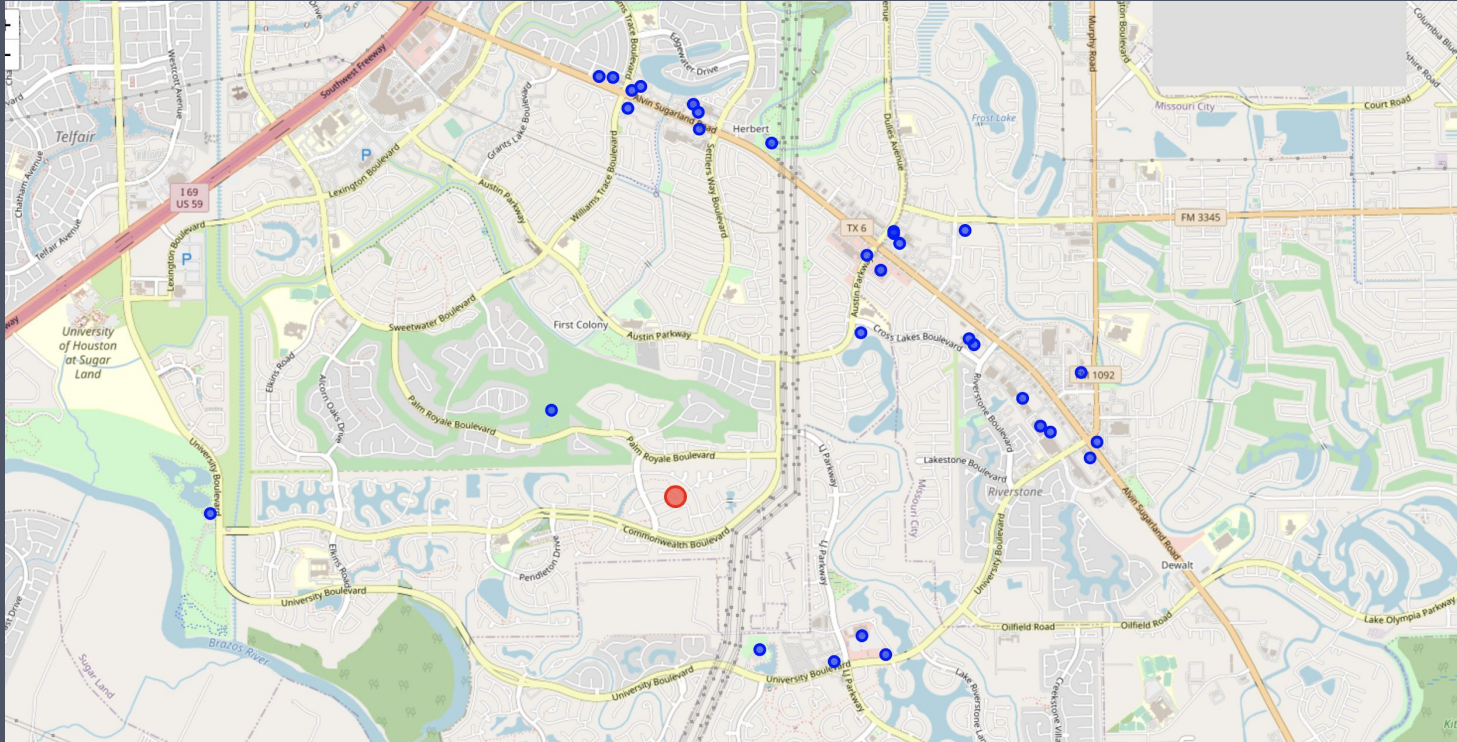
Residual sum of squares:
3868490604.29

Variance score: 0.83
Mean absolute error:
44055.88

Residual sum of squares
(MSE): 3868490604.29

R2-score: 0.77

Neighborhood Information



● is the subject property

● are different venues.

Future direction of the model

I would like to study different neighborhoods to confirm if this model can be generalized in different areas. I also would like to further study if the smaller subdivision or certain street can be extra features to improve the model's performance.

