House Price Analysis by Bruce Pei

Target goal:

1. Find the top features that correlated with the final house price.

2.When is the best-selling month for houses?

3. Houses sold in georgical data visualization by different year?

4. Predict the house price in the next few years on different locations because house price

would be quite different on different locations.

Data Wragling:

1. Filling missing data with a proper meaning either by Na, None or mean of the average data (“Prediction\_Bruce.ipynb”).

Analysis:

1. Build model with machine learning in training data with PCA (R file “PCA analysis for house.R) and correlation (Python file “Prediction\_Bruce.ipynb”) for the top features.

2. Removing the outlier by each of the top feature (Python file “Prediction\_Bruce.ipynb”).

3. Predict the house price with a combination of linear regression and boosting in test data (88% accuracy).

Data visualization:

1. Temoral data:

House price trend in different areas and years (“house\_pred.py”).

House sold in different month (“house\_pred.py”).

2. Geographical data:

Demonstate the relationship between numbers of house sold and location (“Final\_551\_map - Xiangyu Pei.html”).