

# HOW TO SELECT YOUR NEXT HOME IN BEIJING



*Caining Jin*  
*Aug-2020*

# INTRODUCTION

- This project is about how to select the most suitable house in Beijing.
- Beijing, Capital of China.

The house price increased a lot in last ten years

- For self use
- For investment

# INTRODUCTION

- Target Audience:
  - People who want to buy a house in Beijing
  - People who want to invest a house for earning money in Beijing

# DATA REQUIREMENT

- House price
- House latitude and longitude
- House nearby venues
- House detail:
  - Square
  - Price per square
  - Elevator?
  - Next to subway?
  - etc...

# DATA

- Source from:

- <https://www.kaggle.com/ruiqurm/lijia>

df.head()

7]:

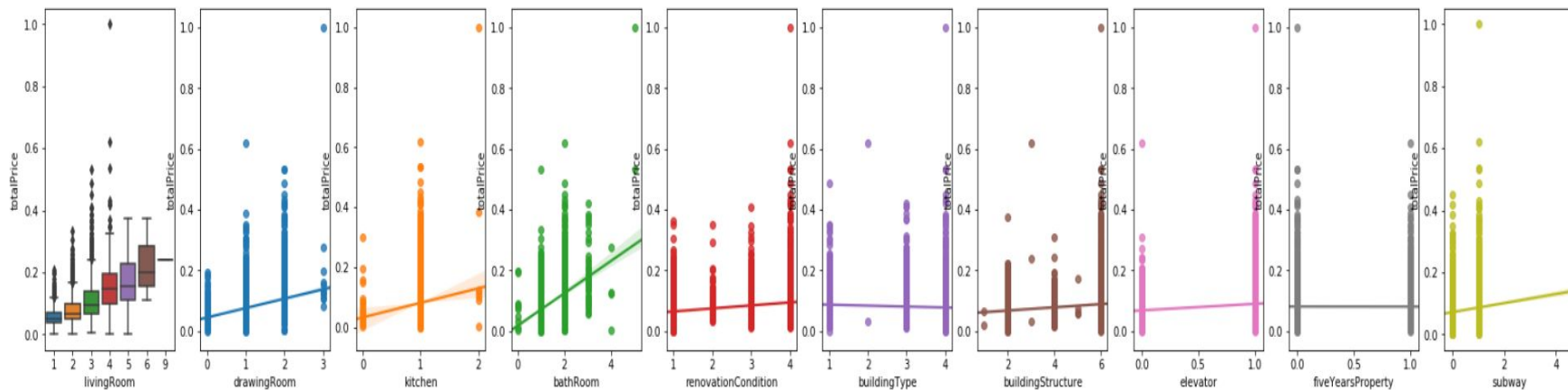
	Lat	Lng	totalPrice	square	livingRoom	drawingRoom	kitchen	bathRoom	buildingType	renovationCondition	buildingStructure	elevatc
0	39.955386	116.355475	0.093925	0.090975	1	1	1	1	4.0	3	2	0.
1	39.919830	116.609957	0.087617	0.150109	1	1	1	1	3.0	3	6	1.
2	39.983371	116.490734	0.051402	0.147903	2	1	1	1	4.0	1	6	0.
3	39.880102	116.376538	0.065421	0.128821	2	1	1	1	4.0	1	2	0.
4	39.839983	116.247907	0.045280	0.142899	1	1	1	1	3.0	4	6	1.

# METHODOLOGY

1. Data cleaning: pre-processing
2. Correlation heatmap
3. Get relationships between each element with the price
4. Use Train\_test\_split divide the dataset for training
5. Get house nearby venue
6. Unsupervised learning: K-mean clustering

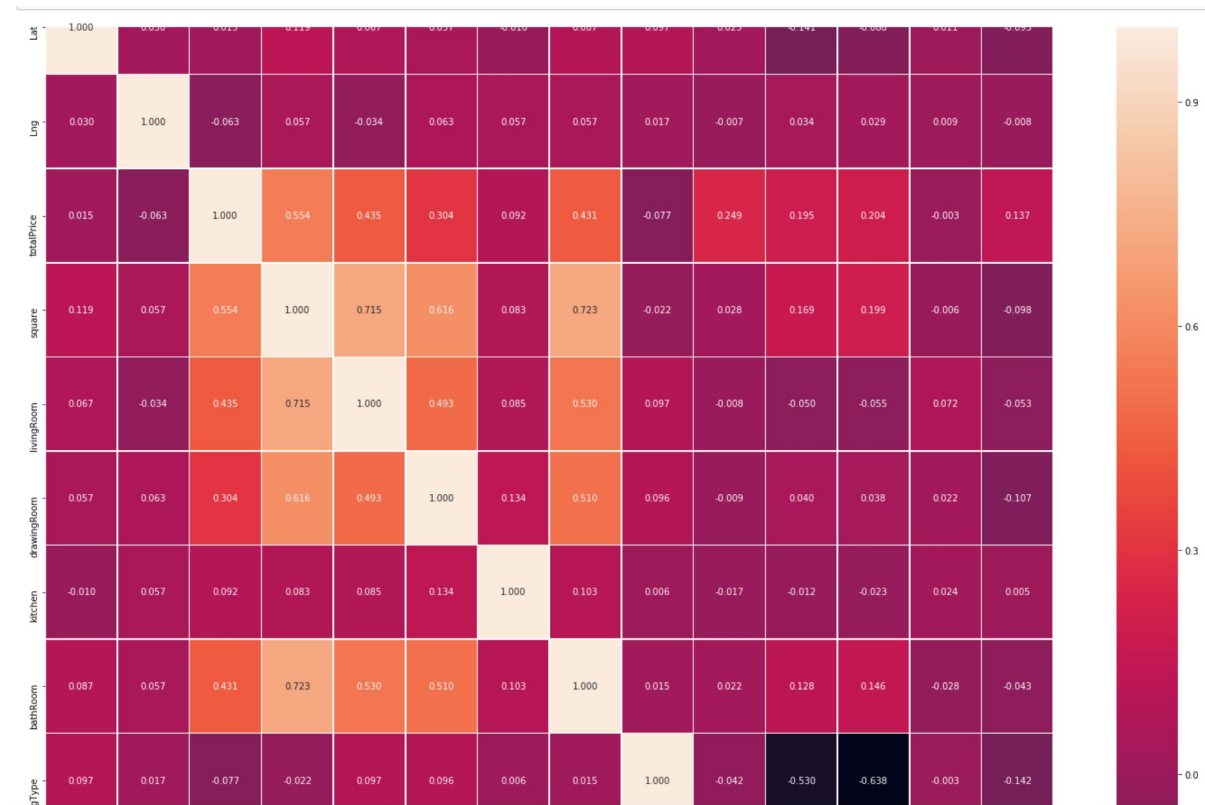
# RESULTS

Correlation:



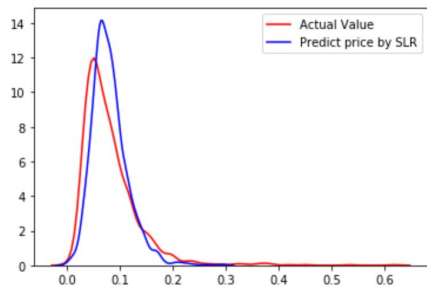
# RESULTS

Heatmap:





# RESULTS

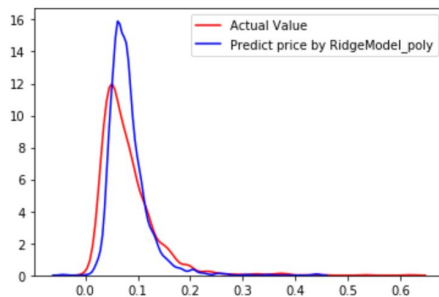


```
5]: ▶ print('MSE for SLR is: ', mean_squared_error(y_test, yhat_lm))  
    print('R score for SLR is: ', lm.score(x_test, y_test))
```

MSE for SLR is: 0.0016599302688888149  
R score for SLR is: 0.4265988630702905

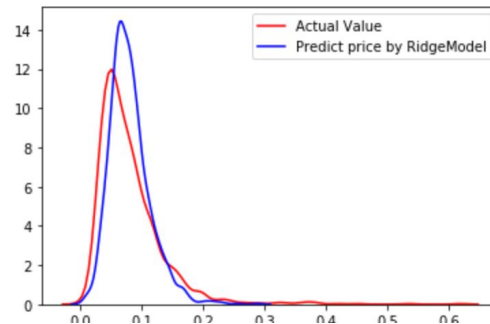
MSE for RidgeModel-poly is: 0.001525068475449515  
R score for RidgeModel-poly is: 0.4731851005381091

```
9]: <matplotlib.axes._subplots.AxesSubplot at 0x1e1aedaf3c8>
```



MSE for RidgeModel is: 0.0016676800561419493  
R score for RidgeModel is: 0.42392180072303576

```
: <matplotlib.axes._subplots.AxesSubplot at 0x1e1b0bbc408>
```



# RESULTS

Top 10 nearby venues:

livingRoom		1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1	Chinese Restaurant	Hotel	Coffee Shop	Fast Food Restaurant	Hotel Bar	Shopping Mall	Restaurant	Grocery Store	Café	Pizza Place
1	2	Chinese Restaurant	Hotel	Fast Food Restaurant	Coffee Shop	Metro Station	Nightclub	Café	Pizza Place	Shopping Mall	BBQ Joint
2	3	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Hotel	Historic Site	Park	Café	Metro Station	Supermarket	Shopping Mall
3	5	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	French Restaurant	Shopping Mall	Pizza Place	Metro Station	Sporting Goods Shop	Bookstore	Korean Restaurant

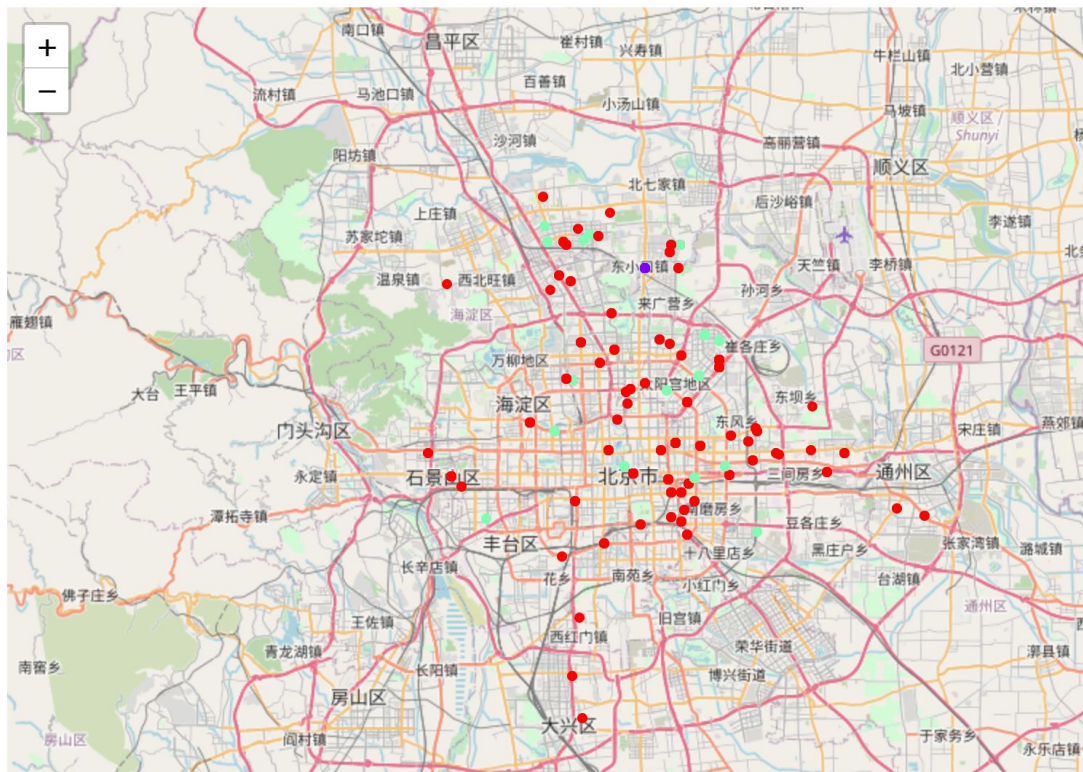
# RESULTS

Difference between clusters:

	HousePrice	square	livingRoom	drawingRoom	kitchen	bathRoom	buildingType	renovationCondition	buildingStructure	elevator	fiveYearsPro
Cluster Labels											
0	0.075643	0.149345	1.675556	0.977778	1.0	1.013333	3.140000	2.575556	4.480000	0.495556	0.71
1	0.094626	0.406182	5.000000	2.000000	1.0	2.000000	4.000000	4.000000	6.000000	0.000000	1.00
2	0.107156	0.220451	3.000000	1.285714	1.0	1.314286	3.142857	2.171429	4.342857	0.421429	0.85

# RESULTS

Display by folium:



# DISCUSSION

- Cluster 0:
  - Low price
  - recommend for young people who just start their career
  - Convenient transportation
- Cluster 1:
  - Square is the largest, medium price
  - No convenient transportation
  - Suggest for people who greater than 50-year old
- Cluster 2:
  - Good transportation
  - Good night life (next to nightclub)
  - High Price

# CONCLUSION

Based on different people's requirements, they should select the most suitable type of house by themselves based on the different cluster of houses.

Thank you!