Housing Price Prediction

Kun He
Chiehen Hung
Ben Liu
Guojun Xiong
Tianxiao Zhang
Xiaohan Zhang

Outline

- Anomaly Detection
- Time Series Forecasting
- Regression
- Clustering
- Classification

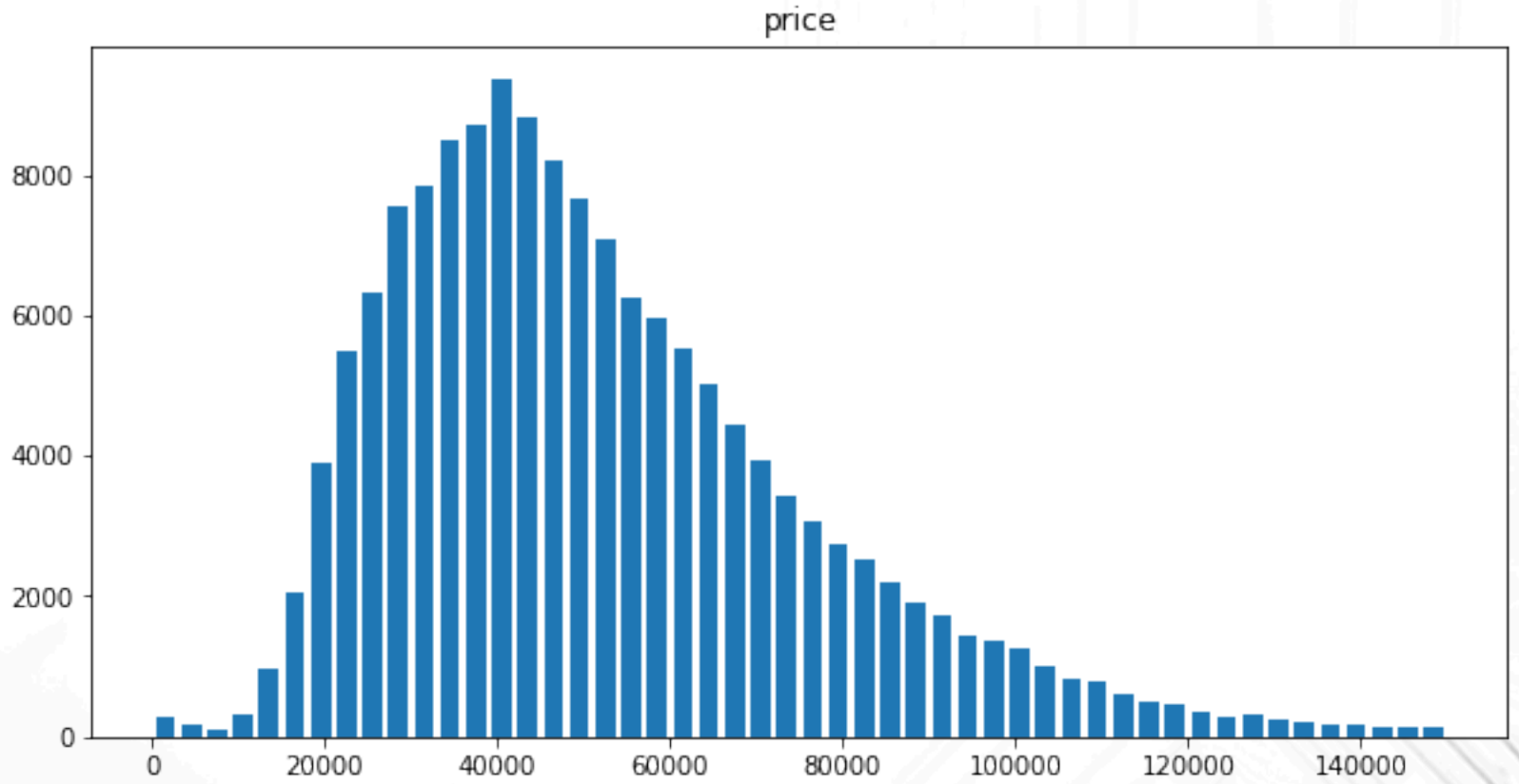
Motivation

- The most expensive item during the whole life for most people
- Investment
- Phrase it as time series, regression and classification, using data of Beijing and **King County**, Seattle as comparison.

Anomaly Detection

- Interquartile Range (IQR)
- Isolation Forest
- One Class SVM
- Local Outlier Factor
- Robust Covariance
- K-Nearest Neighbors (KNN)

Beijing Housing Price



Beijing Housing Price

- Feature Engineering

  Generate Years and Months from date.

- Select Features

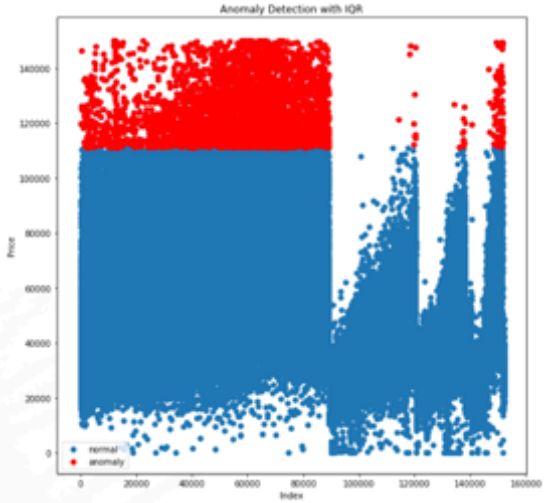
  Such as price, square, kitchen, elevator, subway, years, months, etc.

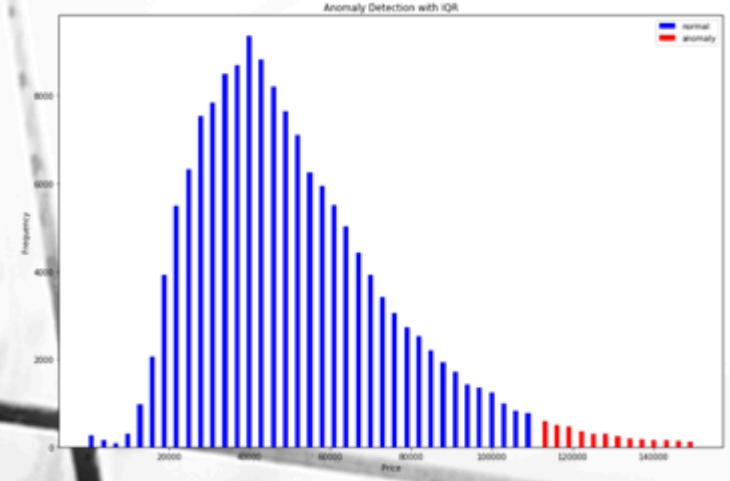
- Principal Component Analysis (PCA)

  Dimensionality reduction with PCA (n_components=3).

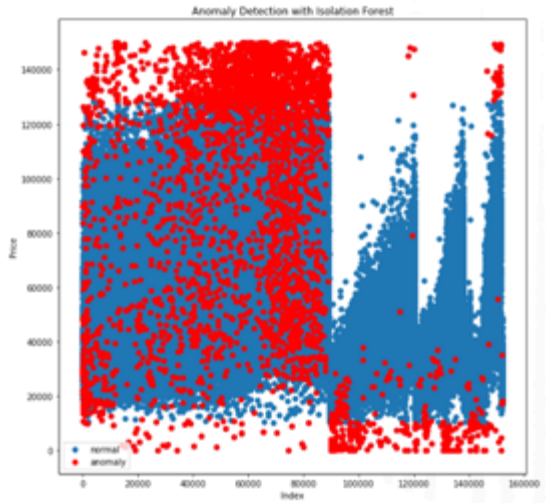
Beijing Housing Price

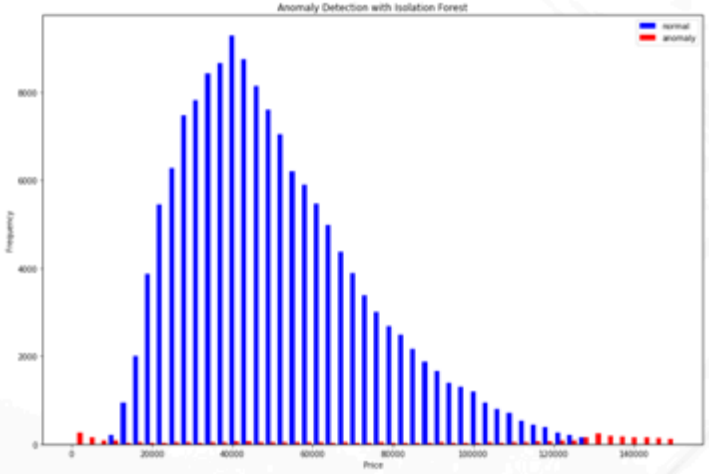
Interquartile Range (IQR)



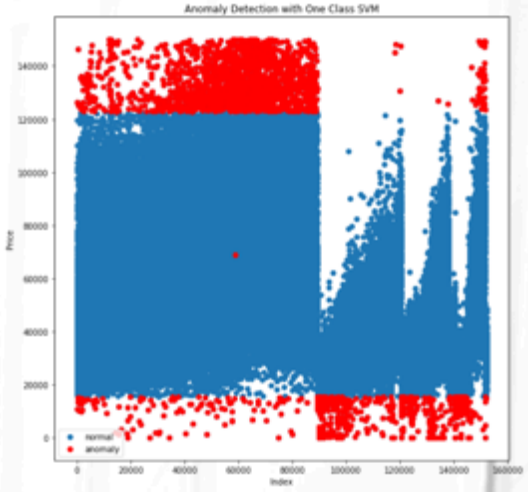


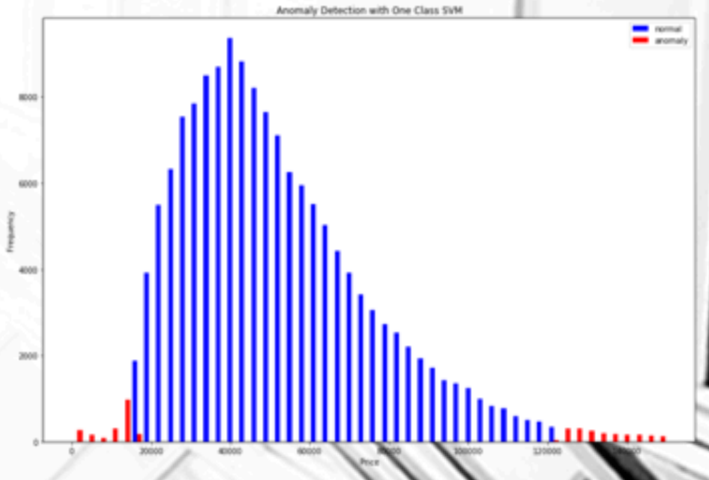
Isolation Forest





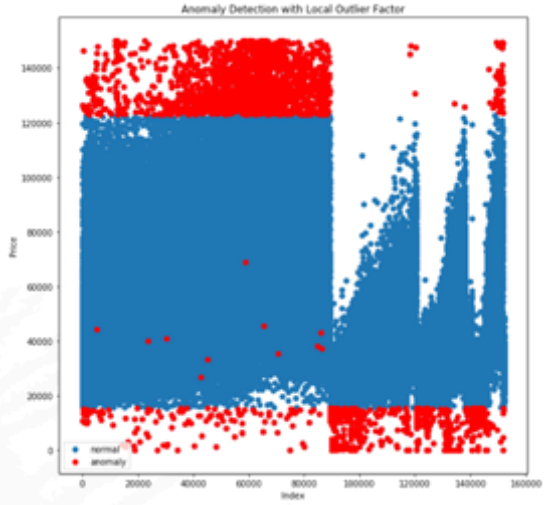
One Class SVM



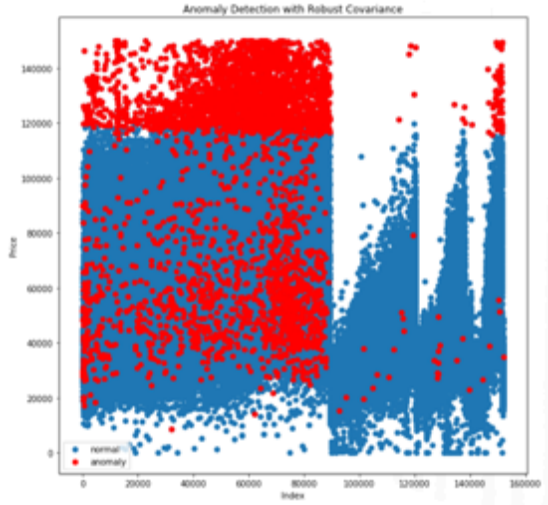


Beijing Housing Price

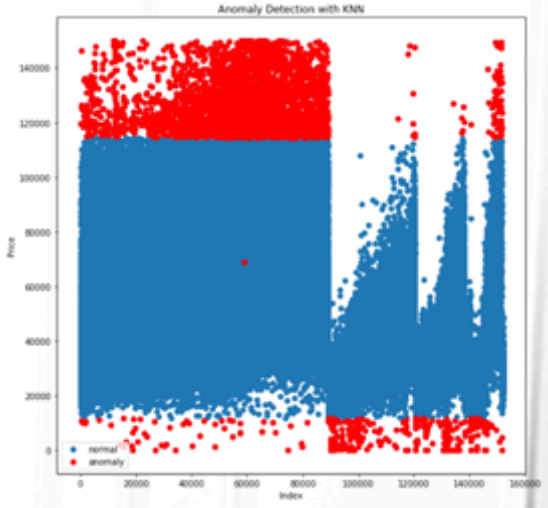
Local Outlier Factor

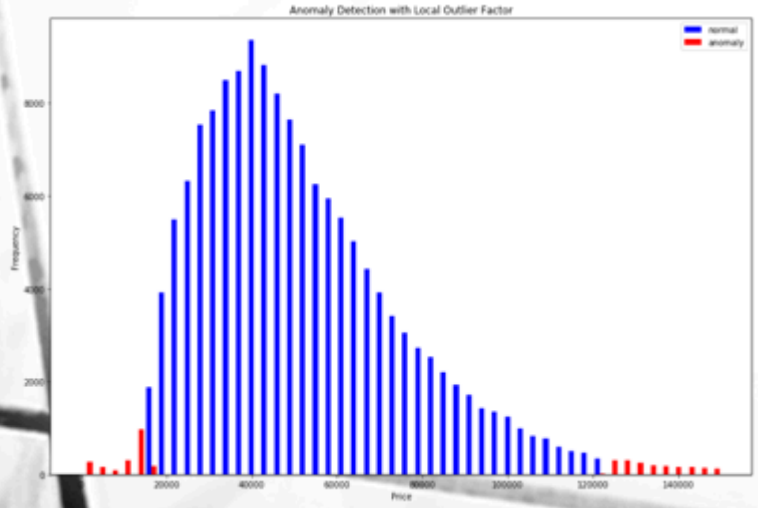


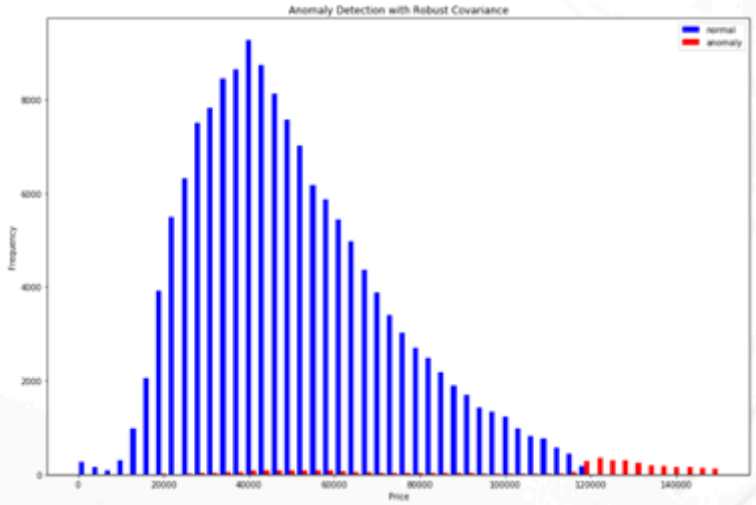
Robust Covariance

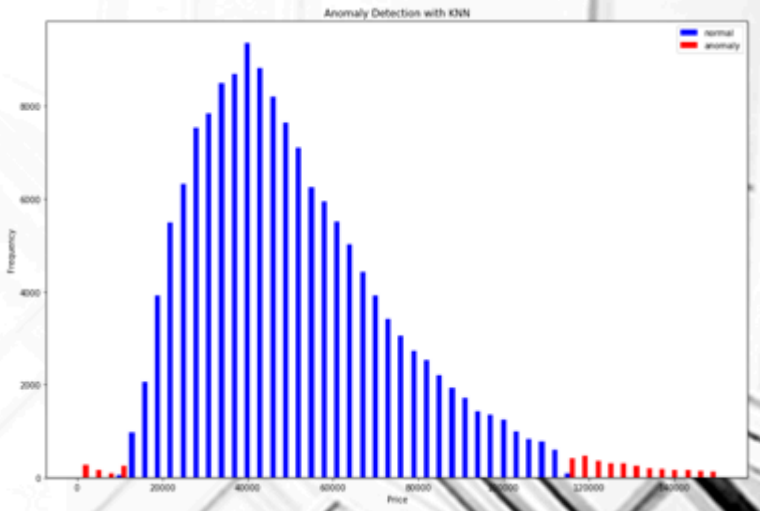


KNN





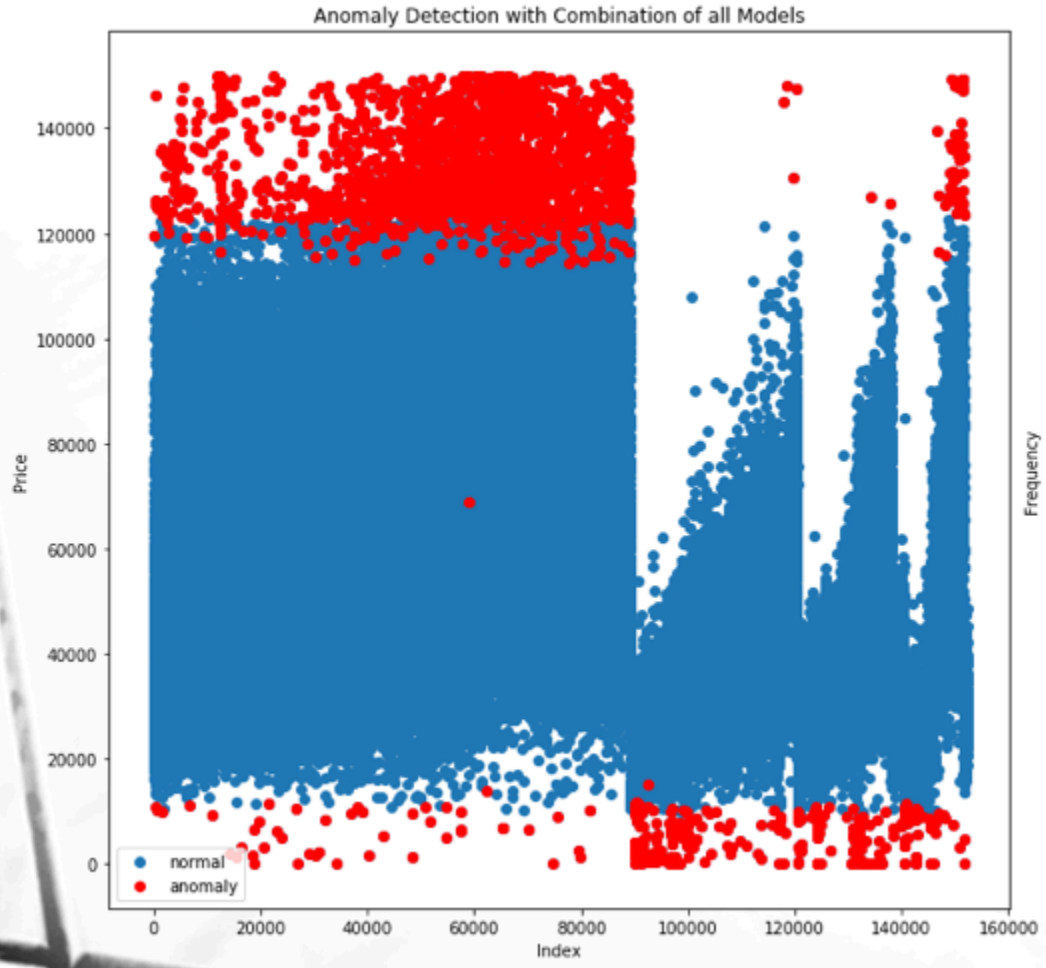


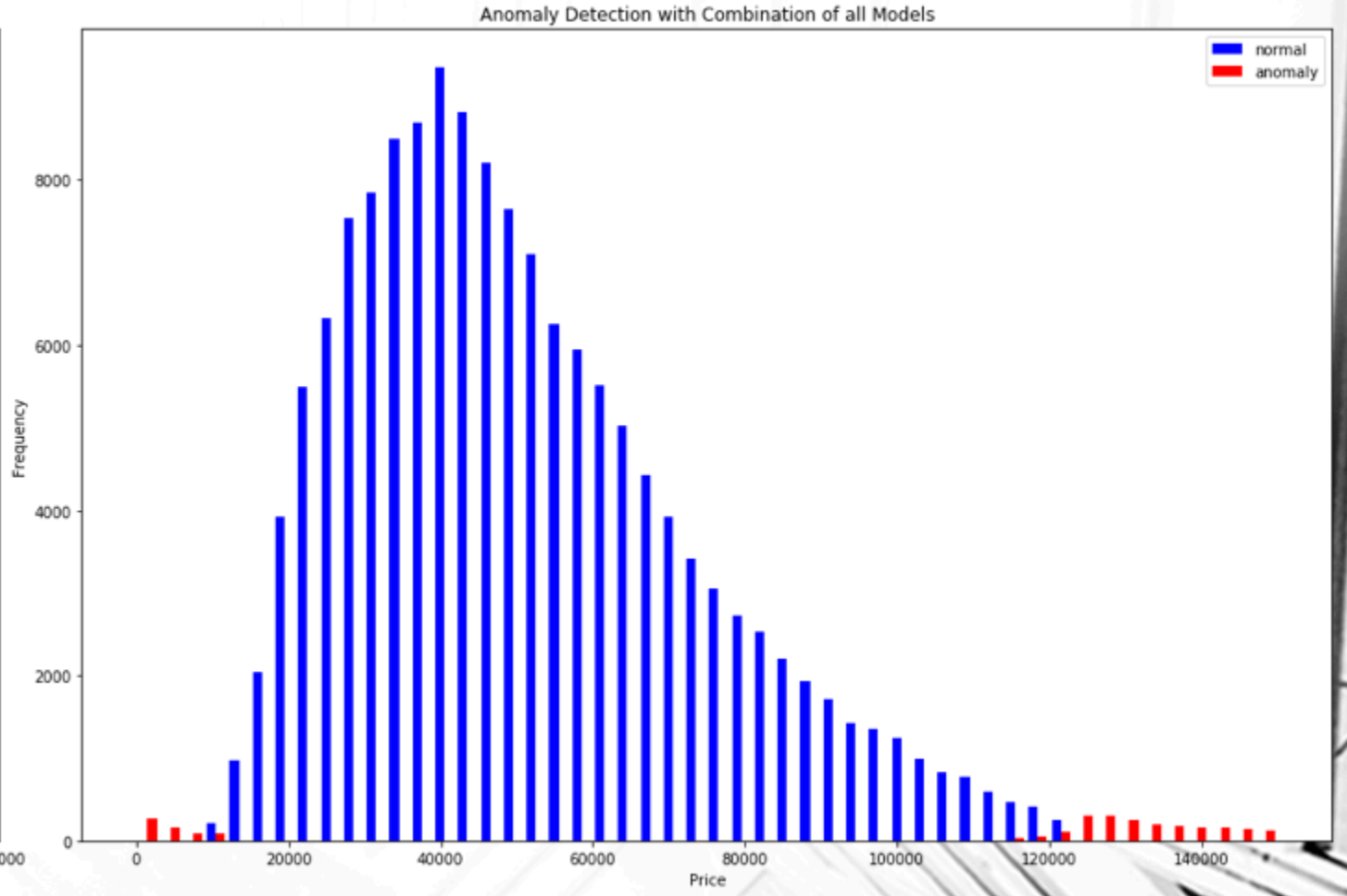


Beijing Housing Price

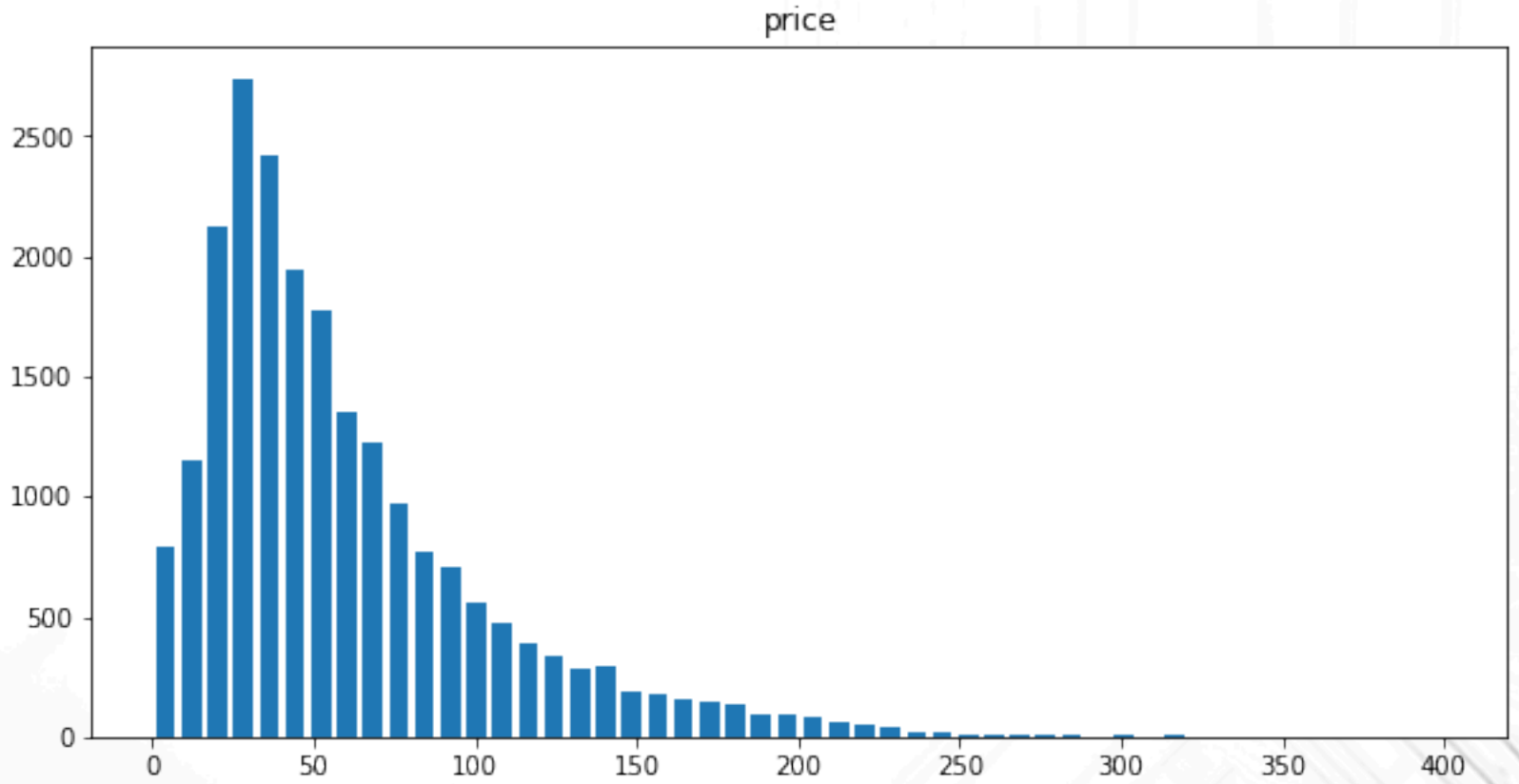
- Determine by voting
- Denote data point as abnormal, if no less than 4 models vote ANOMALY.

Beijing Housing Price





KC Housing Price



KC Housing Price

- Feature Engineering

  Calculate the price per square foot.

- Select Features

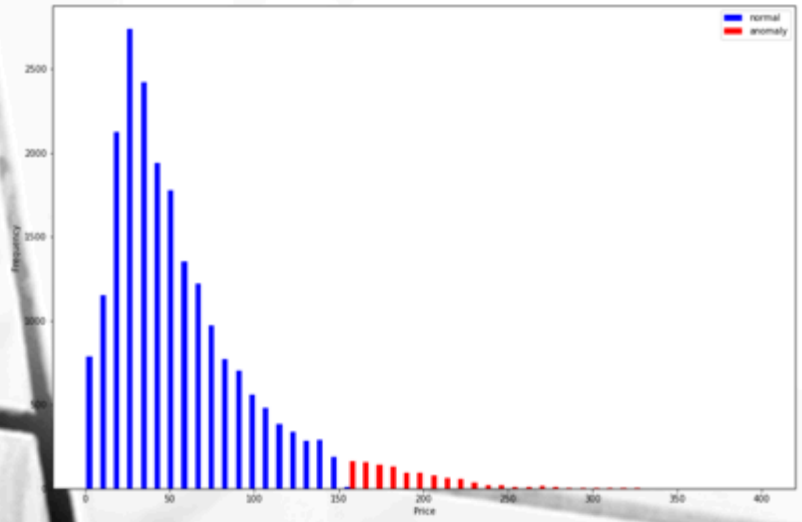
  Such as price, sqft, bedrooms, bathrooms, floors, zip code, etc.

- Principal Component Analysis (PCA)

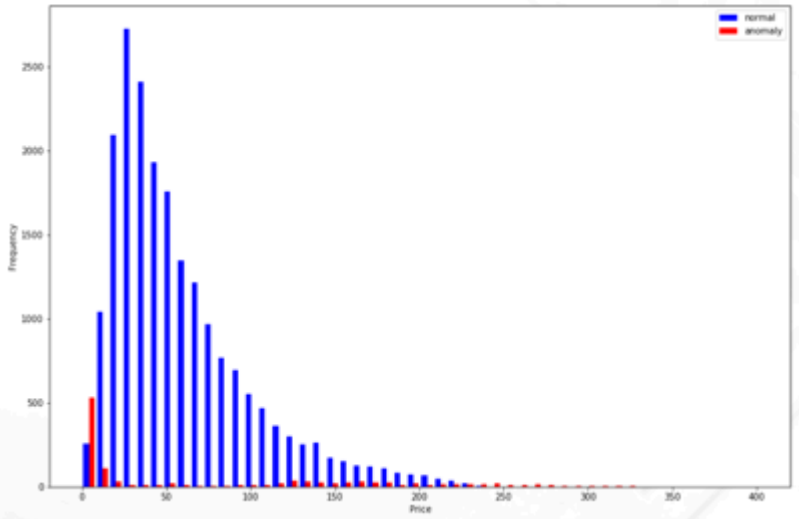
  Dimensionality reduction with PCA (n_components=3).

KC Housing Price

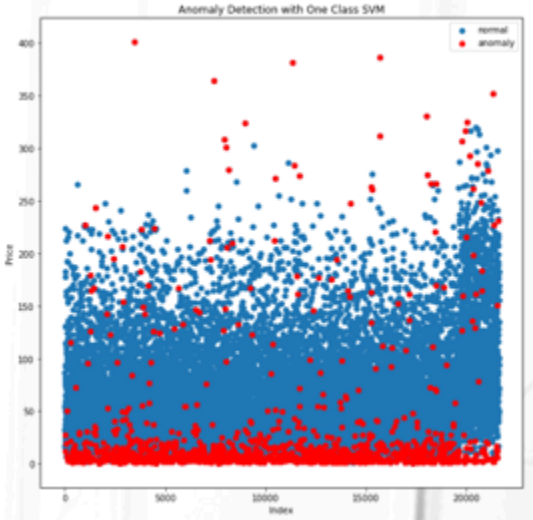
Interquartile Range (IQR)

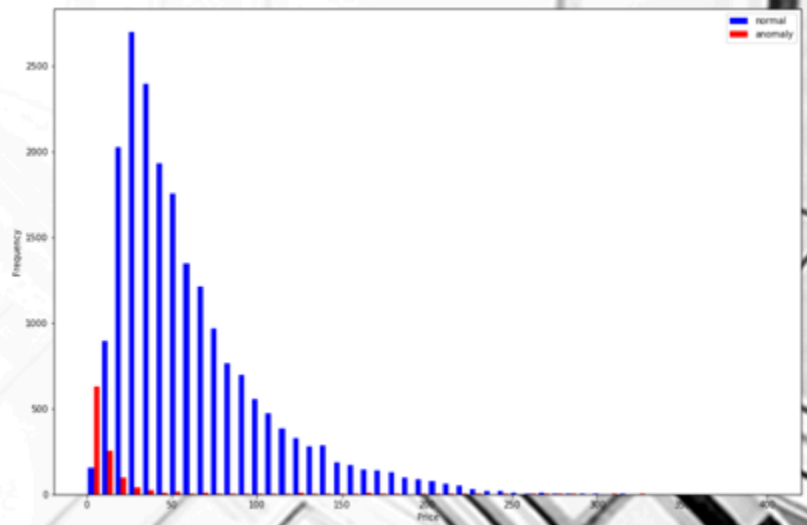


Isolation Forest



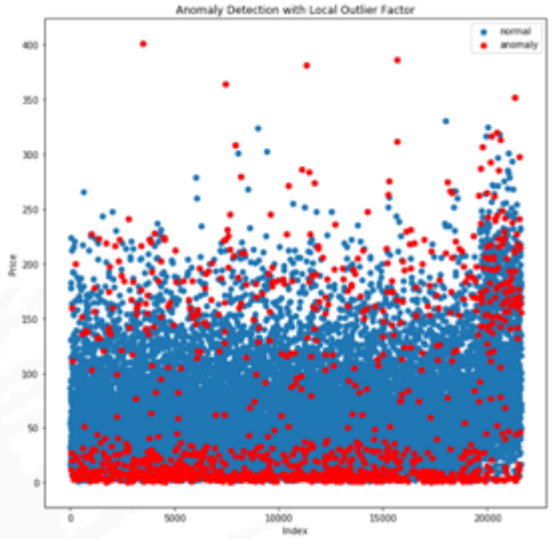
One Class SVM

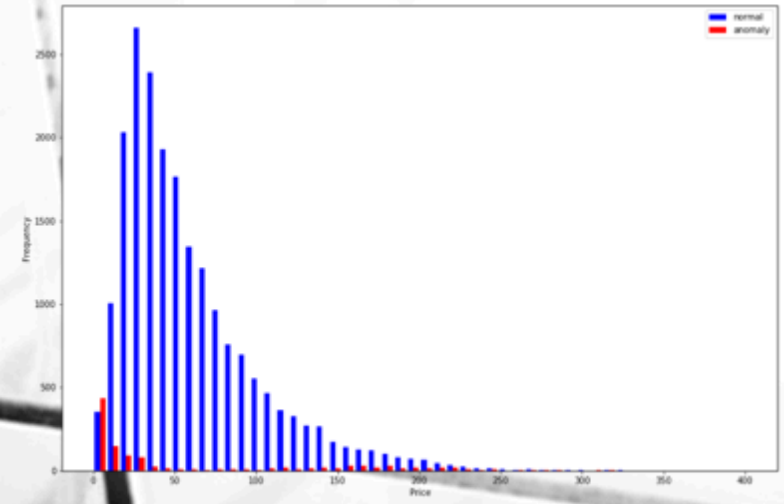




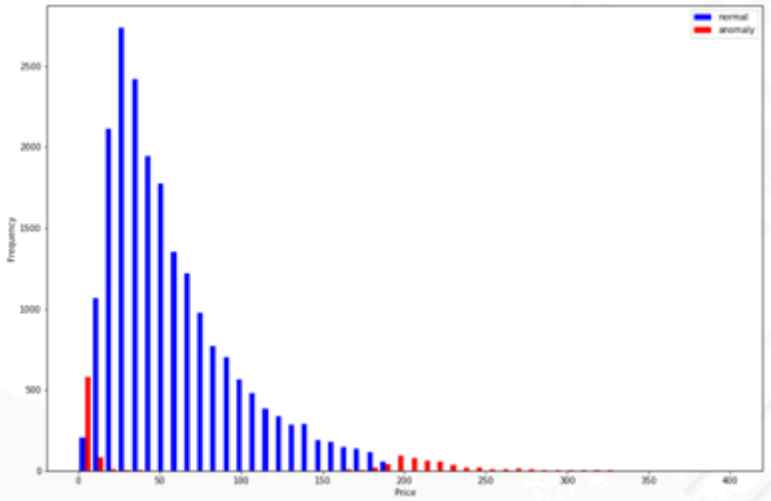
KC Housing Price

Local Outlier Factor

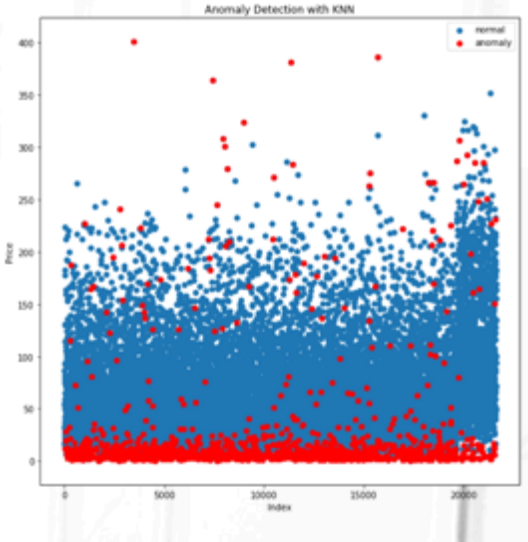


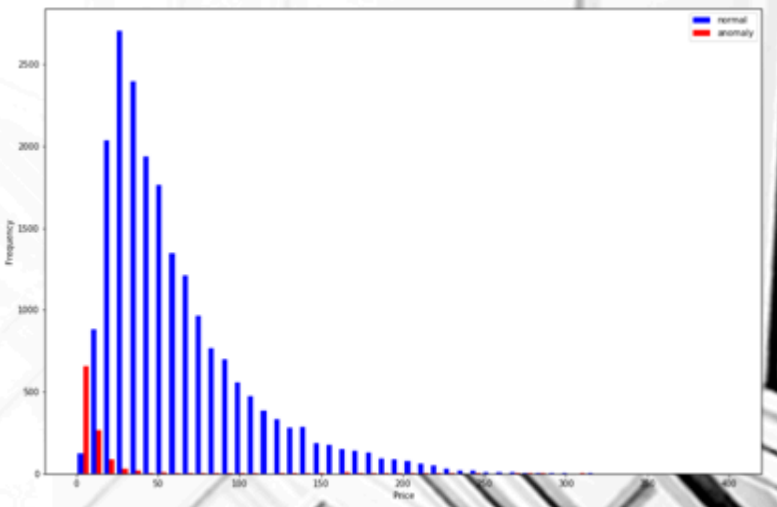


Robust Covariance

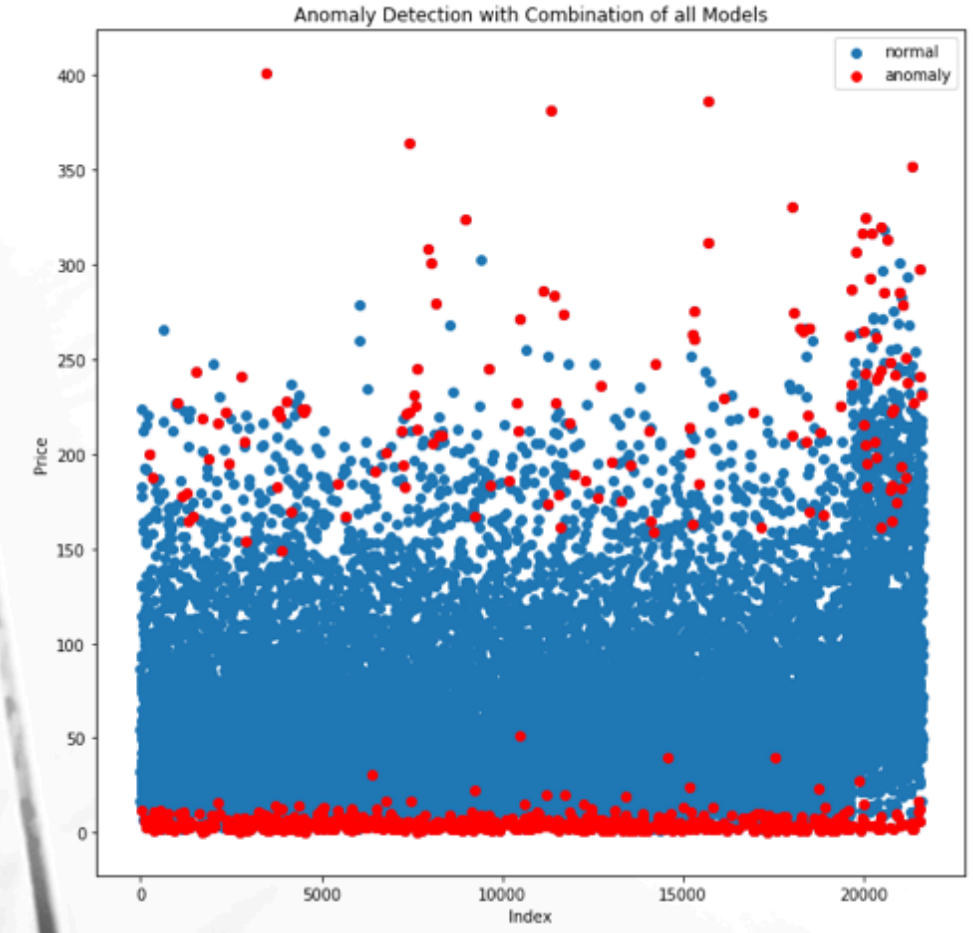


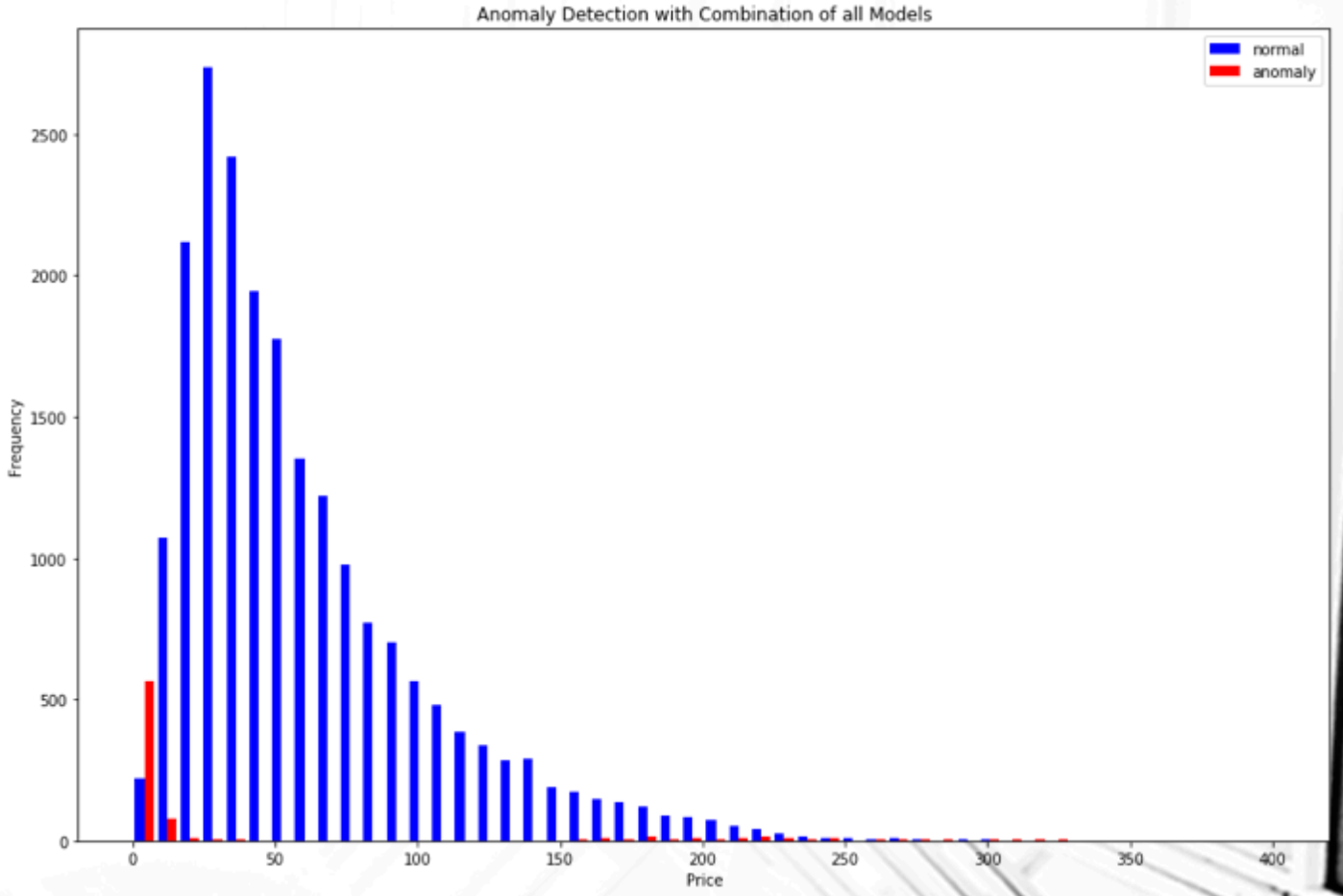
KNN





KC Housing Price



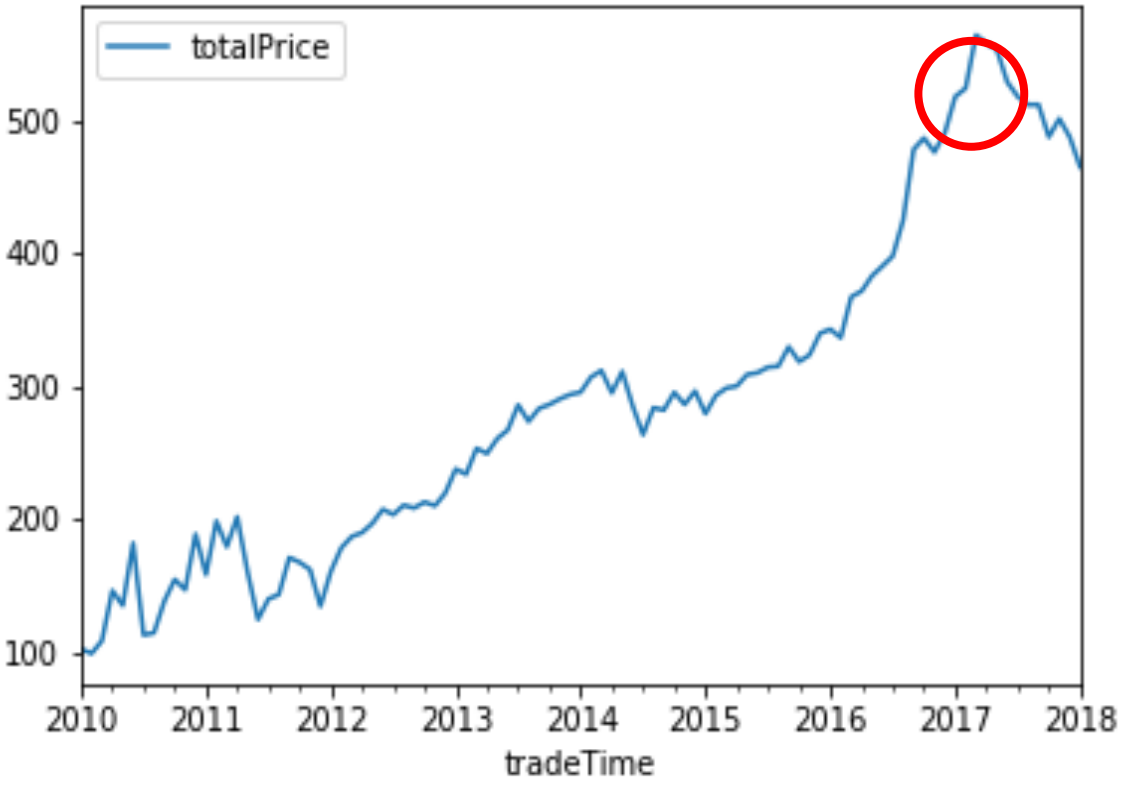


Time series Forecasting

Predict the future trend

Time Trend for Beijing Housing Price (2010-2018)

Total price (10k ￥)

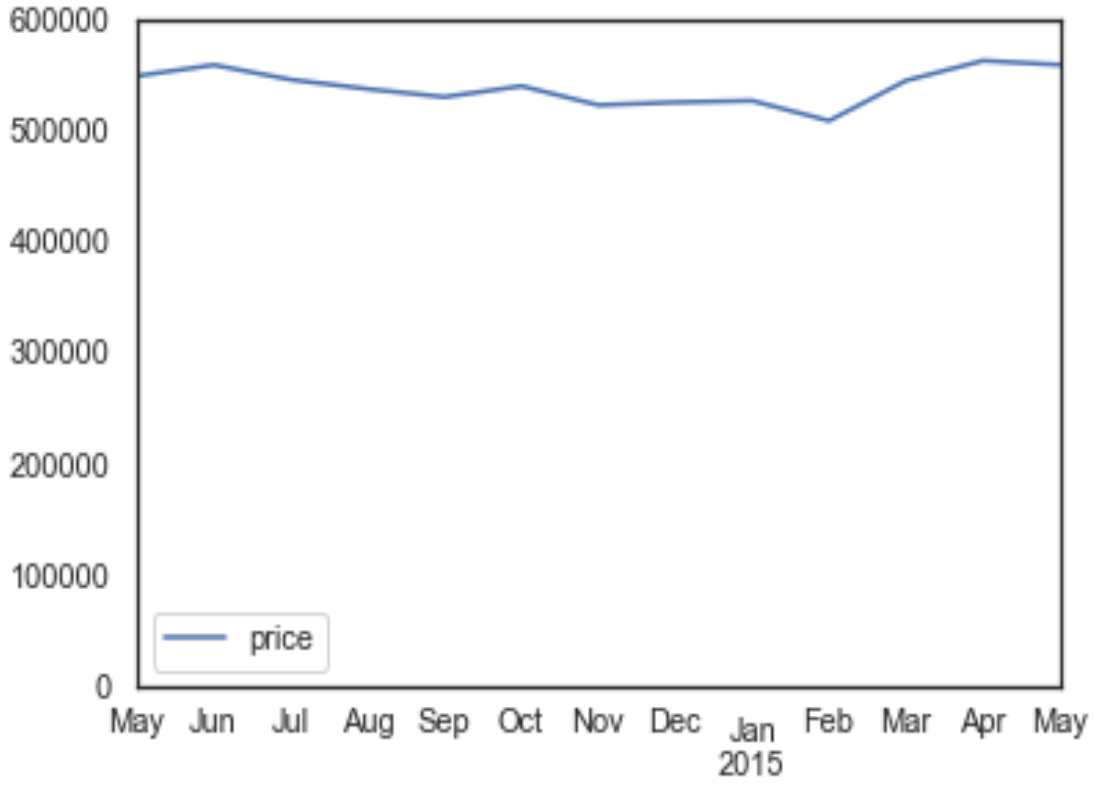


Average price (¥)

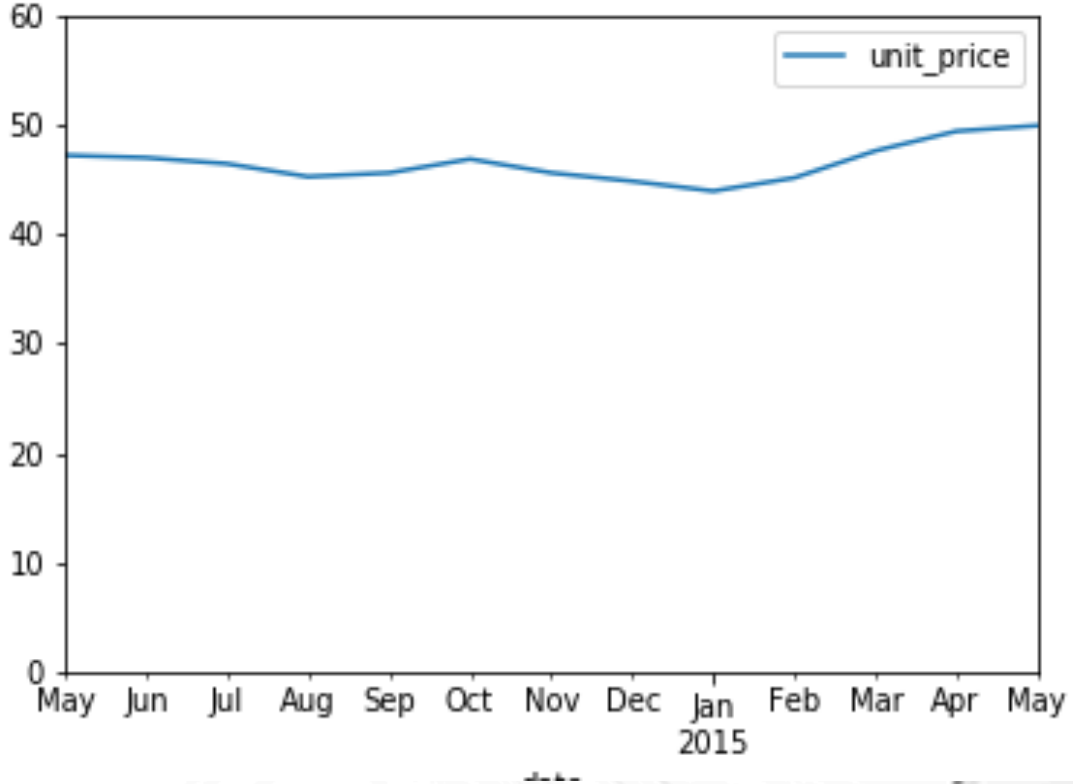
A turning point could observed on total price and average price. The reason for a drop on Beijing house market is a series of regulation released during the first quarter of 2017.

Time Trend for KC Housing Price (2014-2015)

Total price($)



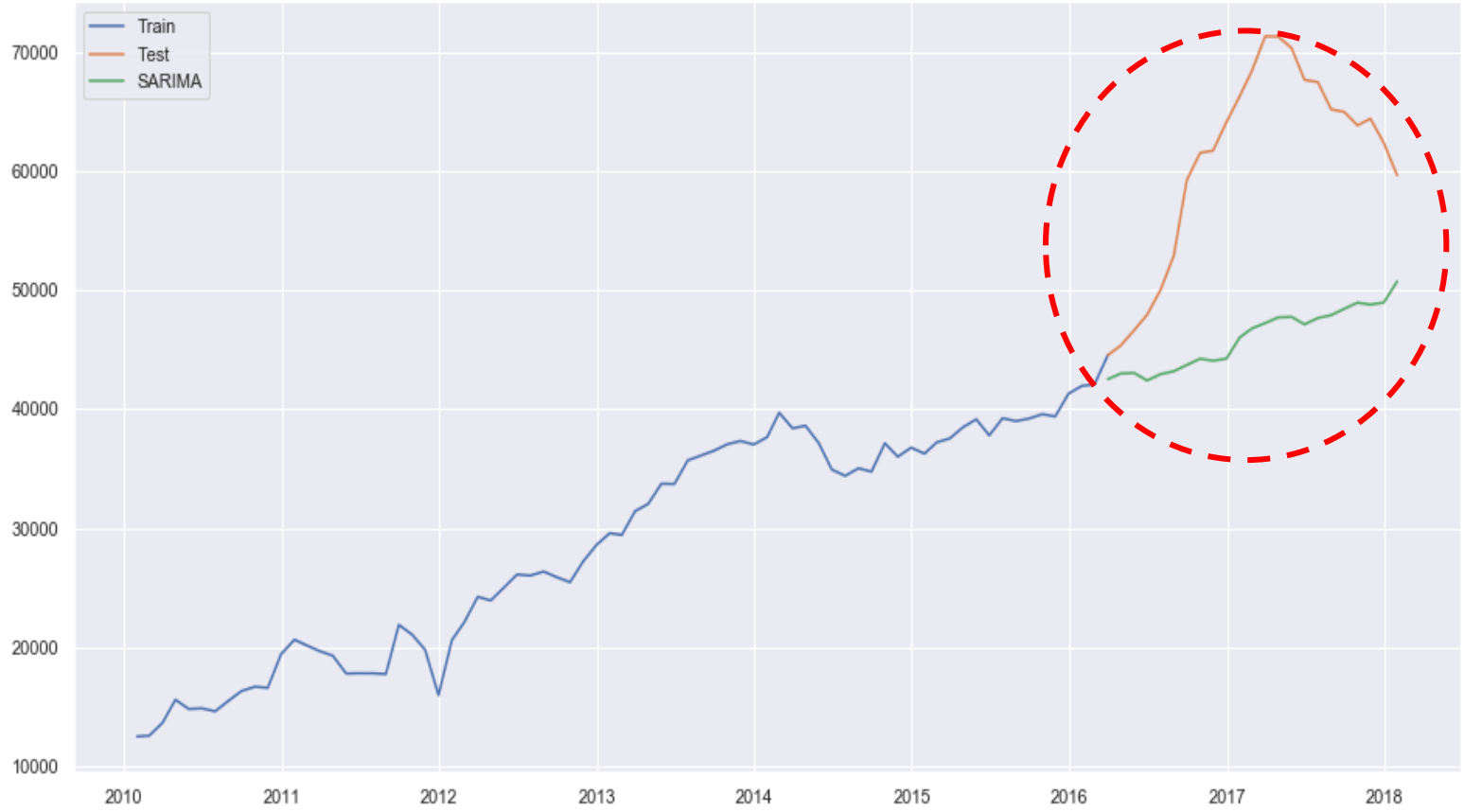
Average price ($)



The housing price is pretty stationary in KC.

Forecasting (ARIMA)

Beijing

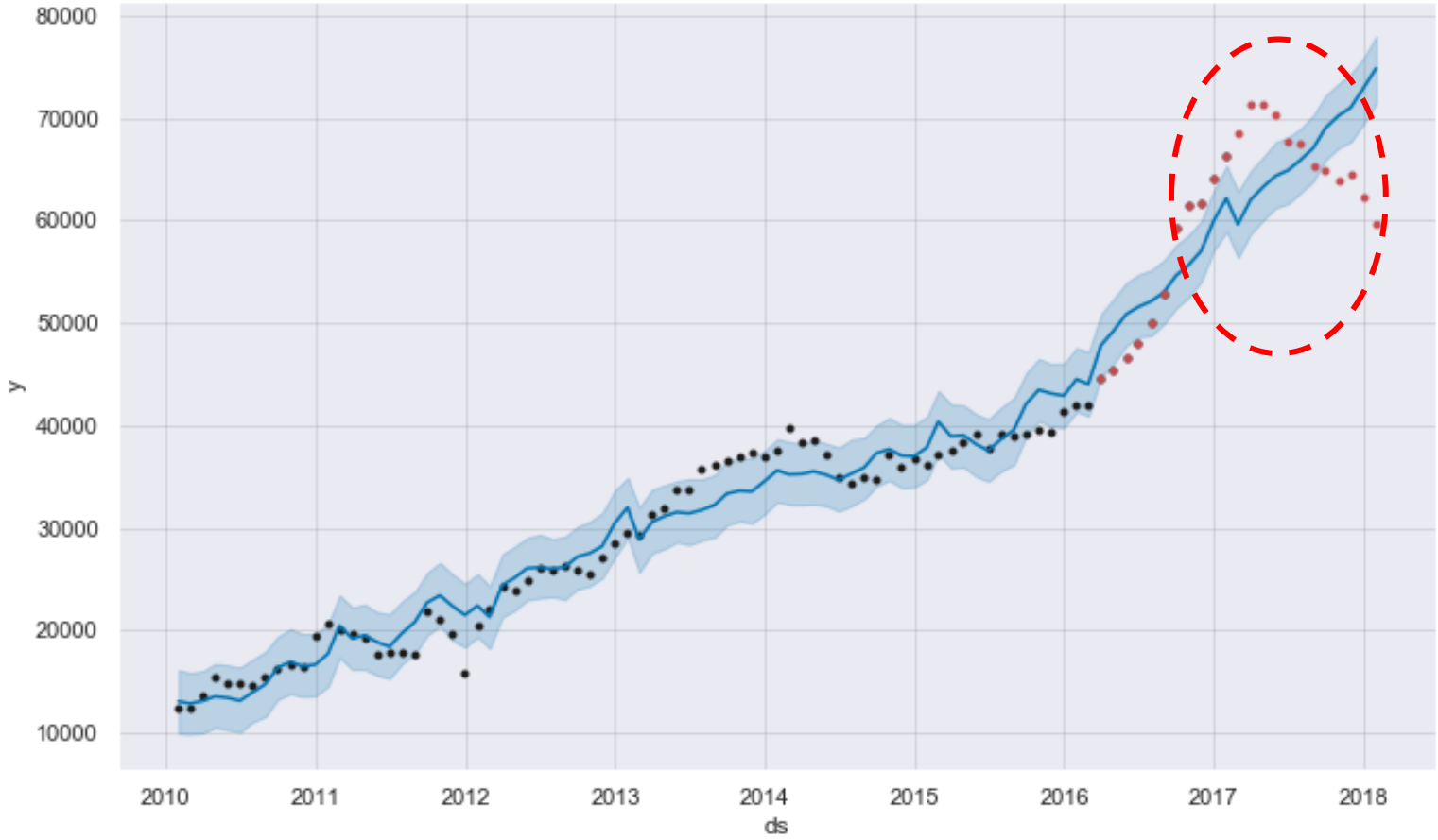


An increasing trend can be observed.

A gap seems to be huge.

Forecasting(Prophet)

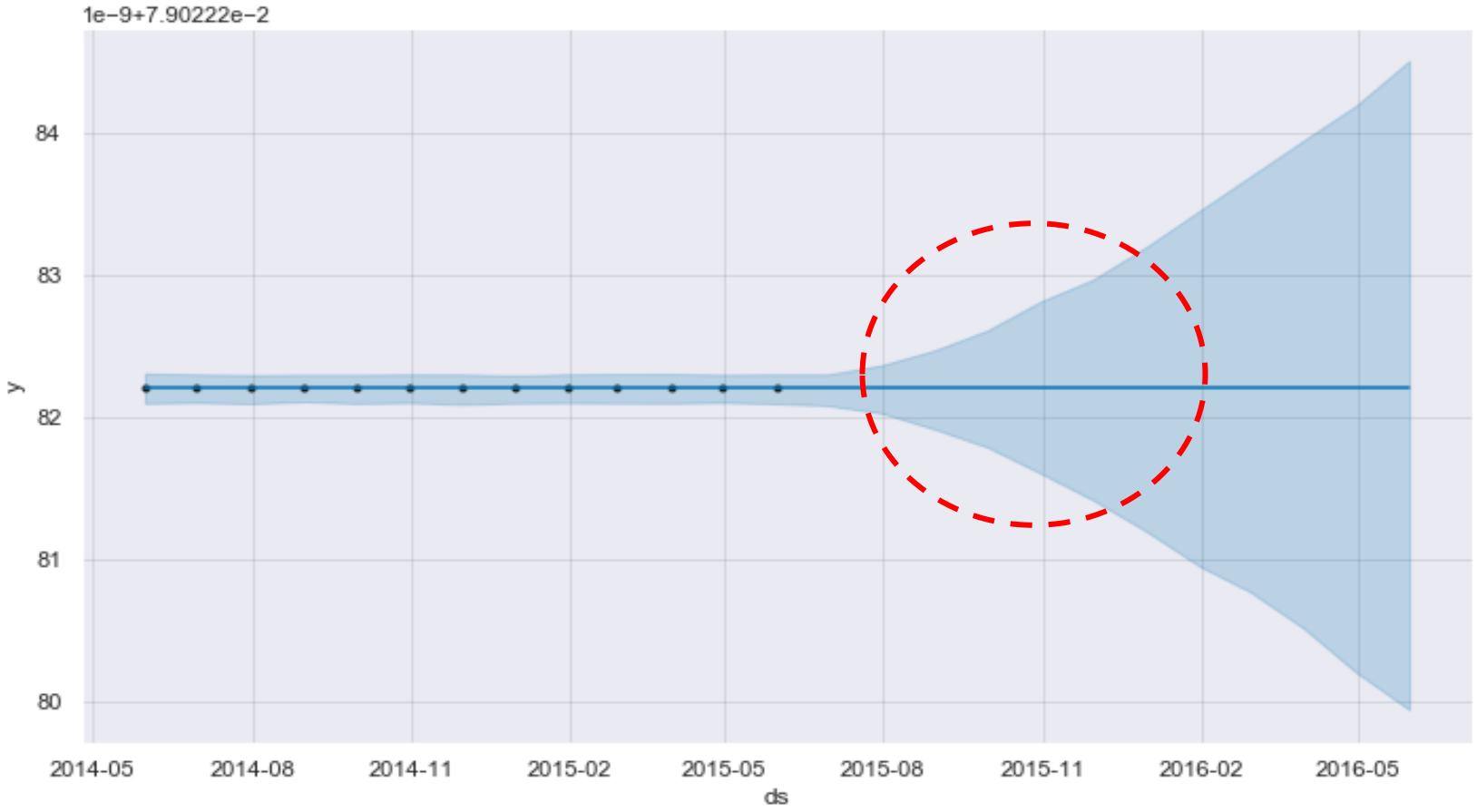
Beijing



The housing price in Beijing has a great potential to increase in the near future. Good investment?

Forecasting(Prophet)

KC

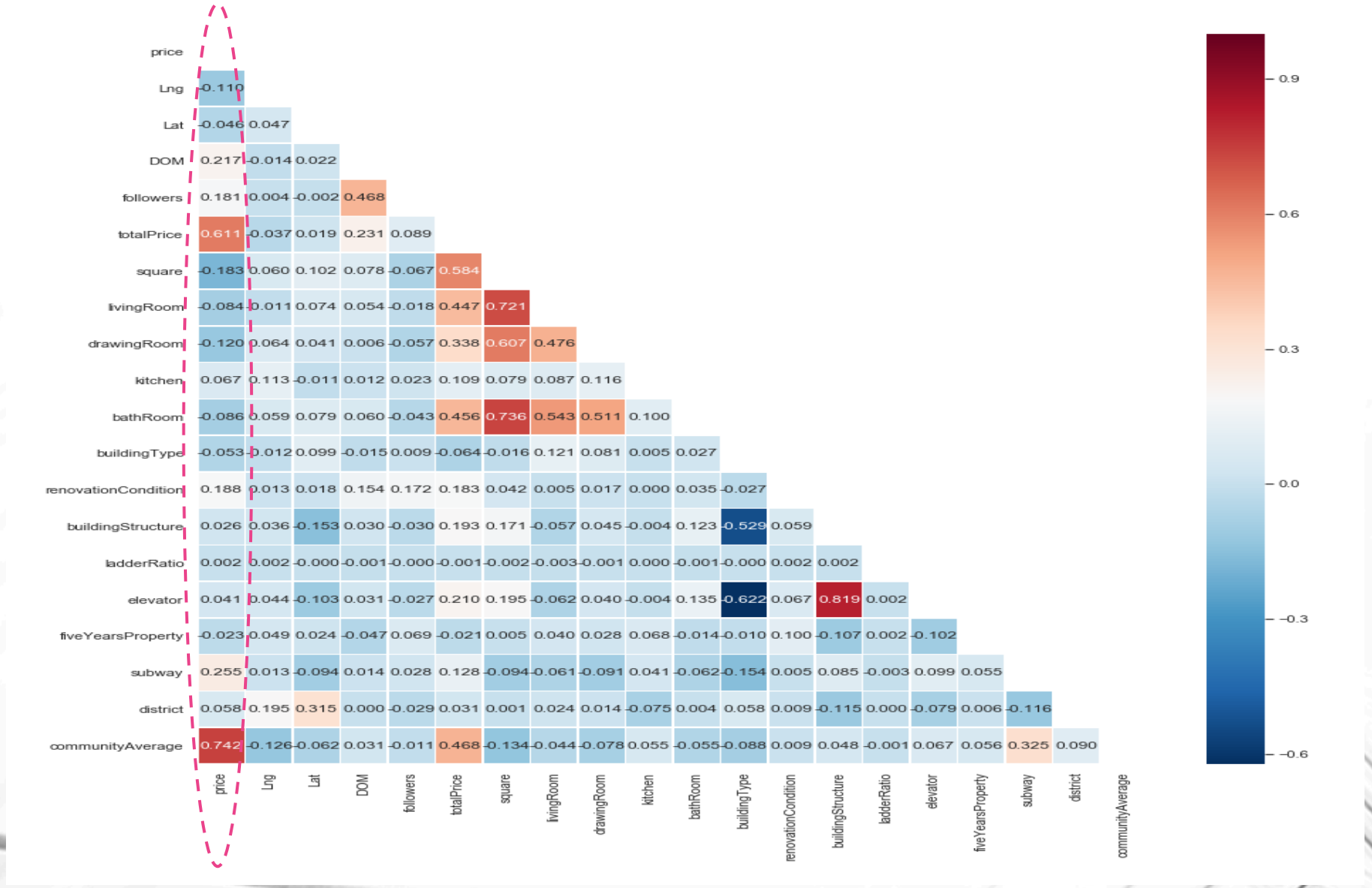


As seen from this predicted price trend, the total price won't change much.

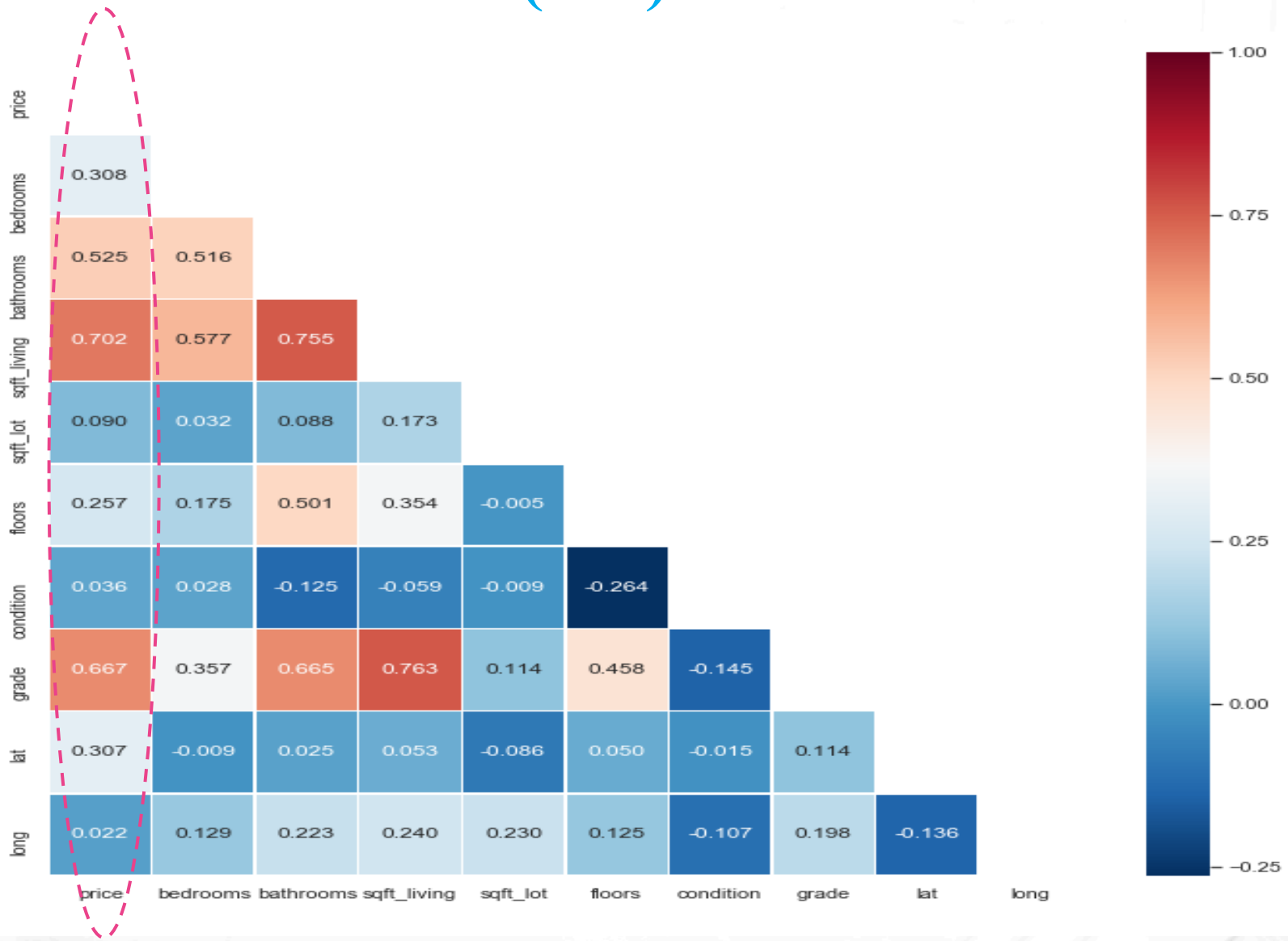
Regression

Predict the housing price

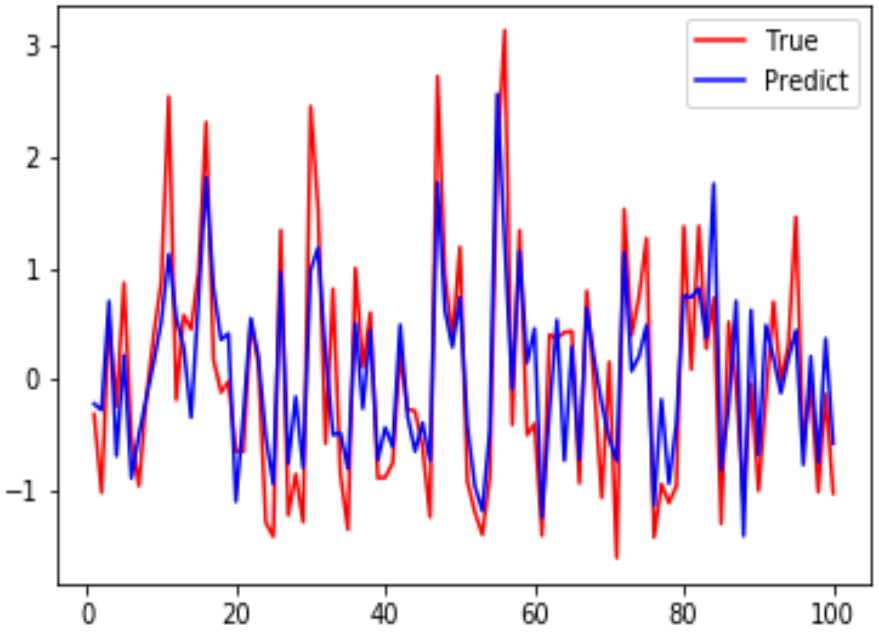
Feature correlation (Beijing)



Feature correlation (KC)

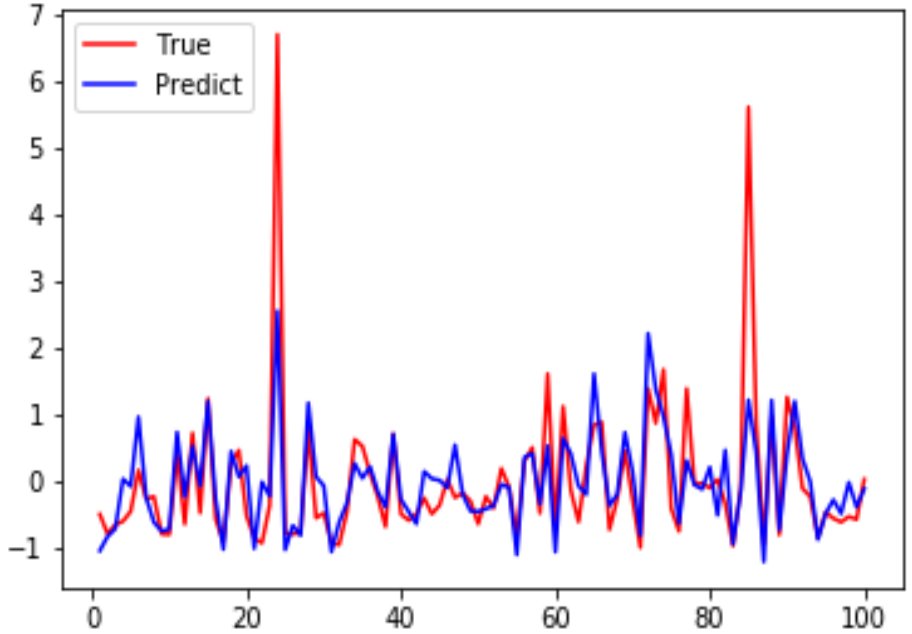


Linear regression



Beijing

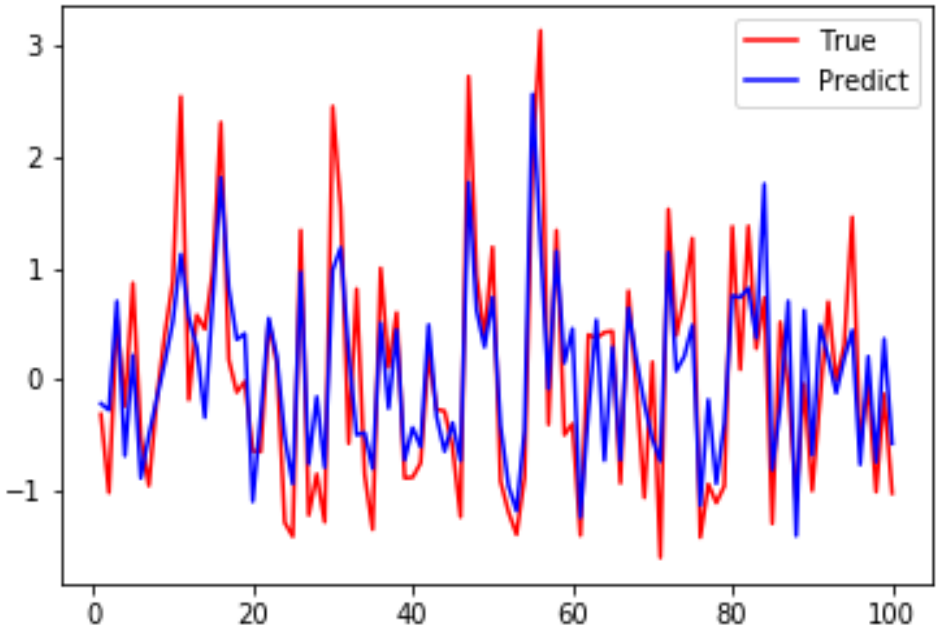
train error: 0.361
test error: 0.365



KC

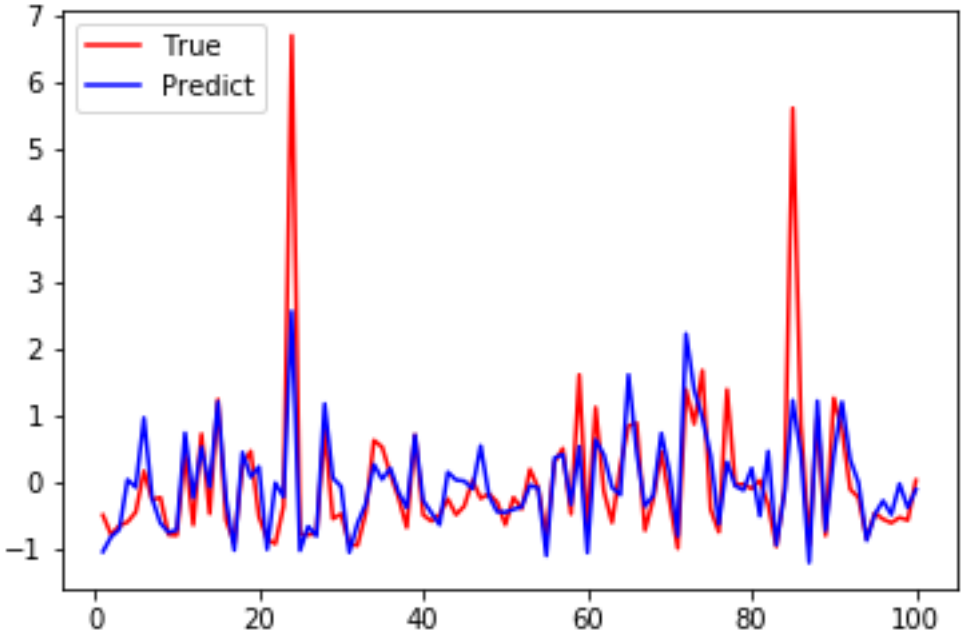
train error: 0.396
test error: 0.440

Linear SVM



Beijing

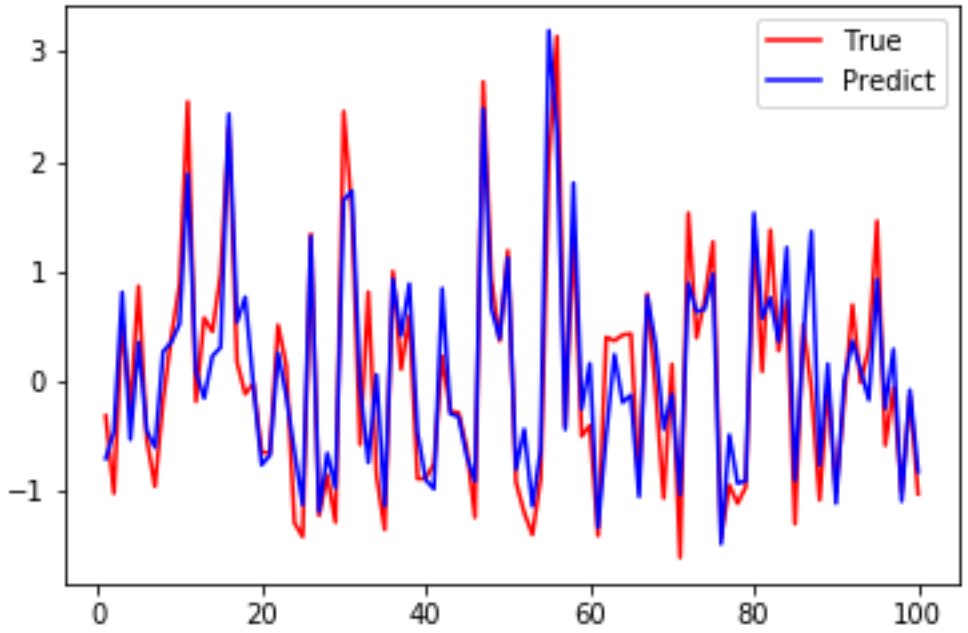
train error: 0.361
test error: 0.365



KC

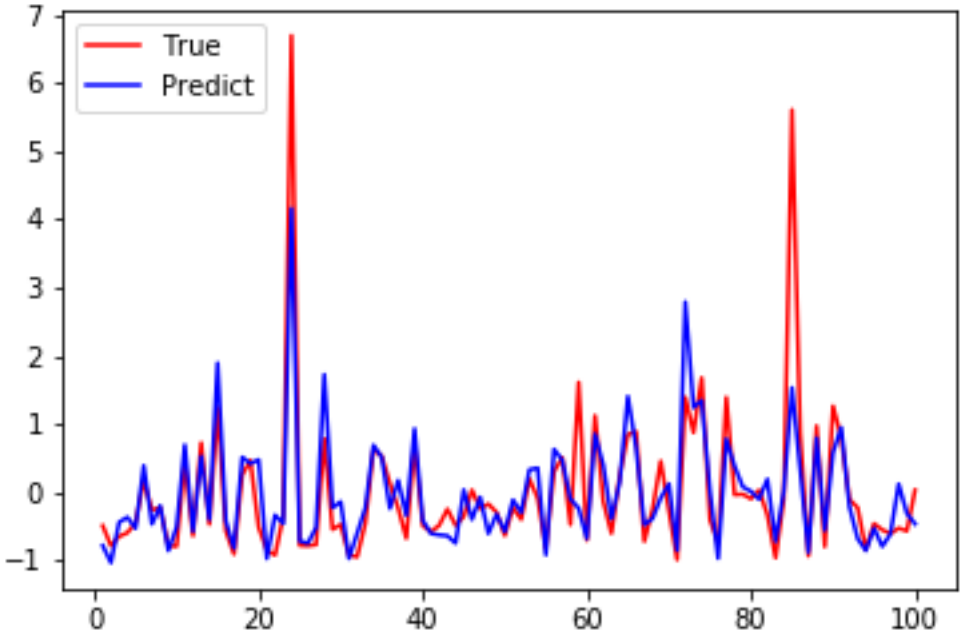
train error: 0.363
test error: 0.385

Neural Network(MLP)



Beijing

train error: 0.234
test error: 0.234



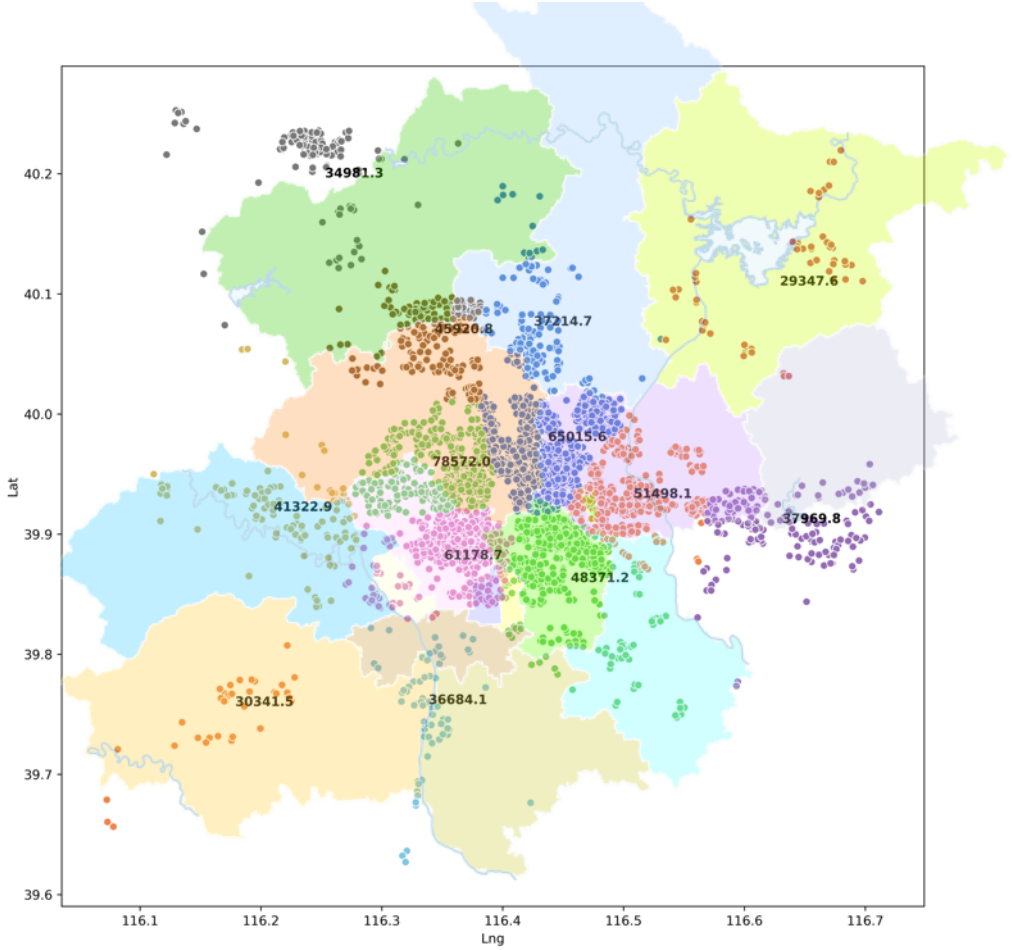
KC

train error: 0.245
test error: 0.275

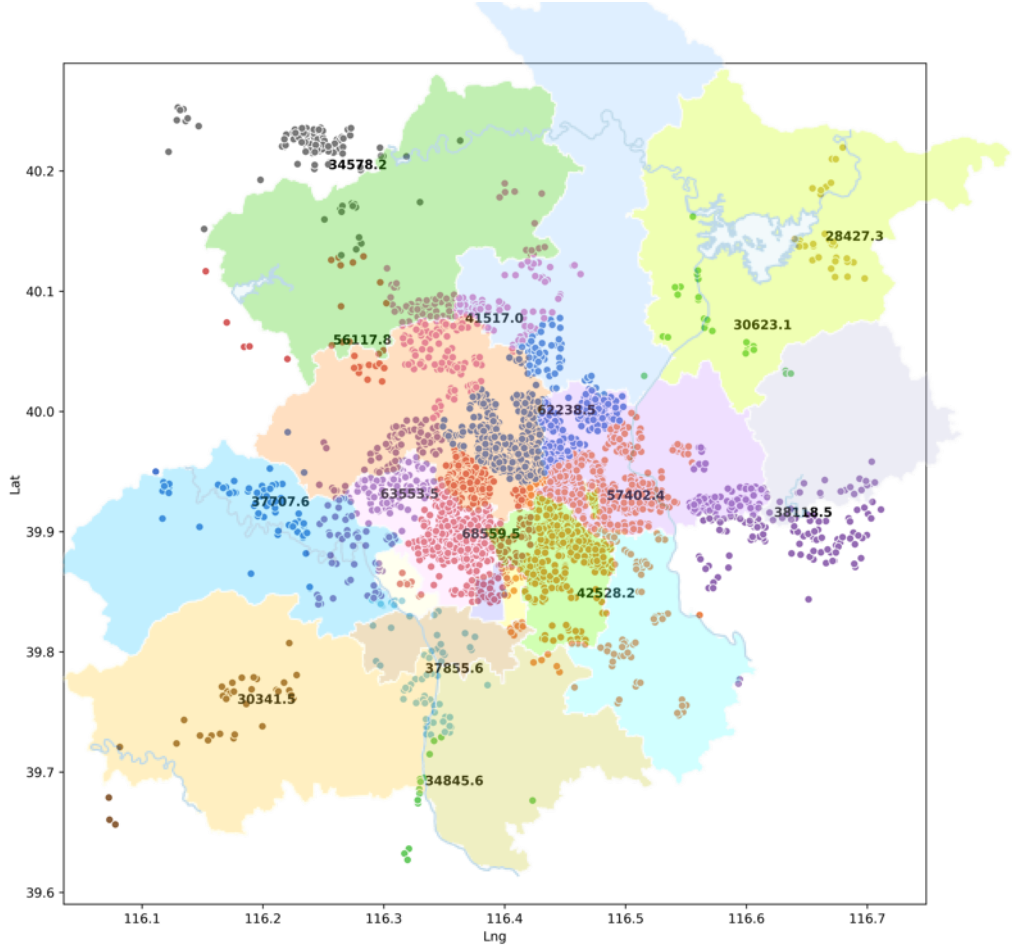
Classification

Predict the range of price

Clustering by latitude and longitude

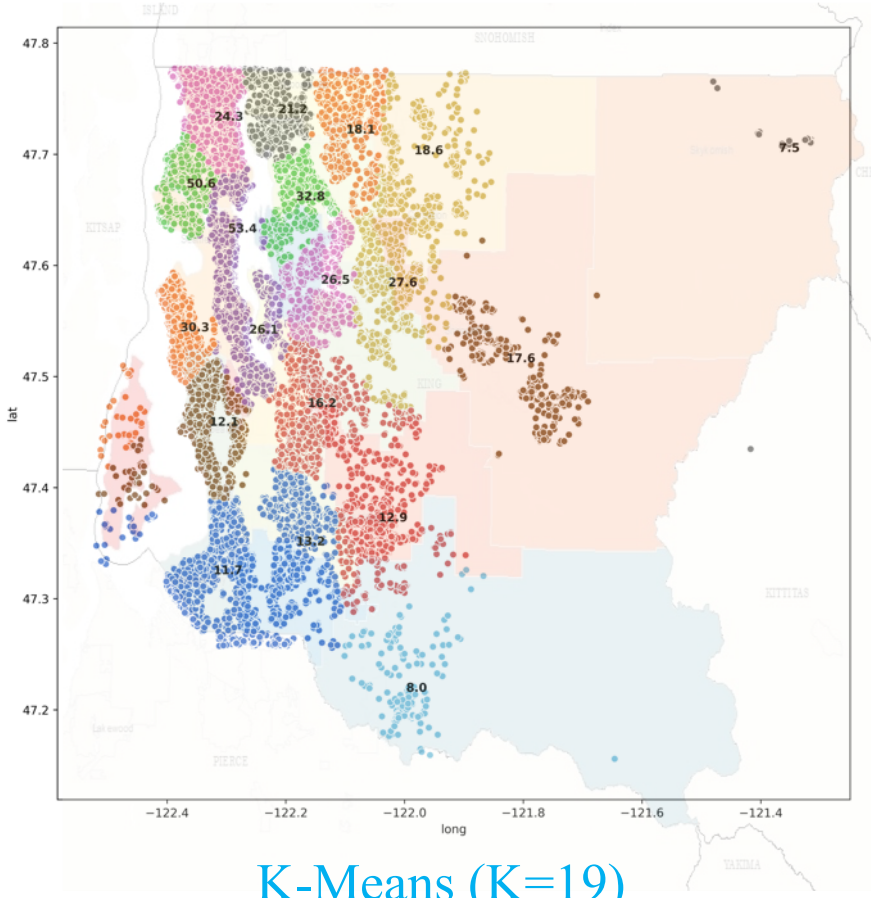


K-Means (K=13)

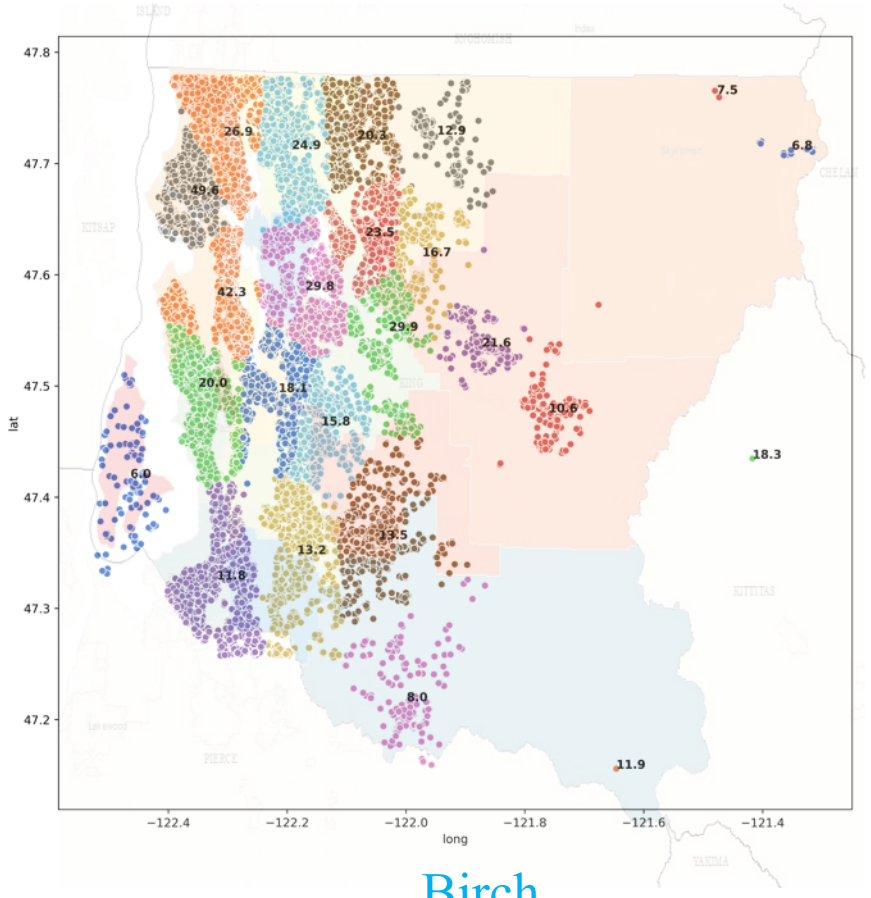


Birch

Clustering by latitude and longitude

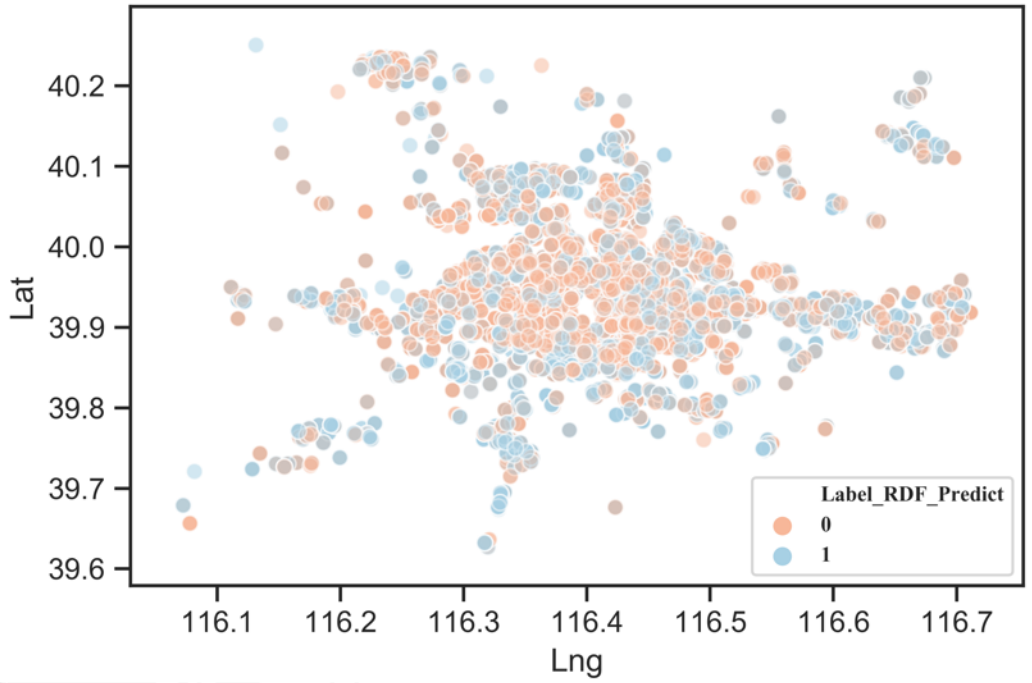


K-Means (K=19)

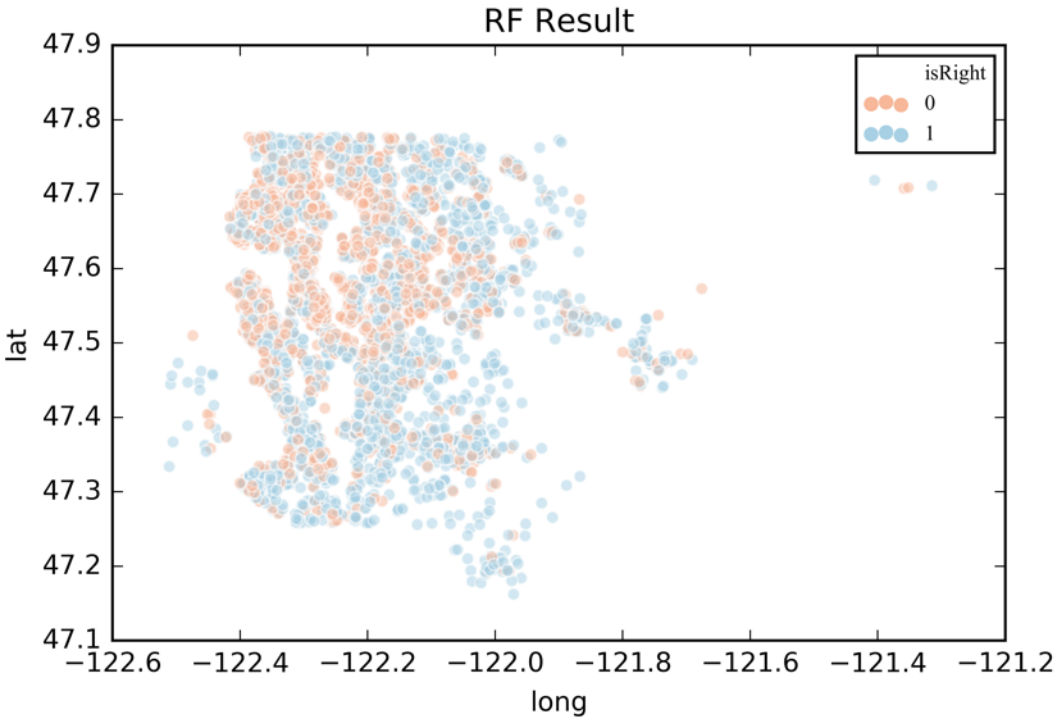


Birch

Random Forest

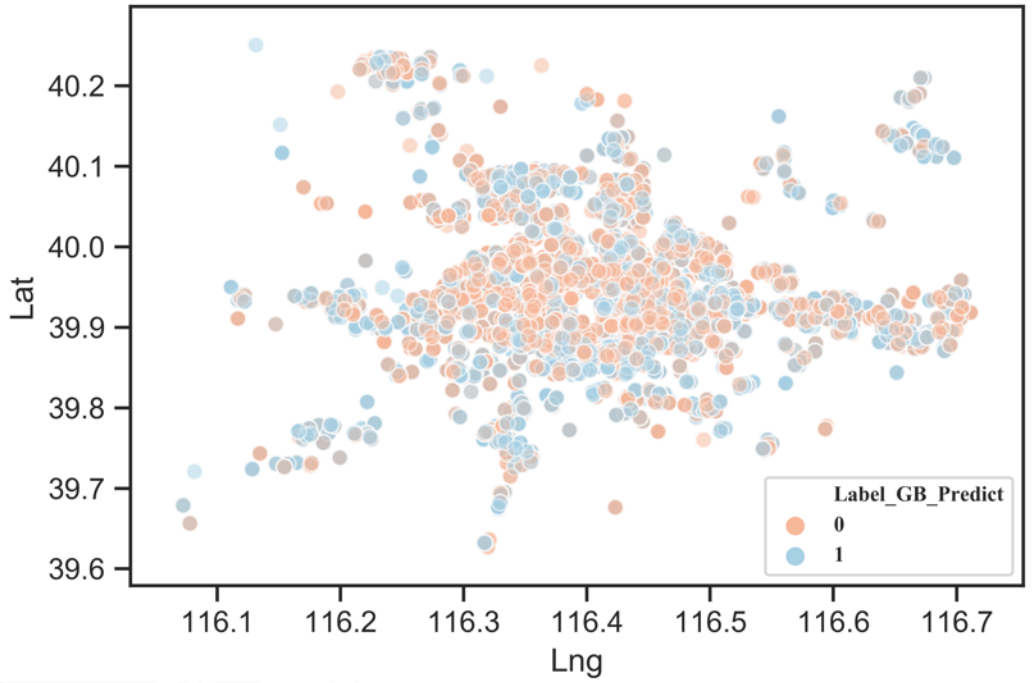


N_Estimator: 100
Beijing: 0.438

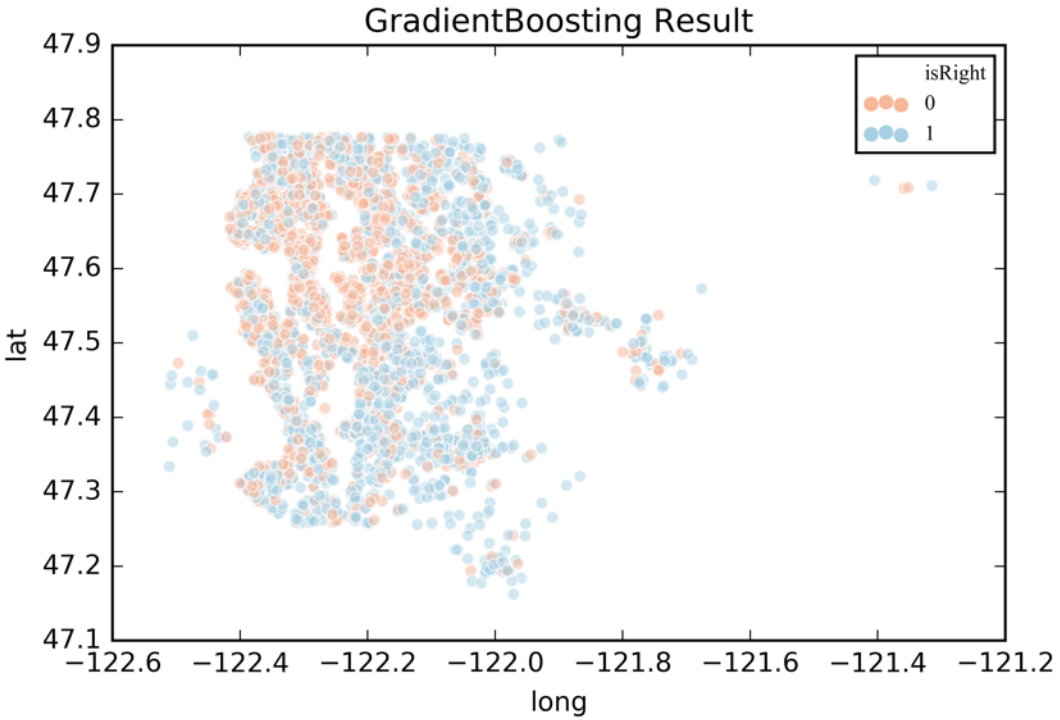


N_Estimator: 150
KC: 0.545

Gradient Boosting

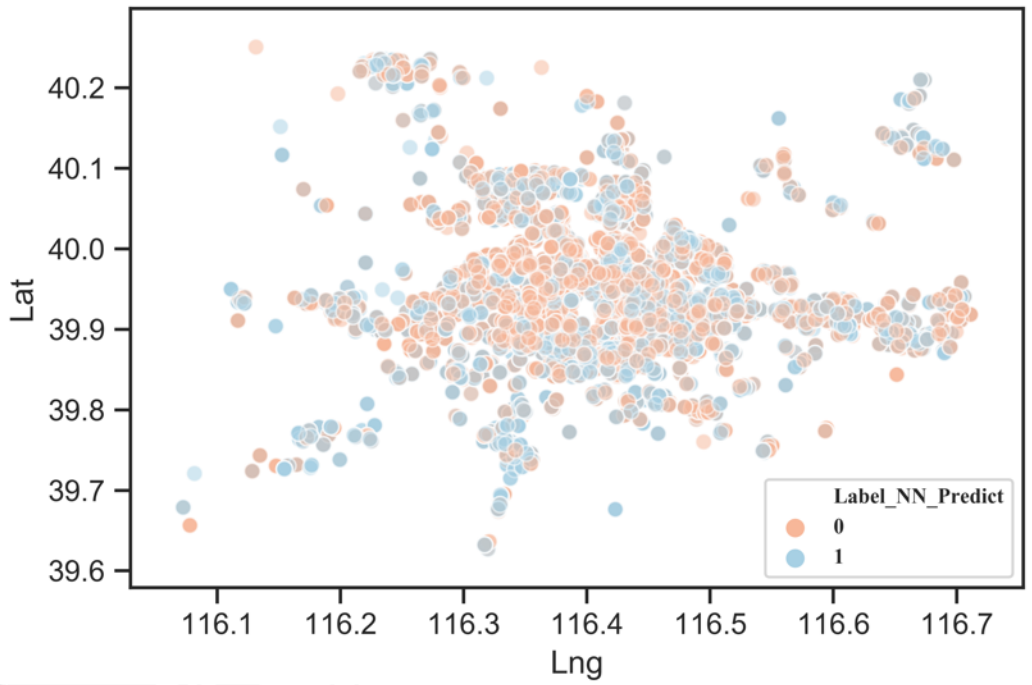


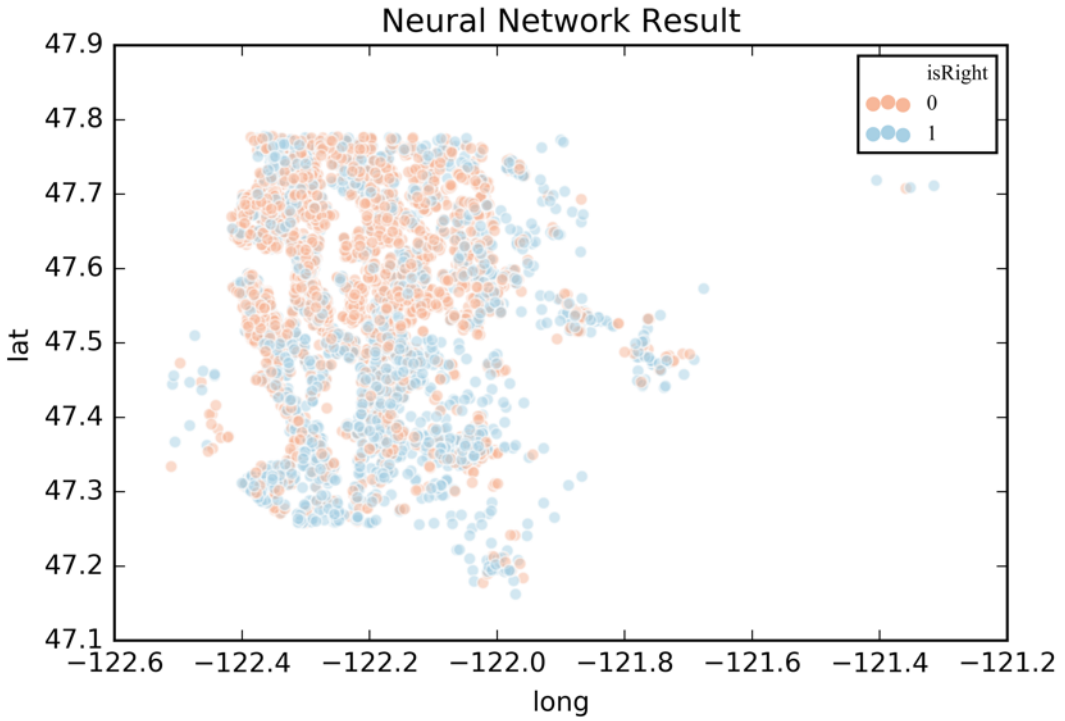
N_Estimator: 160
Learning rate = 0.15
Max_depth = 6
Beijing,: 0.461



N_Estimator: 150
Learning rate = 0.2
Max_depth = 5
KC: 0.527

Neural Network(MLP)

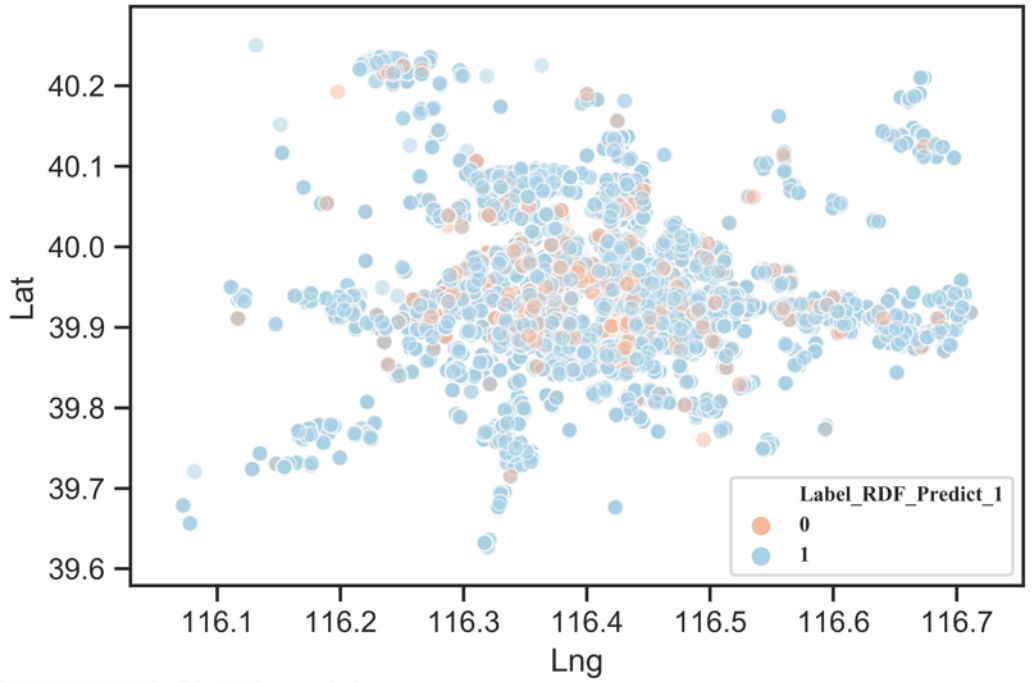




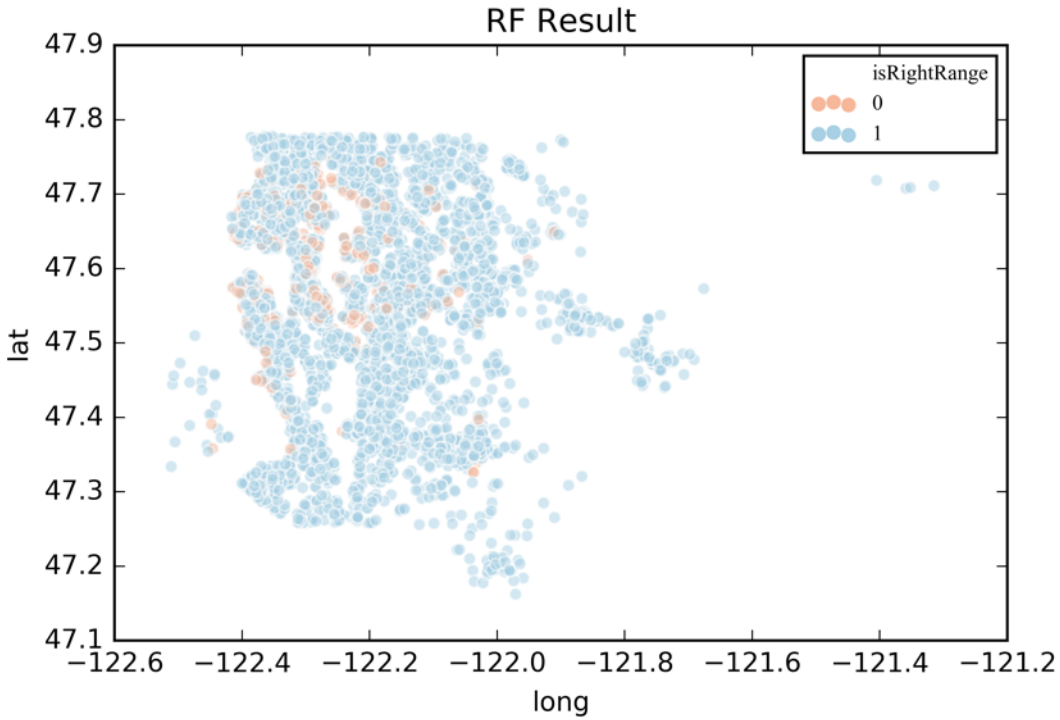
Solver: Adam
Hidden layer size: 163
Learning rate: Adaptive
Learning rate init: 0.01
Beijing: 0.432

Solver: Adam
Hidden layer size: 115, 3
Learning rate: Adaptive
Learning rate init: 0.01
KC: 0.425

Random Forest

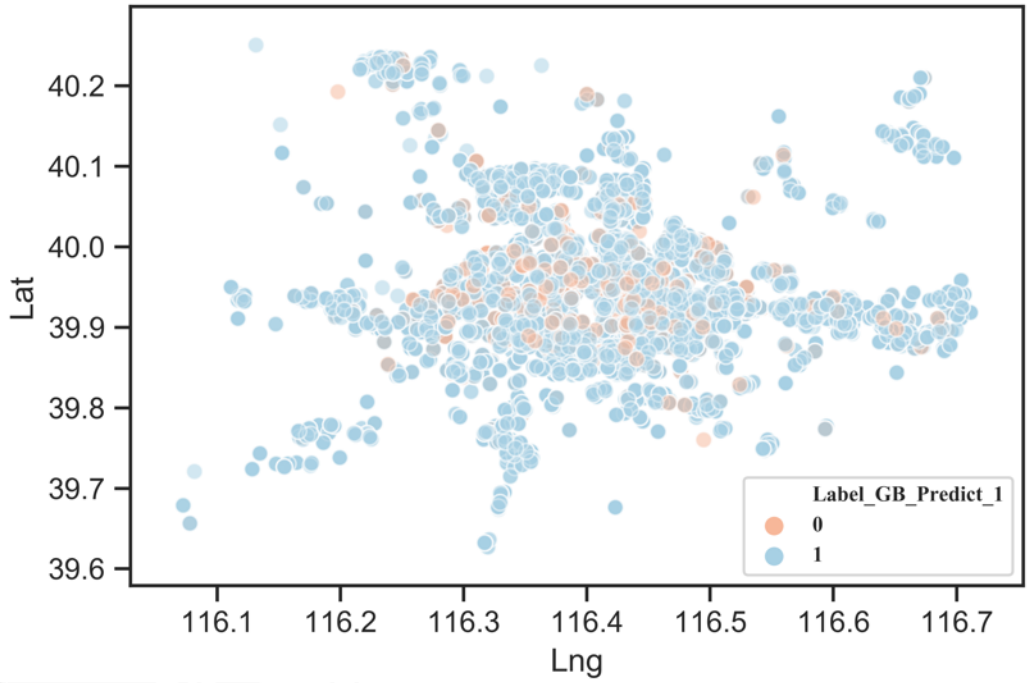


Beijing, 0.848

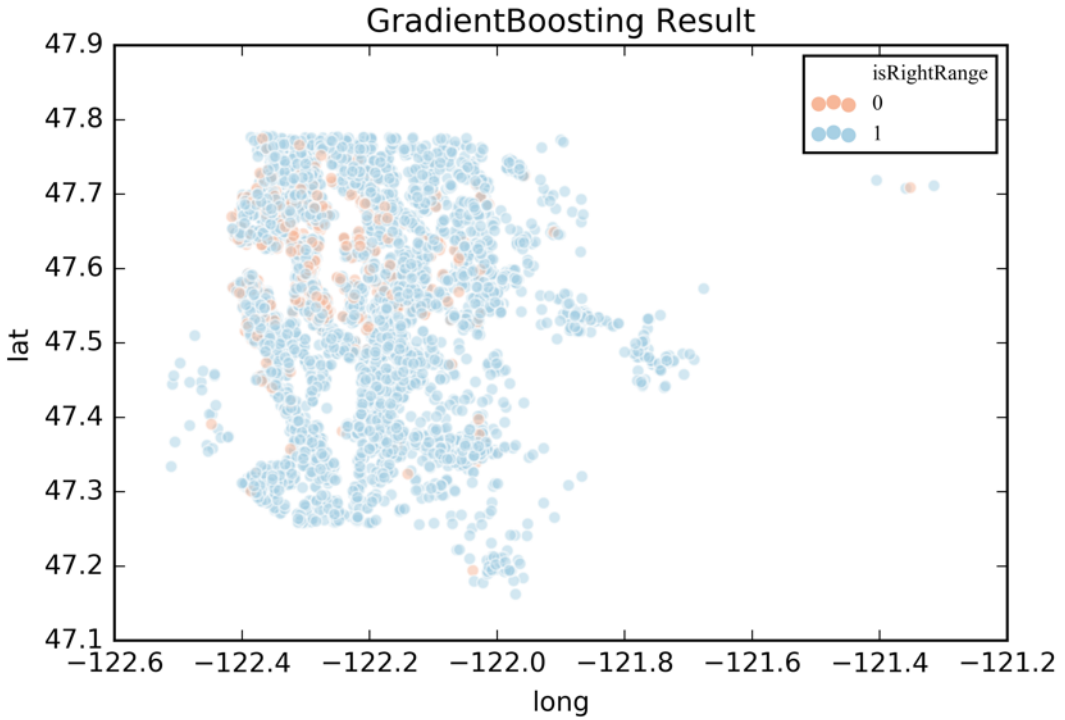


KC, 0.903

Gradient Boosting

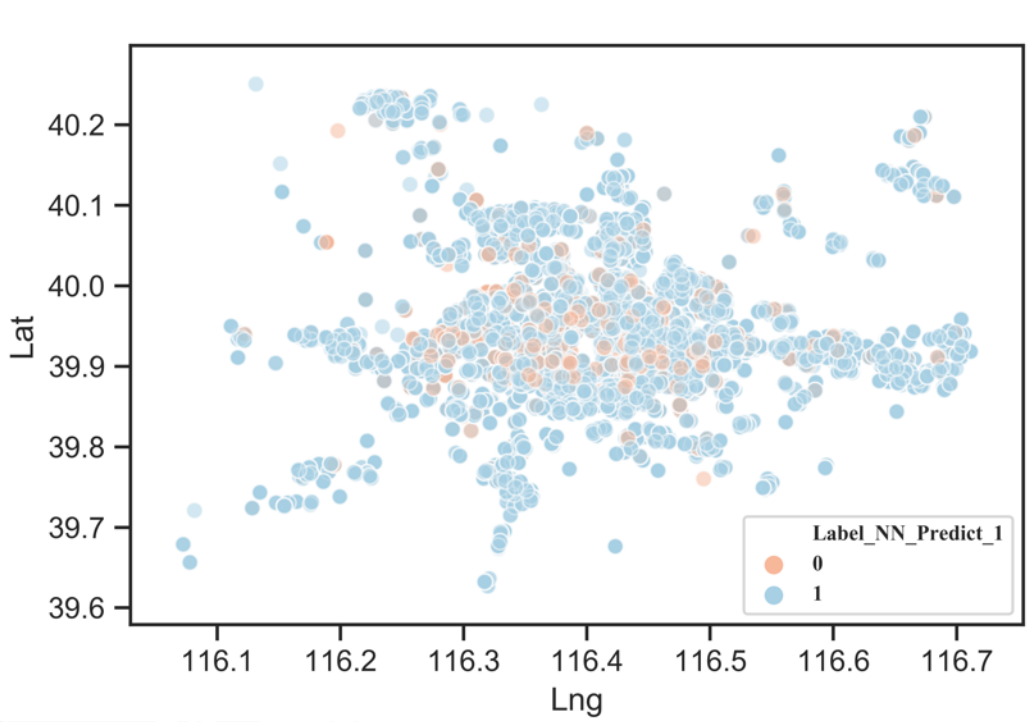


Beijing, 0.870

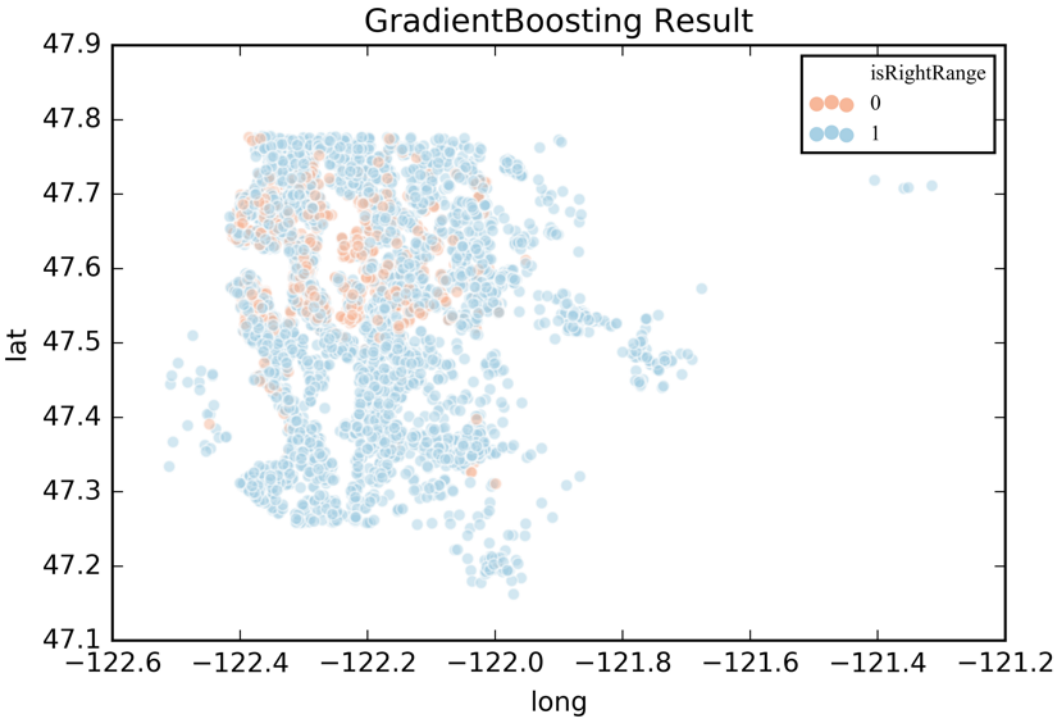


KC, 0.889

Neural Network



Beijing, 0.873



KC, 0.817

Conclusion

- **Anomaly detection** for detecting outliers
- **Time series forecasting** for predicting the housing price trend
- **Regression** for predicting the housing price with given features
- **Clustering** for grouping houses by latitude and longitude
- **Classification** for positioning the house price with given features into particular range

Thanks