

**EXPLORATORY ANALYSIS OF HOUSING
PRICE TREND ACROSS 13 US STATES
BETWEEN 2018 - 2022**

Authors

Raksha Ramaraj

Yogesh Gupta

Introduction

Real estate and the housing market play an important role in the U.S. economy. Housing markets experience substantial price volatility and short-term price change momentum. Real estate prices depend heavily on the market (location, location, location). Even within the same state, numbers can vary widely. Cities that experience new growth can show significant price appreciation, while others can be in decline. With homes getting bigger and inflation adding to the cost of building materials, it is only logical that home prices would rise.

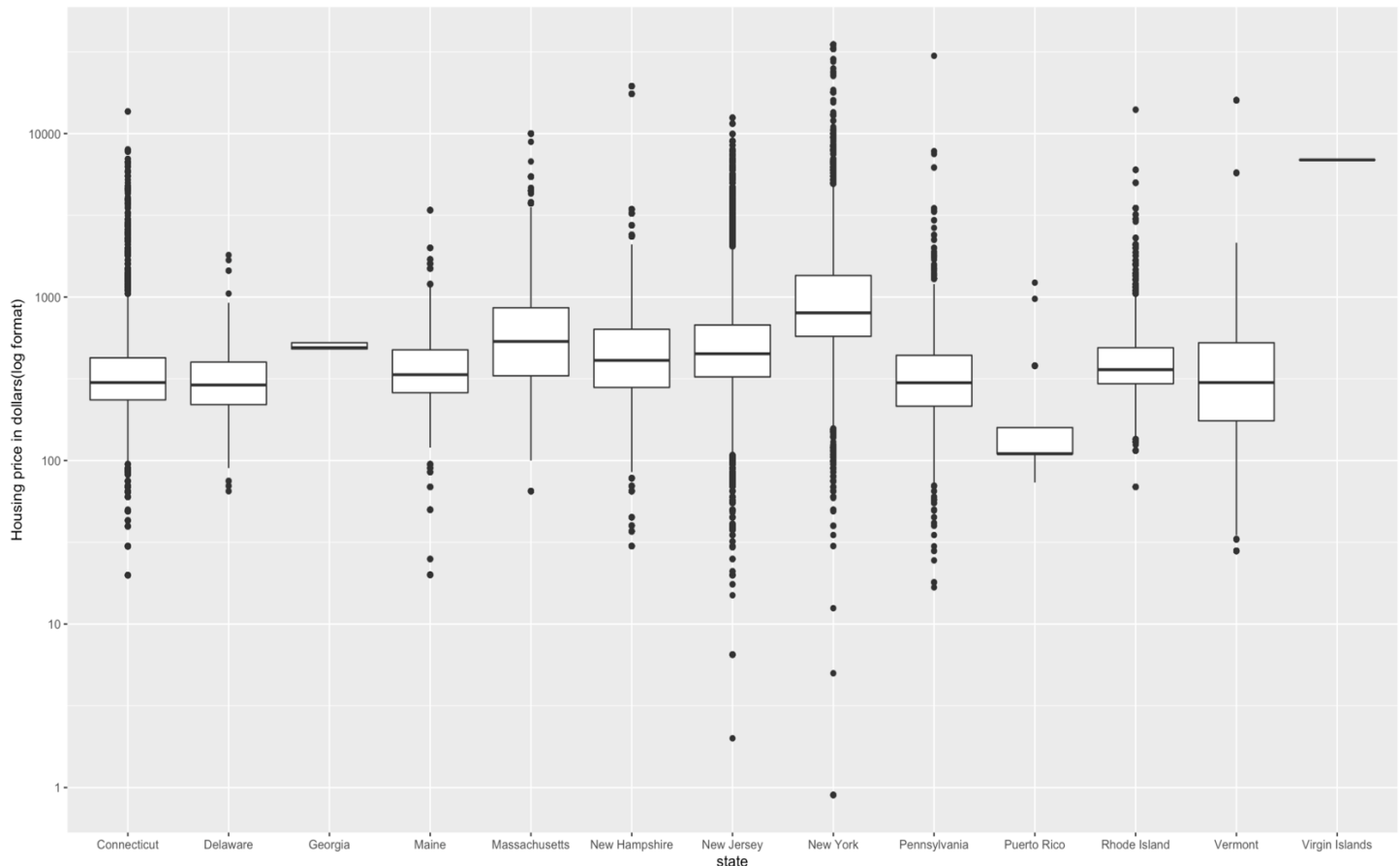
The dataset we have selected for our project looks at the housing market trend across 13 states namely New York, New Jersey, Virgin Islands, Massachusetts, New Hampshire, Vermont, Georgia, Rhode Island, Maine, Connecticut, Pennsylvania, Delaware and Puerto Rico. We have performed exploratory analysis on our dataset to understand the data better and perform some visual interpretation of the same.

Source of the data: The dataset was obtained from Kaggle.

Dimension: 106064 rows and 13 columns.

Variables	Description
1. <i>Status</i>	Status of the house when it was added to market
2. <i>Price</i>	Selling Price of the house in dollars
3. <i>Bed</i>	Number of bedrooms in the house
4. <i>Bath</i>	Number of bathrooms in the house
5. <i>Acre_lot</i>	Lot size
6. <i>Year</i>	Full Address of the house
7. <i>Street</i>	Street location of the house
8. <i>City</i>	City in which the house is located
9. <i>State</i>	State in which the house is sold
10. <i>Zip code</i>	Zipcode of the place where house is sold
11. <i>House_size</i>	Size of the house
12. <i>Sold_date</i>	Date on which the house was sold
13. <i>Full_address</i>	Year of which house was sold

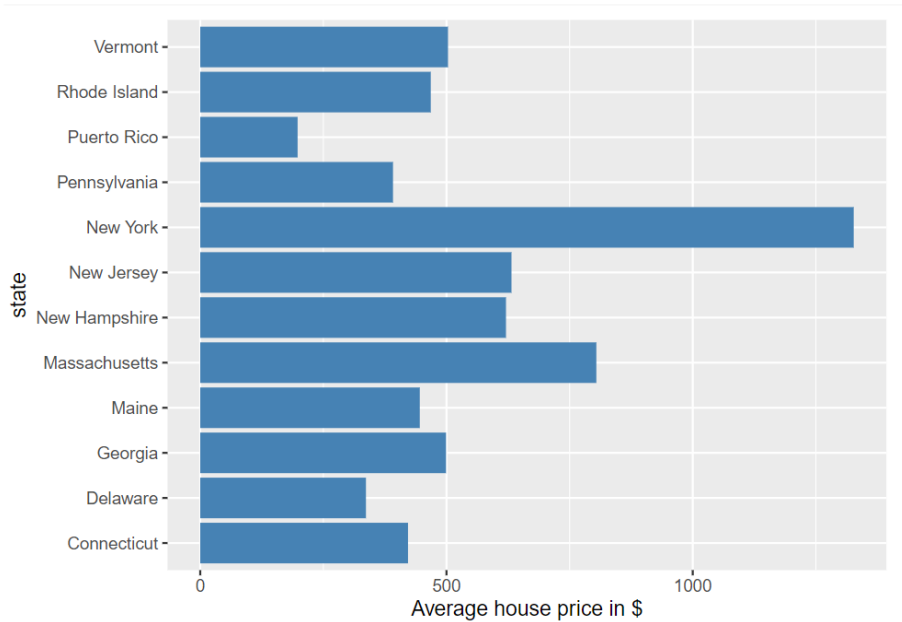
BOX PLOT - HOUSING PRICES v/s DIFFERENT STATES:



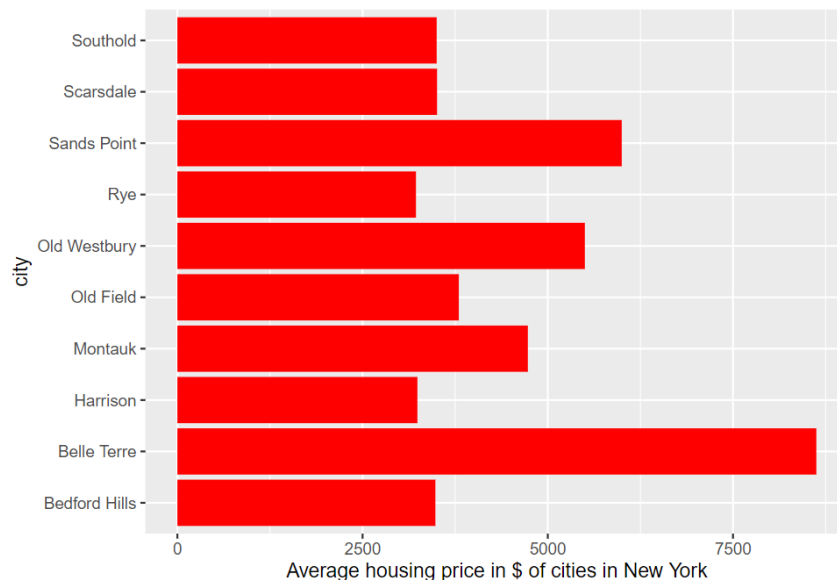
In the above plot we are trying to analyze prices over different states for five years. We can see that most of the house prices are between \$100k to \$1M. We see that New York state has the highest median compared to other states. The least median is observed for the Puerto Rico state.

Plot of average housing prices over different states

The Virgin Islands data has been removed, since it has only three data points and it has the maximum average selling price of \$6.8M. In the below plot, we can see that New York has the highest average selling price.



ANALYSIS OF HOUSING PRICES ACROSS CITIES IN NEW YORK STATE



Based on previous analysis, we noticed that the state of New York had the highest average housing price over 5 years. The above plot shows us the pricing trend in different cities in New York state.

- The city of Belle Terre, though had only two houses sold over 5 years, had the highest average price of \$8.6M across the state.
- Among the above considered cities, Scarsdale had higher number of houses sold – 46, the average price being \$3.5M.

RANDOM FOREST ANALYSIS

Random Forest analysis is used to determine the effect of 7 predictor variables (see VIP plot below for exact variables considered) on the response variable (housing price) across New York state.

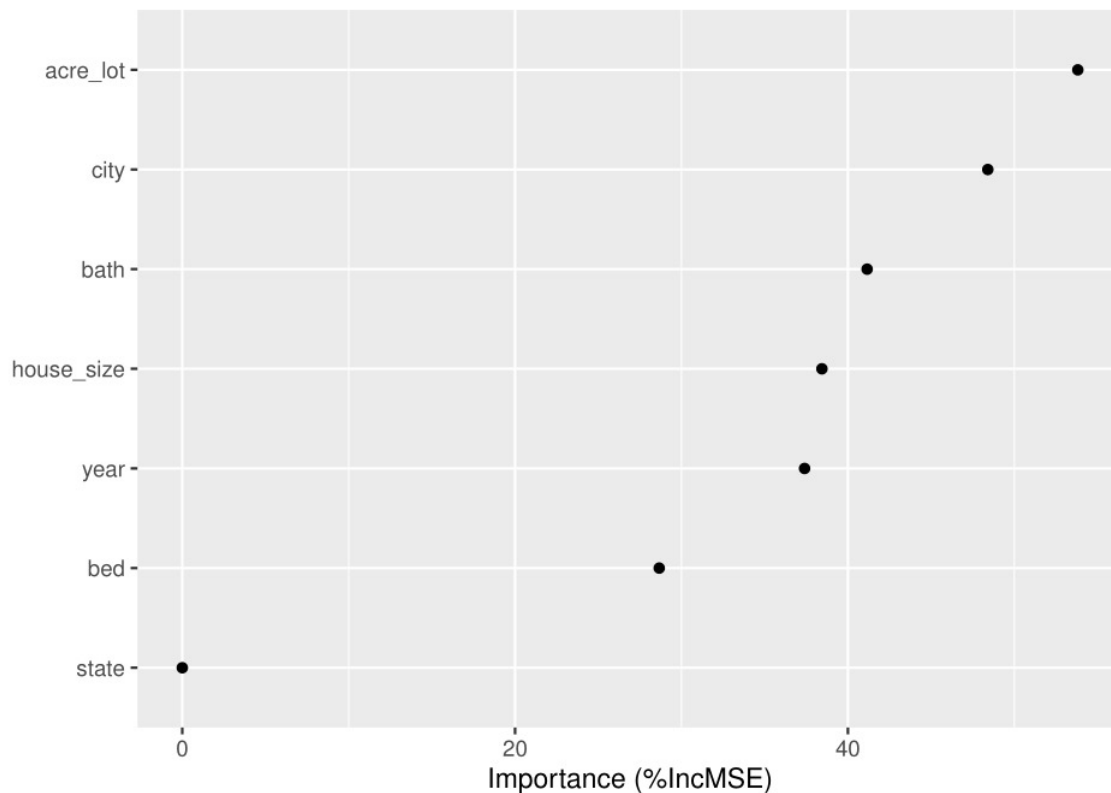
Our Random Forest model shows us that 95% of the variance is explained by our out-of-bag predictions.

```
Call:
randomForest(formula = price ~ ., data = newyork2, importance = TRUE)
Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 2

Mean of squared residuals: 70342497584
% Var explained: 95.1
```

VIP PLOT

This plot shows us the importance of different variables in our model. We see that the size of the lot is of maximum importance, while state seem least important that affects the housing price in state of New York.



CONCLUSION

These are the following observations made during the analysis of the housing prices over 5 years across the 13 states:

- New York had the highest average housing price - \$1.33M.
- New Jersey recorded the highest of houses being sold – 41294
- Looking closely at the state of New York:
 - City of Belle Terre had the highest average housing price - \$8.63M
 - New York city had the highest number of houses sold – 4496
- Random Forest Analysis showed 95% of variance has been explained by this model and using VIP plot, we see that the plot size of the house plays a significant role in it's selling price.