# State and Region Based Housing Price Prediction Report

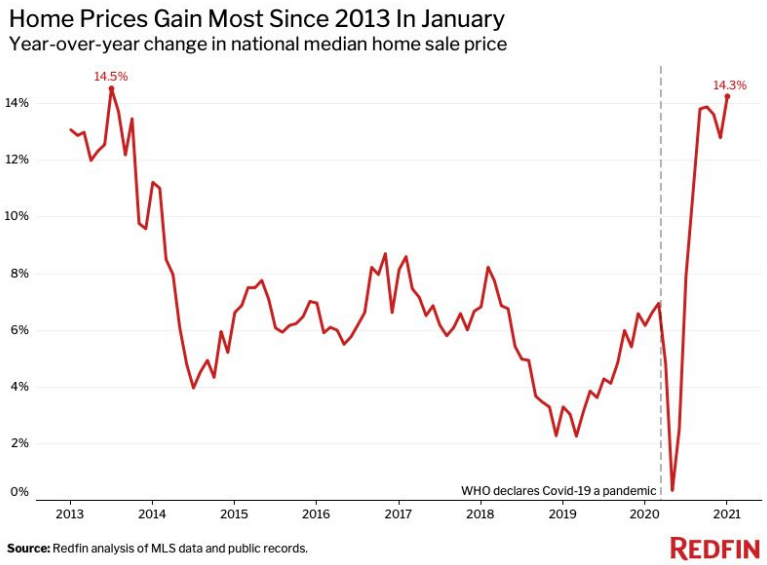
Justin Nichols
Andrew Rodrigue
Zeke Abshire

https://towardsdatascience.com/machine-learning-predicting-house-prices-with-regression-2388bb876a6f

Justin

# Current Problem

- As of late 2021, the house market within the U.S. has become an unlikely beneficiary of the COVID-19 pandemic.
- Presently, the housing market is booming, with properties being sold in record time. With this, home prices have climbed at a fast pace.
- There is a boost in demand, low mortgage rates, supply is shrinking, and prices are increasing.




https://www.redfin.com/news/coronavirus-pandemic-real-estate-impact-charts/

Justin

# Goal and Motivation

- The motivating factor behind the development of this housing predictive model was to provide prospective home buyers transparency in housing prices around the country.

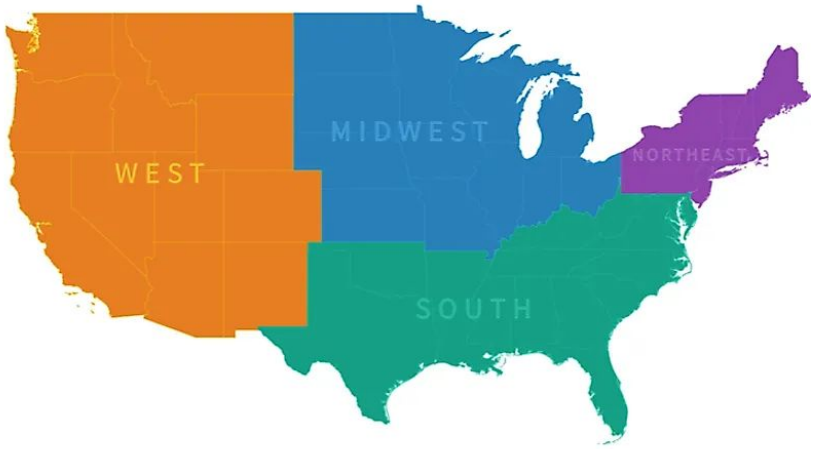

https://www.latimes.com/business/story/2021-04-14/real-estate-bidding-wars-hot-market-above-listing-price

Justin

# Predictive Models

- By providing users the ability to pick a region or state and compare home prices based on their personal information, this model enables them to quickly perform accurate research and make educated decisions on where they want to live.

- Models used include KNN and Linear Regression to predict housing prices.




https://www.worldatlas.com/articles/the-regions-of-the-united-states.html

Justin

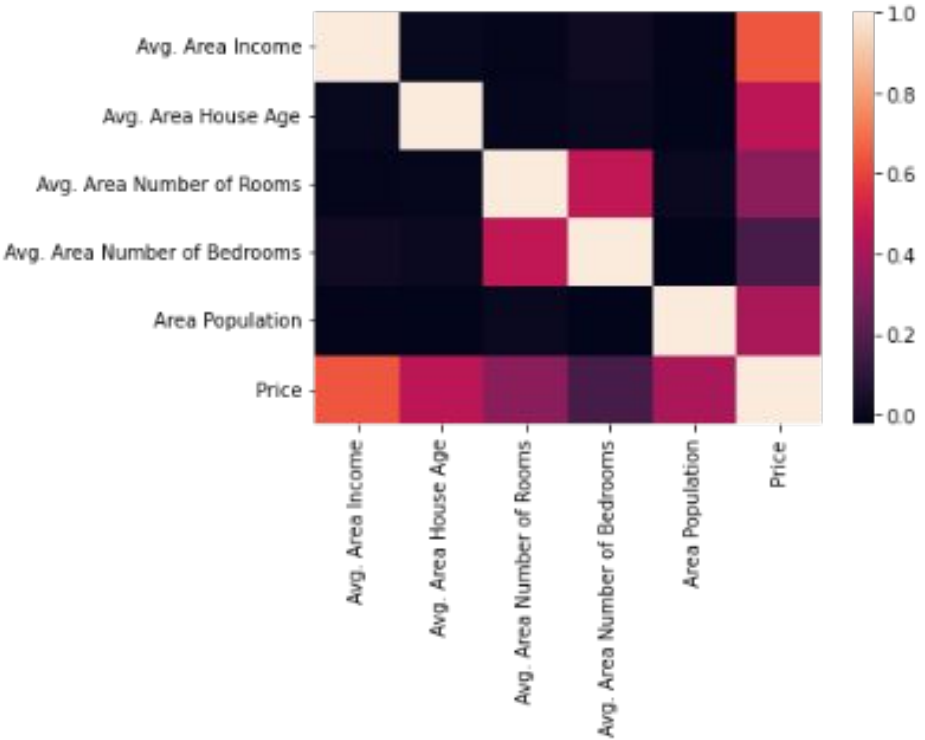
# Dataset

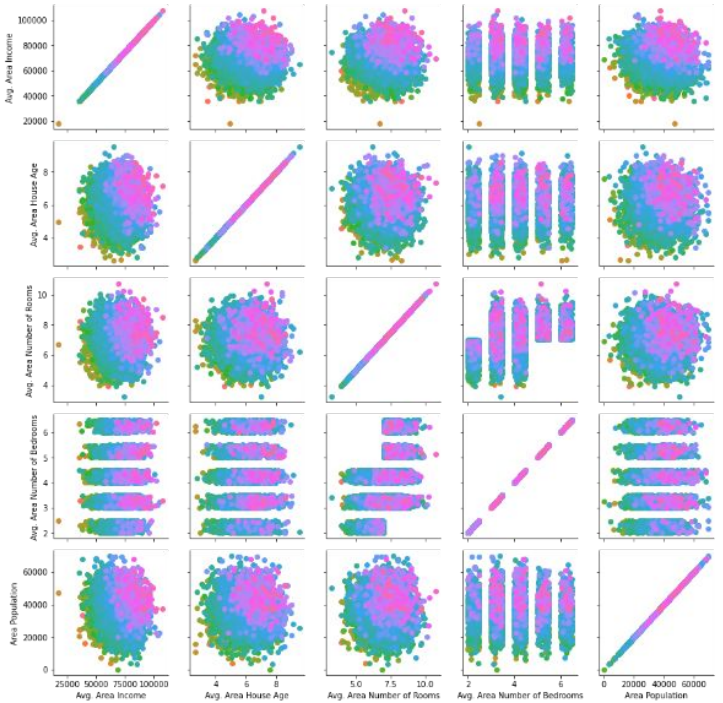
- The data was obtained from Kaggle.
- It is stored in a CSV file format.
- There are 5000 total entries.
- The attributes are area income, house age, number of rooms, number of bedrooms, area population, price, and address.

| | Avg. Area Income | Avg. Area House Age | Avg. Area Number of Rooms | Avg. Area Number of Bedrooms | Area Population | Price | Address |
|---|---|---|---|---|---|---|---|
| 0 | 79545.458574 | 5.682861 | 7.009188 | 4.09 | 23086.800503 | 1.059034e+06 | 208 Michael Ferry Apt. 674\nLaurabury, NE 3701... |
| 1 | 79248.642455 | 6.002900 | 6.730821 | 3.09 | 40173.072174 | 1.505891e+06 | 188 Johnson Views Suite 079\nLake Kathleen, CA... |
| 2 | 61287.067179 | 5.865890 | 8.512727 | 5.13 | 36882.159400 | 1.058988e+06 | 9127 Elizabeth Stravenue\nDanieltown, WI 06482... |
| 3 | 63345.240046 | 7.188236 | 5.586729 | 3.26 | 34310.242831 | 1.260617e+06 | USS Barnett\nFPO AP 44820 |
| 4 | 59982.197226 | 5.040555 | 7.839388 | 4.23 | 26354.109472 | 6.309435e+05 | USNS Raymond\nFPO AE 09386 |

Zeke

# Attribute Correlations



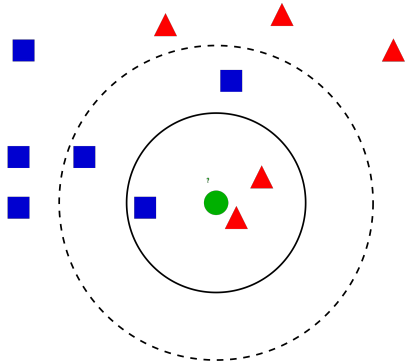

Zeke

# KNN

- The model divides the relevant entries into price categories of $150,000.
- Data is normalized and best value for KFold and KNN is found.
- The user inputs their income and desired house age, number of rooms, number of bedrooms, and population of area they are prospecting.
- The model outputs a predicted price range using this information.

```
KNN: $2100000.0 - $2250000.0 | Classifier output: 14.0
```

https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

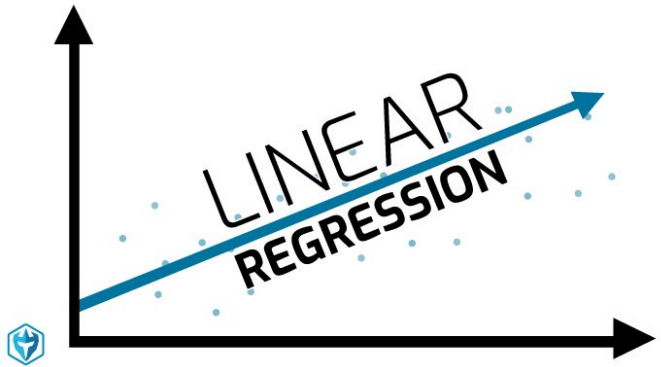


Zeke

# Linear Regression

- The user inputs their income and desired house age, number of rooms, number of bedrooms, and population of area they are prospecting.
- The model outputs a predicted price range using this information.

```
Linear Regression: $1470000
```



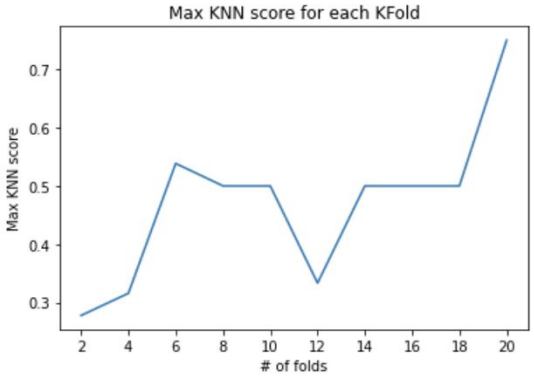

https://www.warriortrading.com/linear-regression-definition-day-trading-terminology/

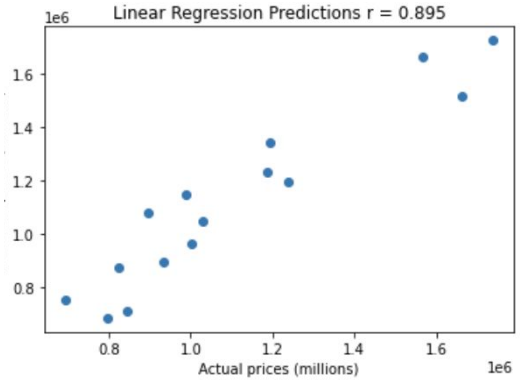
Zeke

# Experiment 1

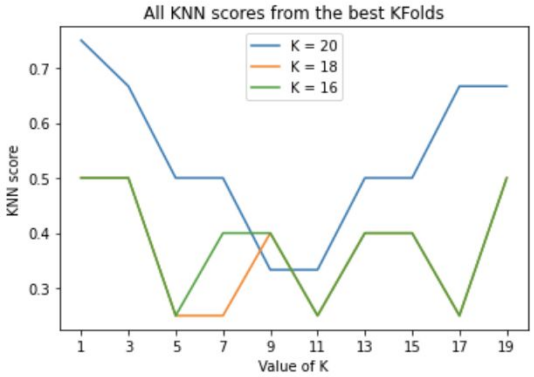
For Louisiana,

```
Obtain housing price predictions.
Enter income: 50000
Enter house age: 8
Enter number of rooms: 5
Enter number of bedrooms: 1
Enter population: 10000
Select an Option:
1.U.S.
2.State
3.Region
2
Input a state abbreviation: la
House Prediction for LA:
```




```
near Regression: $530000
```




```
KNN: $1650000.0 - $1800000.0 | Classifier output: 11.0
```

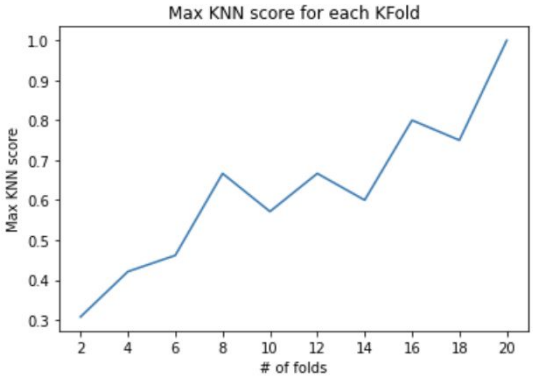
Note, the data used is not based on real prices for the area.

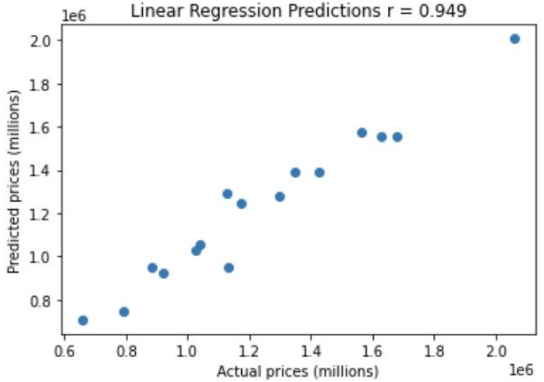
Zeke

# Experiment 2

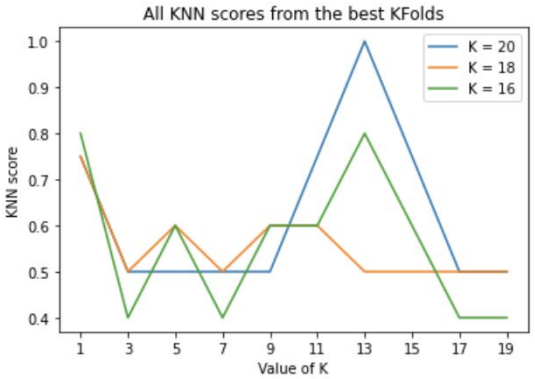
For California,

```
Obtain housing price predictions.
Enter income: 120000
Enter house age: 3
Enter number of rooms: 8
Enter number of bedrooms: 3
Enter population: 50000
Select an Option:
1.U.S.
2.State
3.Region
2
Input a state abbreviation: ca
House Prediction for CA:
```




```
Linear Regression: $2300000
```




```
KNN: $1500000.0 - $1650000.0 | Classifier output: 10.0
```

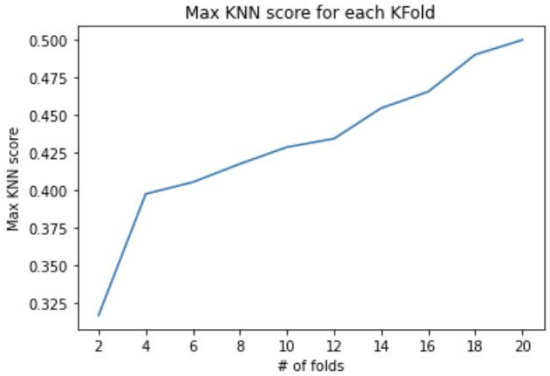
Note, the data used is not based on real prices for the area.

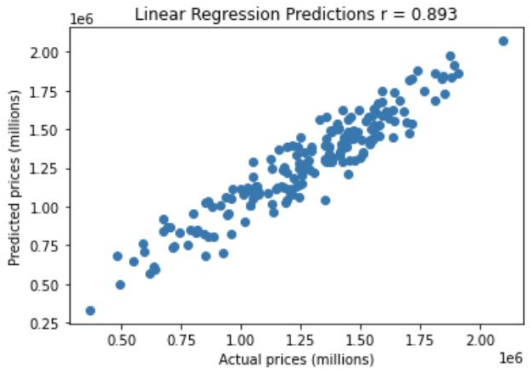
Andrew

# Experiment 3

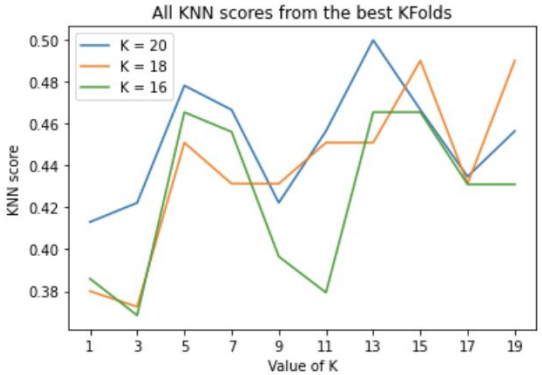
For Northeast,

```
Obtain housing price predictions.
Enter income: 80000
Enter house age: 6
Enter number of rooms: 7
Enter number of bedrooms: 3
Enter population: 35000
Select an Option:
1.U.S.
2.State
3.Region
3
Input a region (northeast, midwest, south, west):
northeast
House Prediction for northeast:
```




```
Linear Regression: $1460000
```




```
KNN: $1800000.0 - $1950000.0 | Classifier output: 12.0
```

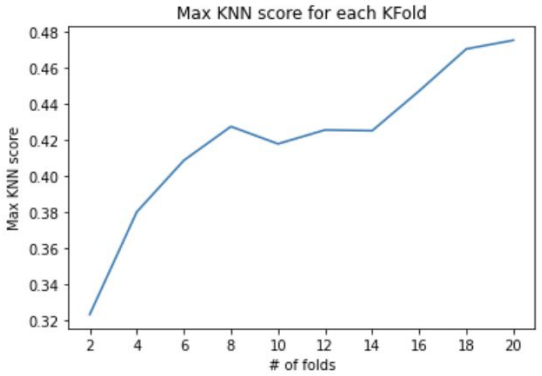
Note, the data used is not based on real prices for the area.

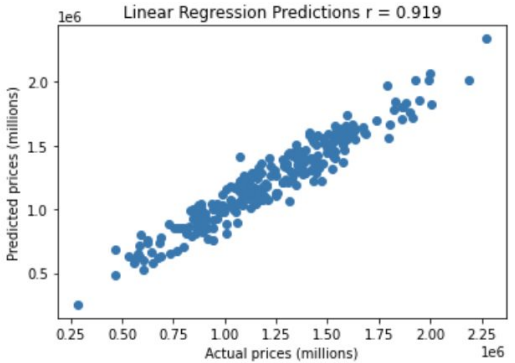
Andrew

# Experiment 4

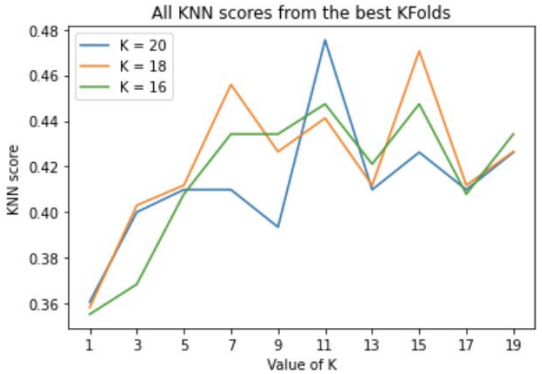
For South,

```
Obtain housing price predictions.
Enter income: 80000
Enter house age: 6
Enter number of rooms: 7
Enter number of bedrooms: 3
Enter population: 35000
Select an Option:
1.U.S.
2.State
3.Region
3
Input a region (northeast, midwest, south, west):
south
House Prediction for south:
```




```
Linear Regression: $1470000
```




```
KNN: $1800000.0 - $1950000.0 | Classifier output: 12.0
```

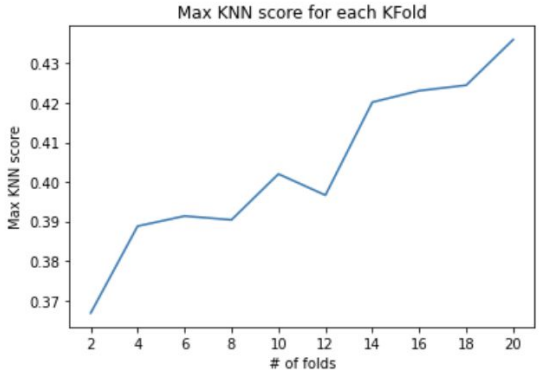
Note, the data used is not based on real prices for the area.

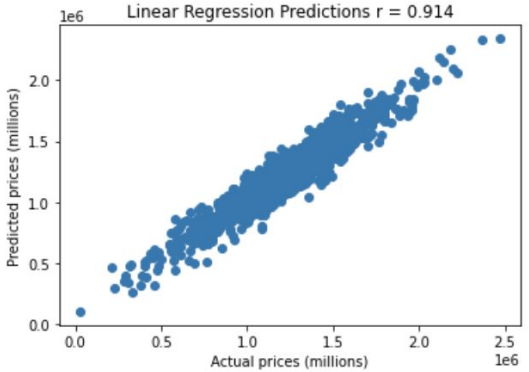
Andrew

# Experiment 5

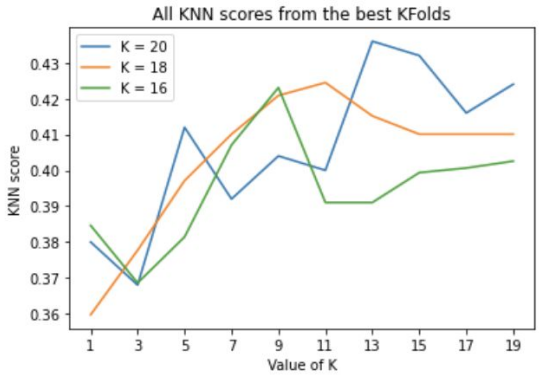
For U.S.,

```
Obtain housing price predictions.
Enter income: 80000
Enter house age: 6
Enter number of rooms: 7
Enter number of bedrooms: 3
Enter population: 35000
Select an Option:
1.U.S.
2.State
3.Region
1
House Price Prediction for US:
```




```
Linear Regression: $1460000
```




```
KNN: $2100000.0 - $2250000.0 | Classifier output: 14.0
```

Note, the data used is not based on real prices for the area.

Andrew

# Conclusions and What We Learned

- Linear Regression was found to be more accurate in predicting the housing price for a certain area than KNN.
- KNN accuracy: low to mid 40 percent for all experiments.
- Linear Regression accuracy: high 80 to low 90 percent for all experiments.
- The insights made from the experiments helped our group to conclude that some states or regions have on average higher home prices.


Andrew