Tax prediction a knn regression approach

Presentation by
Anyinssan Nava

The initial idea was to fit a regression model to predict taxes based on lot acres, year_built, and sqrt_ft.

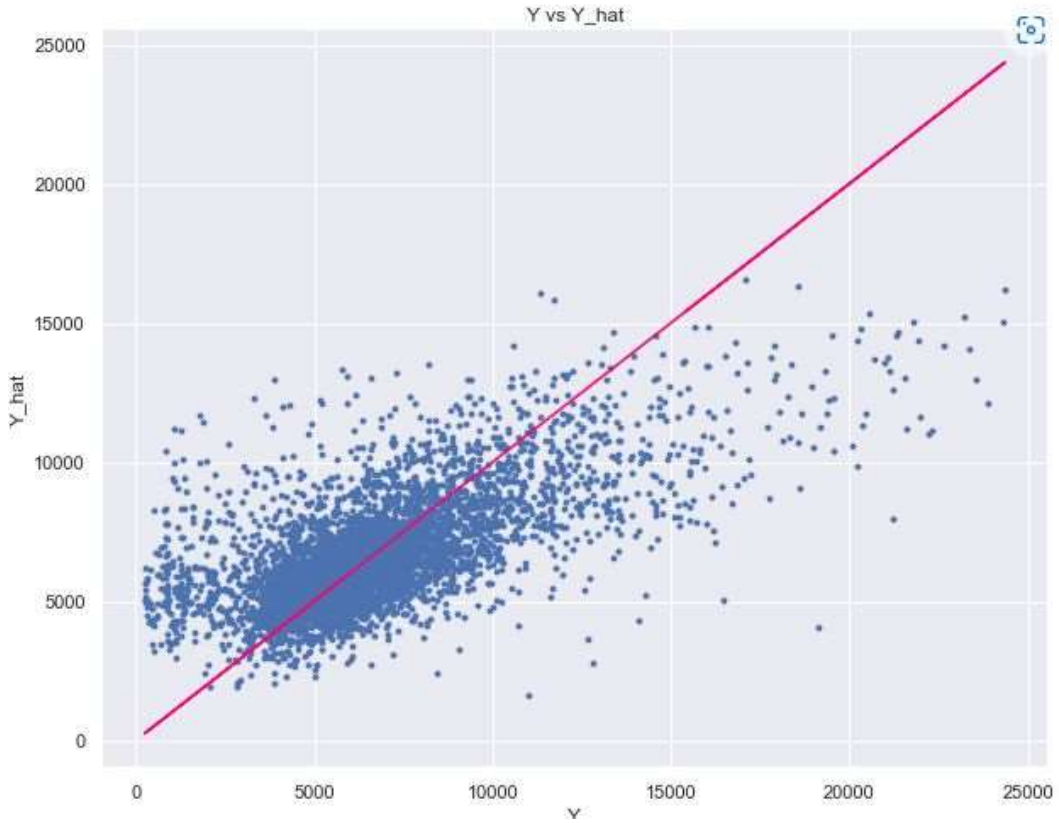
But after removing outliers and fitting the model, a 41% accuracy was obtained.

I decided to use the most correlated variables with the exception of sold_price, after outliers a correlation of 40% was obtained.

- bedrooms
- bathrooms
- sqrt_ft
- fireplaces

After doing some combinations of the explanatory variables I got the highest accuracy of 43%.

- sqrt_ft
- year_built
- fireplaces
- HOA



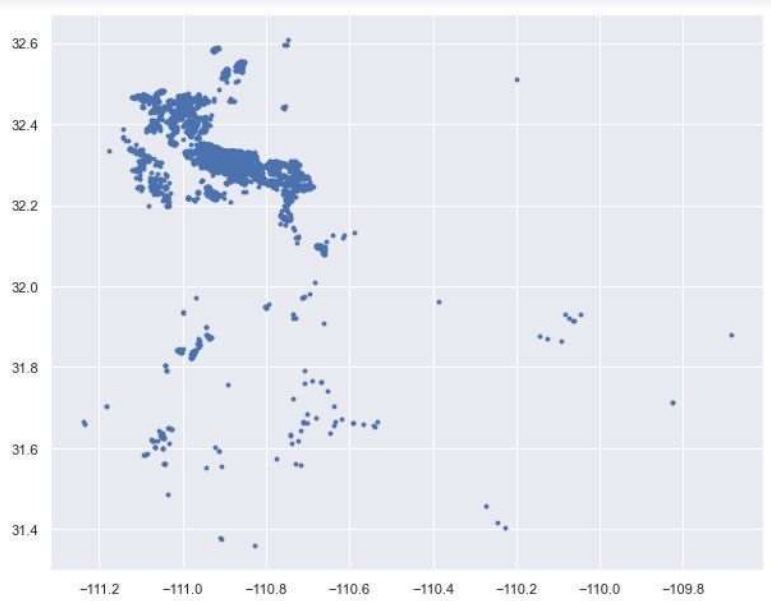
Implement the KNN to try getting a better accuracy

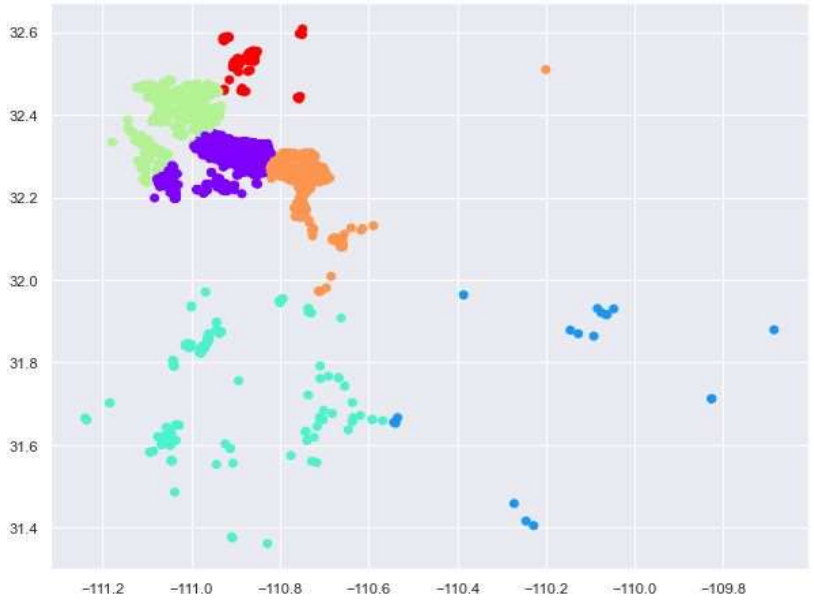
After implementing the algorithm, an accuracy of 62% was obtained.

Longitude and Latitude

Accuracy: 64%

Treat house locations as groups according to their position

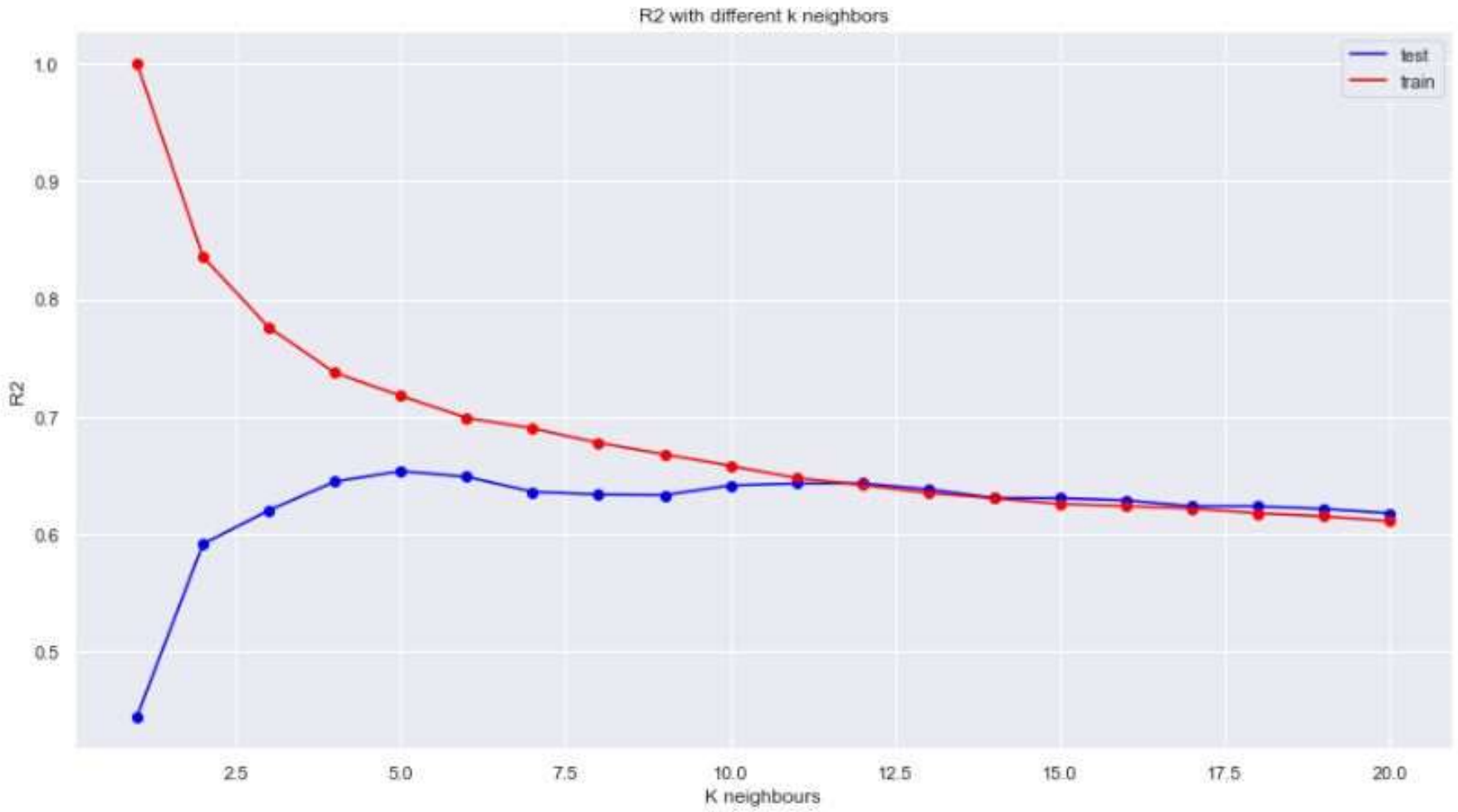




Final precision after entering the locations as groups.

Accuray: 62%

This could indicate that the small increase in precision is due to the increase in variables, although they are not really significant.


R2 with different k neighbors
test
train
R2
1.0
0.9
0.8
0.7
0.6
0.5
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
K neighbours

Conclusions

- The data does not fit properly in a linear way, that is why the regression does not have a good performance. Perhaps more observations are needed.
- A higher accuracy is obtained with knn regression
- No clear relationship was found between taxes and longitude and latitude

Thanks you

for your attention