Business Problem

The king county real estate agency will use this prediction model to give their clients an estimate of the housing price when purchasing or selling houses. The agency will estimate the price based on certain features like the location of the house, the number of bedrooms, and the size of the house.

Data

The king county dataset was provided to me as part of this project by Flatiron School. The dataset consists of 21597 rows, 21 columns with different house features (continuous and categorical). These features will help to understand which factor will affect the selling price. Below is the description of each variable in the data frame:

Analysis

1. Is there any relationship between the house's location and its sale price?

The predicted price will increase with the increase in latitude and decrease in longitude and as the location move to the lower northwest with few scattered houses in the middle to east. These will help the buyer to get an estimate of the housing price range based on the location, and their allocated budget.

House price range based on Location

2. What are the top ten zip codes that have the highest selling houses in King County?

After looking up the corresponding cities to each zip code, the top ten selling cities in terms of the price mean are Bellevue, Seattle, Mercer Island, Cottage Lake, Maltby, Union Hill-Novelty Hill, Sammamish.

The highest price Zip codes

3. What are the top ten affordable zip codes in King County?

After looking up the corresponding cities to each zip code, the most affordable cities in terms of the price mean are Tukwila, Auburn, Numclaw, Wabash, Birch, Krain, Cumberland, Bayne, Osceola, Maywood, Upper Mill, Bayne Junction, Boise, Veazie, Naco, Stampede,Kent, Lakeland North, Black Diamond, Franklin, and more

The top ten affordable Zip codes
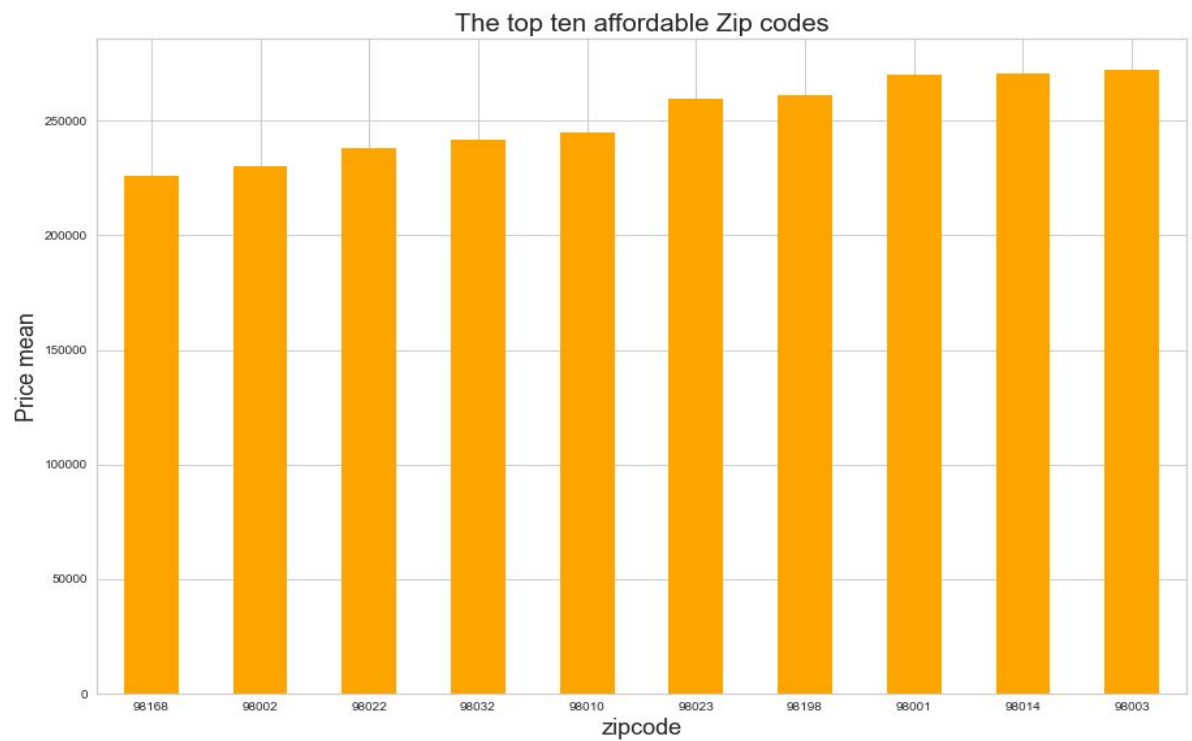
4. Which features are important to predict the the price of the houses ?

- Grade will increase the predicated price %180
- latitude will increase the predicated price %147
- Waterfront will increase the predicated price %77
- floors will increase the predicated price %6
- Bathrooms will increase the predicated price %2
- basement will increase the predicated price %3
-

Conclusion

- 15 features were included in the final model to get the best prediction The following findings are from the features with the highest coefficients:
- The price of the houses is highly affected by its location.
- Houses with larger living space, bigger basement, and more bathrooms have higher predicted price.

  -The houses with waterfront have higher selling prices than the ones without one.

- Each increase of the grade will increase the price, with grade 11 in the top

Limitation

The size of the dataset, a lot of features don't have a linear relationship with the targe. Maybe a different non-linear model would work better.

Future Work

Use APIs to get King country school district data and link it with the zip codes.

Thank you !

Please feel free to ask any questions now

You may also reach me via email:

Ahmet Karaoglan

karaglanahmet@hotmail.com