



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

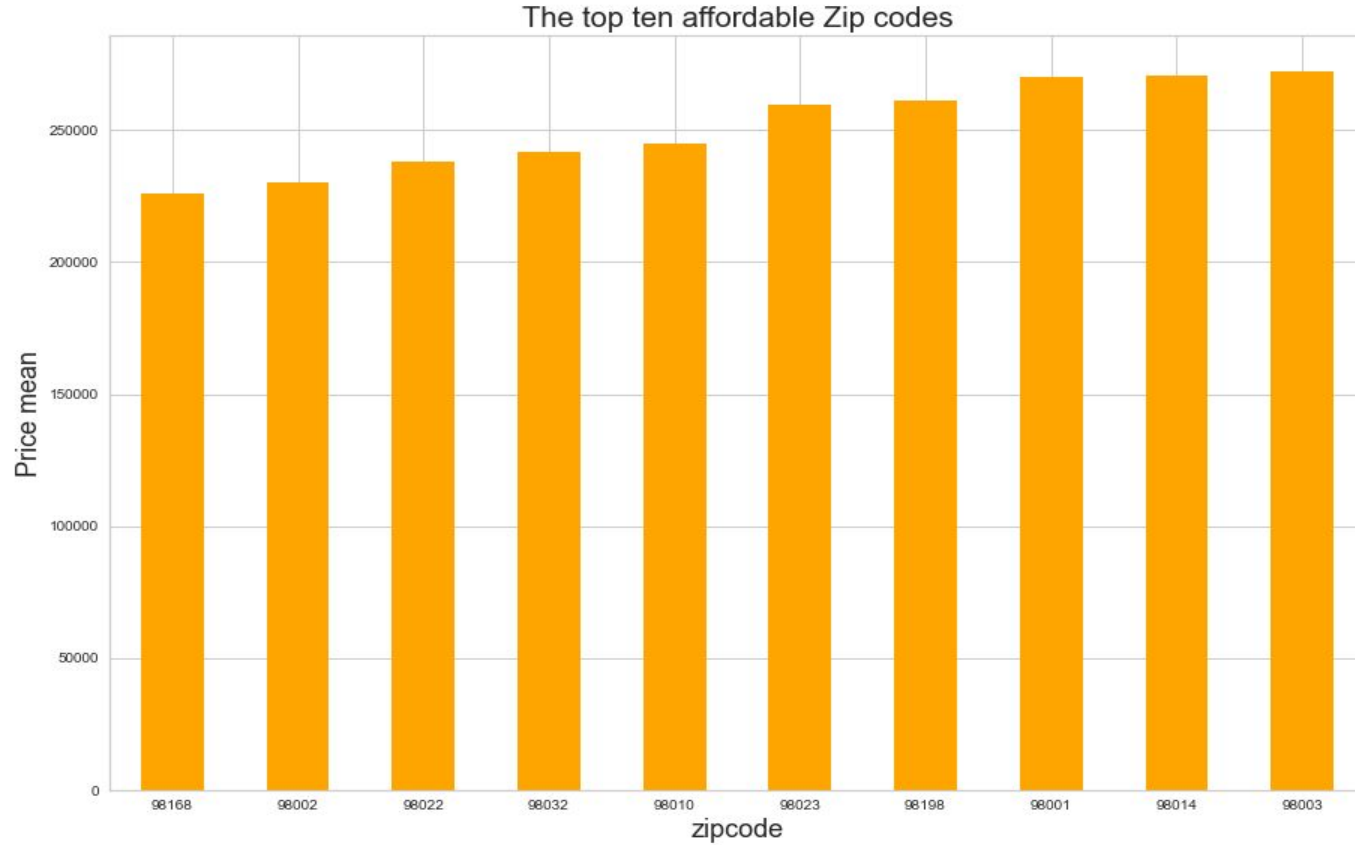
1. Is there any relationship between the house's location and its sale price?



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



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



4. Which features are important to predict the the price of the houses ?

- Longitude will increase the predicated price by \$233
- Square footage of the interior housing living 15 will increase the predicated price by \$171
- Latitude will increase the predicated price by \$56
- Square footage of the basement will increase te predicated price by \$22
- Square footage of the interior living space increase the predicated price by \$20

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 renovated houses selling price is higher than non-renovated one
- 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 target. Maybe a different non-linear model would work better.

Feature Work

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