King County Housing Analysis

By: Dan Tong

Summary

I aim to help a real estate agency identify features that most impact house sale pricing.

This presentation will cover:

- Business Problem
- Data Understanding
- Modeling
- Results
- Recommendations

Business Problem

The **real estate** market often changes and both buyers and sellers look to maximize their sale or purchase.

Real Estate **agencies** aim to help their clients both buy and sell their homes and advise them on housing prices.

There may be key indicators for determining housing prices: House Size, Location, Construction Quality, Number of Bedrooms, etc.

Data Understanding

The King County dataset contains information on house sales in King County, Washington.

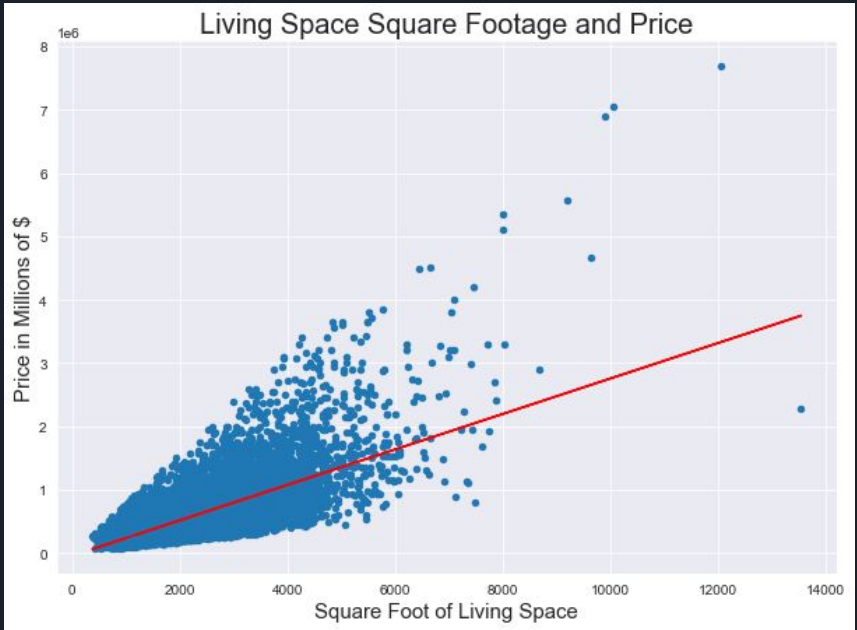
I found that the features that correlated most with sale price were:

- Square footage, # of bedrooms, # of bathrooms, grade of a house, and location

Many features similarly describe the size of a house like bedrooms, bathrooms, and living square footage.

- ID: Unique identifier for a house
- Date: Date house was sold
- **Price**: Sale price (prediction target)
- Bedrooms: # of bedrooms
- Bathrooms: # of bathrooms
- **Sqft_living**: Square Footage of living space in the home
- Sqft_lot: Square footage of the lot
- Floors: # of floors (levels) in house
- Waterfront: Whether the house is on a waterfront
- View: Quality of view from house
- **Grade**: Overall grade of the house. Related to the construction and design of the house.
- Sqft_above: Square footage of house apart from basement
- Sqft_basement: Square footage of the basement
- Yr_built: Year when house was built
- Yr_renovated: Year when house was renovated
- **Zipcode**: ZIP Code used by the United States Postal Service
- Lat: Latitude coordinate
- Long: Longitude coordinate
- Sqft_living15: The square footage of interior housing living space for the nearest 15 neighbors
- Sqft_lot15: The square footage of the land lots of the nearest 15 neighbors

Modeling



My initial findings found Living Square Footage to have the strongest correlation to Price.

As you can see, there is a positive relationship between Price and Living Square Footage.

I used Living Square Footage to create my baseline model and added more variables to improve the model.

Results

My **Baseline Model** used only Living Square Footage as the predictor variable. It revealed a 46% variance in price.

R-squared: 0.455

To improve this, I added Grade as a second predictor variable. It increased the variance to 54%.

R-squared: 0.537

Next I added ZIP code as a third predictor variable. It increased it the variance to a strong 84%.

R-squared: 0.839

Results

My model shows:

- For every % increase in Living Square Footage, Price will increase by 0.559%
- For every % increase in Grade, Price will increase by 0.867%
- ZIP codes where houses were priced higher on average than other ZIP codes:
  - 98112: Houses were priced 104% higher on average
  - 98004: Houses were priced 117% higher on average
  - 98039: Houses were priced 141% higher on average

Recommendations

Living Square Footage
The larger size of the house's living space, the higher it sells

Construction Grade
The better the quality and design of the house, the higher it sells. Prospective buyers and sellers should be mindful of contractors' quality

Location
Based on my findings, ZIP code has the largest impact on price. Certain neighborhoods and areas perform better than others

Next Steps

- Continue collecting data as more houses are purchased and sold.
- Improve modeling by using a larger dataset, or different datasets entirely.
- Observe more spacious opportunities for building larger homes.
- Work with reputable contractors to ensure quality of construction when building, buying, and selling homes.
- Examine resident demographics of homes within the best ZIP codes and see how they compare with others. Look for possible correlations.

Thank You!

Dan Tong
Email: dtong136@gmail.com
Github: @dtong136