

Kings County Housing Project

Jessica Rodriguez
Flatiron Self-Paced
January 2022



Summary

The goal of this project was to create a linear regression model that predicts housing prices in Kings County, Washington.

Housing data spanning from 5/2014 - 5/2015 was provided. Exploratory Data Analysis was used throughout this project. Statistics and and visualizations were used to identify and analyze linearity assumptions and data distributions.

The main question addressed was: ***What attributes of the data contributed to higher housing prices?***



Presentation Outline

- Business Value
- Methods
- Results
- Next Steps



Business Value

- This project's predictive model can be used by real estate agents and brokers, in addition to their domain-specific knowledge, to predict trends and housing prices.
- The model can also be used by home-buyers and sellers in their decision-making processes.

Methods

- Exploratory Data Analysis (EDA) was used for this project.
- 'Price' was our target variable. We compared this variable to various housing features in the data set (see descriptions)
- Special consideration was given to location, square footage, condition, grade and amount of bedrooms/bathrooms.

Column Names and descriptions for Kings County Data Set

#	name	description
1	id	unique ID for a house
2	date	Date day house was sold
3	price	Price is prediction target
4	bedrooms	Number of bedrooms
5	bathrooms	Number of bathrooms
6	sqft_living	square footage of the home
7	sqft_lot	square footage of the lot
8	floors	Total floors (levels) in house
9	waterfront	Whether house has a view to a waterfront
10	view	Number of times house has been viewed
11	condition	How good the condition is (overall)
12	grade	overall grade given to the housing unit, based on King County grading system
13	sqft_above	square footage of house (apart from basement)
14	sqft_basement	square footage of the basement
15	yr_built	Year when house was built
16	yr_renovated	Year when house was renovated
17	zipcode	zip code in which house is located
18	lat	Latitude coordinate
19	long	Longitude coordinate
20	sqft_living15	The square footage of interior housing living space for the nearest 15 neighbors
21	sqft_lot15	The square footage of the land lots of the nearest 15 neighbors

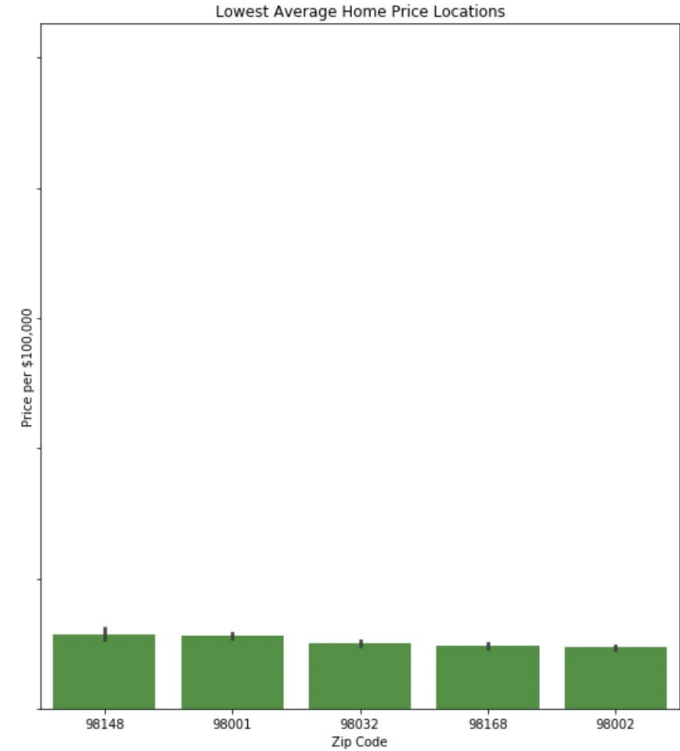
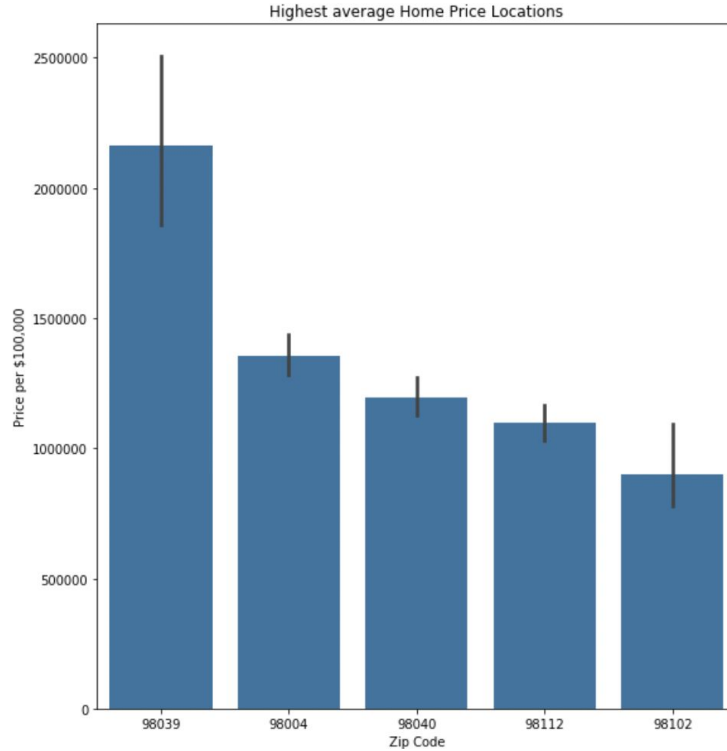
Modeling:

The following models were created for this project:

- **A simple linear regression model** compared our target variable (price) to the following continuous features: square footage, flooring, and year built. Administering a recursive feature eliminator and KBest test, had a negative impact on the model.
- **A polynomial regression model** where degree = 2.

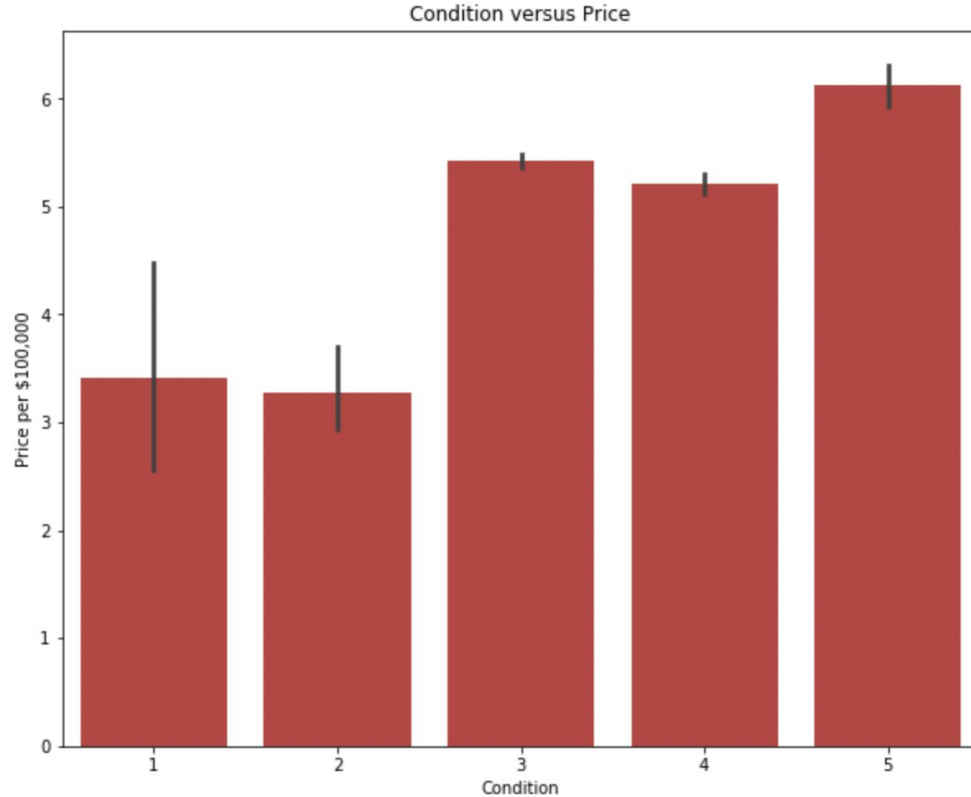
The polynomial regression model had the best R-squared testing result and lowest mean-squared error, and therefore **was saved as the best fit model**.

Results: Location



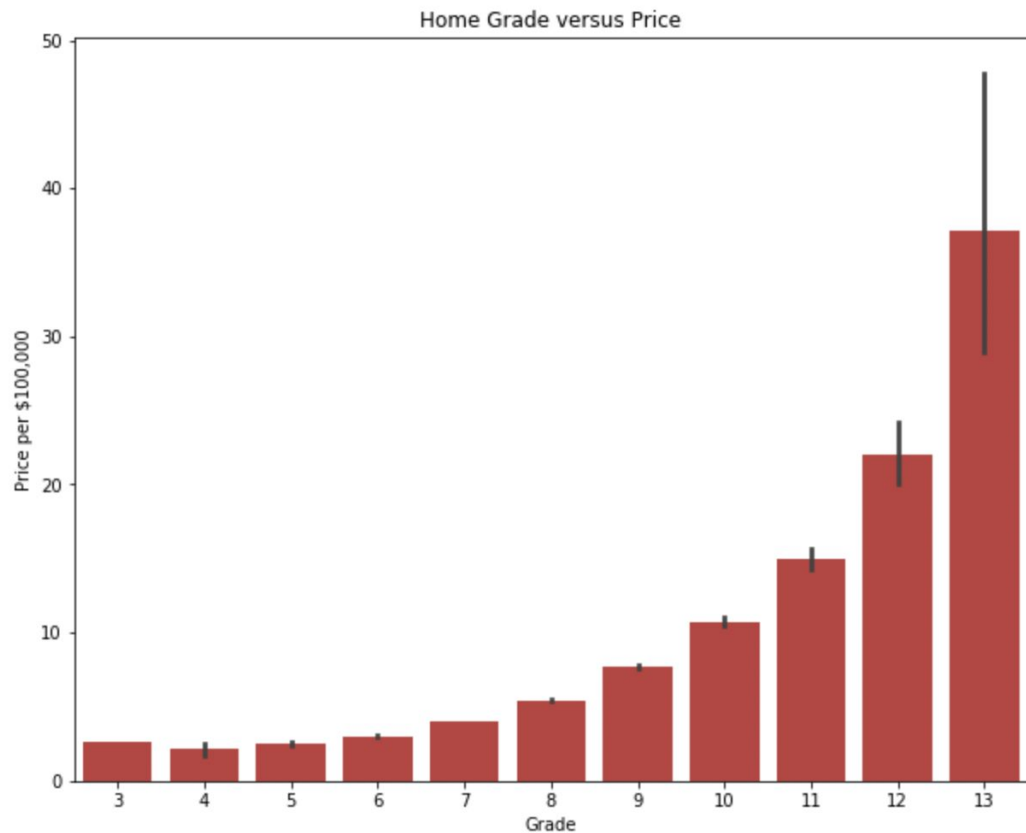
- Home locations were highly correlated with price.
- Highest-priced zip codes: Medina, Bellevue, Mercer Island, Capital Hill
- Lowest-priced zip codes: Normandy Park, Burien, Auburn, Kent, Seatac

Results: Condition



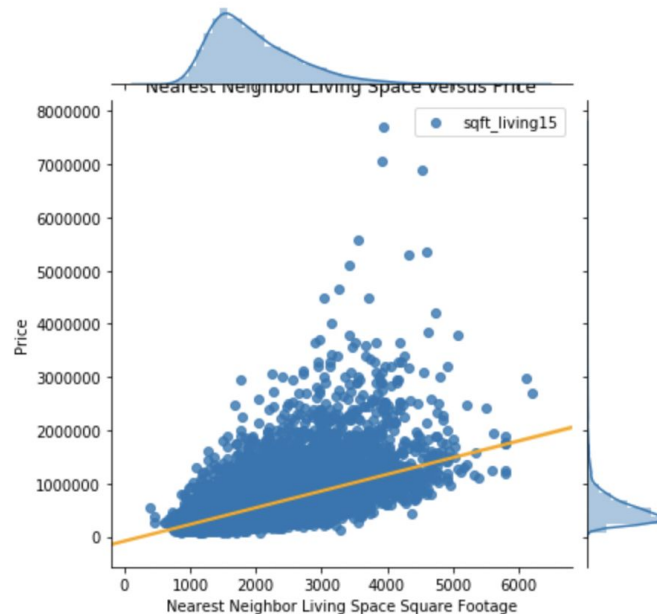
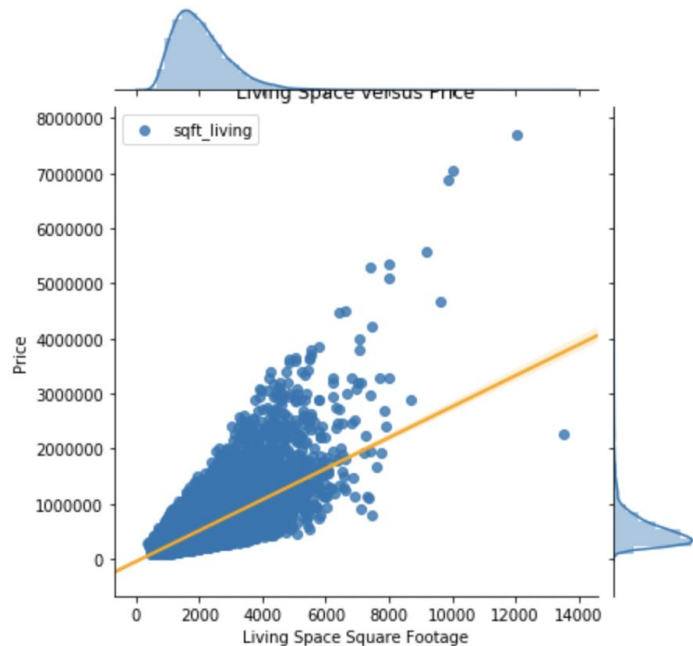
- Price and Home condition are highly correlated.
- Housing prices increase significantly after condition hits a score of 3.

Results: Grade



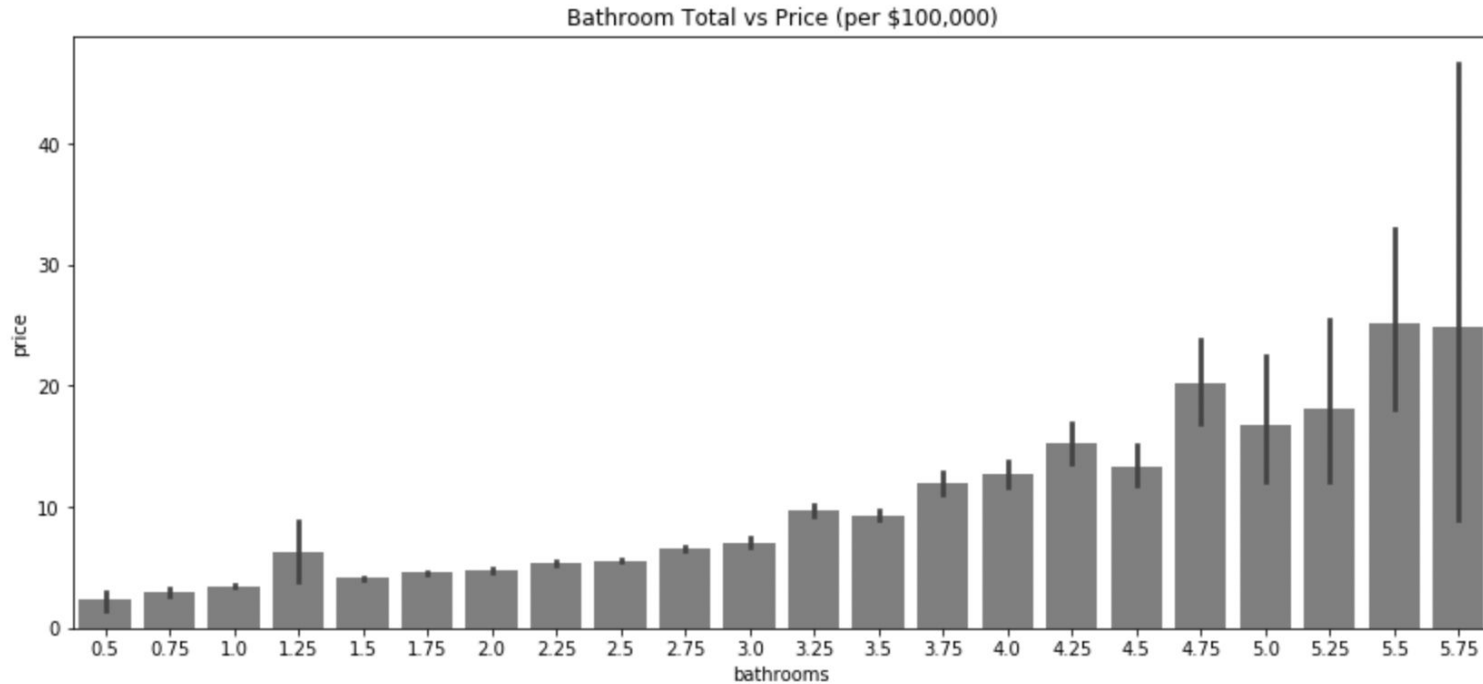
- Price and Grade are highly correlated and show a positive linear relationship.

Results: Square Footage



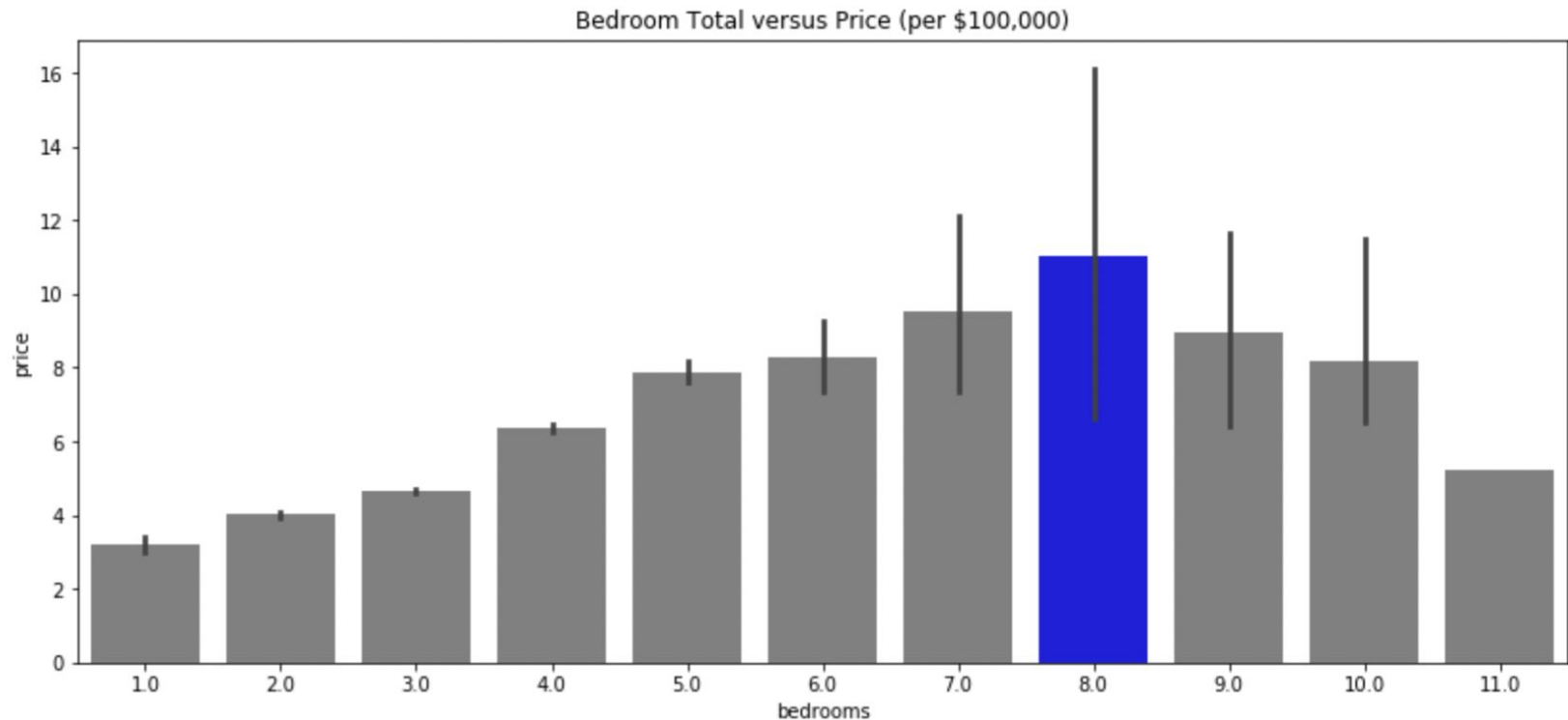
- Home square footage is highly correlated with price.
- There is also a high correlation between price and the square footage of a home's 15 nearest neighbors.

Results: Bathrooms



Bathroom totals are highly
correlated with price

Results: Bedrooms



Bedroom totals are highly correlated with price.

Next Steps

For next steps, I would analyze the detailed geography of the Kings County.

I would specifically look at the proximity to certain zip codes to schools, amenities, public transportation, etc.

I would also analyze census data to see how demographics affect housing prices.

Thank You!

Email: jessicallowell@gmail.com

GitHub: [@jesrodriguez0816](https://github.com/jesrodriguez0816)