Modeling King County Home Prices

Objective

To determine the best features to use to create a model that will predict the sale prices of King County homes

The Dataset

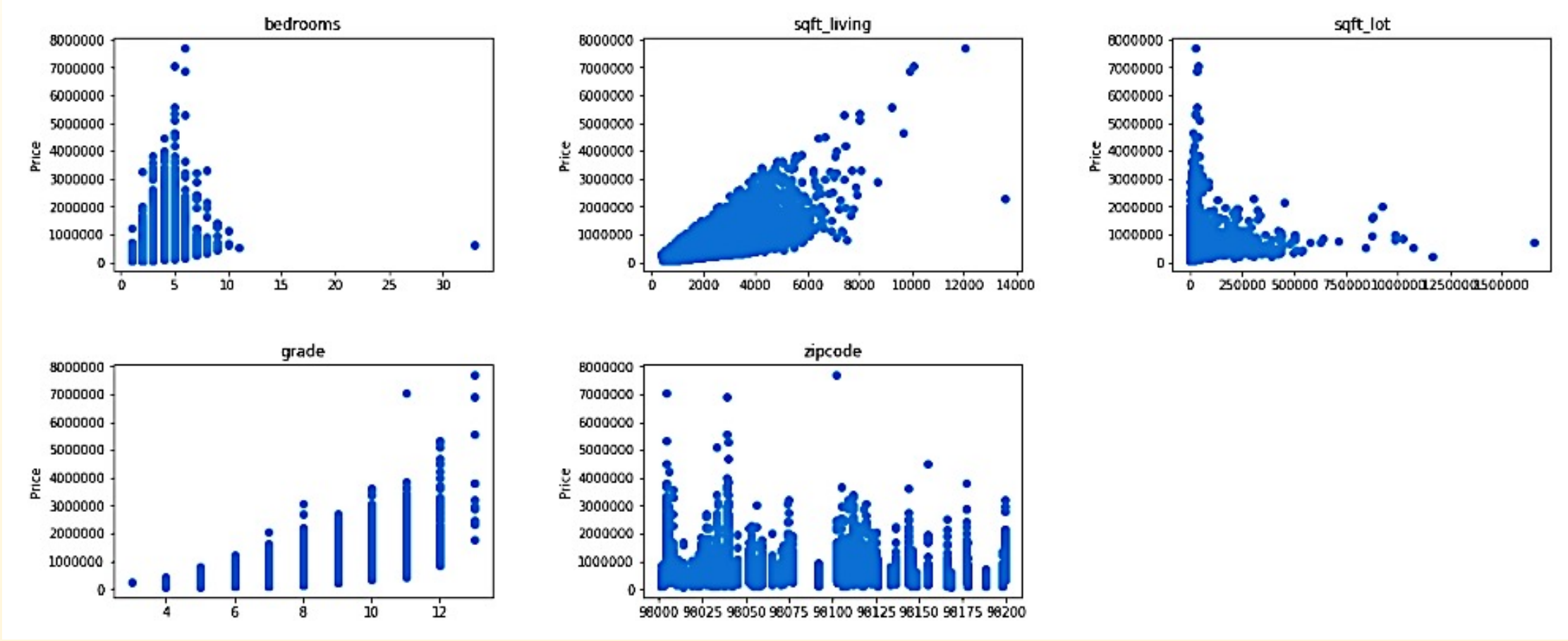
A database of homes sold in King County in 2015 which includes:

Sale Price
Sale Date
Bedrooms
Bathrooms
Living Square Feet
Lot Square Feet
Total Floors
Waterfront presence
View presence
Condition
Grade
Square Feet Above Ground
Square Feet Basement
Year Built
Year Renovated
Zipcode
Latitude
Longitude
Living Square Feet of Neighbors
Lot Square Feet of Neighbors

Selected Features

- Living Square Feet
- Lot Square Feet
- Bedrooms
- Grade
- Zipcode

Plots of Considered Features



Data Processing

- Removed outliers
- Log transformed and normalized square feet features
- One-hot encoded categorical features

Model Building

Used the Python statsmodel library to create an OLS model

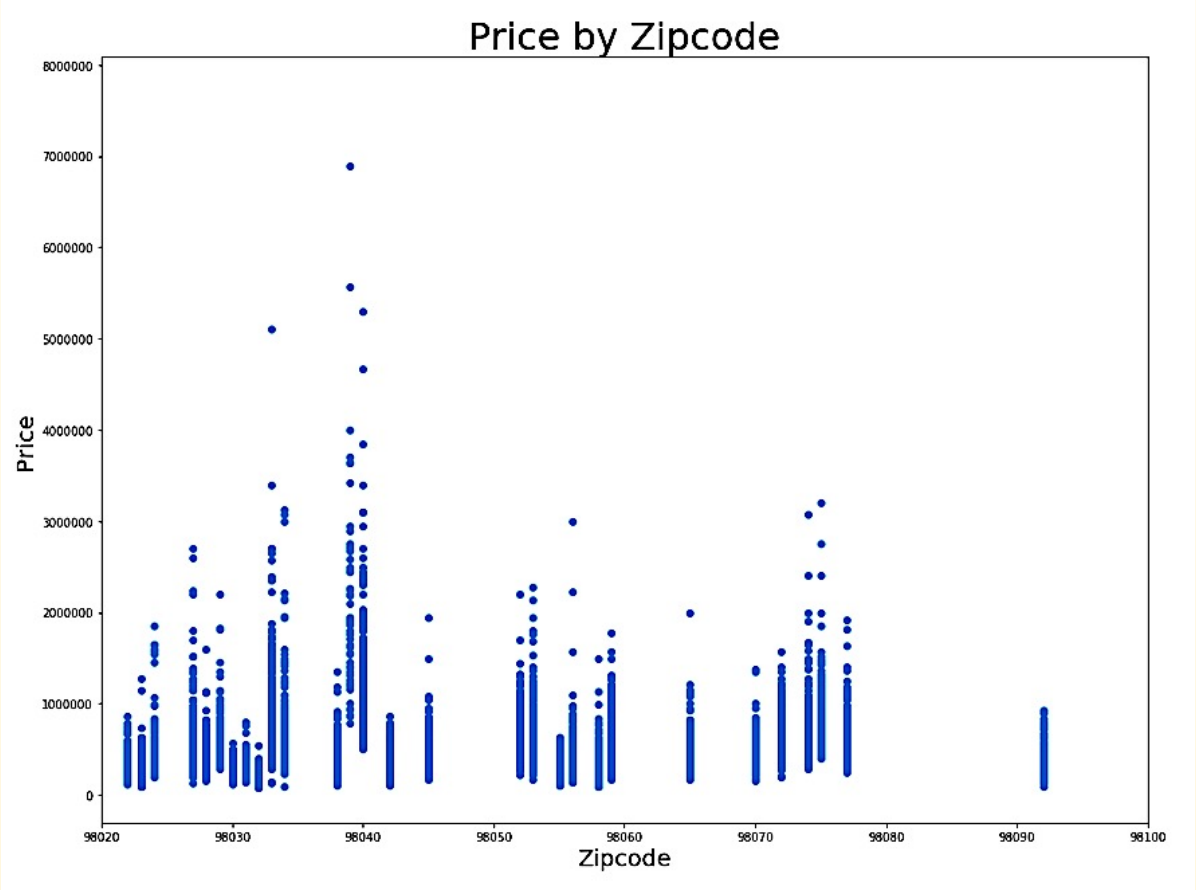
Results

	R2	Adj R2	p-values
• Model with Living Square Feet	0.374	0.374	0.000
• Model with all five features	0.767	0.767	<=0.723

High p-values were restricted to seven zipcodes.

Affected Zipcodes

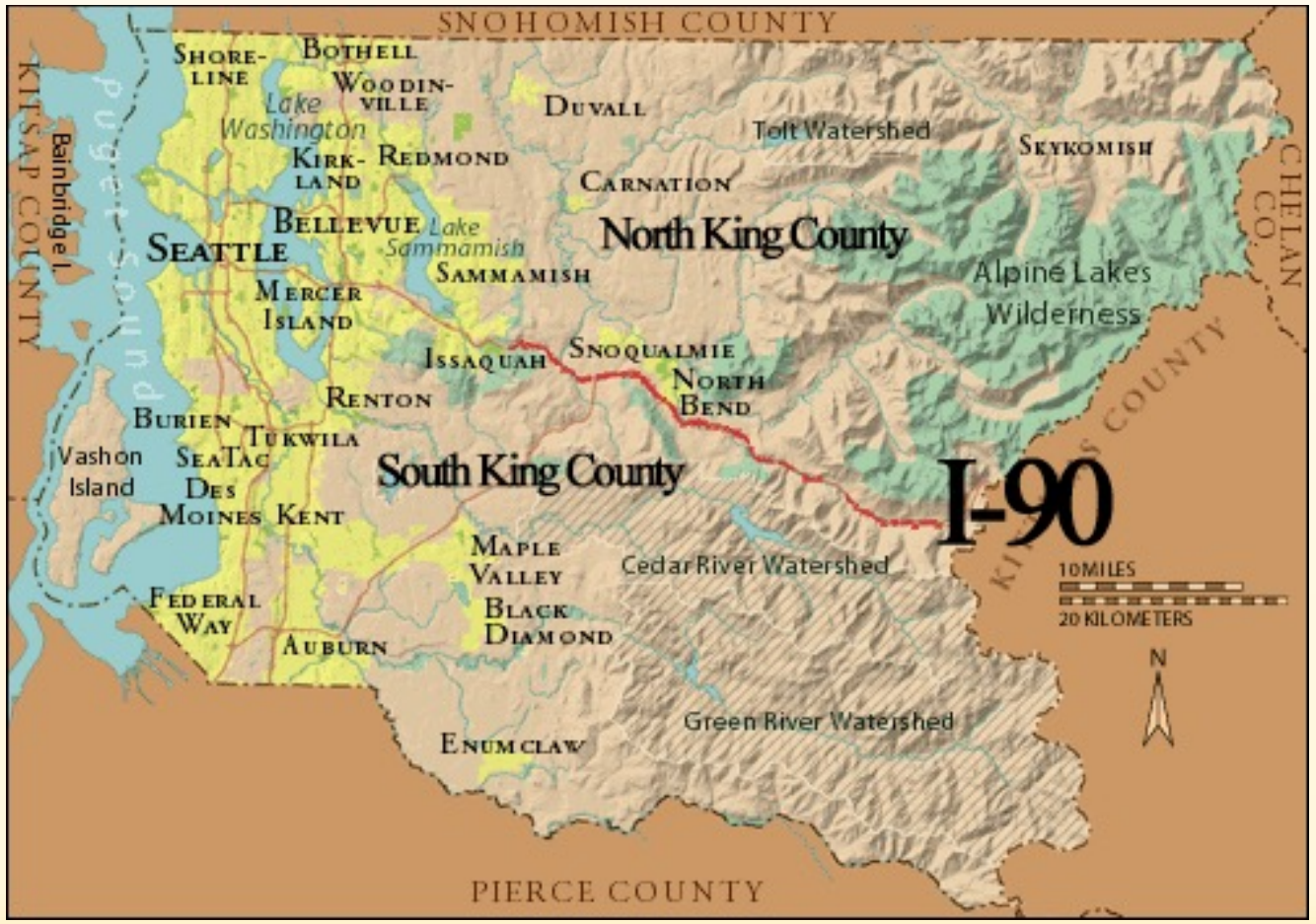
- Enumclaw – 98022
- Kent - 98030, 98032, 98042
- Federal Way - 98023
- Auburn – 98092



Why are these zipcode different?

For that we need to consider the geography of King County.

Map of King County



Enumclaw is the most isolated city in King County.

Kent, Auburn, and Federal Way are all along major commuting corridors and close to Tacoma.

Results

	R2	Adj R2	p-values
• Model with Living Square Feet	0.374	0.374	0.000
• Model with all five features	0.767	0.767	<=0.723
• Model with selected zipcodes removed	0.761	0.760	<=0.049

Conclusions

- Living Square Feet, Lot Square Feet, Bedrooms, Grade, and Zipcode are important modeling features
- This model is a starting point for setting a listing price, but that price should be adjusted based on a realtor’s understanding of other features.
- Federal Way, Kent, and Auburn should be modeled separately.