Module 1 final project: King County house sales

Laura Lewis

Outline

Project aim and data source

Exploring the data - three examples of insights from the data

Modeling - predicting house prices

Interpretation and conclusions

Project aim and data source

Aim: to build a model to predict the sale price of houses using a dataset from King County, Washington, USA

Dataset of 21,597 houses sold in the area between May 2014 and May 2015

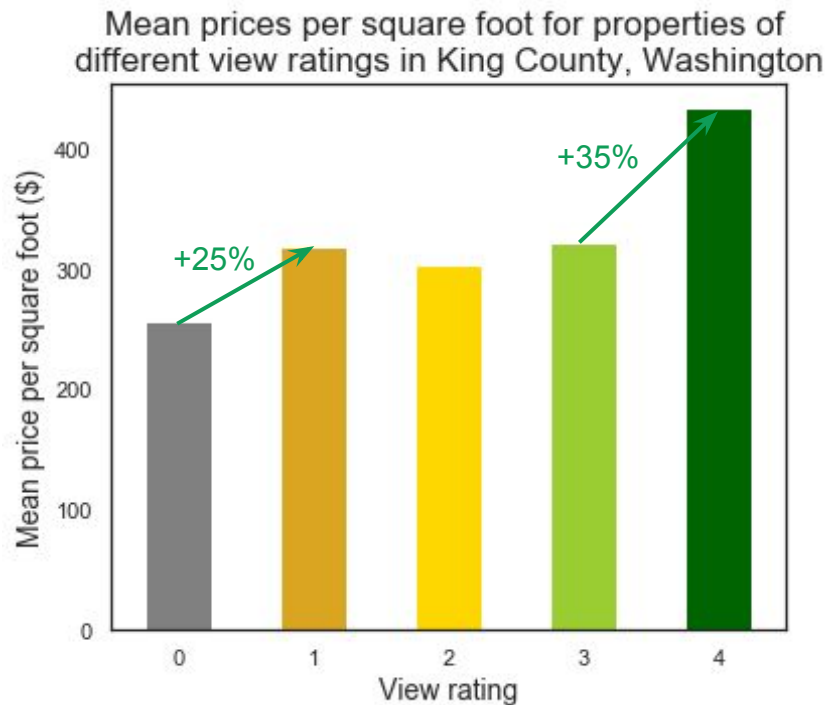Data cleaning to ensure the data is of sufficient quality for modeling

Exploring the data - sudden increase in house prices Feb-April 2015

Question: how have house prices changed over time?

Change in average house price in King County, Washington, May 2014 - May 2015
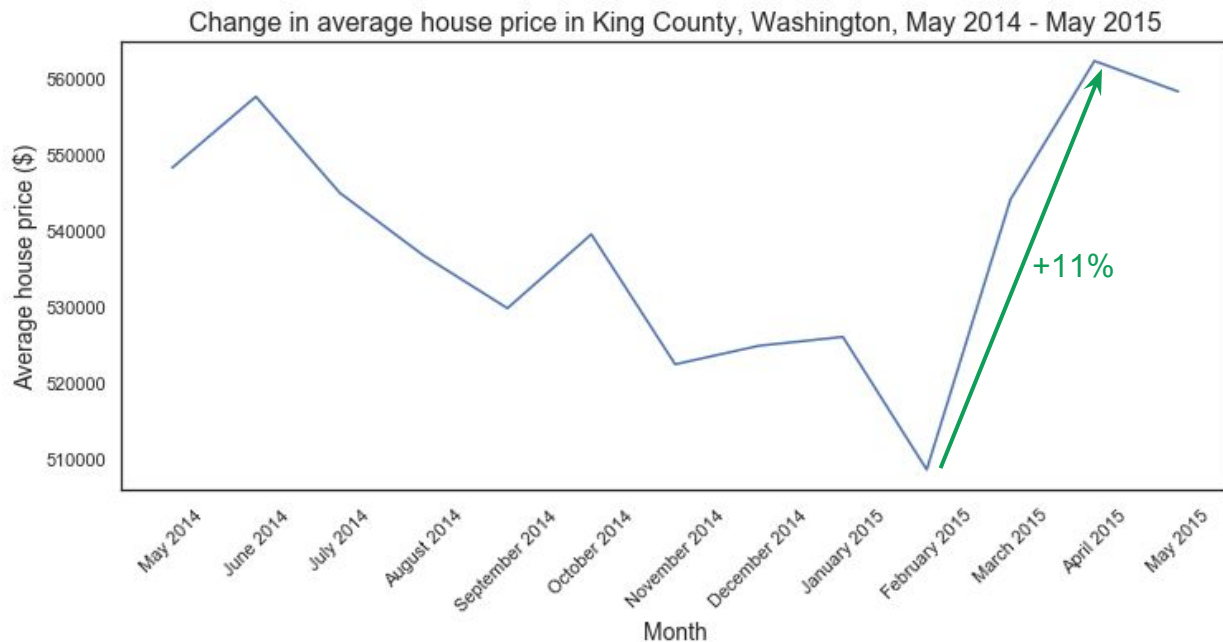
Exploring the data - people will pay for the best views

Question: how much does having a good view increase house prices by?

Mean prices per square foot for properties of different view ratings in King County, Washington

+25%

+35%

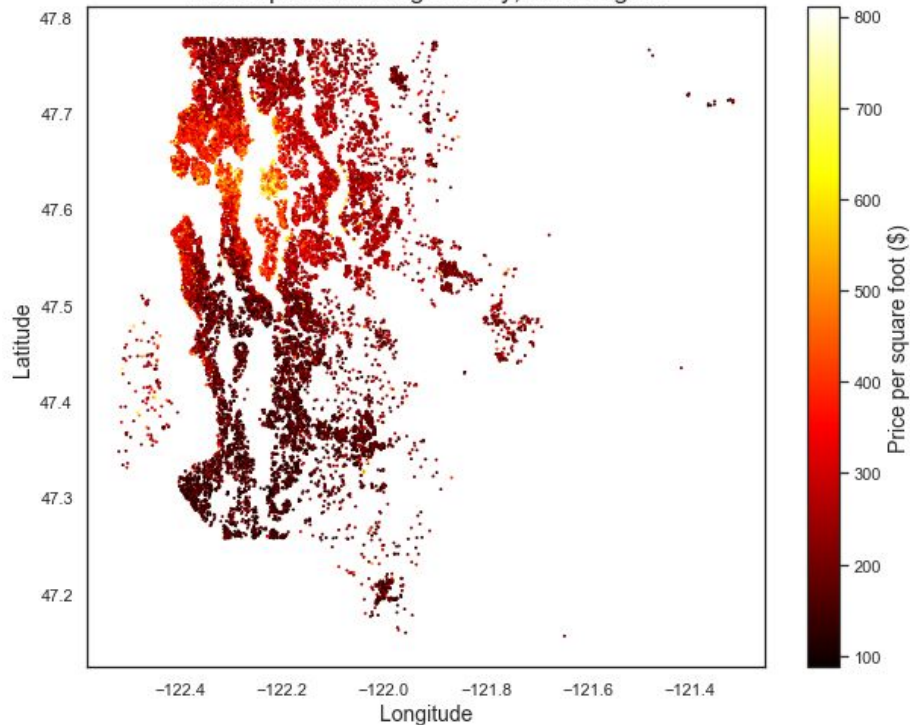Mean price per square foot ($)

View rating

Exploring the data - geographical hotspots

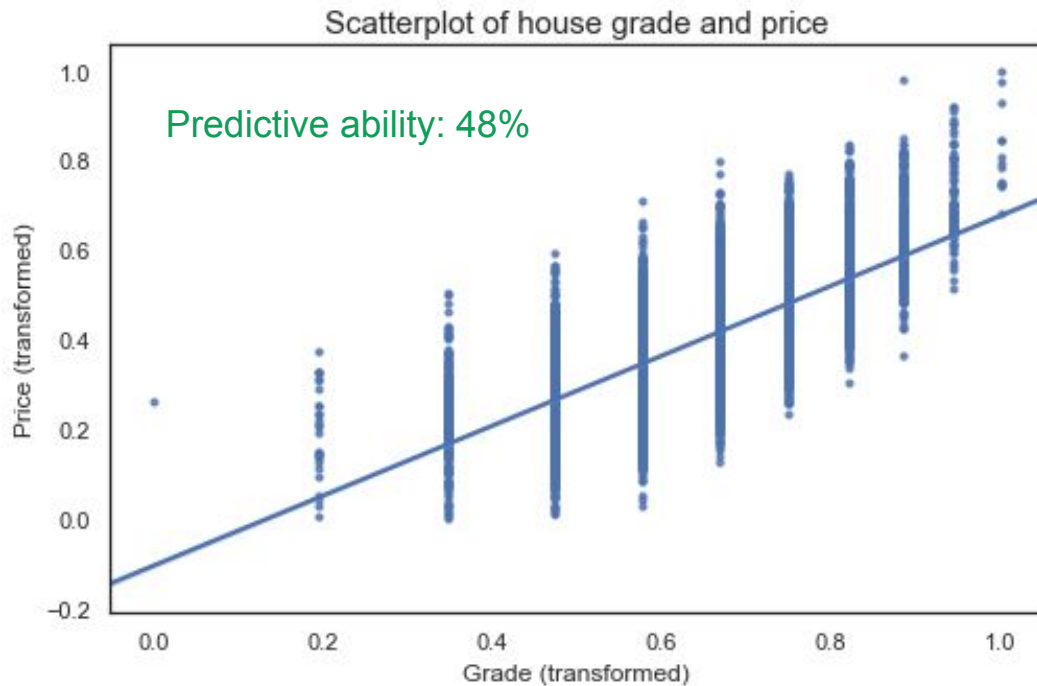*Question: How much does house price vary by location?*

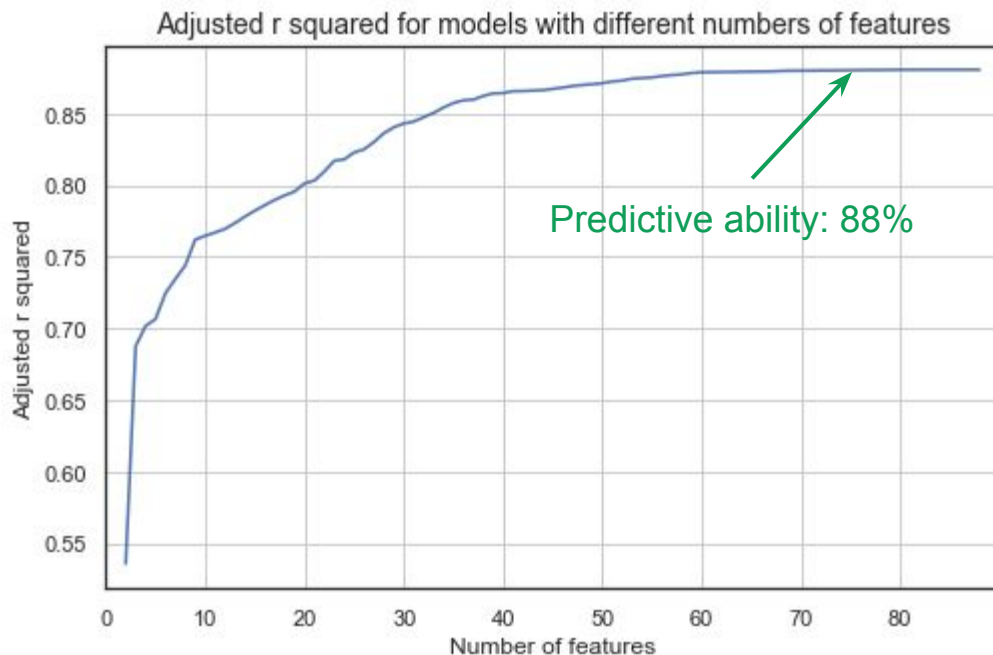Number of properties in each area of King County, Washington
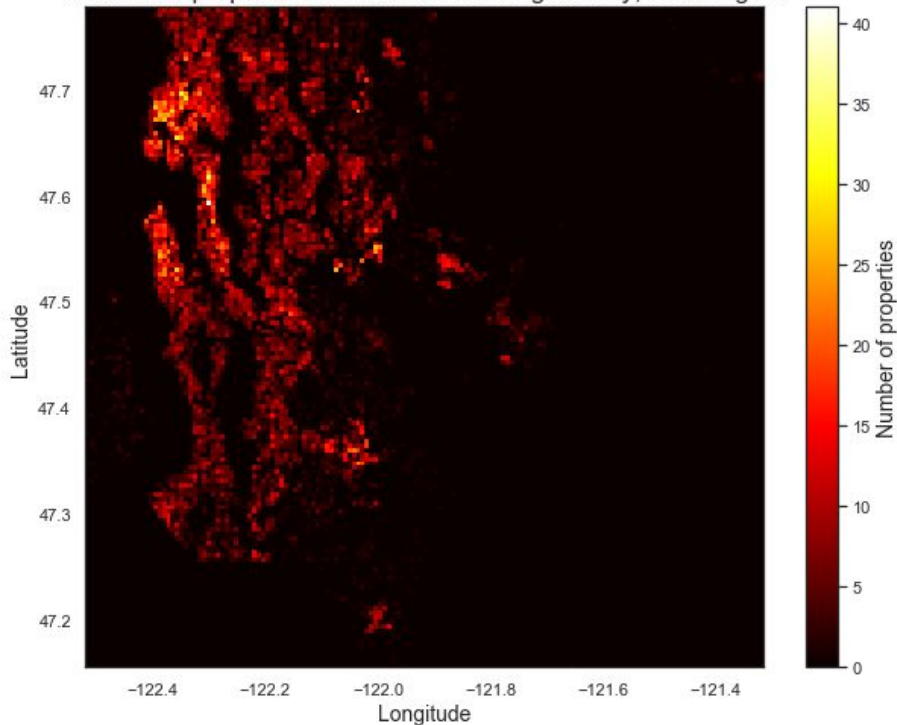
House prices in King County, Washington

Building a model - using a single feature to predict house price

Scatterplot of house grade and price

Predictive ability: 48%

Final model - predicting 88% of variation in house prices using multiple features

Adjusted r squared for models with different numbers of features

Predictive ability: 88%

Interpretation and conclusions

The final model explains 88% of variation in house prices

The results of the final model also reveal the features that have the biggest impact on house price:

- Grade
- Latitude
- Square footage of the house
- Being in Bill Gates' zipcode (98039)

Recommendations for increasing sale price:

- Increase building quality
- Extend the size of the living area

Questions?