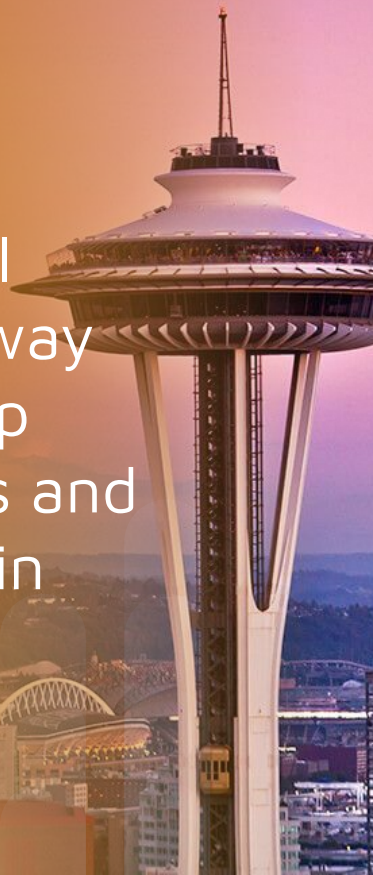
PHASE 2 FINAL PROJECT

EDA & Linear Regression of King's County Housing Data

Presented by Wolfgang Wallach
Full-time Student at Flatiron School

Business Problem:

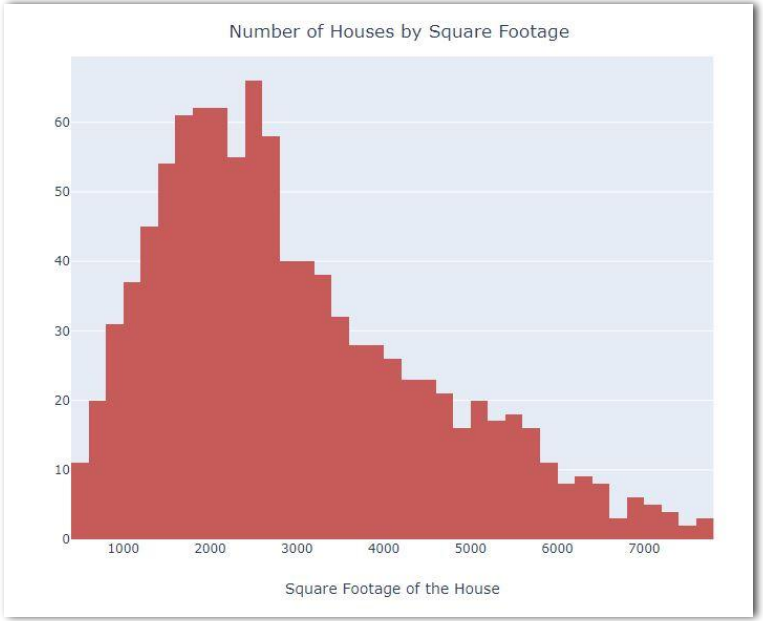
Keller Williams of the Greater Seattle area wants to create a data driven online portal that will draw prospective home buyers away from their competitors. This portal will help home buyers be realistic about their needs and budget, as well as taking some of the strain off the Keller Williams realtors.

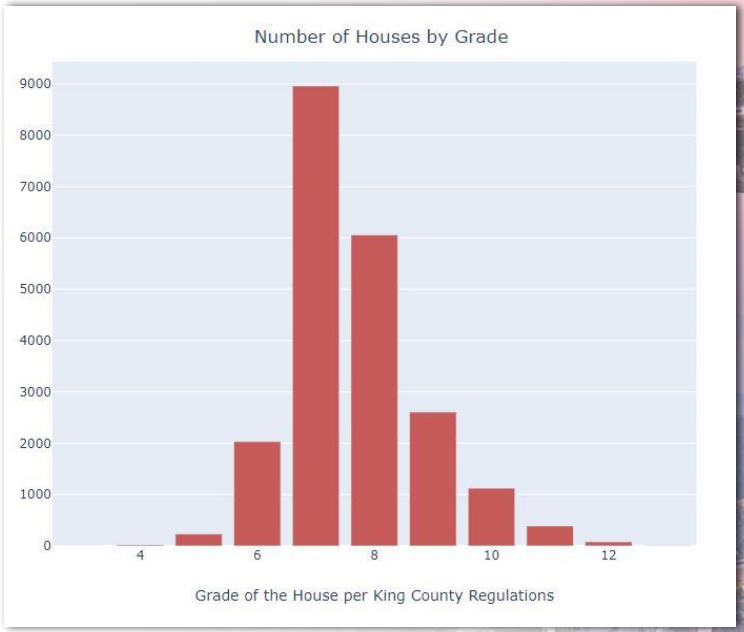


Project Goals:

To create a regression analysis of King's County housing data so as to benefit Keller Williams, such that they can begin to construct their portal. This regression should identify the most important features of a home in regards to price.

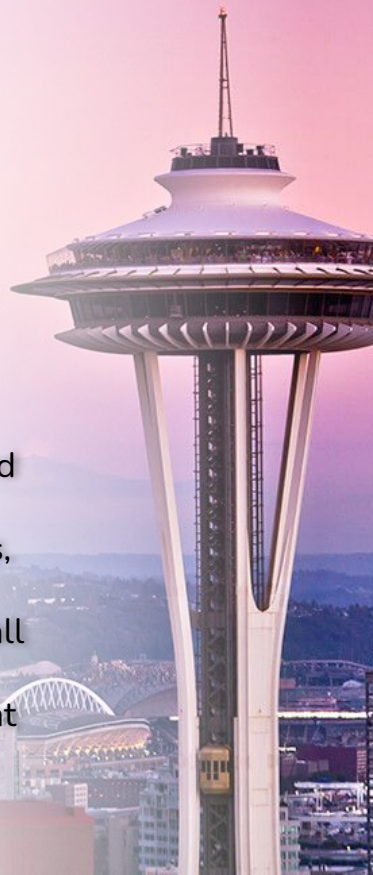
Distribution of KC House Grade and Square Footage of Home





Overview of Grading System

Grade 1-3:	Falls short of minimum building standards.
Grade 4:	Generally older low quality construction. Does not meet code.
Grade 5:	Lower construction costs and workmanship.
Grade 6:	Lowest grade currently meeting building codes.
Grade 7:	Average grade of construction and design.
Grade 8:	Just above average in construction and design.
Grade 9:	Better architectural design, with extra exterior and interior design and quality.
Grade 10:	Homes of this quality generally have high quality features. Finish work is better, and more design quality is seen in the floor plans and larger square footage.
Grade 11:	Custom design and higher quality finish work, with added amenities of solid woods, bathroom fixtures and more luxurious options.
Grade 12:	Custom design and excellent builders. All materials are of the highest quality and all conveniences are present.
Grade 13:	Generally custom designed and built. Approaching the Mansion level. Large amount of highest quality cabinet work, wood trim and marble; large entries.



Grade of Home and Square Footage in regards to Sale Price of Homes

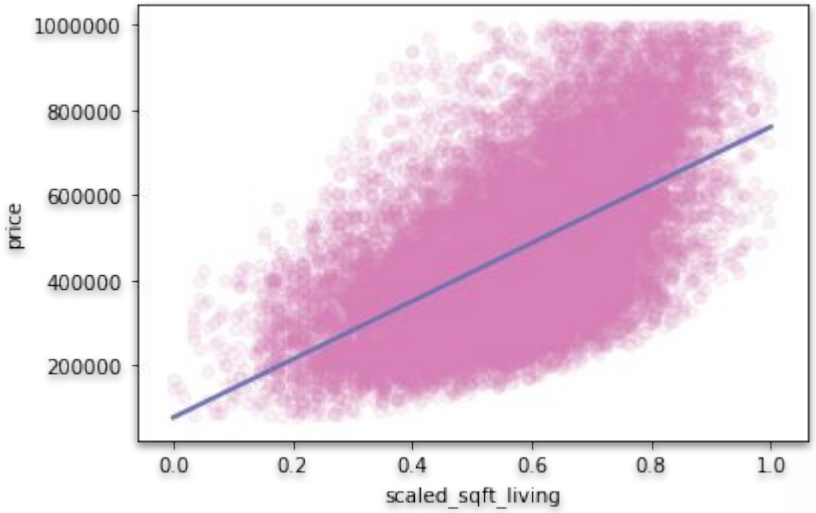
The Regression Model

- Dropped all data with excessive values:
  - Homes greater than 5000 square feet
  - Home Price of over $1,000,000
  - Four or more bathrooms
  - Six or more bedrooms
- Change latitude and longitude to Geohash buckets

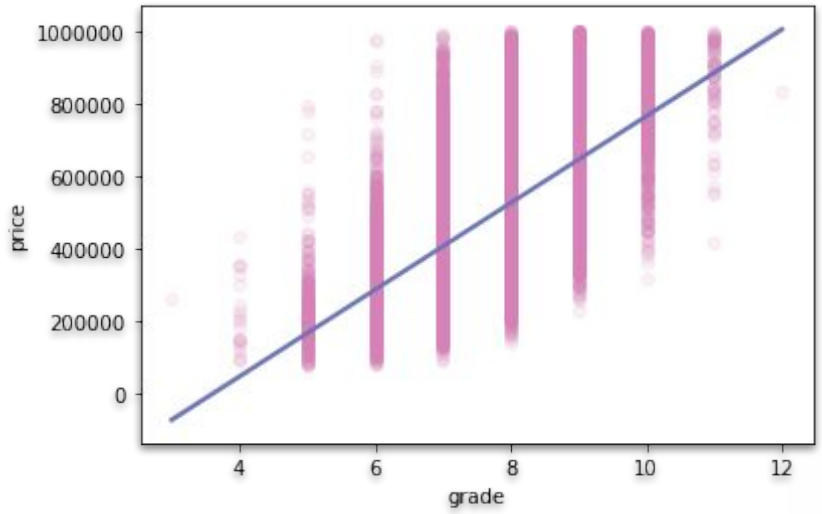
Remaining Features

- Square footage of the home
- The geohash coordinates that the home falls within
- The condition of the home
- The grade of the home
- The year the home was constructed
- How many floors the home has
- How many bathrooms the home has
- How many bedrooms the home has

Square Footage of Home vs Regression Line of Home Price



Grade of Home vs Regression Line of Home Price



Conclusions

- Square footage of home and county grade are the best predictors for portal purposes
- Clients may be interested in bedrooms and bathrooms, but the layout, size, and shape of those can be important

Future Work

- Look at bedrooms and bathrooms in more depth; floor plan may be important for a more robust model
- Location was tricky, the model could benefit from something like a feature stating distance from POI

THANK YOU!

Additional questions can be addressed to:

Wolfgang Wallach
wwallach95@gmail.com
github.com/wwallach95