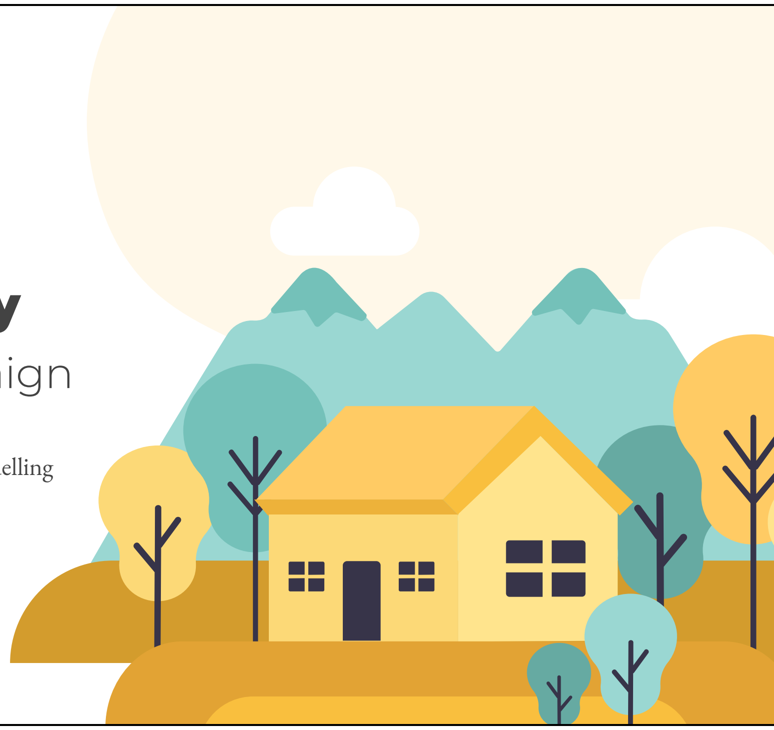
King County

Marketing Campaign

Data-driven insights and predictive modelling



Problem Statement

Advertising Campaign to encourage sales in King County

- Data-driven recommendations
- Model to predict house prices

PropertiesInc
REAL ESTATE

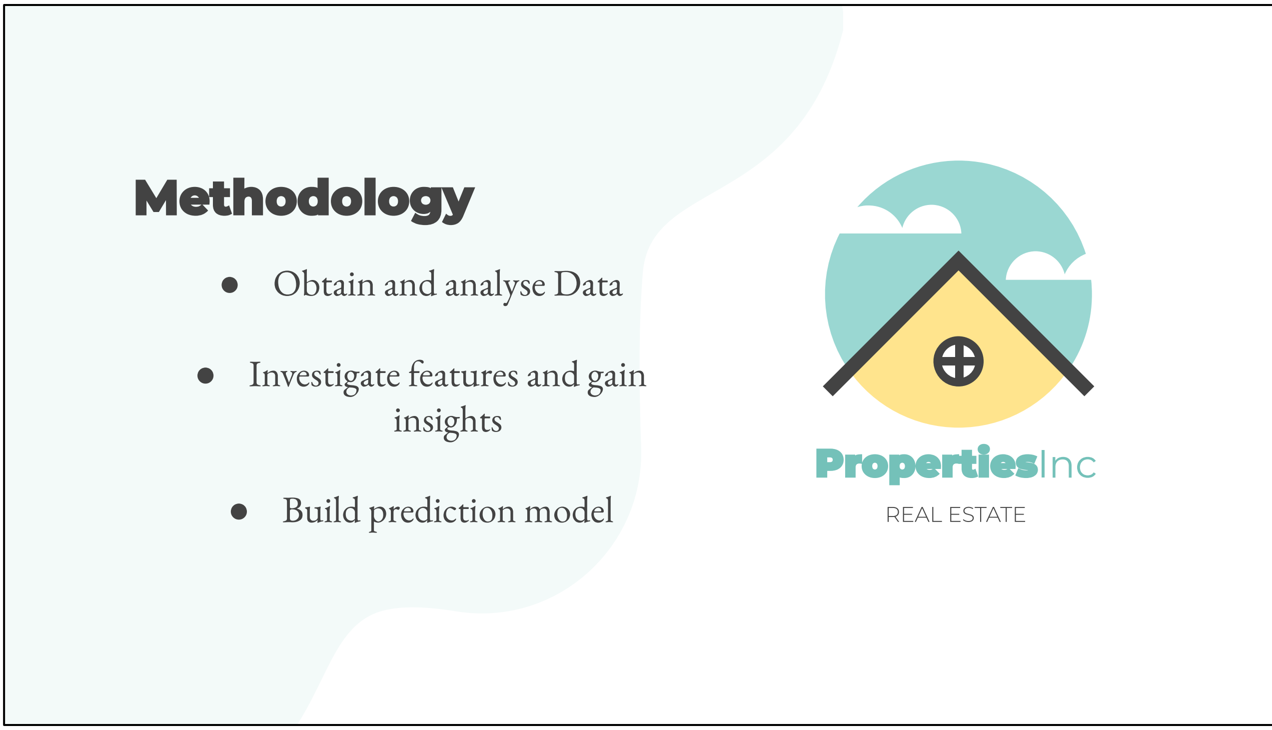
Intro: I’m a junior data scientist here at PropertiesInc Real Estate.
I’m here to share the result of our project supporting the upcoming marketing campaign focused on home owners in King County who may be interested in selling.
We will provide recommendations to shape the campaign based on data and our model for predicting house prices in the area

Business Value

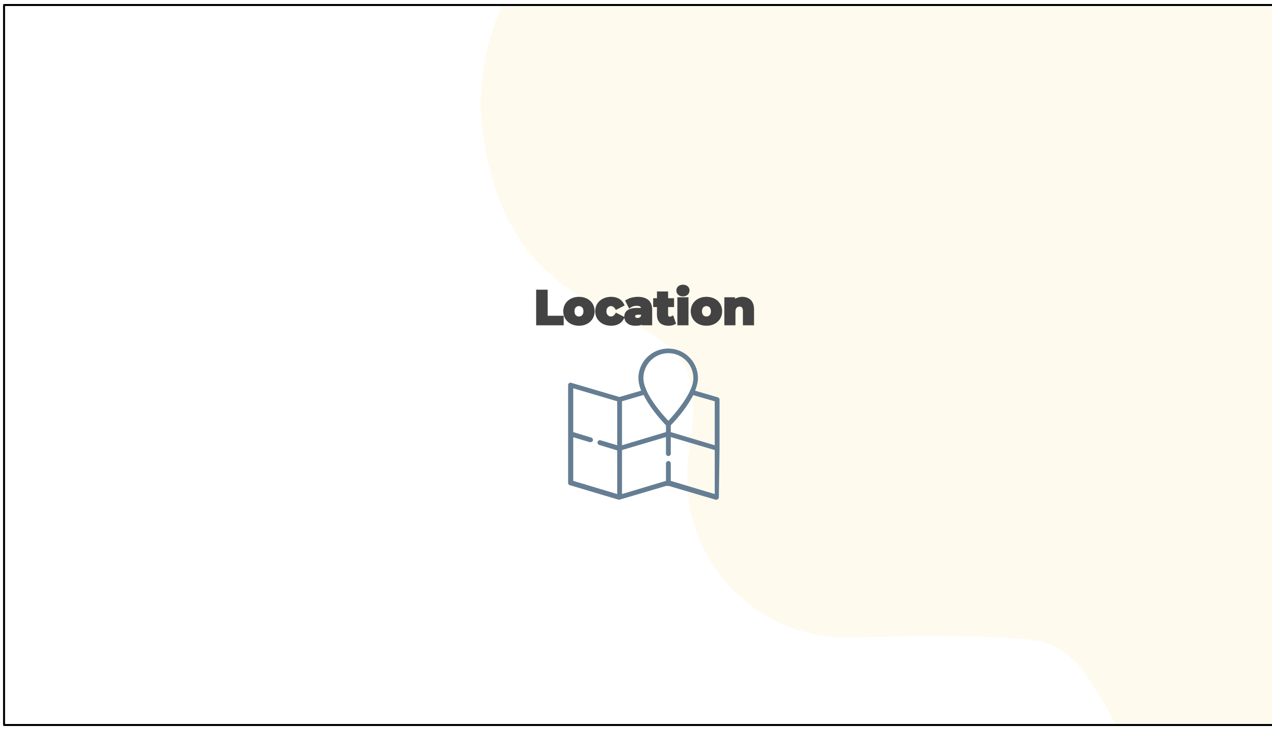
- Generate revenue via successful marketing campaign
- Better understanding of house buyer preferences
- Assist Valuations Team by providing price predictions

PropertiesInc
REAL ESTATE

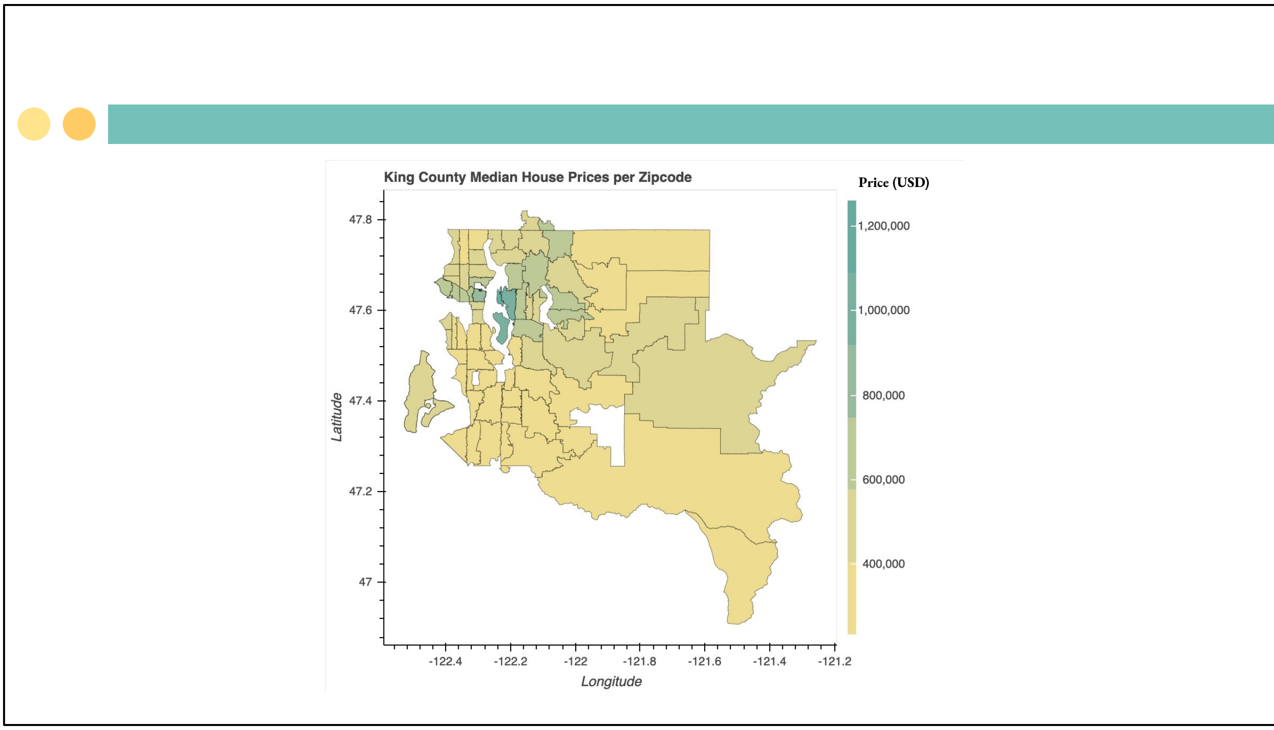
The business value of our project can be understood to be the following. Firstly, it goes without saying that a successful campaign will bring in revenue for the firm, and as such by helping shape the campaign we are indirectly involved in revenue generation. This project also helps understand buyer preferences and improves colleagues' in the Sales teams market knowledge. Finally by having a model capable of predicting house prices, we will help the Valuations team by providing initial estimates and combining domain expertise and AI, we can improve on our suggested sale prices.



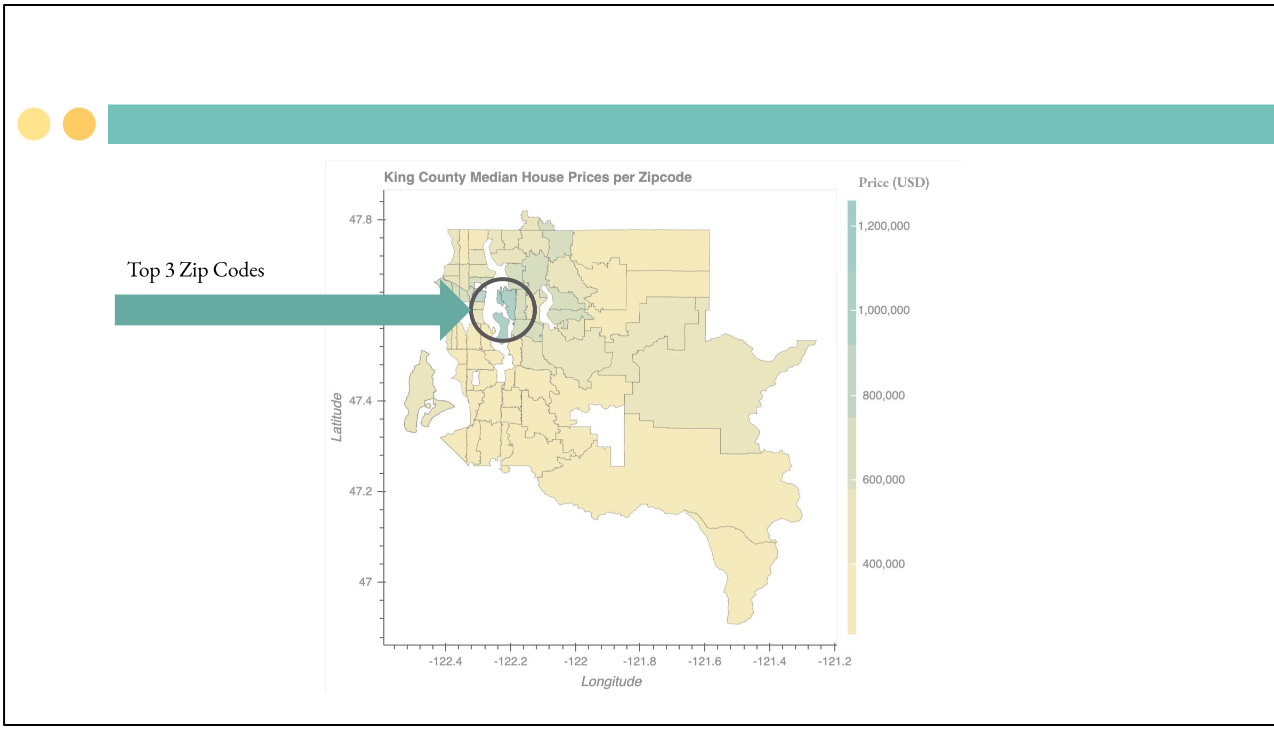
The Data used for this project consists of around 20,000 prices for house sales which occurred between May 2014 and 2015. The first step was to clean this data and ensure fit for purpose. We then looked at which features would provide interesting insights and answered questions which would help the marketing campaign. Finally we looked into building a model capable of predicting house prices.



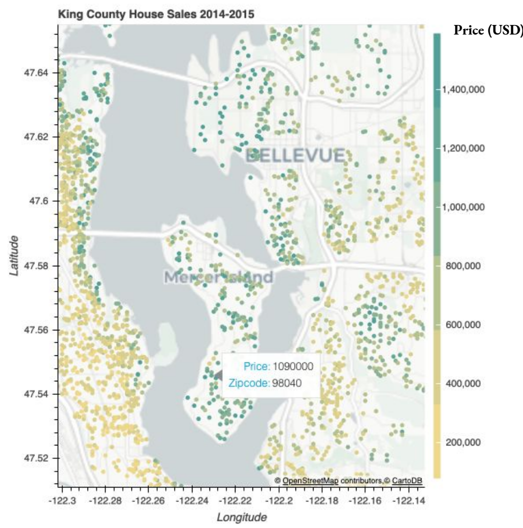
The first key feature we explored was location.
We want to understand which areas have the highest house prices and should be the focus of our campaign.



We looked at the median house price per zip code. There were 70 different zip codes in our dataset.



We can see that the top 3 zip codes are in the NW area



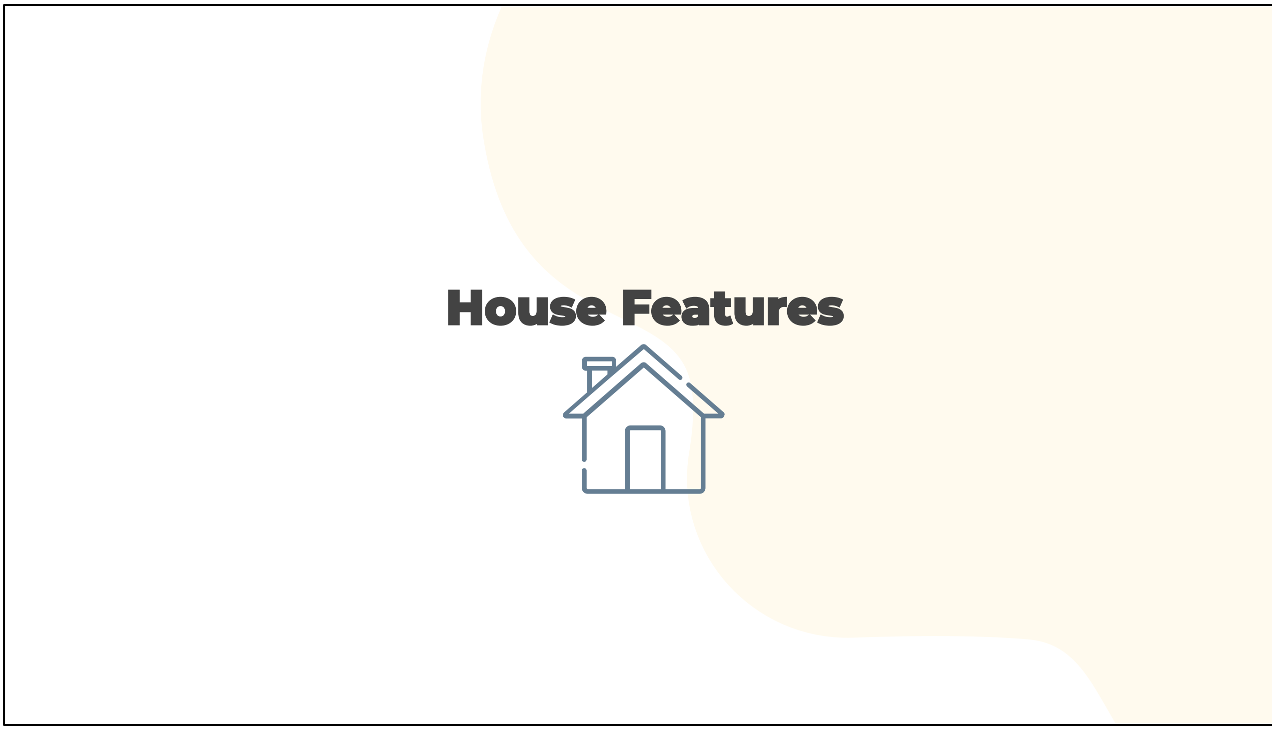
We then zoomed in and saw that being on or close to the waterfront is key.
The most expensive houses are in Mercer Island and Bellevue

Recommendations

Location

- Importance of waterfront
- Bellevue and Mercer Island areas

So as a first recommendation, we would advise focusing on homes on/around the waterfront and in particular those two neighbourhoods.



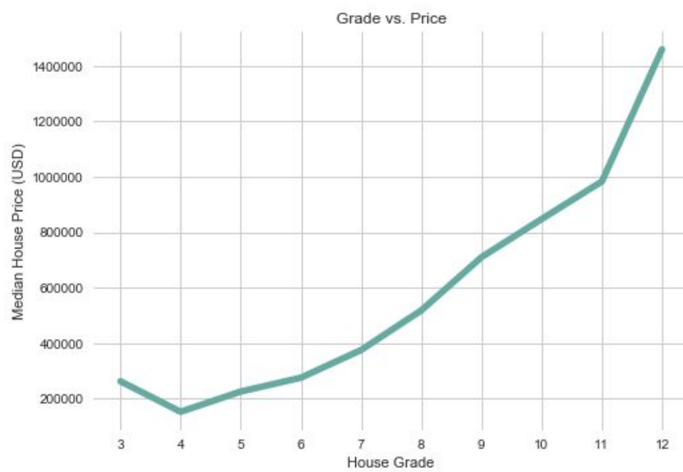
We then looked at house features to see which factors drive the price up and also which factors are the best predictor of price

The most obvious feature is sqft of living space. We would expect a larger house to be more expensive.
We see broad increase, more sqft is more expensive but not clear cut

House Features - Living Area

Here is a house with small sqft, living area less than 2000 sqft but with a price of around USD 1.5m

On the other side of the spectrum, here is a house with a large sqt, over 4000 sqt but less than USD 500k.



Better predictor is grade, construction value from King County which ranges from 1 to 13 (though here we only had 3 to 12)
For the campaign we would recommend looking at houses with a grade of 10 or above. This would have a starting median price above $800,000. From the description we note that "Homes of this quality generally have high quality features. Finish work is better and more design quality is seen in the floor plans. Generally have a larger square footage."

Other House Features

Was viewed

Square Feet of Lot

Recently renovated

Condition

Year Built

Bedroom count

Bathroom count

Waterfront view

Presence of Basement

Number of floors

Other features we investigated briefly

Recommendations

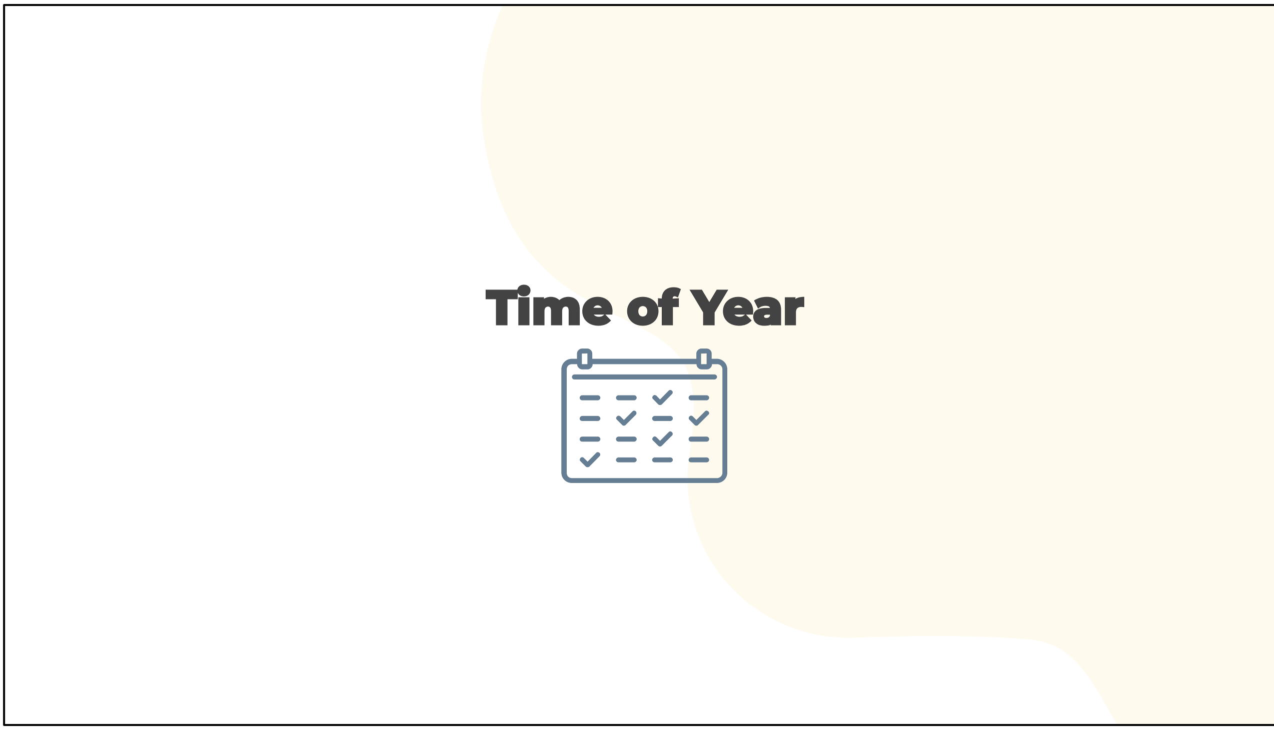
Home Features

- Larger living area (sqft)
- Grade 10 and above
- Grade better predictor

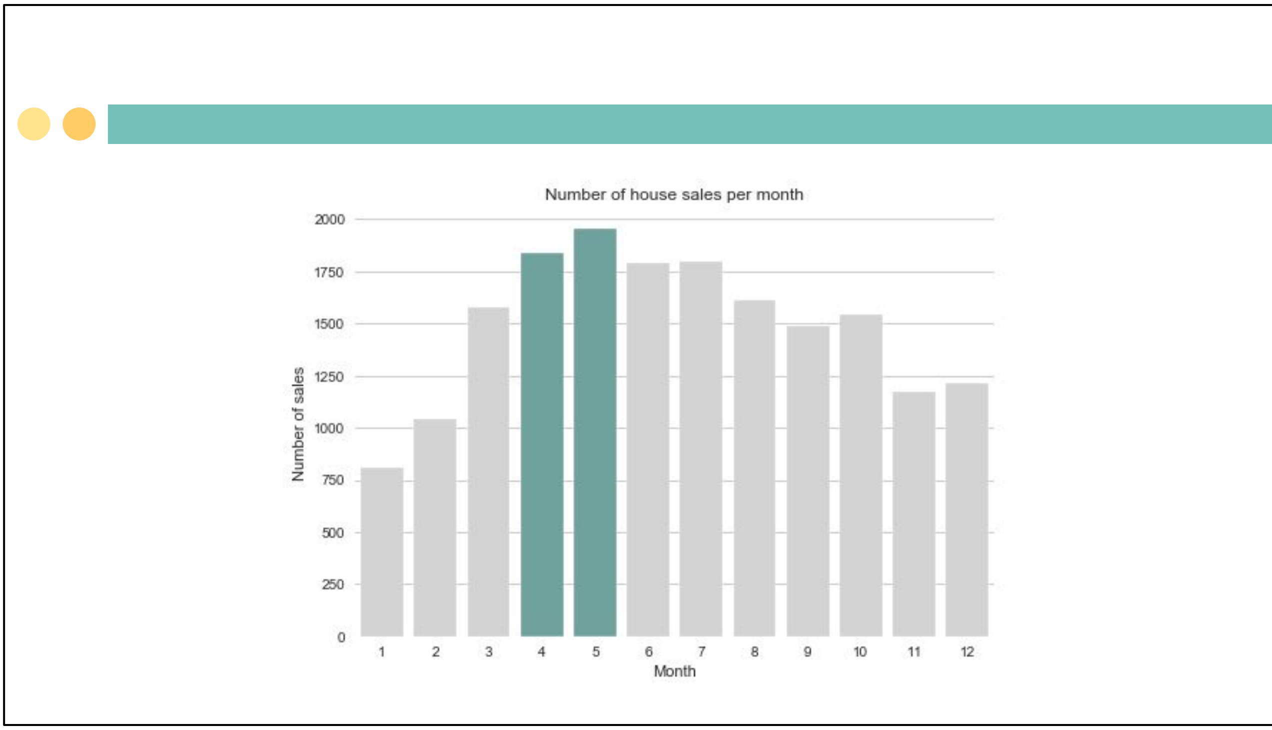
Location

- Importance of waterfront
- Bellevue and Mercer Island areas

So for our second recommendation, we would advise looking at grade and focusing on grade 10 and above



We investigated whether the median price was higher at a certain time of the year and whether there are any trends.
Firstly we saw that the median price was almost the same throughout the year, so no influence.



However the volume of sales varied throughout the year.
April and May are the most popular months for house sales. In contrast, January and February have the lowest number of sales.
We recommend launching the campaign in March/April

Recommendations

Home Features

- Larger living area (sqft)
- Grade above 10
- Grade better predictor

Location

- Importance of waterfront
- Bellevue and Mercer Island areas

Time of Year

- Constant median house price
- April/May highest sales volume
- Campaign in March/April

Recommend March/April for marketing campaign, with a view of sale closing in Q2

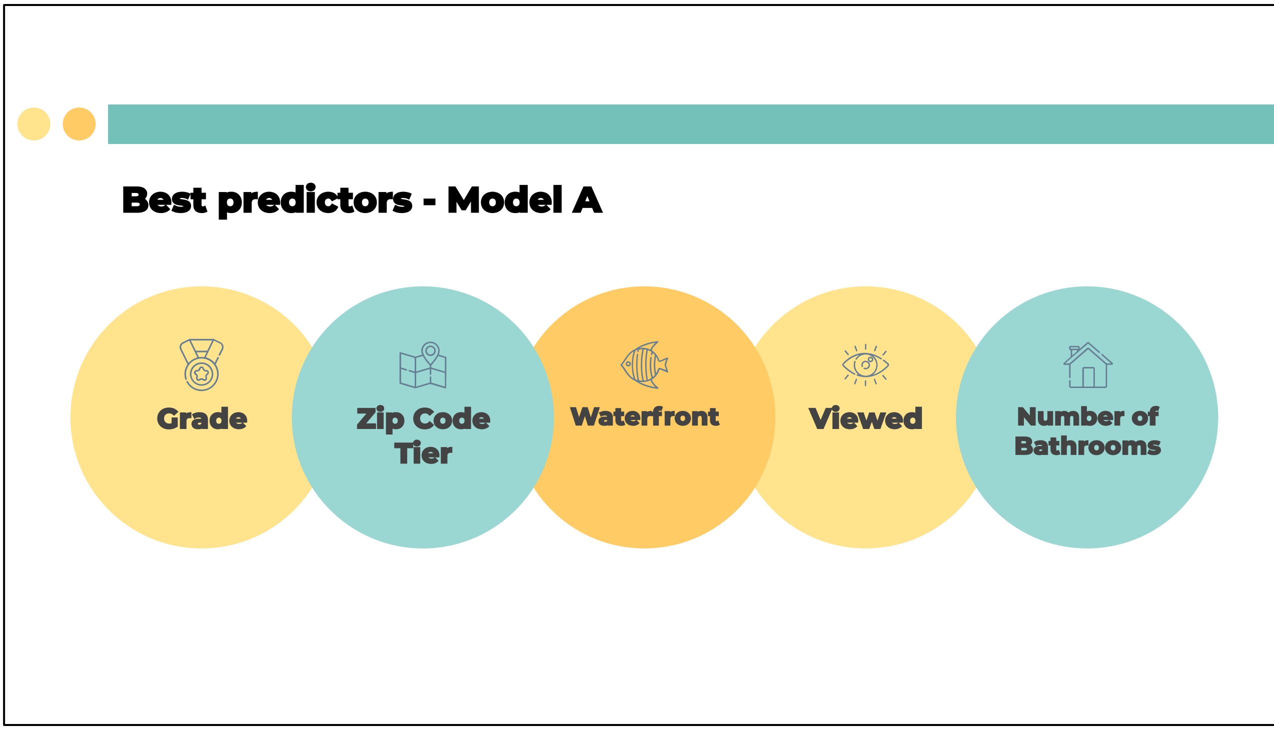
Predicting House Prices

Let us now discuss our model for predicting house prices. In fact we will present two models which serve different purposes.

Predicting House Prices

	Model A	Model B
Features	17	87
Pros	Easy to interpret, generalises well	Performance
Cons	Less accurate	Uses exact zip code data
Score	0.70	0.83
Mean Error	USD 132,444	USD 99,654

- Features is number of attributes used to predict
- Score is a value between 0 and 1 with 1 being a perfect fit
- Mean error is the amount in USD on average over or under
- 70% of variations in price can be explained by model A vs 83% for model B



Let’s look at the top predictors based on Model A. Whilst it looked like it had 17 features, in fact only 5 due to our the computer processes. So here are the key features. These have the strongest influence on house prices.
We can also quantify the monetary value of a house displaying a certain feature.
E.g. being grade 12 is worth $52,000 more than grade 11
E.g. being on the waterfront is valued at USD 277,442.
More details in the appendix

Further work

School Ratings Data	Proximity to a good school is likely to increase sale price
Commuting Time Data	Homes better connected to downtown Seattle are likely to be more valuable
Longer Time Scale	See which areas show signs of growth/ decline

Here are some of the things from a non-technical perspective that would be worth exploring further.
We only had data for one year, so couldn't establish any trends but with more recent data, would be able to see which locations are worth targeting now.
From a data science point of view, we would look to use more advanced models if we wanted to improve performance. However such model wouldn't be as easy to interpret.

THANK YOU

Nadine Amersi-Belton
nzamersi@gmail.com
https://github.com/nadinezab

APPENDIX

Model A Coefficients

	Feature	Coefficient
1	grade_12	515,536
2	grade_11	463,199
3	zip_tier_1	435,924
4	grade_10	327,375
5	zip_tier_2	313,144
6	waterfront	277,442
7	zip_tier_3	250,036
8	zip_tier_4	248,289
9	grade_9	233,634
10	zip_tier_5	168,374
11	viewed	153,790
12	zip_tier_6	133,920
13	bathrooms_4	105,063
14	zip_tier_7	99,454
15	grade_8	88,713
16	grade_5	-94,930
17	grade_4	-136,094