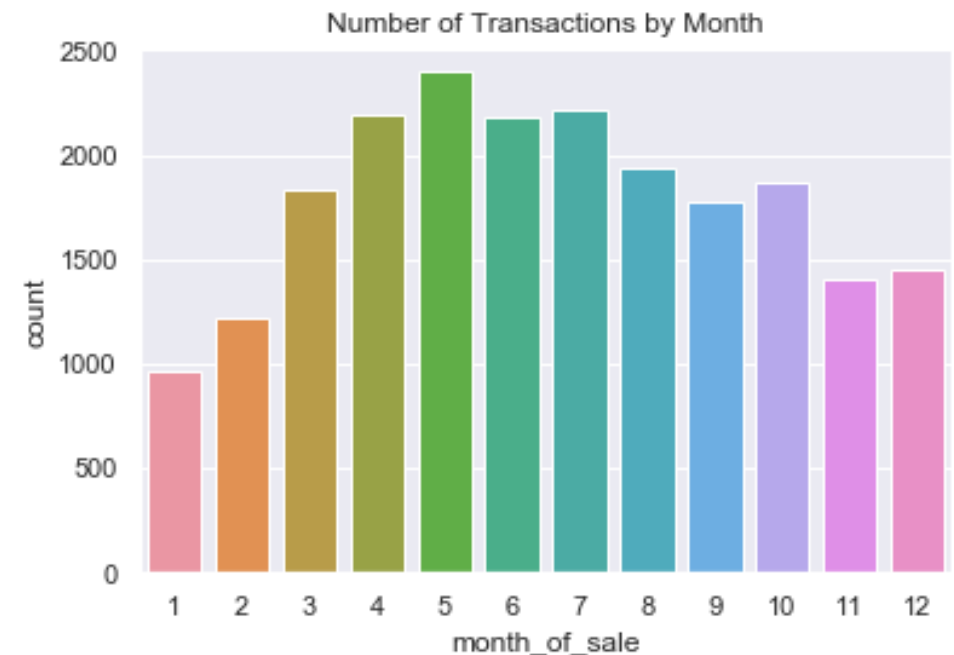# MOD 2 PROJECT

VITTORIO SCACCHETTI & LUC BATTY

SCOPE OF THE PROJECT

Detecting trends in seasonality of the King County housing market

Investigation into the difference in house price based on location

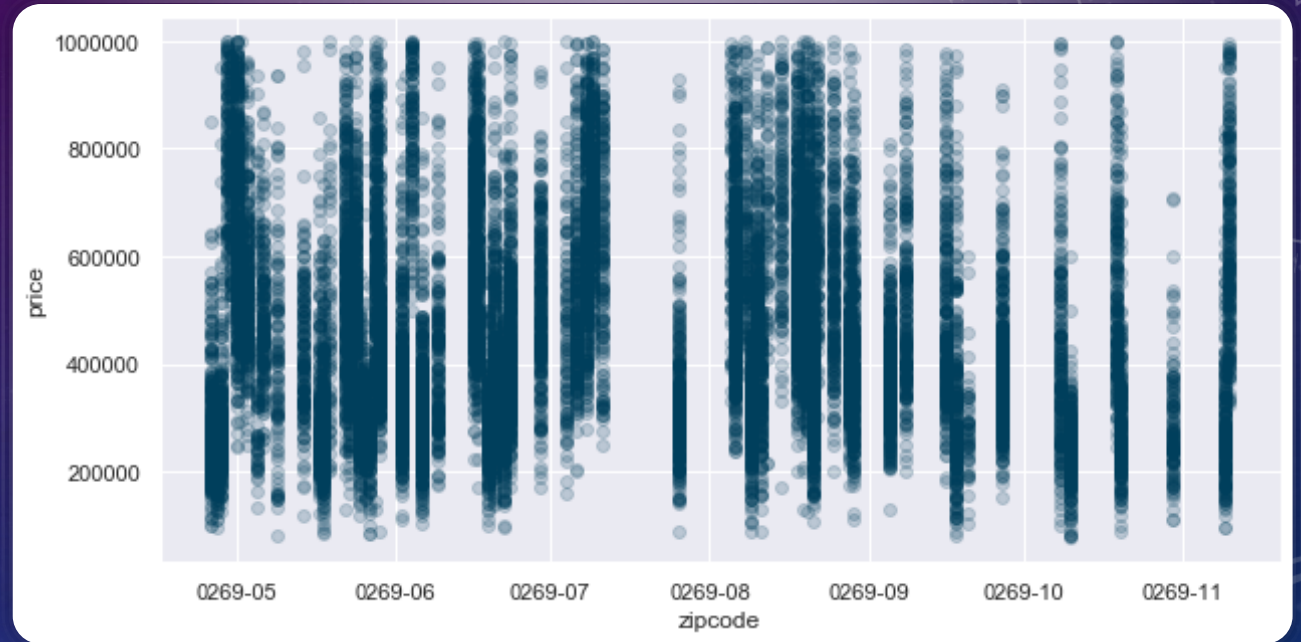Prediction of house price based on proximity to water

UNCOVERING THE SEASONALITY OF THE MARKET

Number of Transactions by Month

EFFECT OF LOCATION ON PRICE

Seattle city mean price: $544k

Seattle vicinity mean price: $530k

PREDICTING PRICE BASED ON PROXIMITY TO WATER

Washington State has an idyllic coastline. It would not be surprising if houses carried a premium for being on the waterfront.

In fact, the waterfront houses in King County are more than three times more expensive on average than non-waterfront houses.

But is there a relationship between the price of the house and how close it is to the sea?
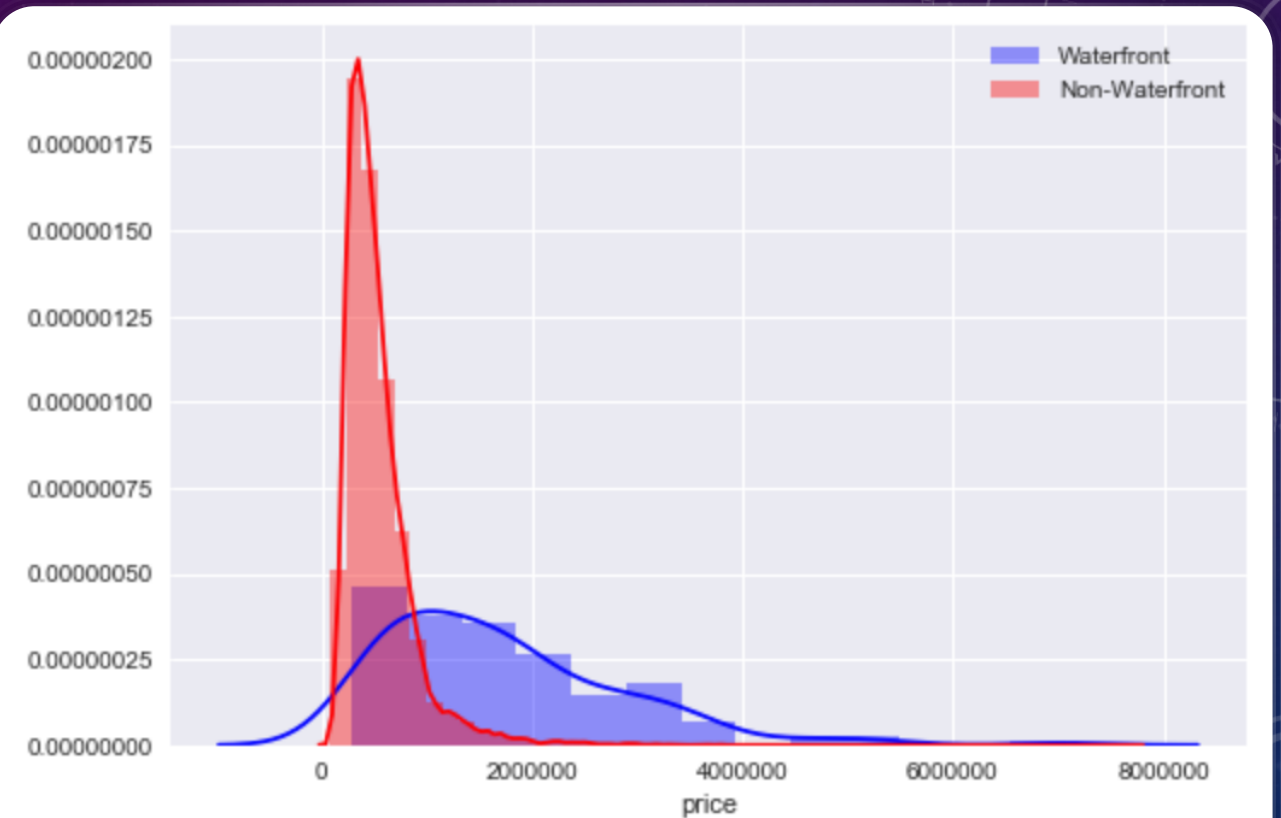
The mean price of a waterfront house is $ 1717215
The mean price of a non-waterfront house is $ 532560

# QUESTION: WHAT IS THE PREMIUM FOR PROXIMITY TO THE COASTLINE?

Step 1: Feature Engineering

- We derived an approximate value (in km) for the distance to the coastline for each house in the dataset.

- This was done by triangulating the coordinates for each house to the nearest house one the waterfront.

Based on the grouping and the high-price outliers, we supposed that some of the houses may be very large and / or very luxurious.

We therefore considered another variable: the house grade.

The aim was to construct a multivariate regression analysis that would predict the price of a house based its proximity to the sea, given its grade.

Fitted values versus z_distance_to_water_km
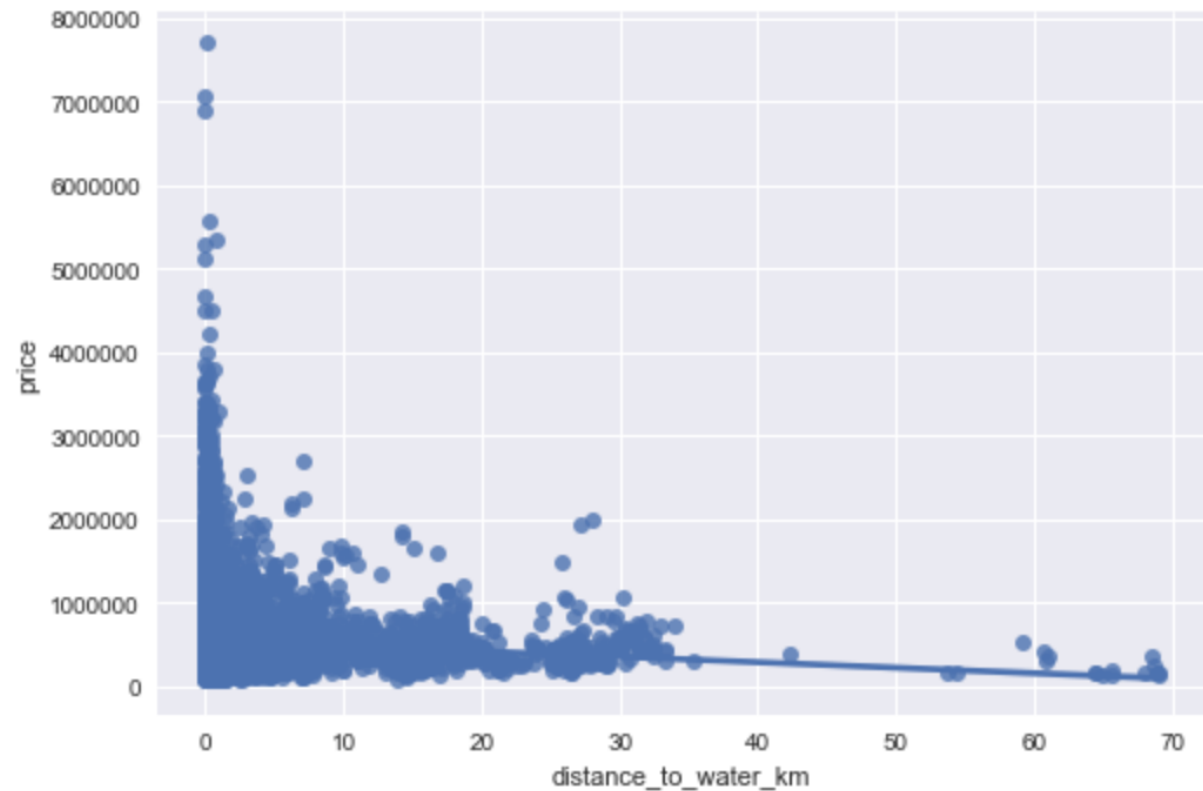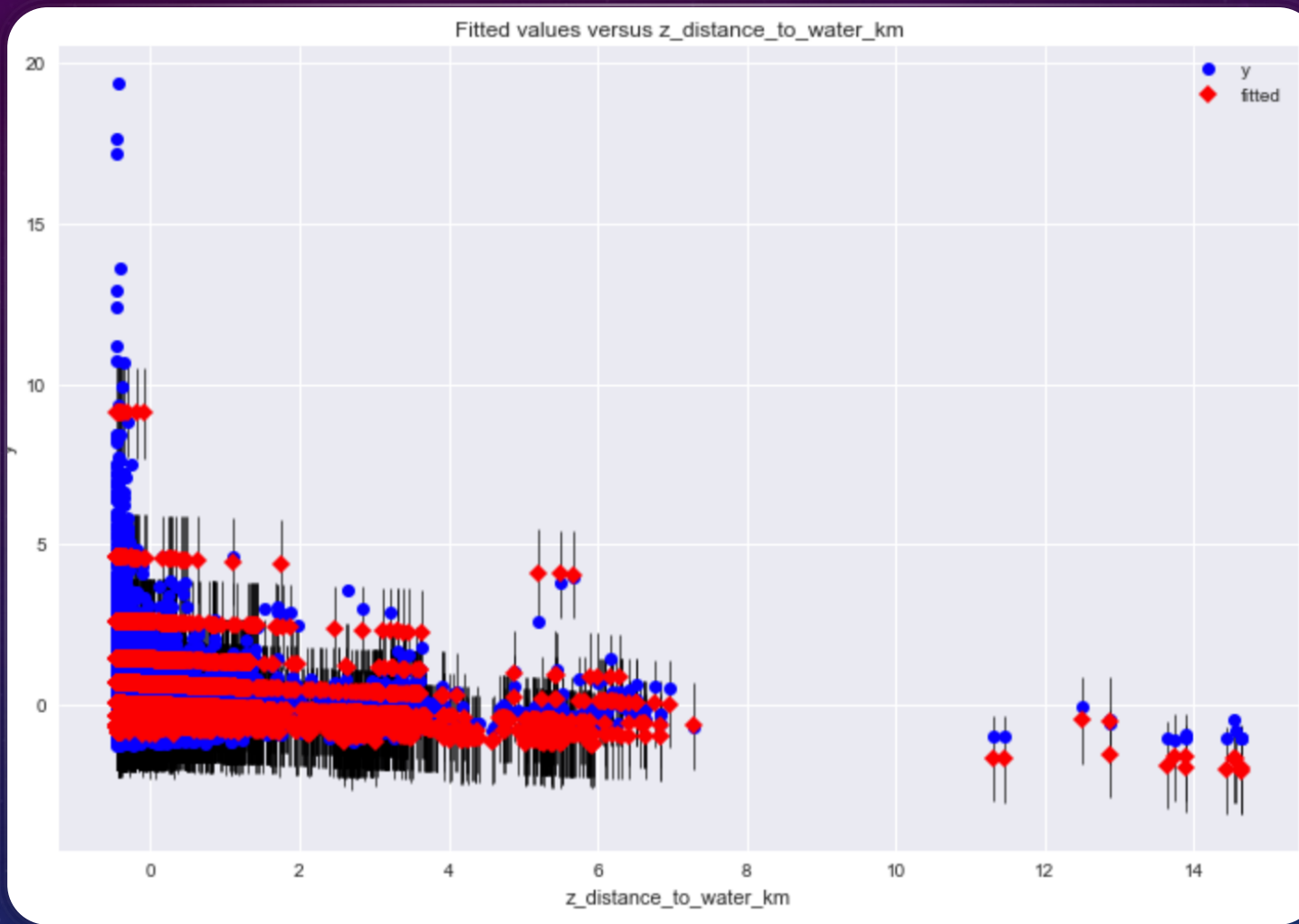
During the training phase, the model was deemed adequate based on a consideration the actual vs. fitted values.

The model performed relatively well, with an R^2 value of 0.529.

This suggests that approximately 53% of the variation in price is explained by the variation in grade and proximity to the coastline.

CONCLUSION

THE MODEL GENERATED THE FOLLOWING INSIGHTS:

- As the grade band increases, the loss in value decreases.

- All else being equal, the value of the house decreases by approximately $8,200 for every kilometer from the coastline.

- The P-Values suggest that the probability of the coefficients being zero is very low, and therefore that we can reject the null hypothesis: that there is no relationship.

Feature	Loss of value
Constant (initial value)	$3,058,213.49
Grade 3	$0.00
Grade 4	-$2,796,257.72
Grade 5	-$2,806,812.65
Grade 6	-$2,739,424.35
Grade 7	-$2,635,540.02
Grade 8	-$2,489,964.94
Grade 9	-$2,285,851.95
Grade 10	-$1,945,105.00
Grade 11	-$1,524,451.31
Grade 12	-$870,999.49
Distance to water	-$8,227.47 per km