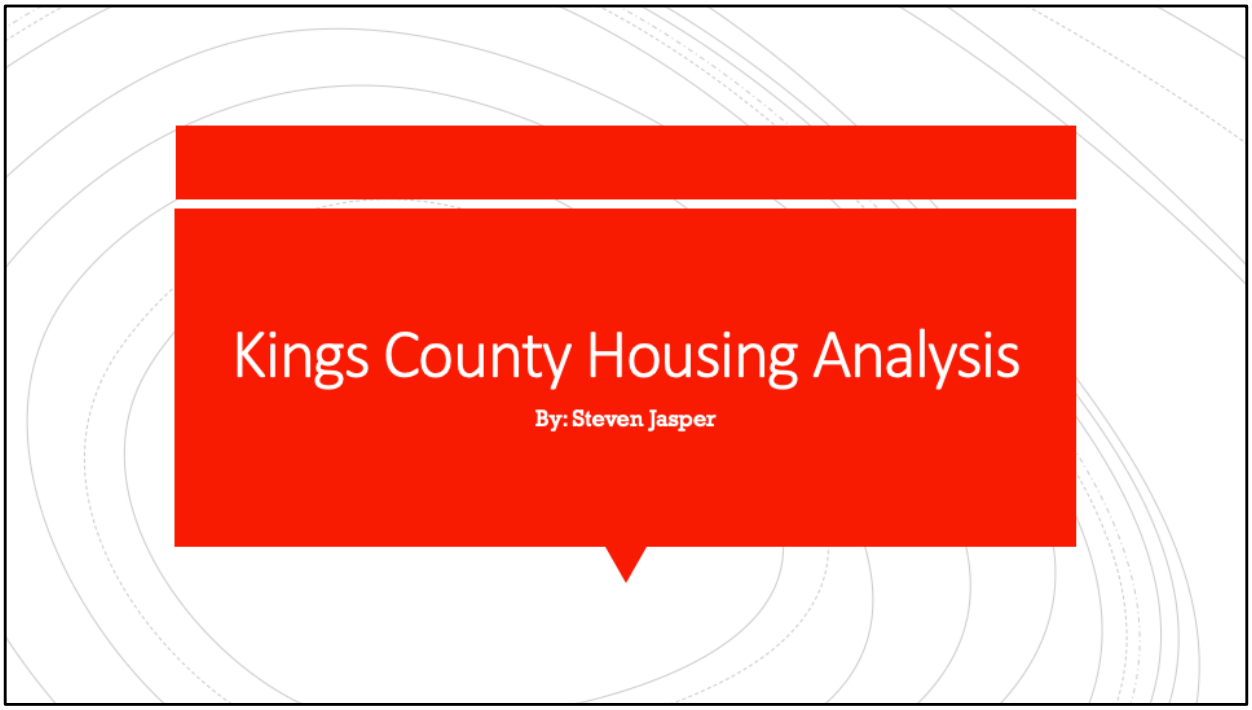
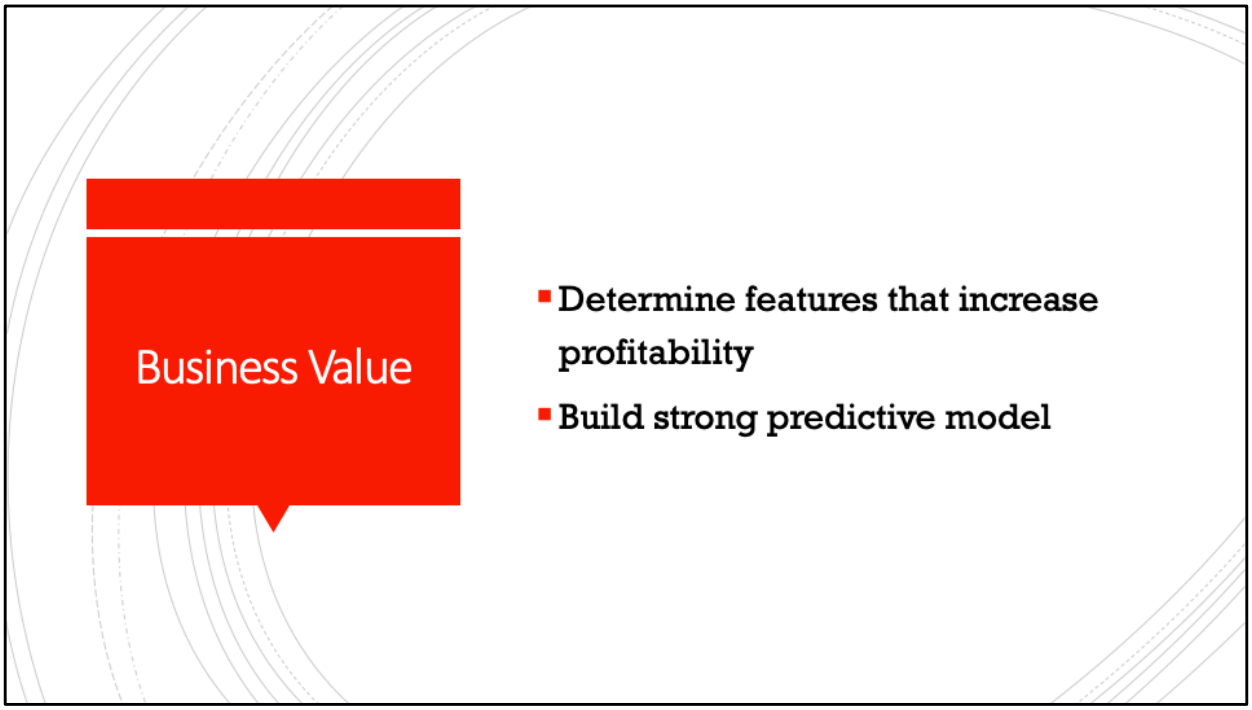


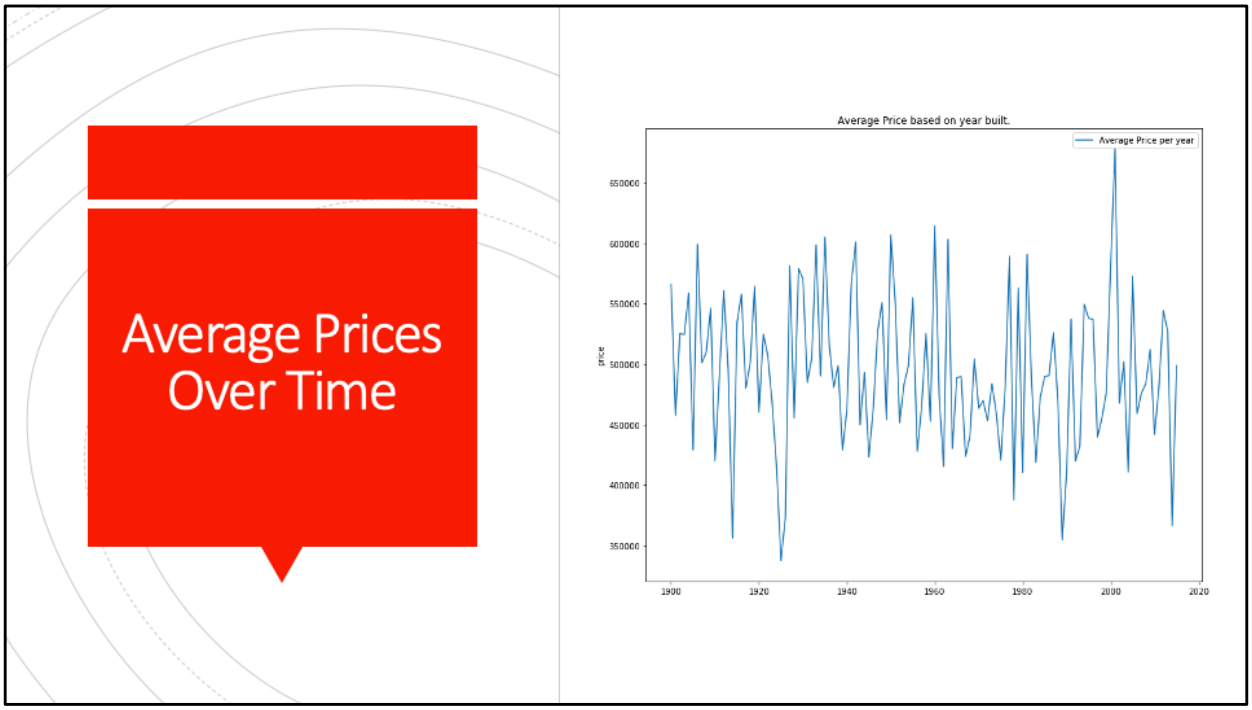
Here we have the presentation for the analysis conducted on the Kings County Housing Dataset.

Our Problem statement: We would like to maximize our ROI within Kings County, Washington. We will explore a number of features from our data that will effect our prices.



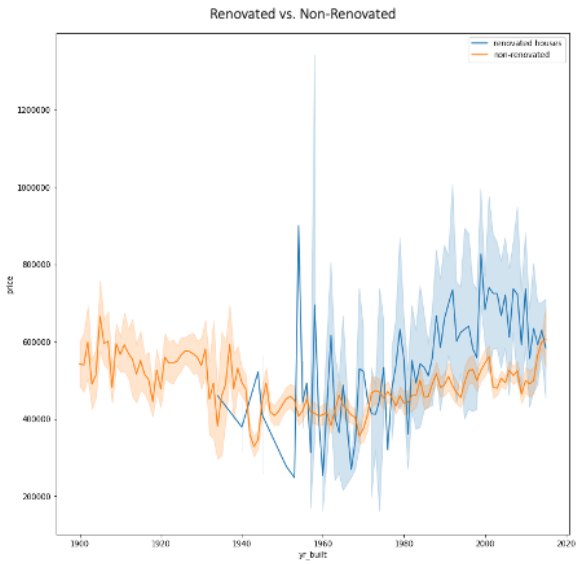
The business value that this analysis will add: We will determine features that increase the prices of the houses to allow us to increase ROI in the area, and create a strong predictive model for future investments.

Our methodology for this analysis is to utilize the OSEMN model building workflow as a framework. We will handle outliers utilizing z-score analysis, and use specific significant features to train and develop our model.



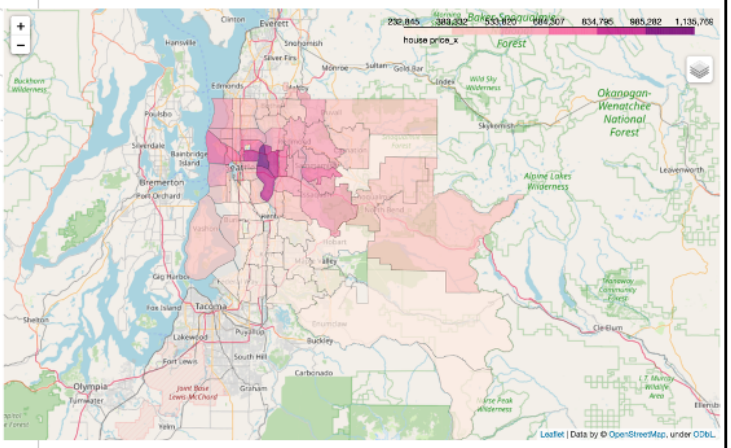
This graph demonstrates the average prices over time. You can see that this seems to be highly dependent on the global economy at the time. You can see shortly after 1930, housing began to fall, this is the time around the Great depression, also you can see similar behavior around 2008 where the economy also crashed. We would want to do an economic analysis prior to investing.

Renovated vs. Non-Renovated

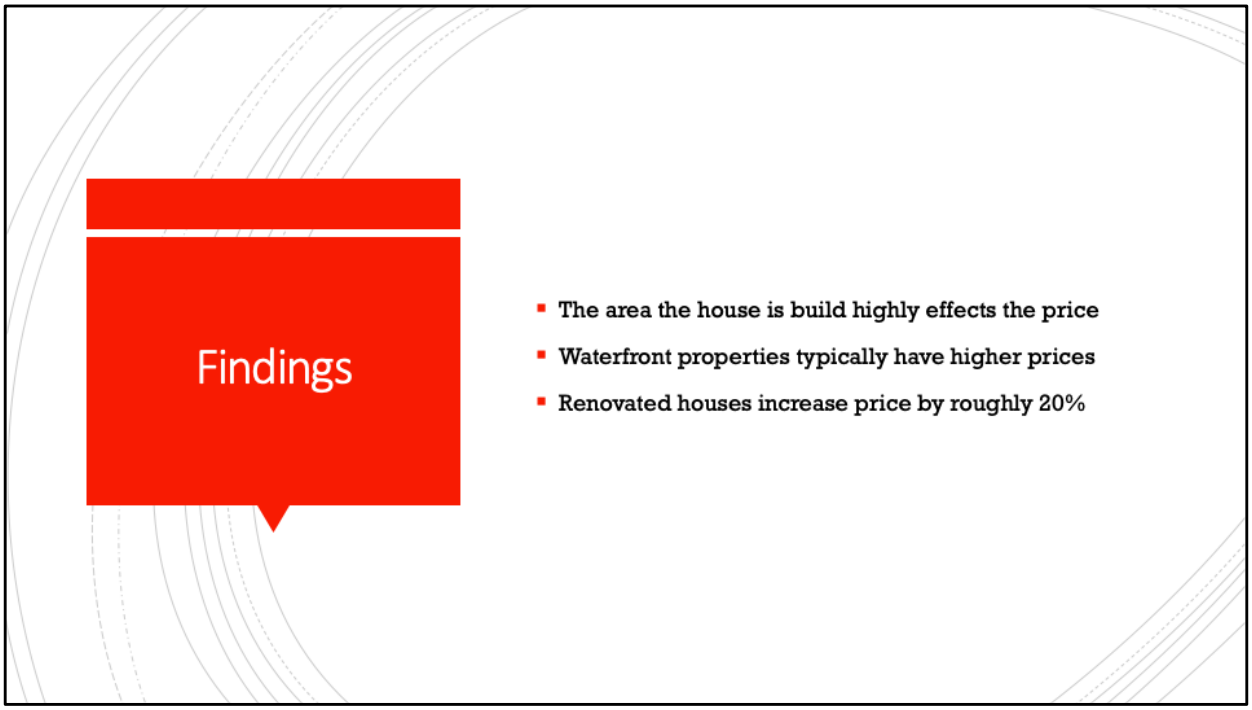


Here we will see the comparison between the price for renovated houses and non-renovated houses. The blue line demonstrates the year that the house was renovated whereas the orange line is houses that have not been renovated and the year they were built. As you can see houses renovated in the 80s and early 2000s are priced higher than non-renovated newer housed. However, more recently we see that new houses out-sell renovated houses the closer to the current year.

Housing Prices per Zipcode

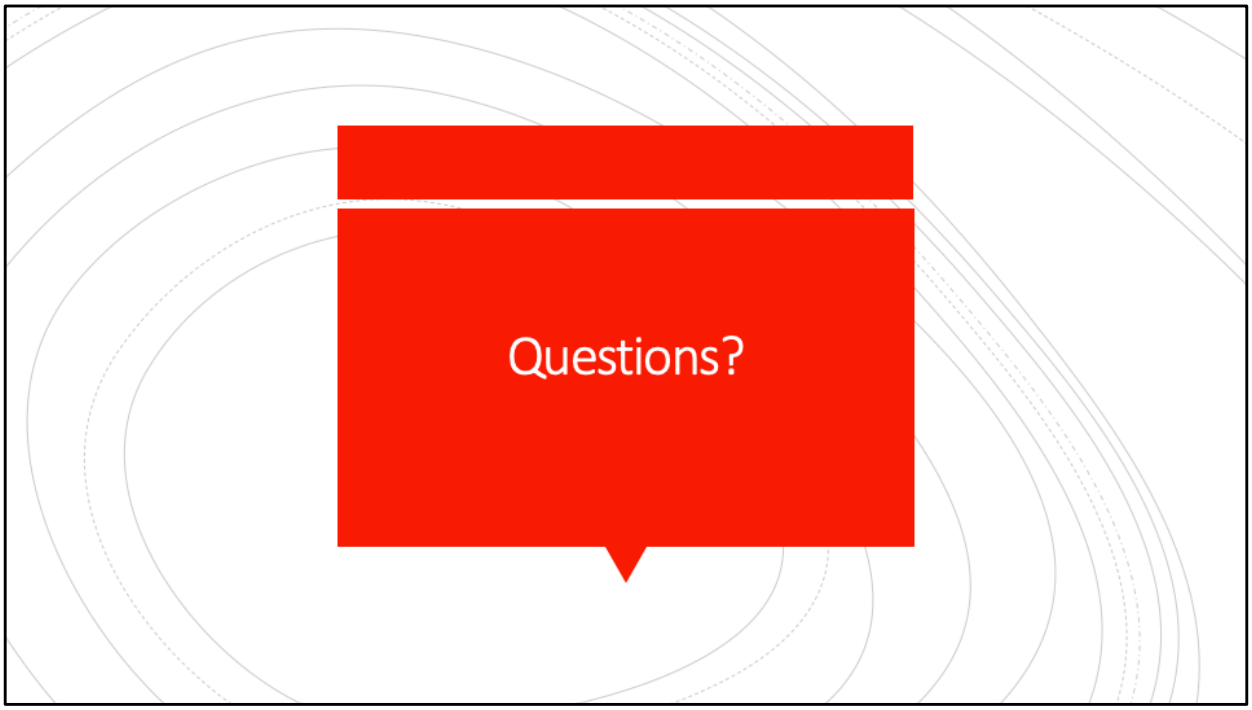


Here you can see a breakdown of kings county by zip code. The darker colors are higher average price for that area. As you can see the Bellevue, Medina, Mercer Island, and Madison Park areas have a much higher average housing price.



Our findings are that we want houses in high price areas, near the water that are either brand new, or renovated in the last 40 years.

For future work I would like to explore additional features such as housing prices based on local population and the strength of local economy. Are there more grocery stores in the higher price areas?
I would also like to incorporate these features into the model to increase the accuracy beyond the current 84%.



Do I have any questions?