

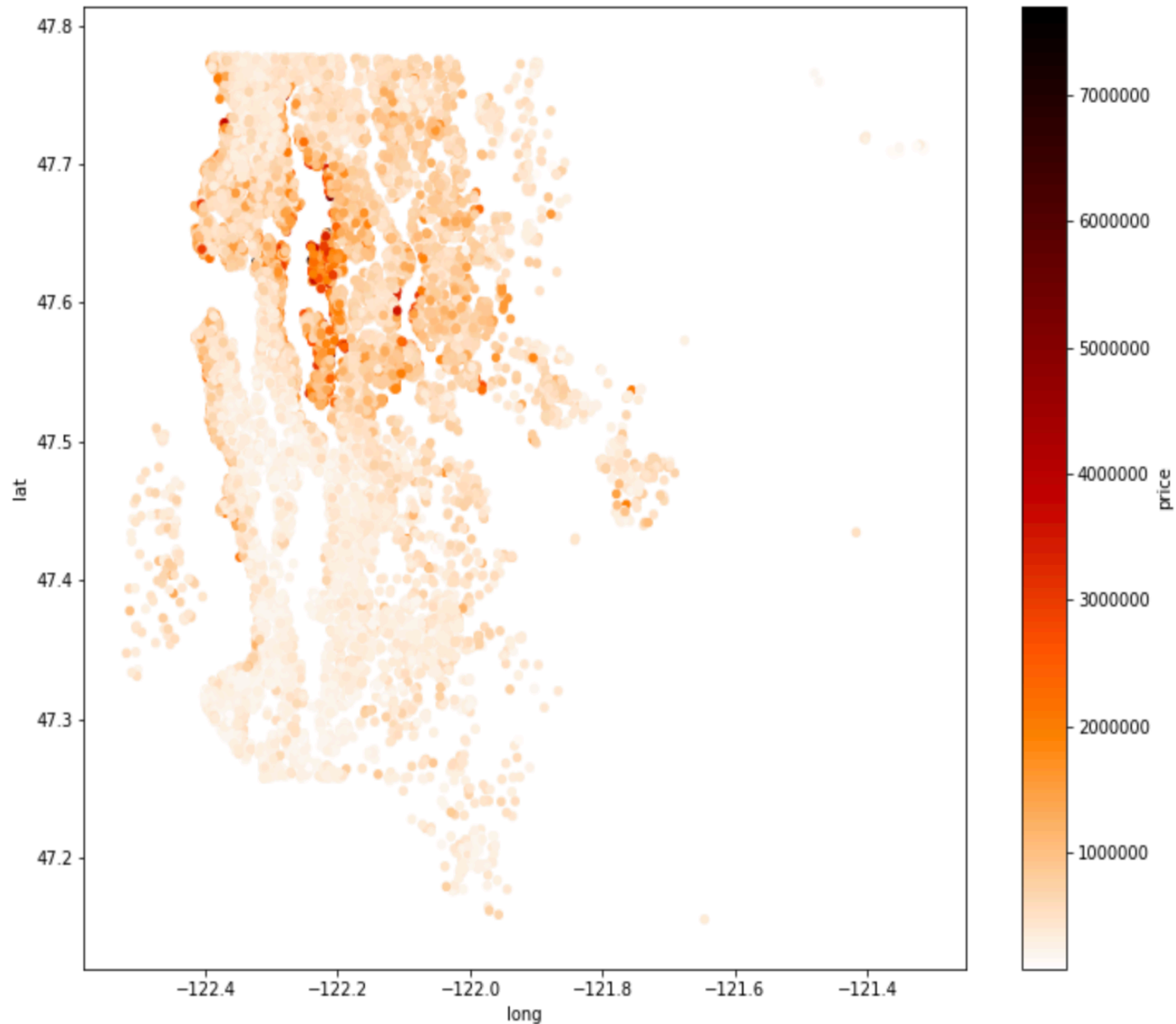
King County House Data Analysis

Esra Guzel

Modul 2 Final Project Presentation

DATA UNDERSTANDING

- Dataset contains information for more than 21000 house sold during 2014 and 2015 in King County
- For every house, there are over 20 attributes ranging from location, living square foot, grade and renovation year to build year etc.

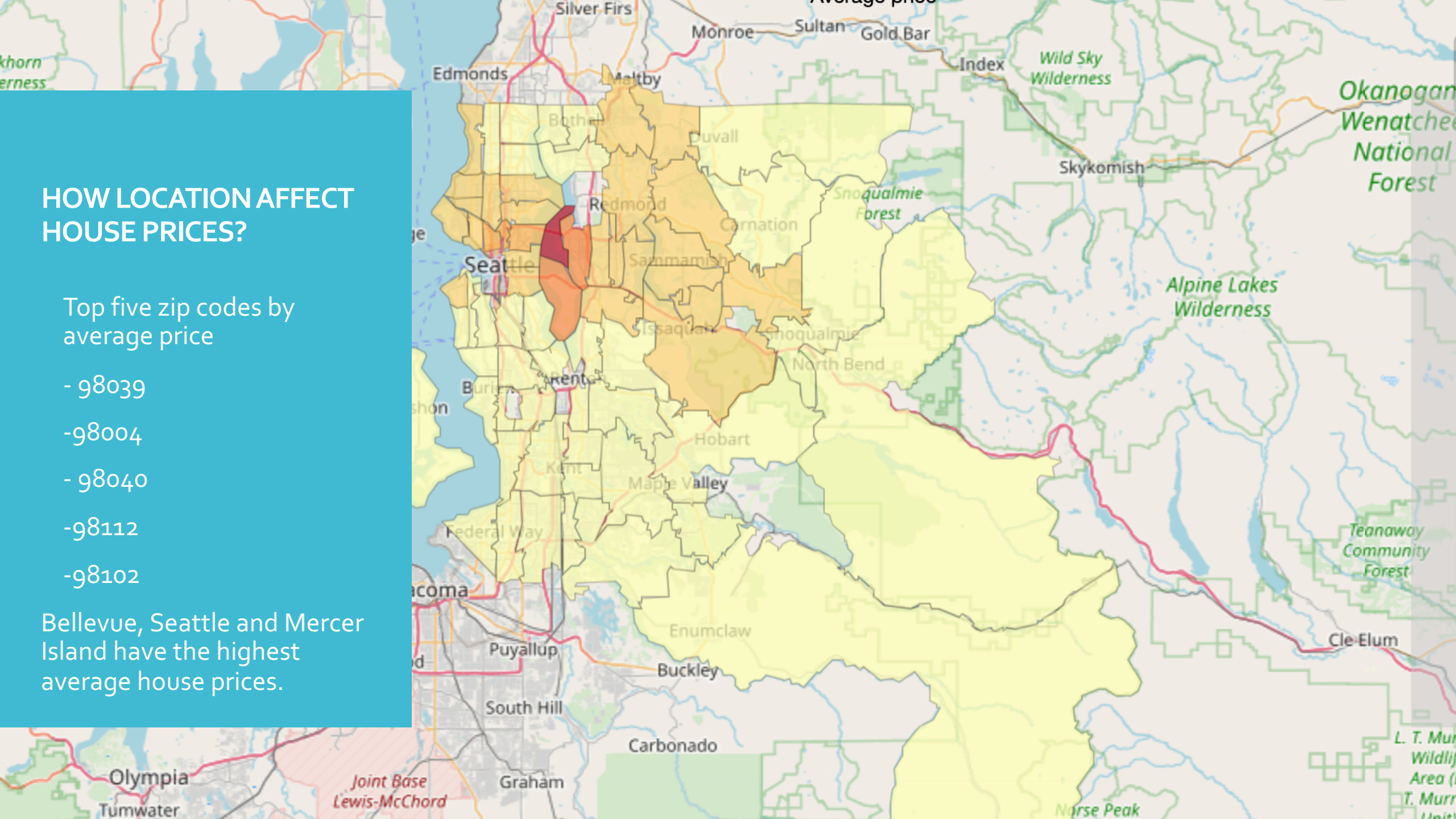


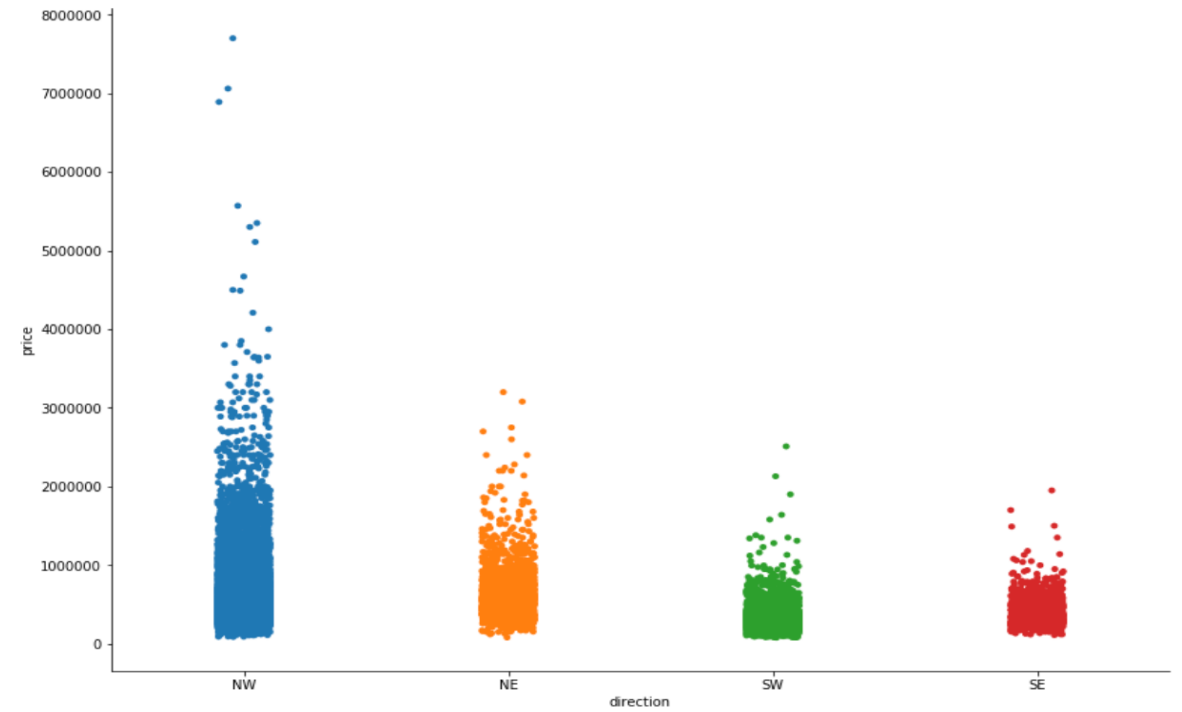
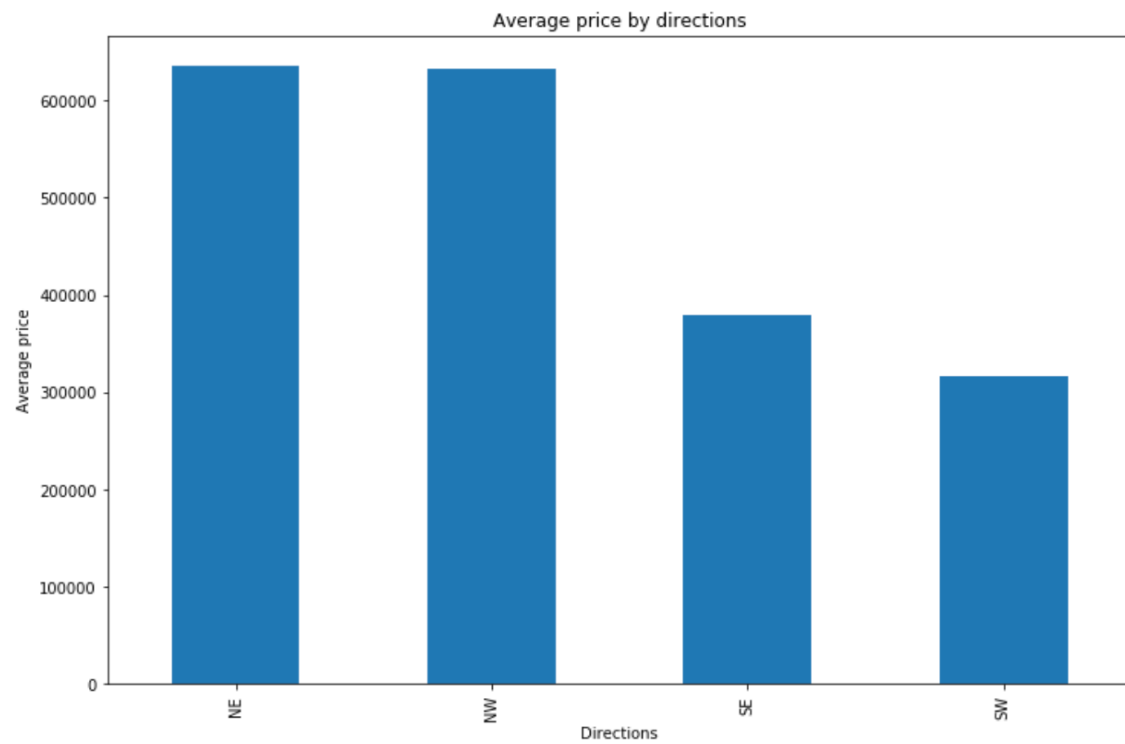
HOW LOCATION AFFECT HOUSE PRICES?

Top five zip codes by average price

- 98039
- 98004
- 98040
- 98112
- 98102

Bellevue, Seattle and Mercer Island have the highest average house prices.

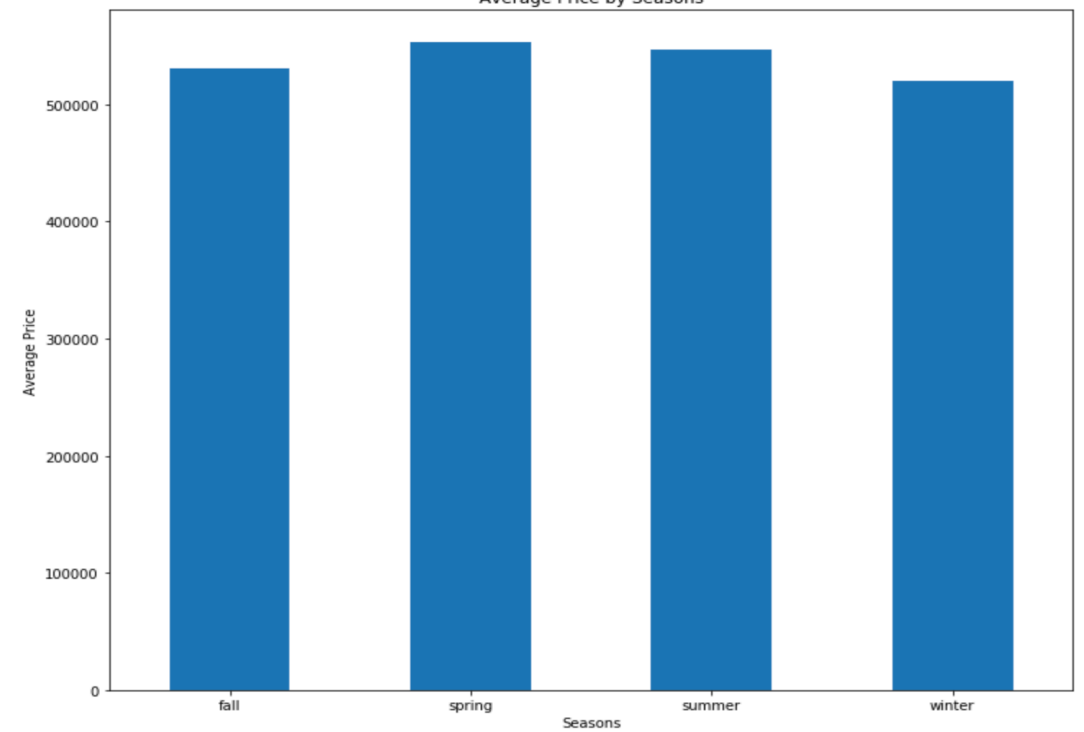
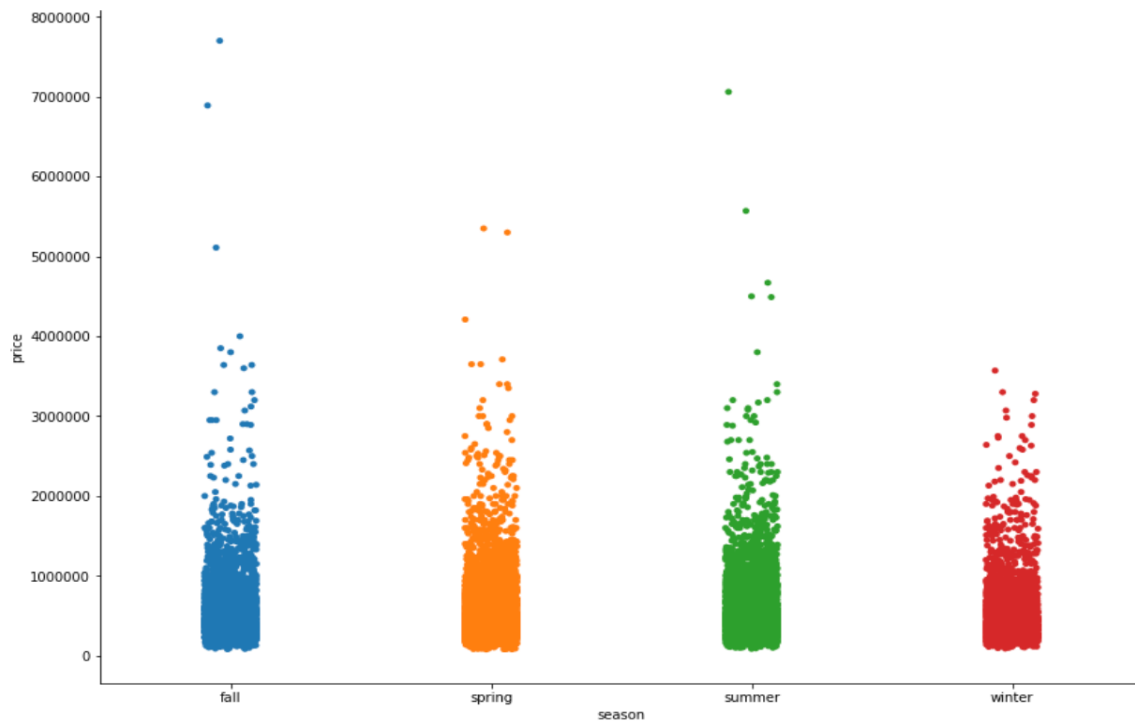




LOCATION EFFECT

- Northwest part of the county has the highest prices and shows high variance
- Northeast is the second in terms of highest price and variance
- Southeast average prices are higher than Southwest
- Average price for Northeast and Northwest houses are close

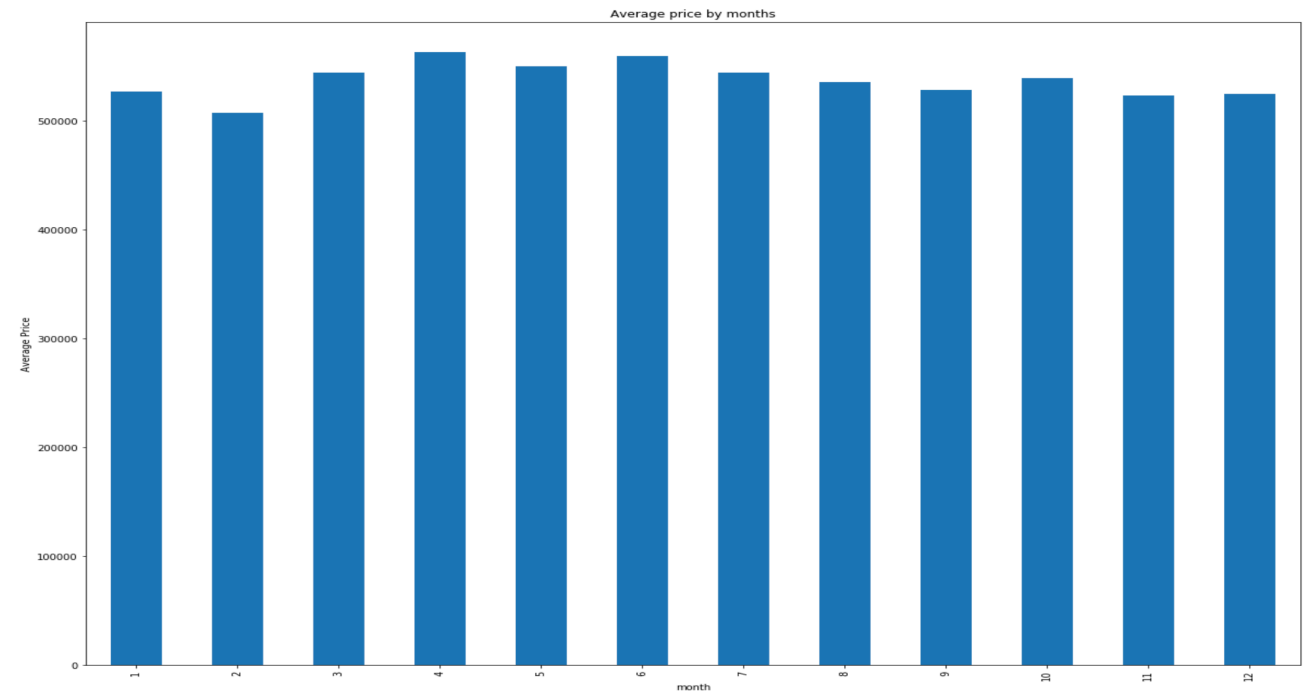
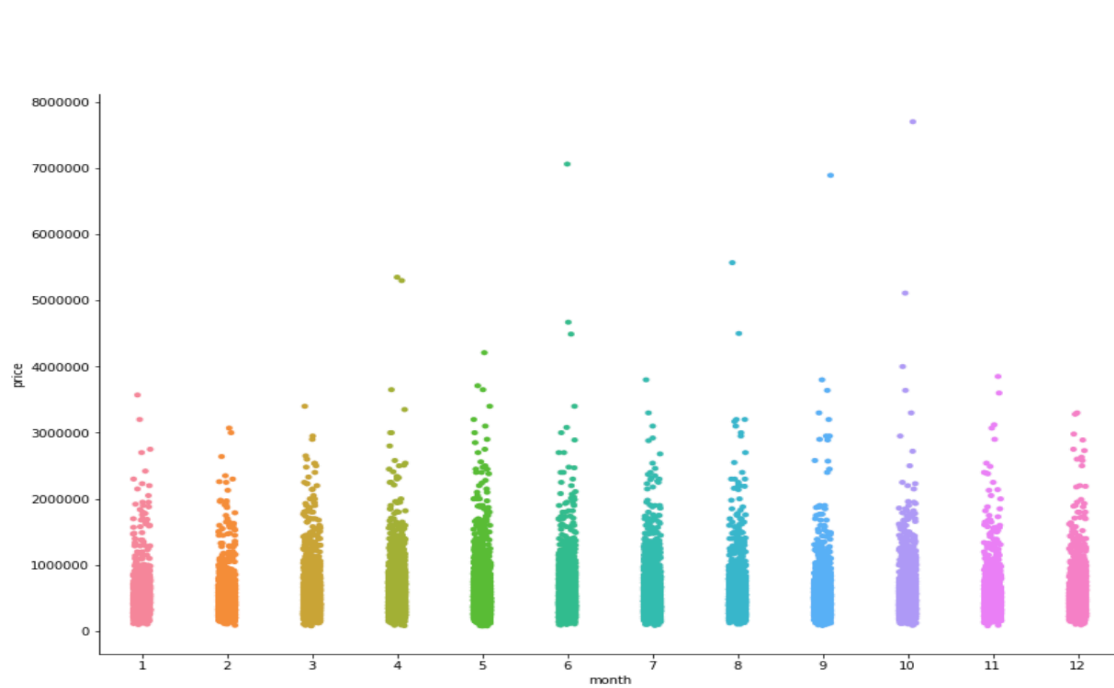




HOW SEASONALITY EFFECT HOUSE PRICES?

SEASONS:

- In general during winter, houses were sold for relatively lower prices although this is not a major indicator for the price.
- Fall shows high variance for price range followed by summer and spring.
- Spring has the highest average price followed by summer and fall.
- The average prices for each season are very close.
- Overall season is not a major driver of the price.



HOW SEASONALITY EFFECT HOUSE PRICES?

MONTHS:

- February has the lowest, April has the highest price average but the difference between them isn't major.
- While October has the highest variance, February has the lowest variance.
- There isn't a big difference between average price for months.
- Month is not a good predictor for house prices.

DRIVERS OF PRICE AND FORECASTING

While predicting the house prices, not every attribute is equally important as a predictor. Some of the most important features are:

Zip codes

Grades

Bedrooms

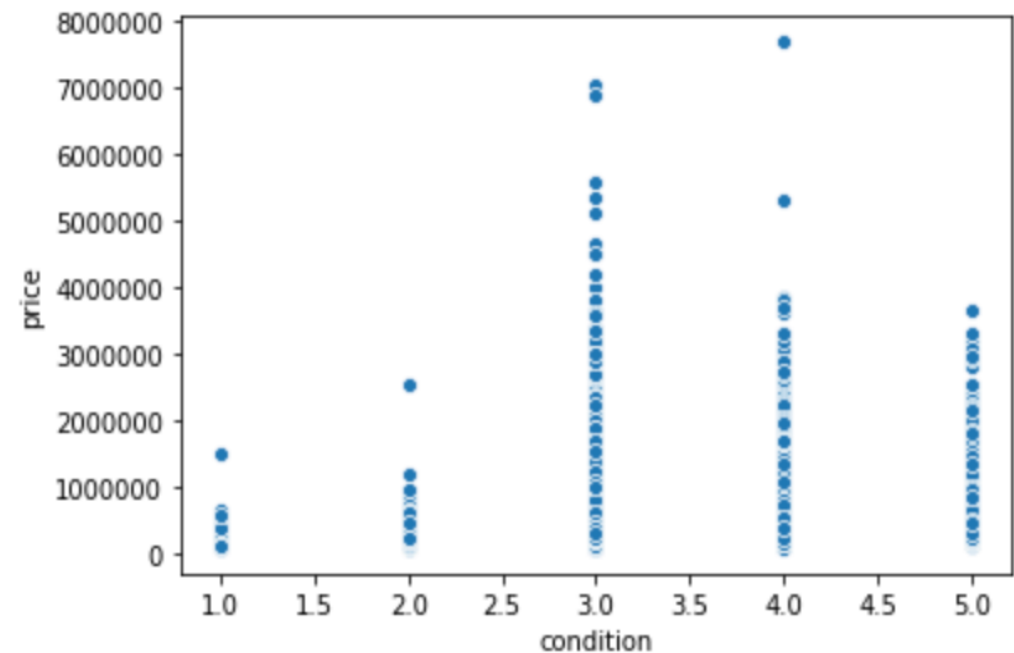
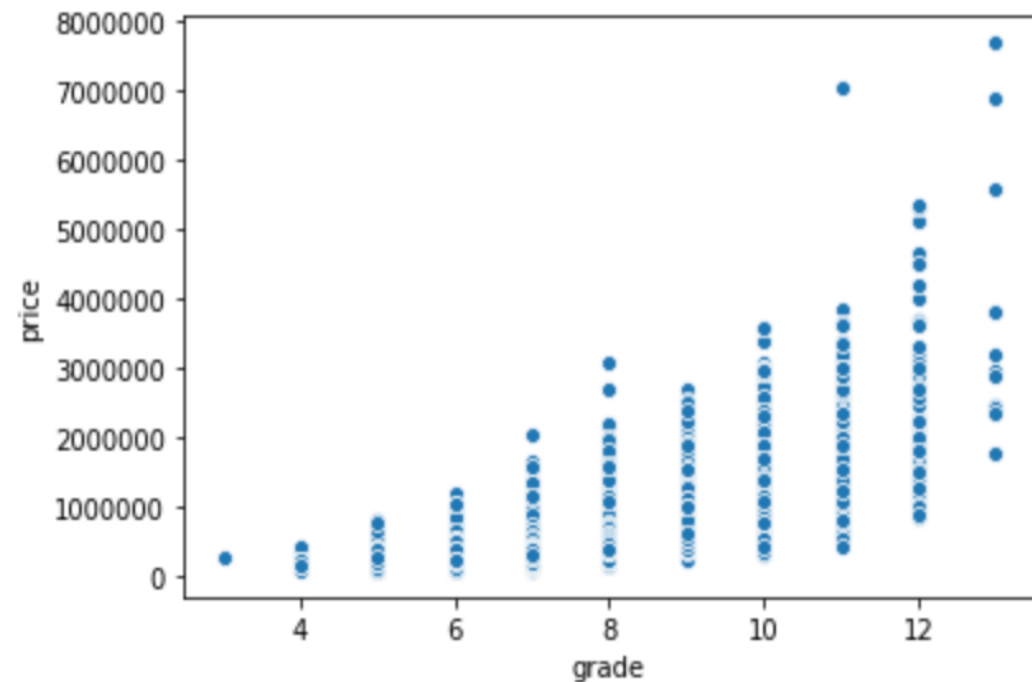
Weak features are distinguishable by their relatively lower correlation output. Additive value of these features as predictors are diminishing, therefore they provide less support to a robust model. Some of the weak features are:

Square living space for the nearest 15 neighbours

Condition

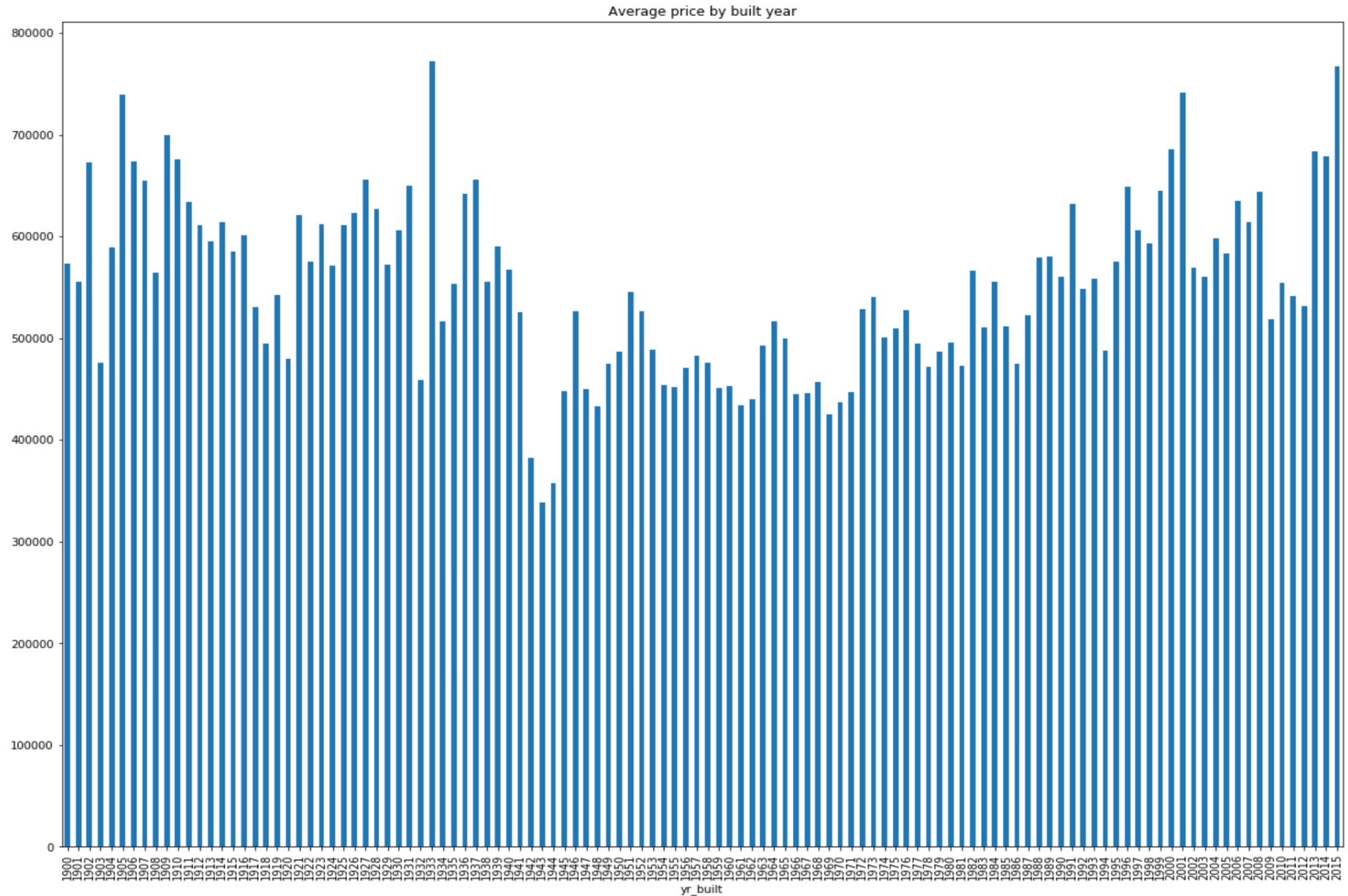
View

The model explains **85 %** of the variance in house prices.



TRENDS

- Decreasing trend from 1900's until 1940
- Increasing trend from 1940's to 2015
- Average house prices by built year is cyclical
- Time series analysis may reveal more insight about the trends in house pricing.



The slide features a light gray background. On the left side, there is a teal-colored geometric shape that is a parallelogram with a slanted right edge. On the right side, there is a larger teal-colored geometric shape, also a parallelogram with a slanted left edge, which appears to be a continuation of the design on the left.

THANK YOU