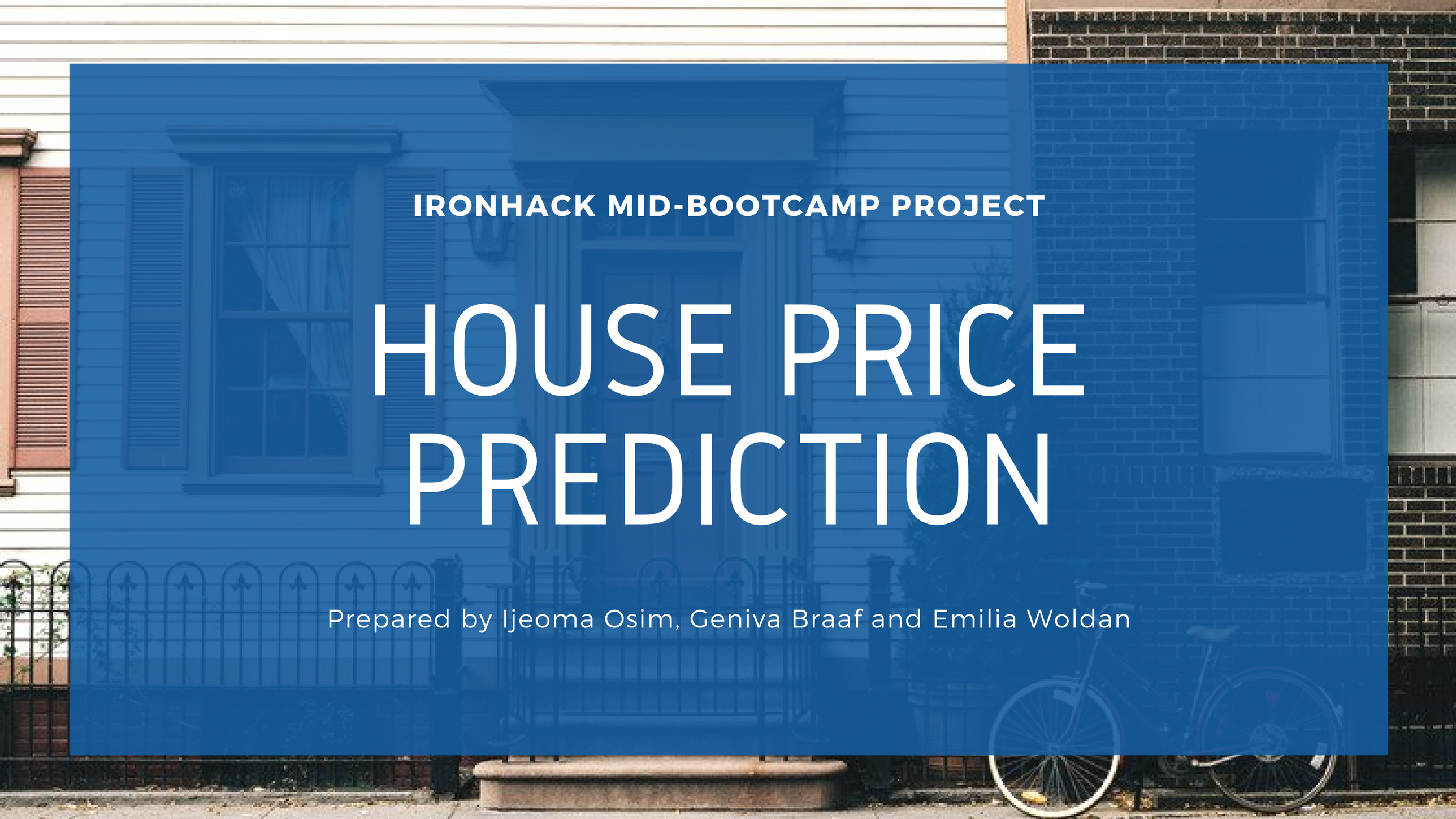


IRONHACK MID-BOOTCAMP PROJECT

HOUSE PRICE PREDICTION

Prepared by Ijeoma Osim, Geniva Braaf and Emilia Woldan

TODAY'S DISCUSSION

OUTLINE OF TOPICS

Intro to project
Tableau & SQL insights
Machine learning & Results
Lesson learned
Questions

INTRO TO PROJECT

Using the dataset that consist historic information of houses sold between 2014 to 2015 in King County, Washington with over 20K transactions describing 19 unique features about the transaction, the building, the property, and the location we will present a machine learning model using a regression algorithm.

PROJECT GOALS

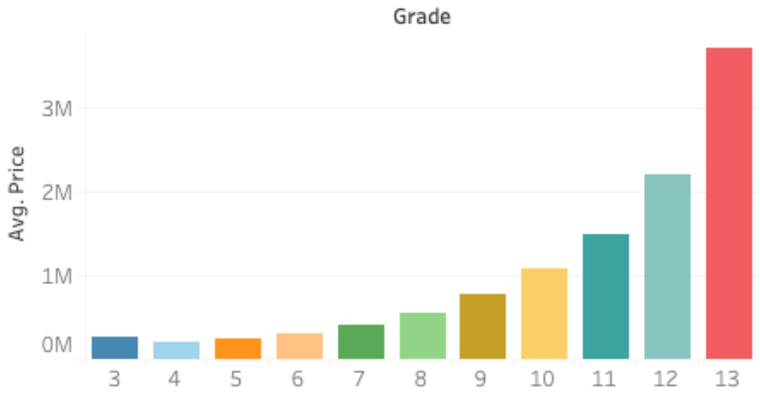
STEP 1
CLEAN DATA

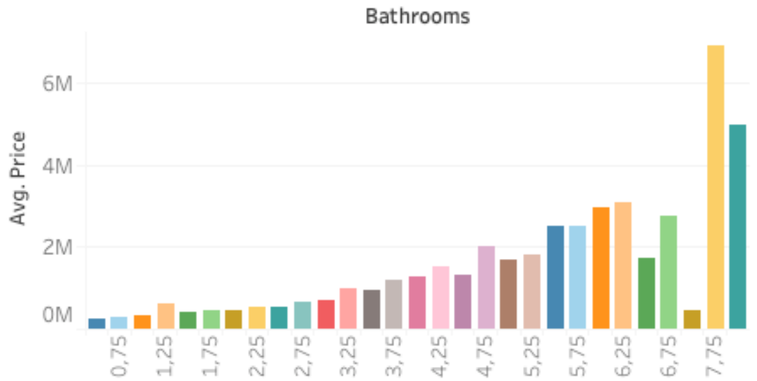
STEP 2
EXPLORE DATA

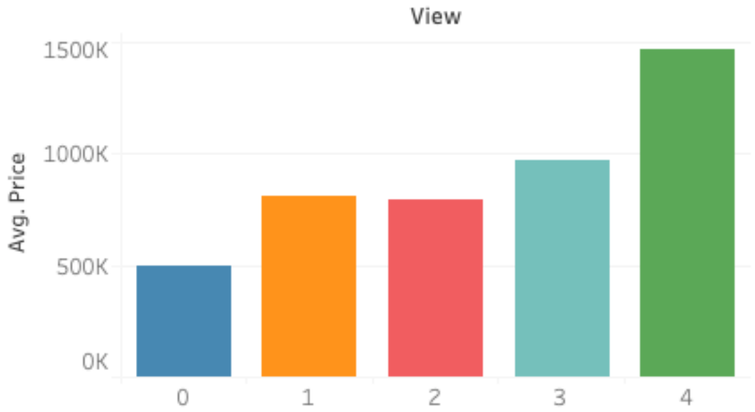
STEP 3
MODEL

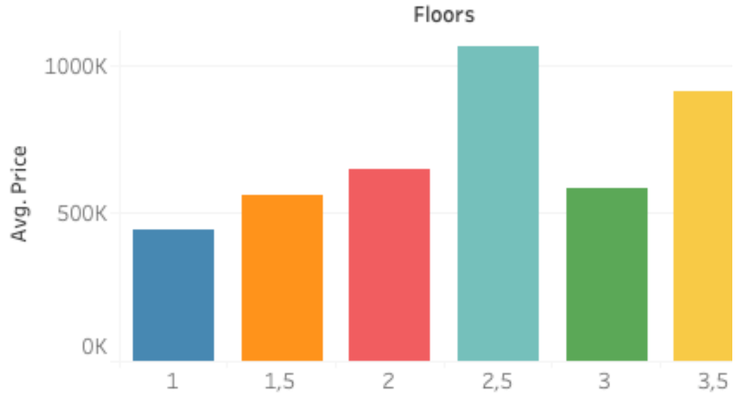
STEP 4
PRICE PREDICTION

Tableau & SQL insights









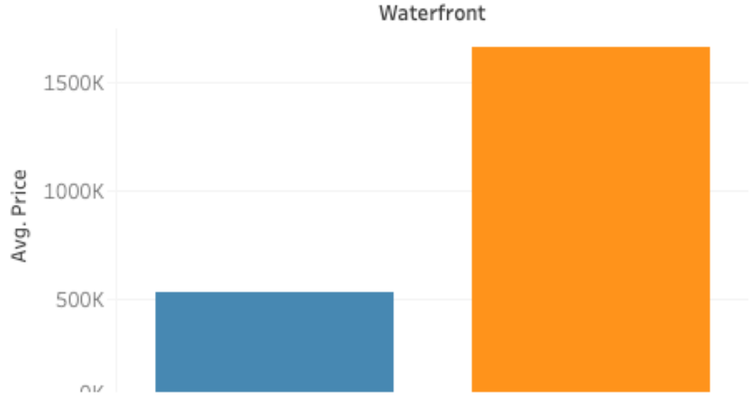
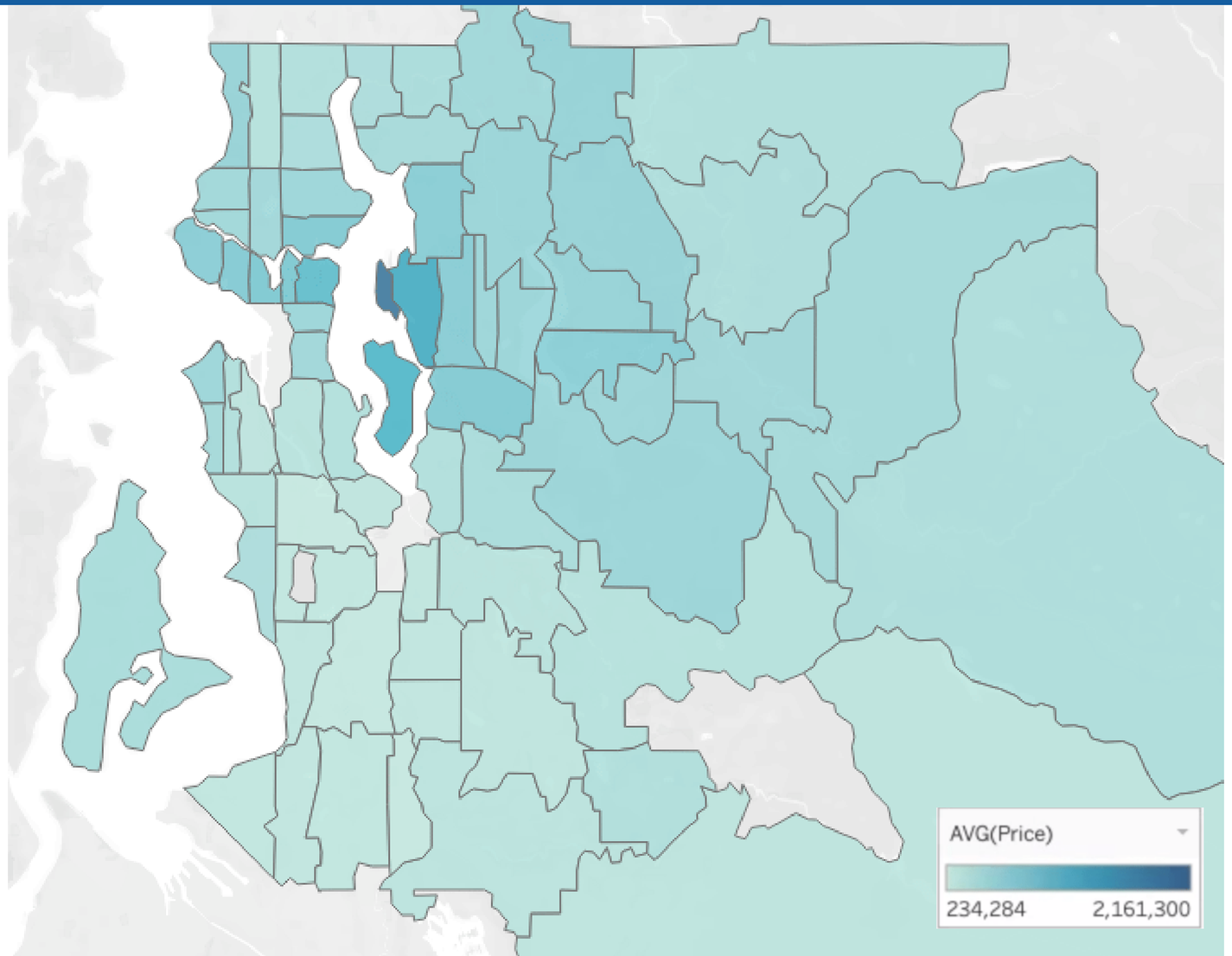


Tableau & SQL insights



MACHINE LEARNING & RESULTS

FEATURE ENGINEERING

- changing zip code to categorical and getting dummies
- removing outliers from price and scaling it with log10
- extracting additional information from data

RESULT ANALYSIS

- final model has a score of 84%
- house price is correlated with a location, which was not obvious at the beginning
- model performs well in predicting houses above $650K
- model is not good enough in predicting house price below $100K

NEXT STEPS

- removing outliers from specific features
- finding patterns between high errors features

LESSON LEARNED

LESSON 1

Features in given data may have some hidden information. So always look deeper beyond the surface value of the columns.

LESSON 2

Additional features can be extracted from already existing features, which can impact on your model performance.

LESSON 3

This project forced us to pay more attention to scaling techniques and be more sensitive when and where to apply them.

LESSON 4

Checking out the relationship between the features and target variables on Tableau can be very helpful in understanding your data set.

QUESTIONS