

# PREDICTING HOUSE PRICES IN KING COUNTY

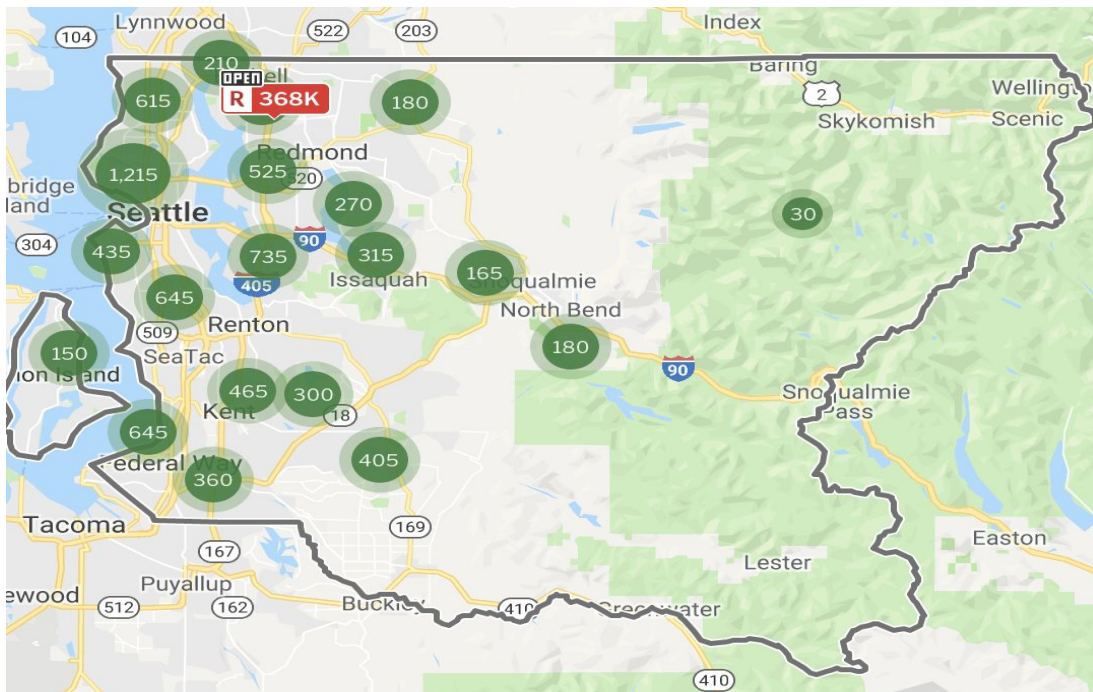


---

-HIRANYA KRISHNA KUMAR

# OBJECTIVE

# STREAMLINE YOUR HOUSE SEARCH BASED ON LOCATION AND BUDGET



# DATA COLLECTION



Scraped features related to the type of house like bedrooms, bathrooms, sqft-living, lot area, School rating, View etc



**BestPlaces**

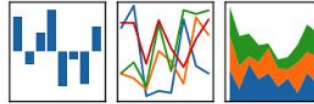
Scraped features for median income, commute time, median age by Zip Code of the house

# TOOLS

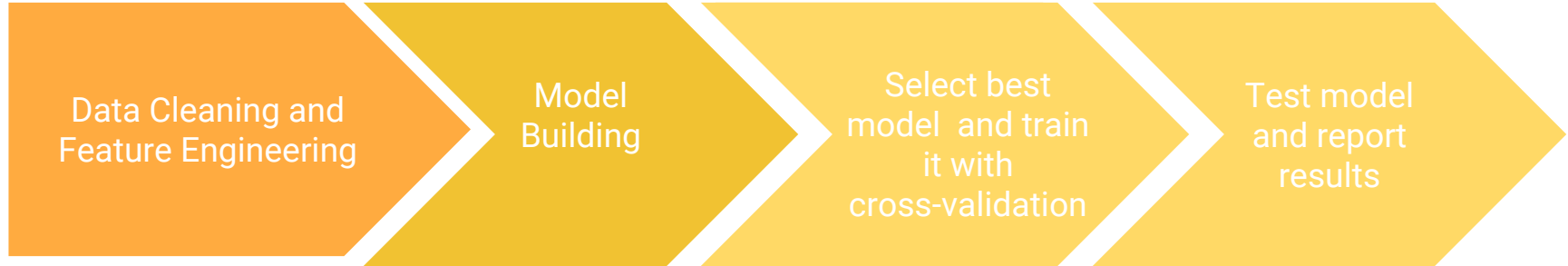


pandas

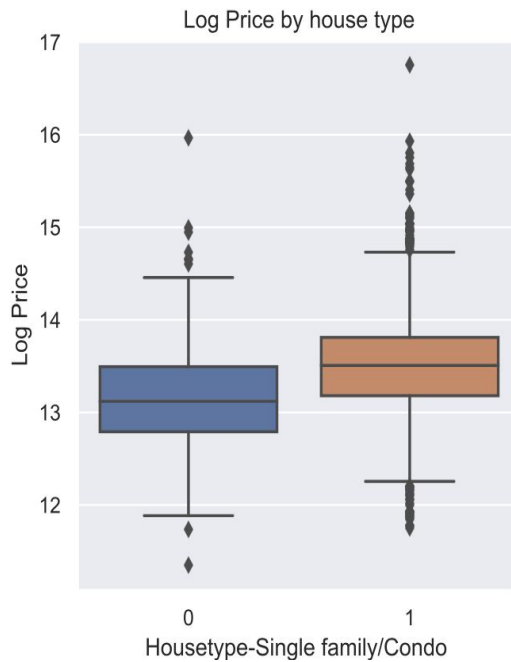
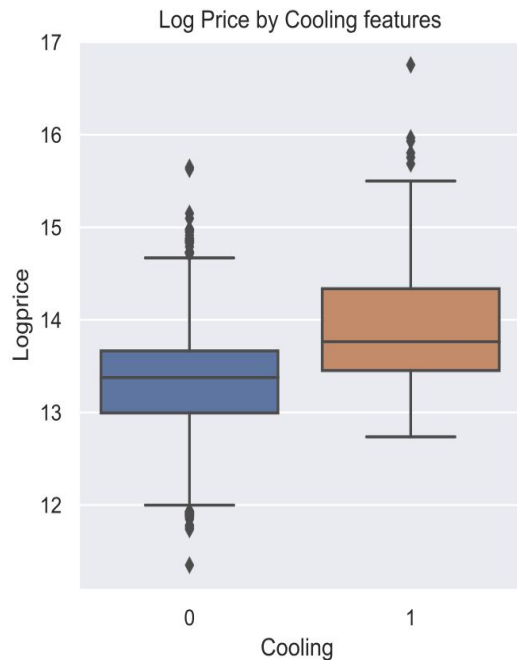
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# METHODOLOGY

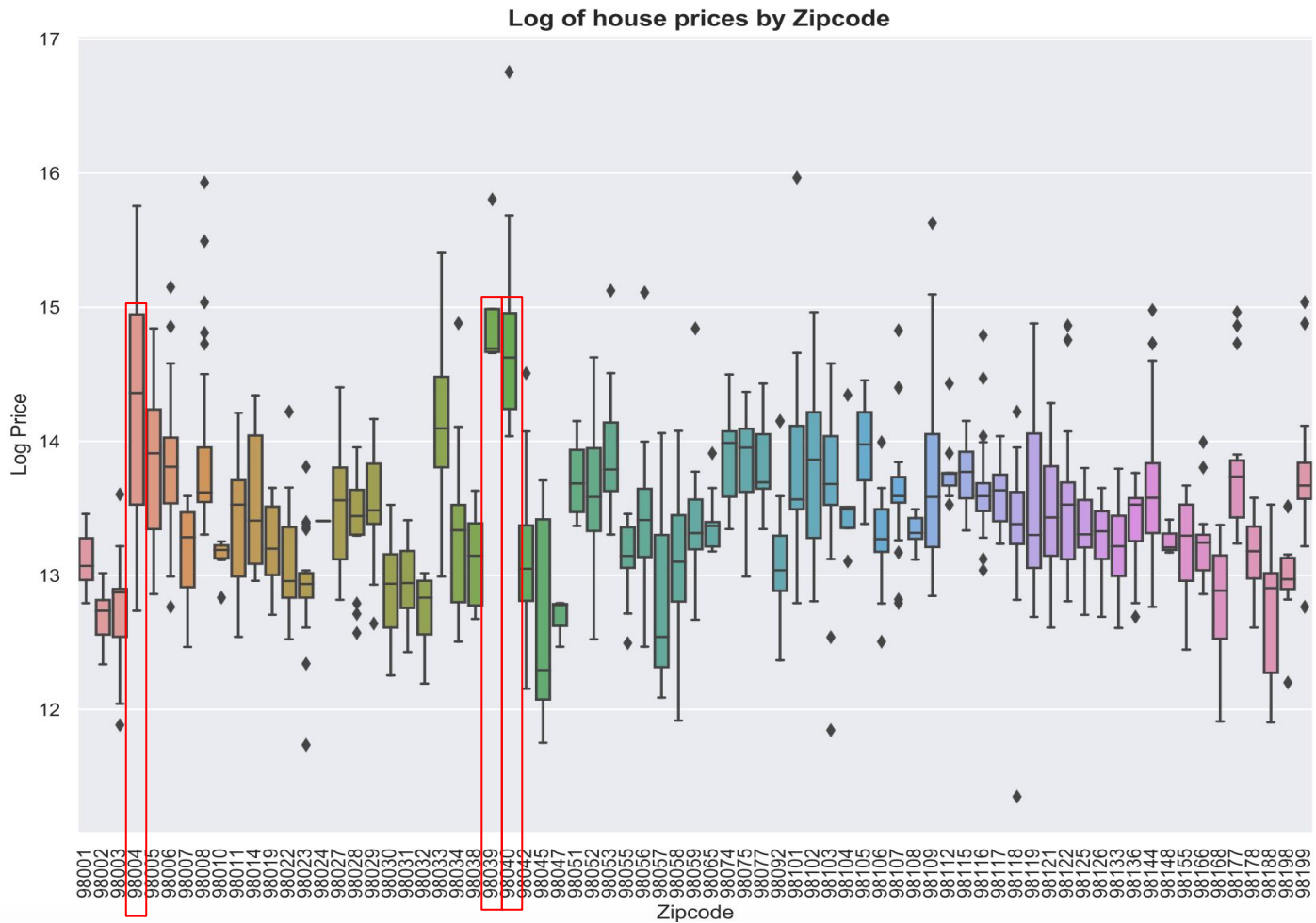


# EDA-CATEGORICAL FEATURES

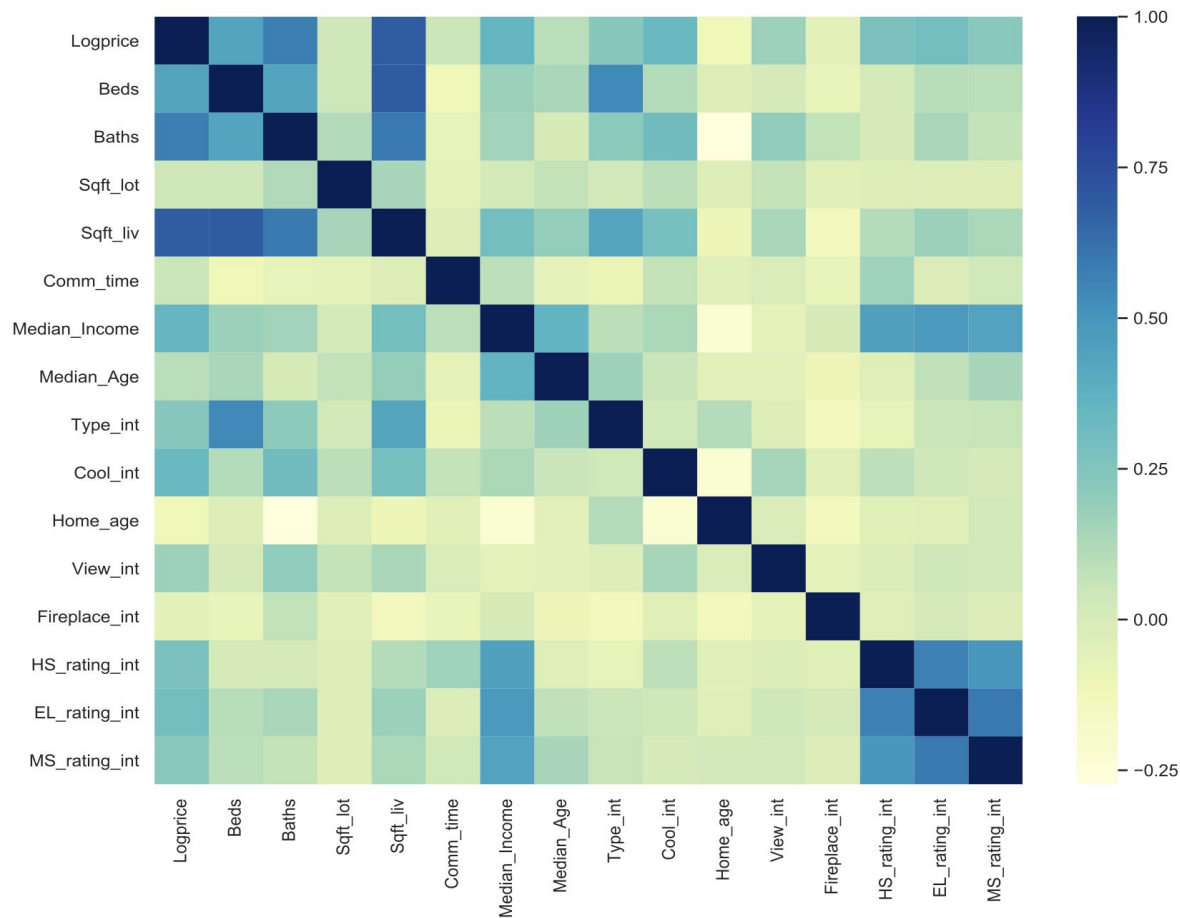


# EDA-CATEGORICAL FEATURES

Highest price  
Zip-code-  
98039 in  
Medina,  
followed by  
Clyde Hill/  
Bellevue  
and Mercer  
Island



## FEATURE SELECTION-CORRELATION



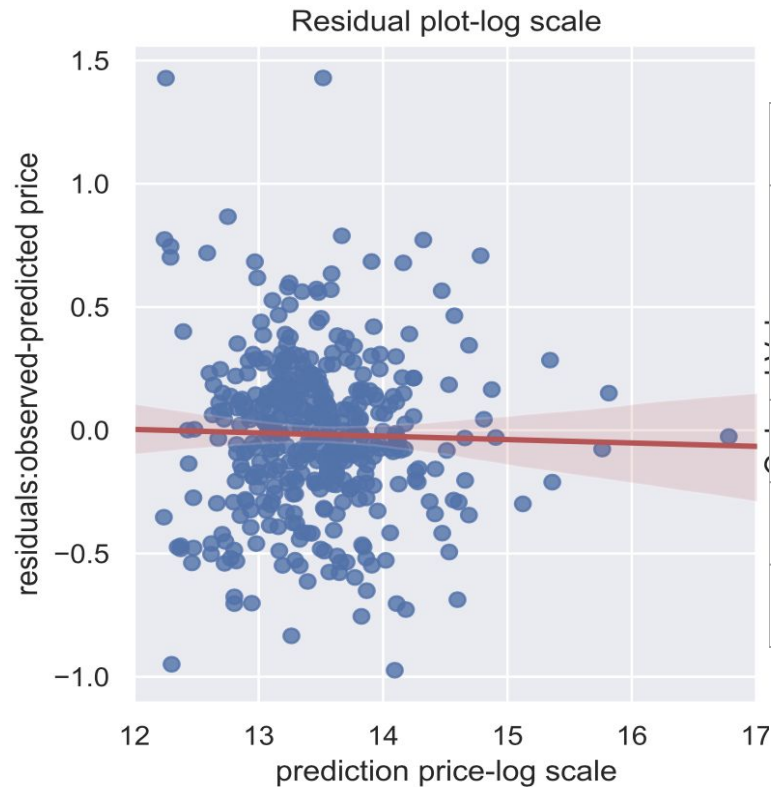


# RESULTS

FINAL MODEL: LASSO REGULARIZATION, 3-FOLD CV, STANDARD SCALER, 2ND DEGREE POLYNOMIAL,  $\text{ALPHA}=1\text{e}^{-5}$



# RESULTS



PARAMETERS	VALUE
TEST MSE (Actual price= $1.055 \times$ Predicted price)	1.114
RMSE	1.055
TEST R-Square	0.751
DEVIATION IN MSE	2.67%

# FUTURE WORK

Add more features like crime data by Zip-code, Condition of house, Use Zip-code as a feature

Categorize features by Type Condo/ Townhouse/ Single-Family/ Apartment more effectively

Get more data for houses with prices between 2 million-20 million

Better imputation of missing data

THANK YOU!

# FEATURE ENGINEERING

