Module 1 Project
Kings County House Price Regression Model

Hyungjun Kang

About the project

- Objective

  -Predicting house price of Kings County using observed data.
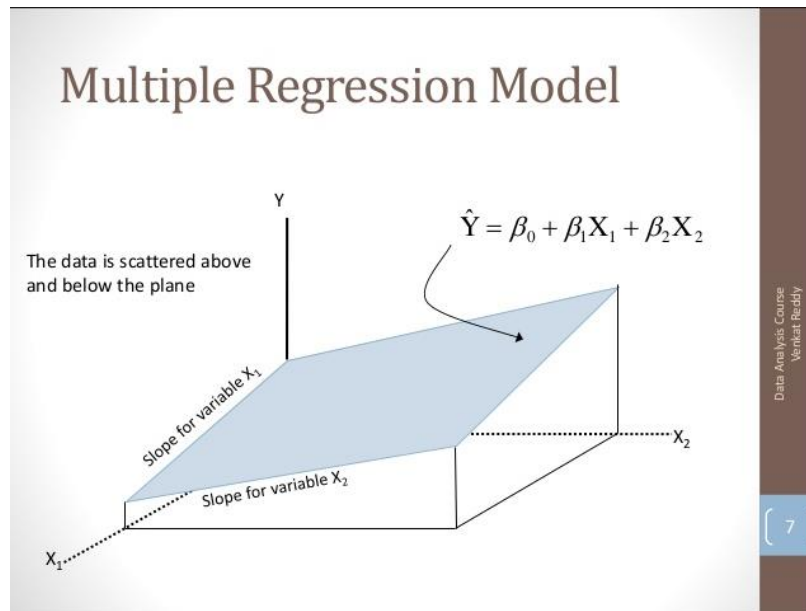
  1. Methodology
  2. About the Data
  3. Important Features
  4. Recommendations

Methodology

• Multiple Regression Model

-Uses collected/observed data to predict the outcome and trends by using best fitting line/plane.

Multiple Regression Model

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

The data is scattered above and below the plane

Slope for variable X_1

Slope for variable X_2

Y

X_2

X_1

Data Analysis Course
Venkat Reddy

7

The Data

- Data from year 2014 & 2015

- 21420 Data Entries

- 19 Different key features
  - Size, Location, Conditions

- Average price of $540,506

- Average size of sqft living 2083 feet squared

The Data

- Data from year 2014 & 2015

- 21420 Data Entries

- 19 Different key features
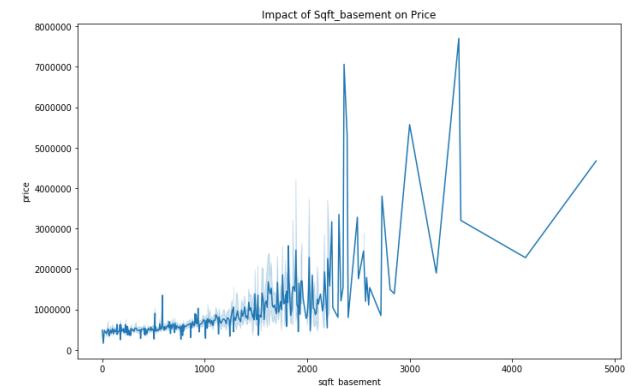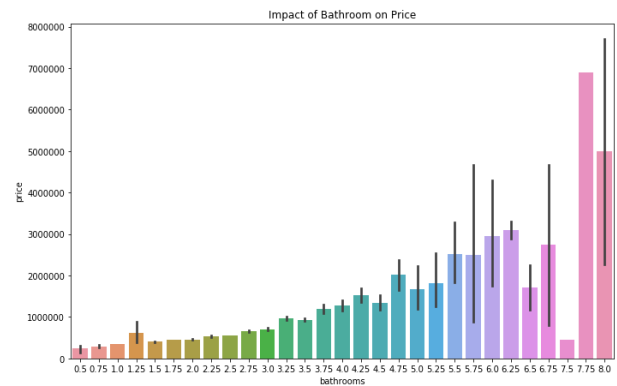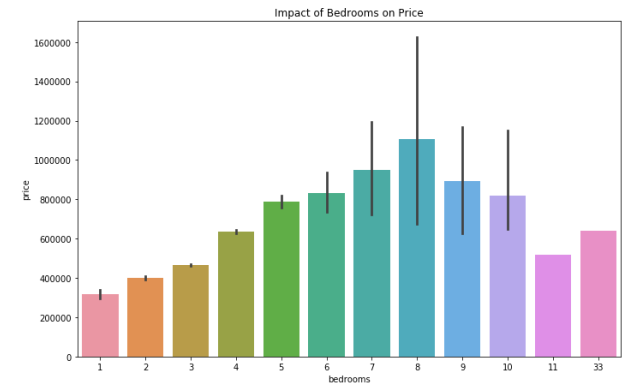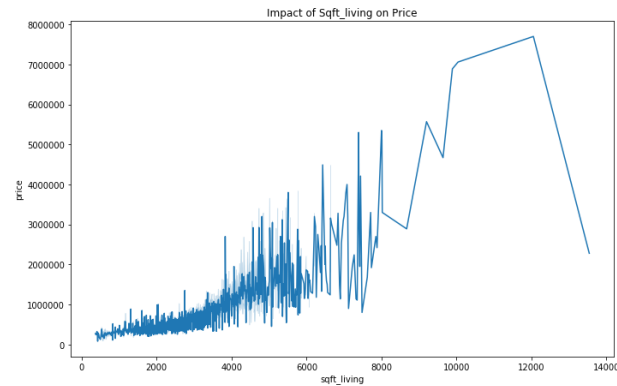 - Size, Location, Condition

Features

- ## Location, Location, Location.. = #1 Feature
    - Magnitude of 0.1186 for zipcode[2]
    - Waterfront with Magnitude of 0.08
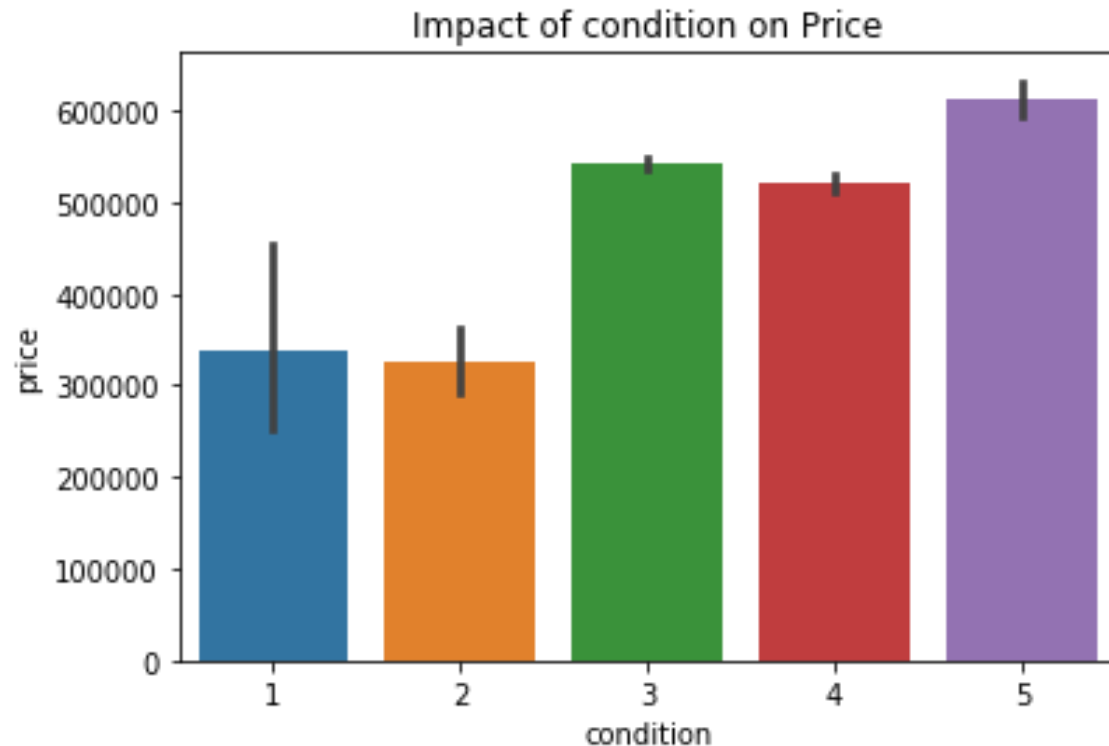   - Top 5 features were all from the Location feature

Locations and Prices

Impact of Zipcode on Price
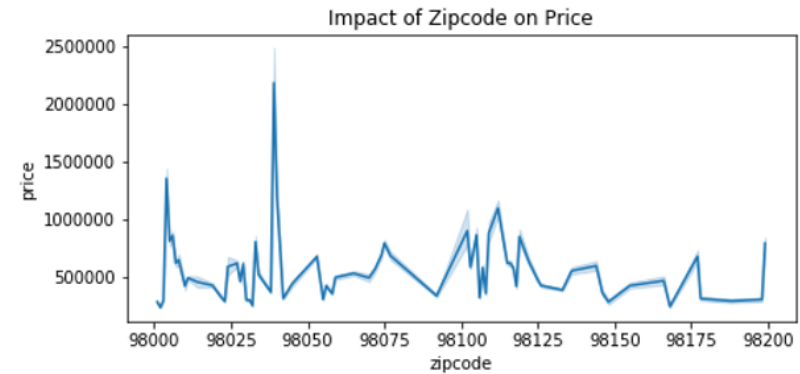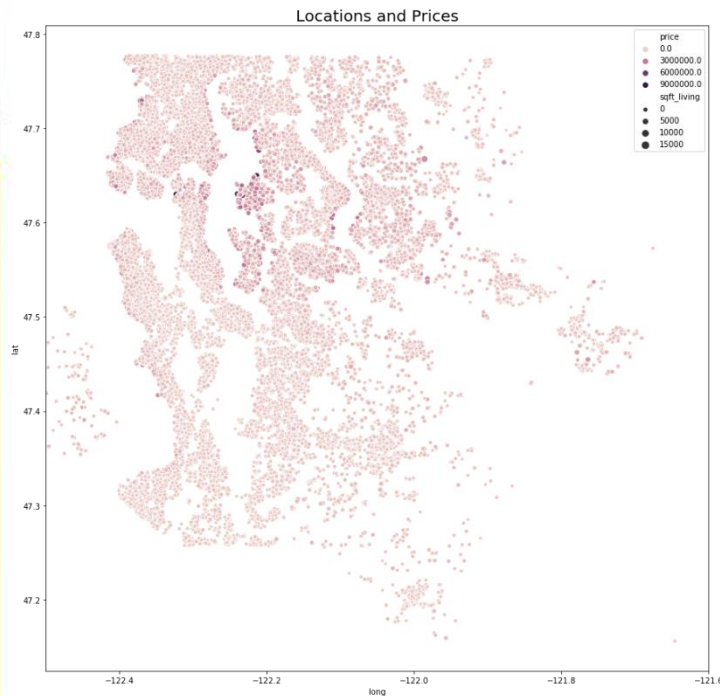
Features continued.

- Bedroom was magnitude with -0.0024
- Bathroom with 0.0065
- Sqft_living with -0.0060
- Least important feature

Features continued.

- Condition
  - Magnitude of 0.123
  - 2nd most important feature out of Location, Condition, and size

Impact of condition on Price

Final Summary

Most important factor

1. Location
2. Condition
3. Size

Thank you!