CALIFORNIA HOUSE PRICE PREDICTION

A new start-up company is planning to create real estate guidance for real estate agencies and consumers. They will prepare a guide for the California region in 1990 for testing purposes at the beginning. They will predict average block value by using location, room number, household average income and number of people who live in block. If the model that they created give 70% accuracy they will start guidance in other years and states.

Feature Space

.

Dependent Variable

Median House Value($): (for households within a block)

Continuous Variables

1. Longitude

2. Latitude
3. Housing Median Age
4. Total Rooms: (within a block)
5. Total Bedrooms: (within a block)
6. Population: (within a block)
7. Households: (a group of people residing within a home unit, for a block)
8. Median Income: for households within a block of houses (measured in 10K$)

Categorical Variable
Ocean Proximity (island 1, near bay 2, near ocean 3, less than 1 hour ocean 4, inland 5)

Box Plot shows range, median, the smallest and largest value and outliers. Total rooms, total bedrooms, population, households and median income include to many outliers. It means that average of those values are pulled toward the large values (outliers).

Housing prices in California can be predicted with our model. We can explain 67% of price distribution.
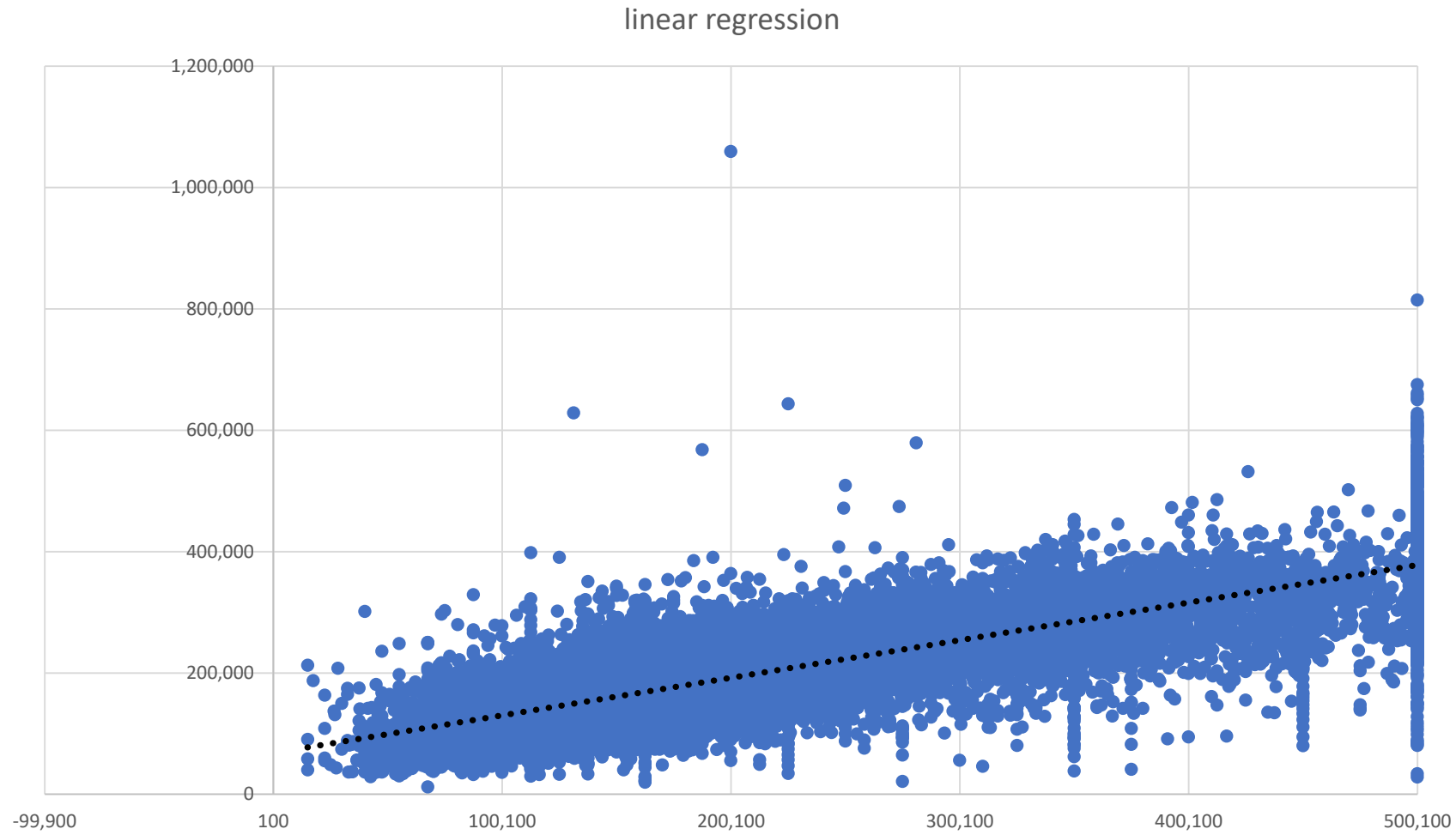

linear regression

Our regression model takes the below mentioned features into consideration and for each feature $value of one unit change as shown here.


$ value of each observation feature
(Coefficent)
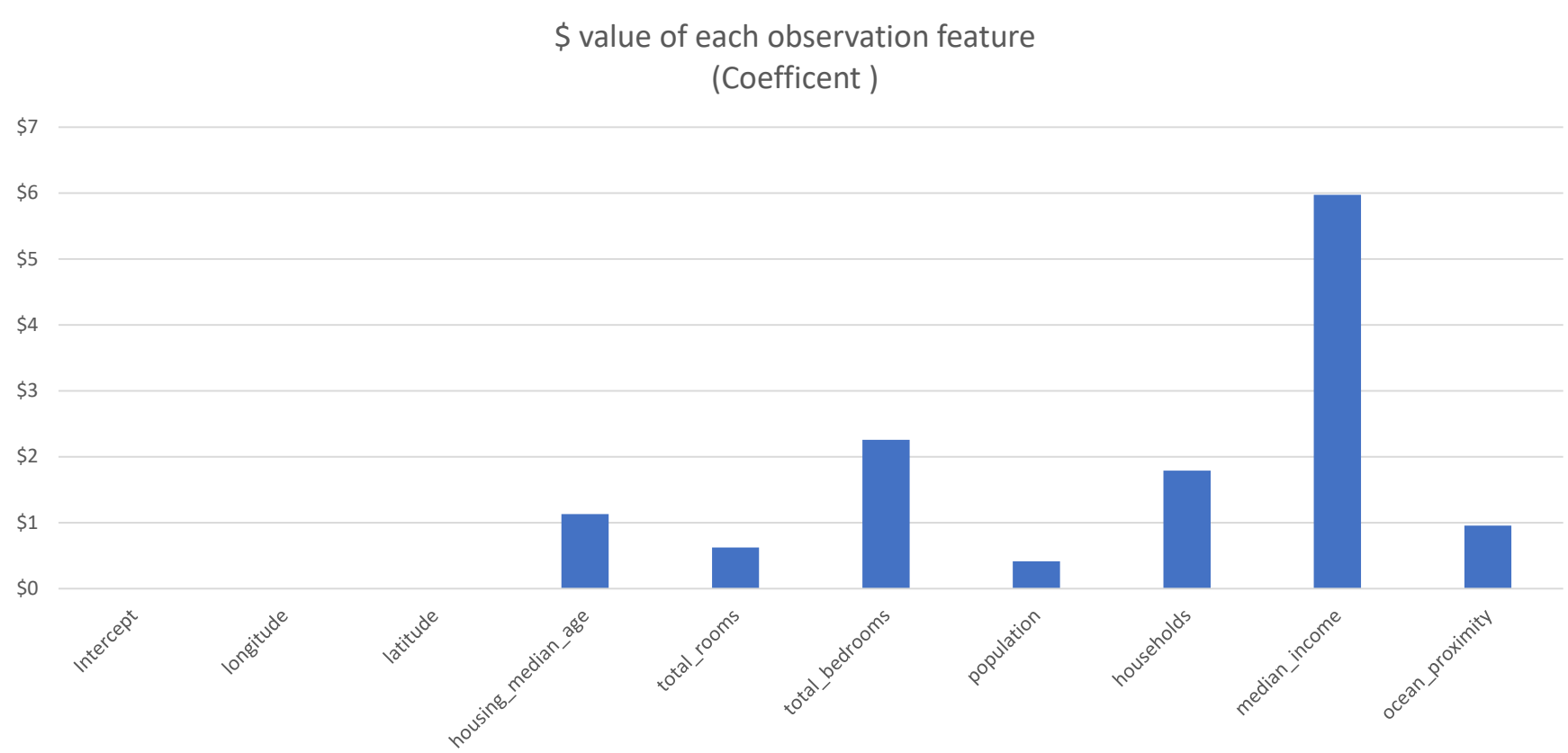
Neighborhood income has positive effect on the house price. For one unit increase ($10K) has approximately $6 positive effect on price.
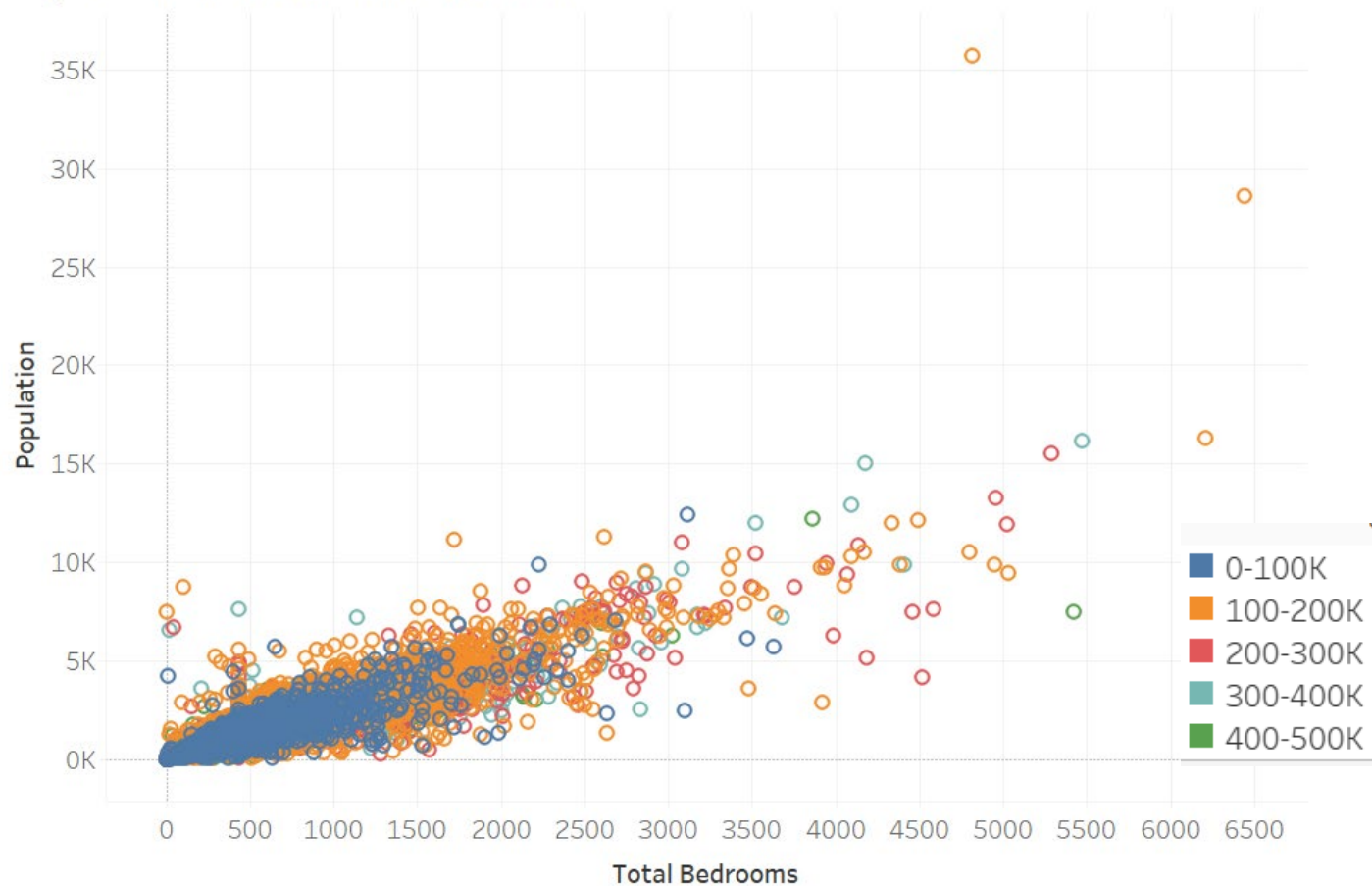
MEDIAN INCOME VS MEDIAN HOUSE VALUE

y = 2E-05x
R² = 0.1048
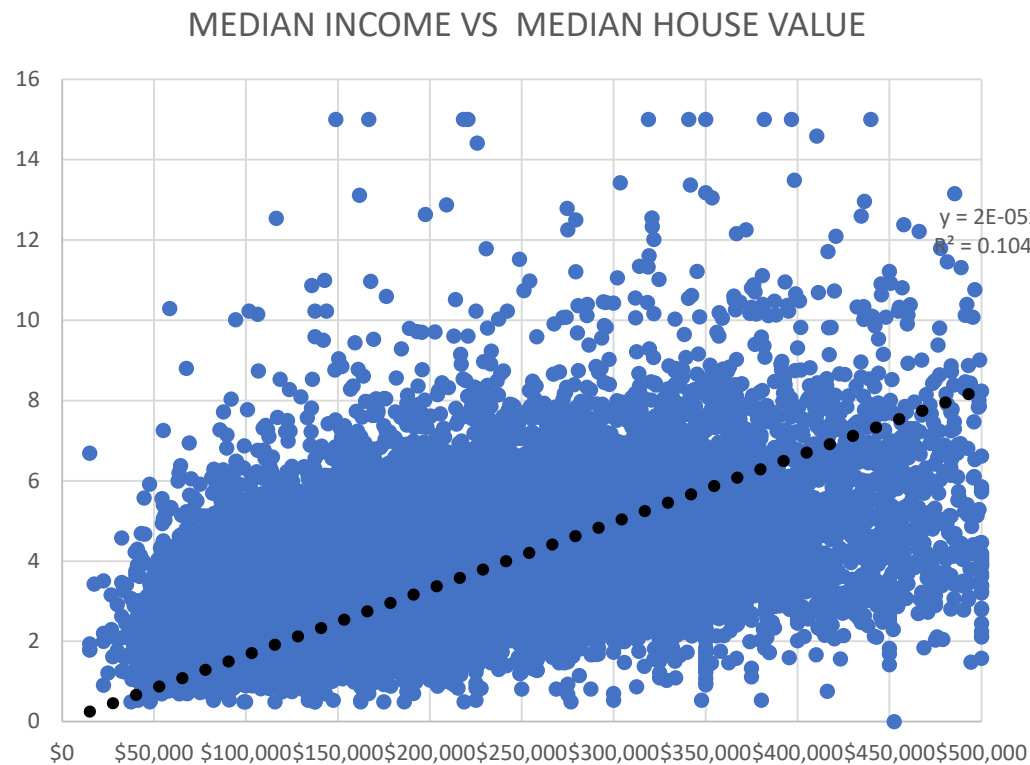
median income vs median house value
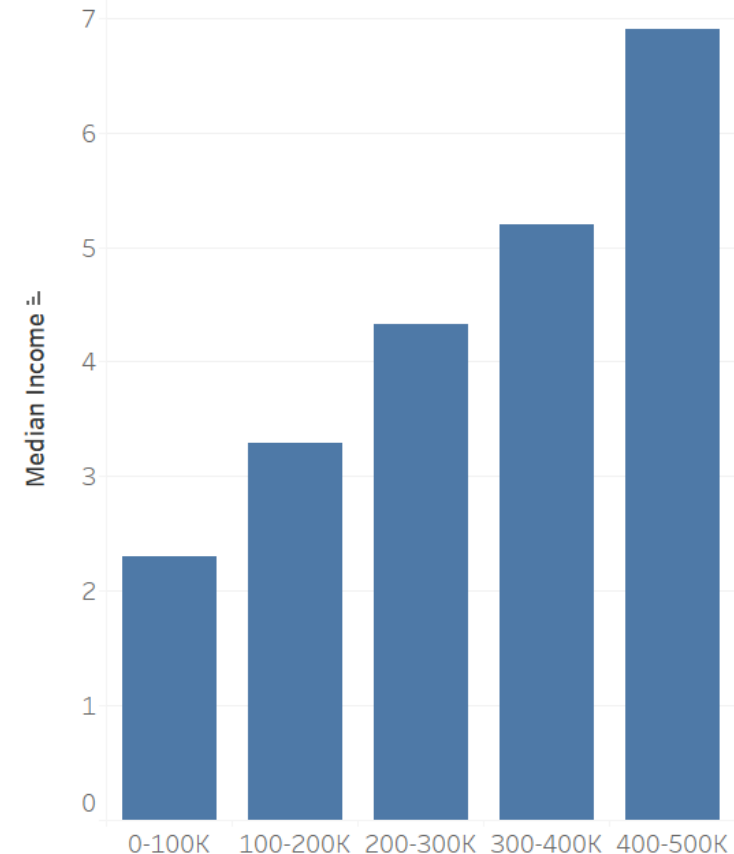
Approx combine effect of total bedroom and household population is $4.

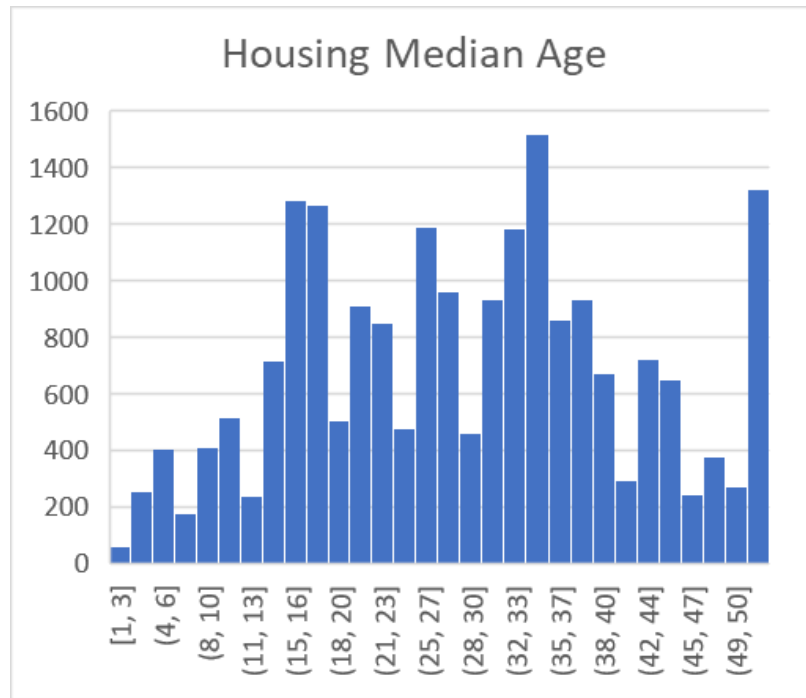Households vs Total Bedrooms
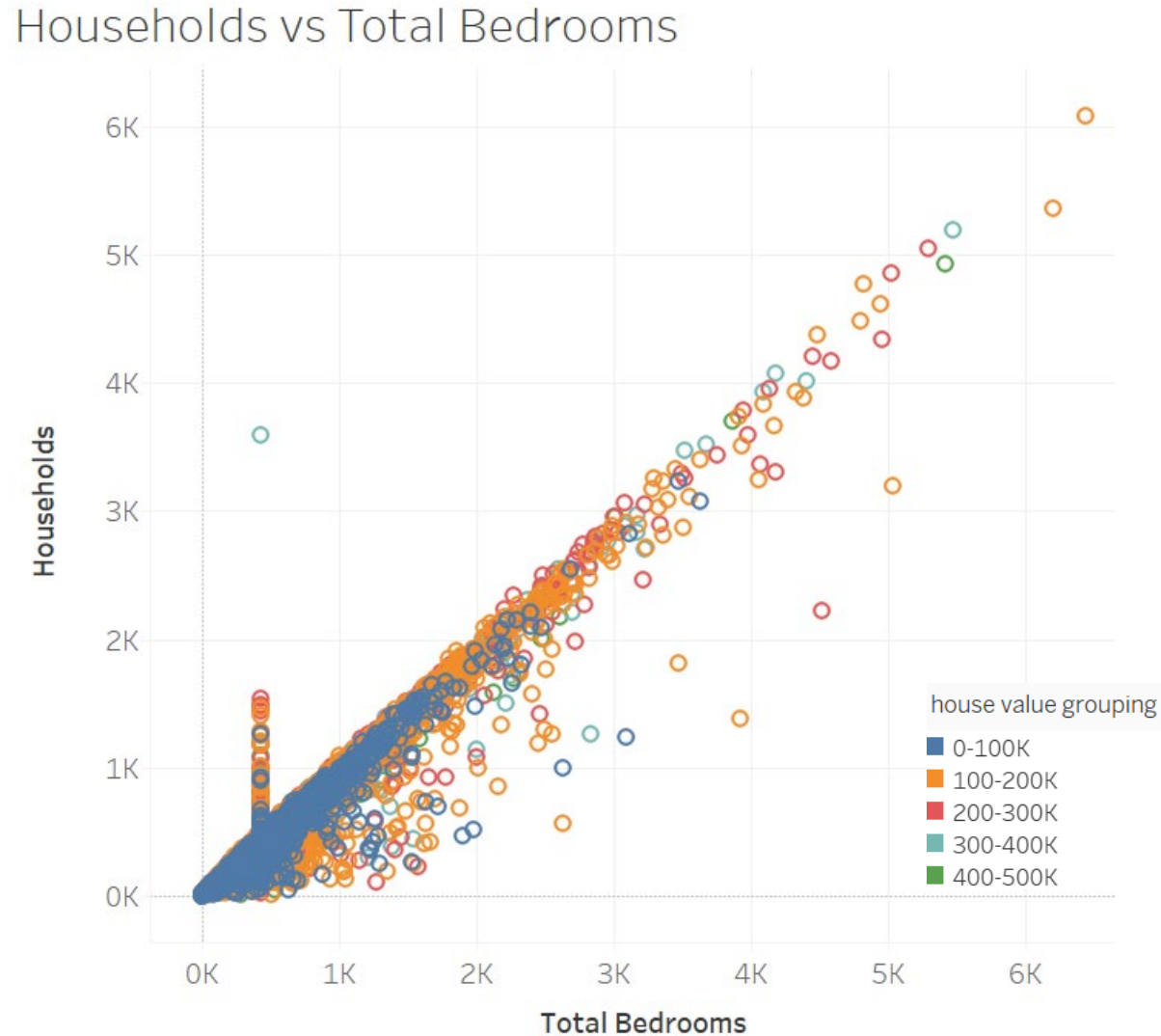
Approx combine effect of total bedroom and population is $3.


Population vs Total Bedroom

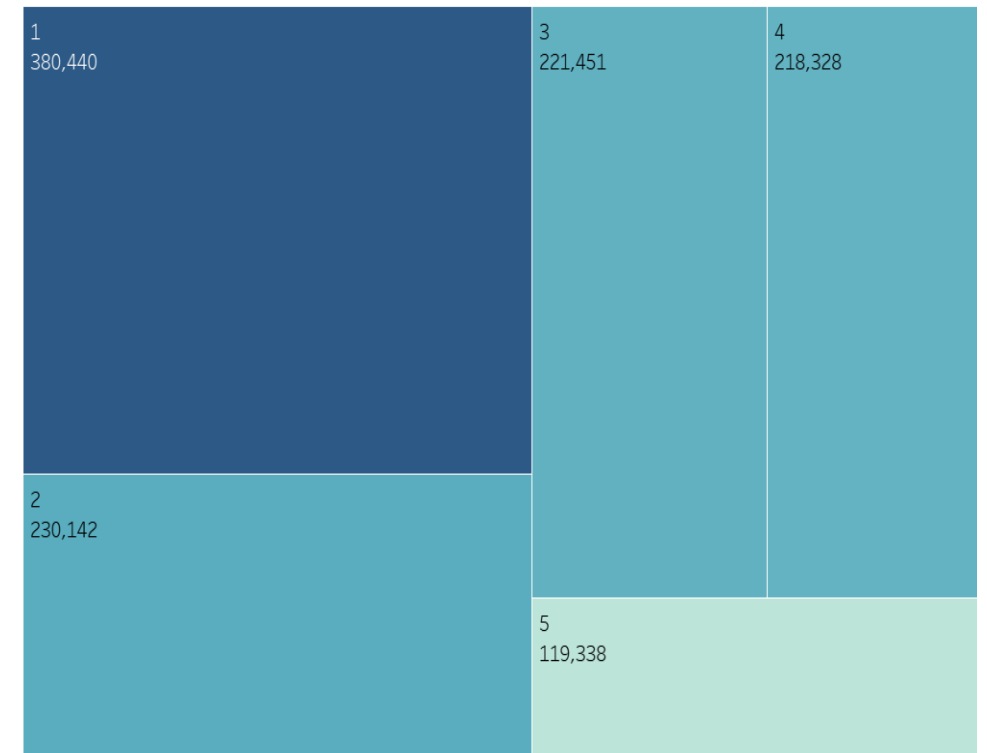# 1 Year increasing house age has more than 1$ affect on median house value.

Housing Median Age

- The age distribution of the California Houses are;
  - Youngest %5 is less than 8 yrs
  - Median house age is 29
  - Oldest %5 is more than 52 yrs

5%	8
10%	13
25%	18
50%	29
75%	37
90%	46
95%	52

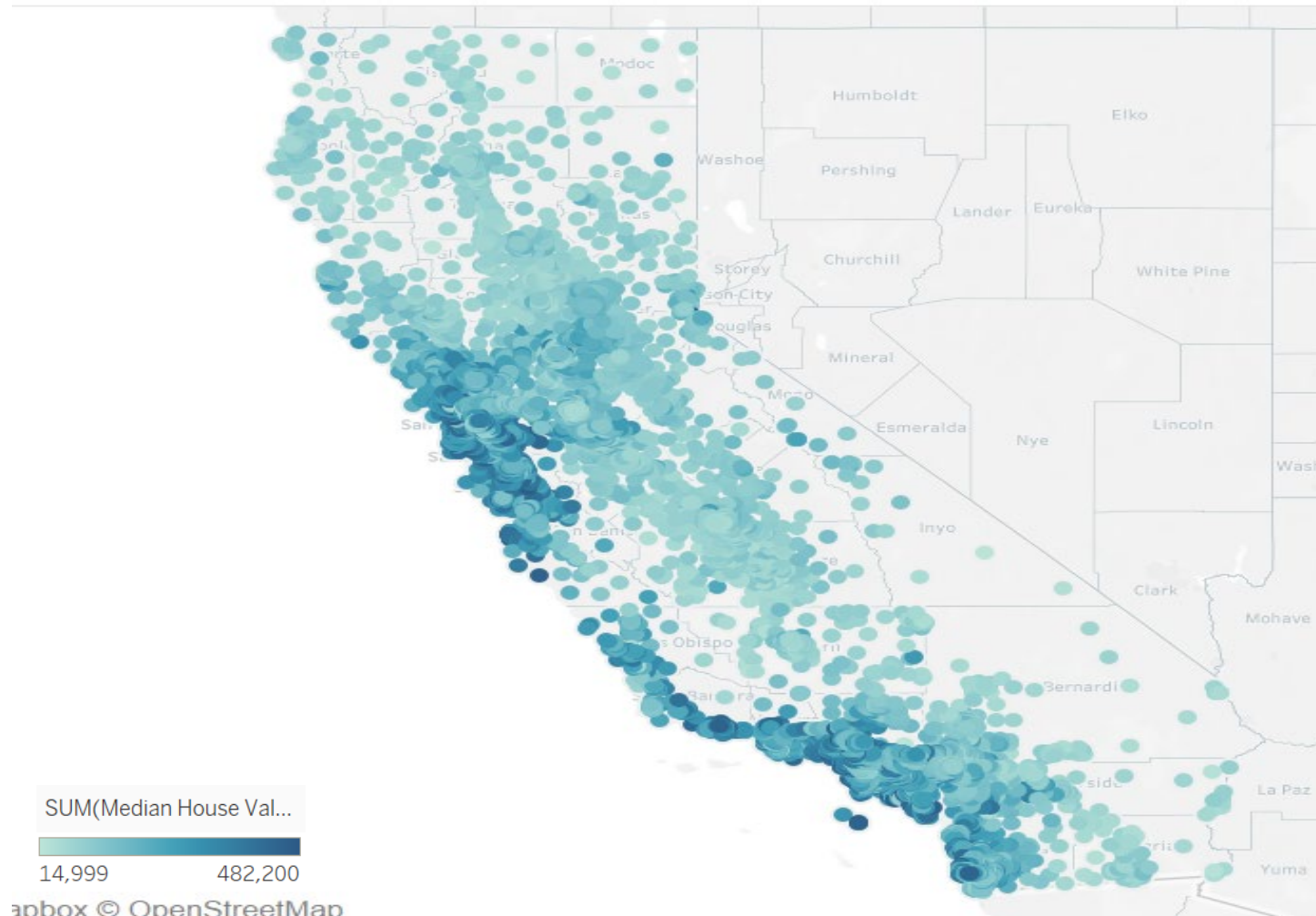# Colors indicates the price on the tree map, this shows ocean proximity can categorize price band on the houses.
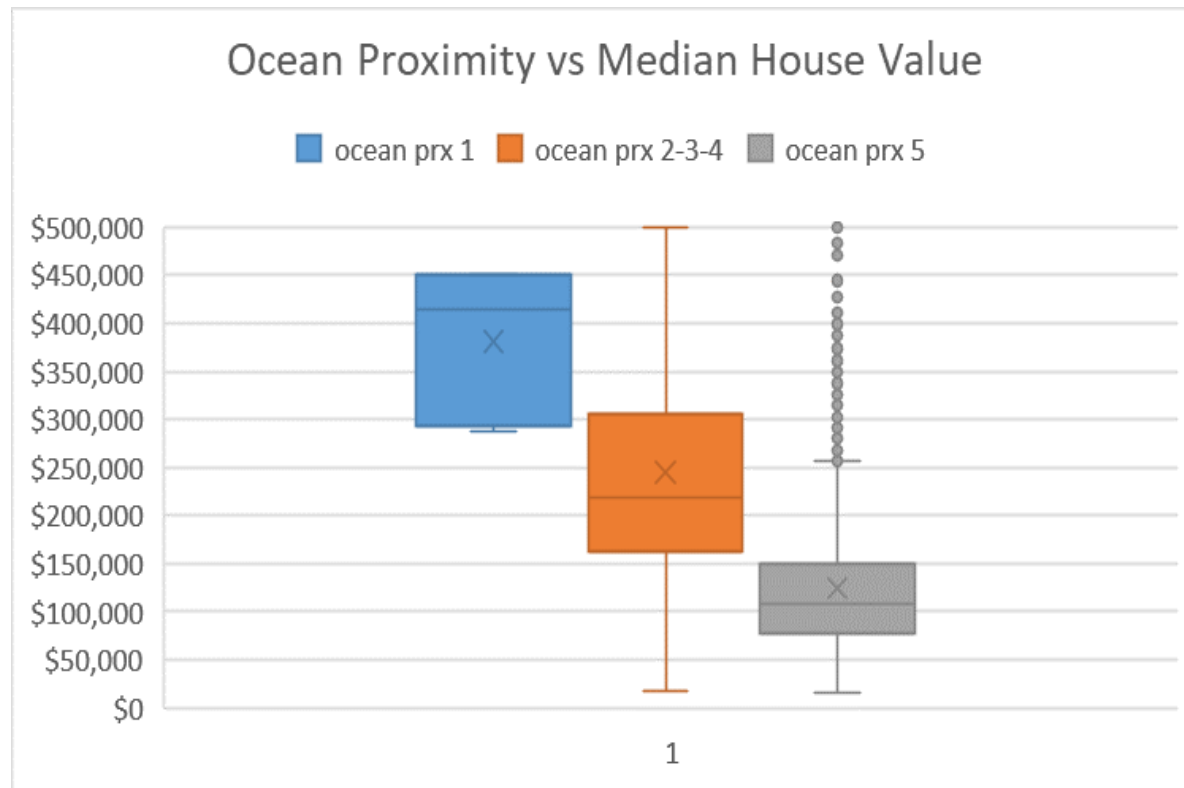
Ocean and big cities are effecting the house prices.

Housing is scattered all over the state but at the same time major cities are densely populated areas with high price fluctuation.

CROSSCHECK THE MODEL

As explained above income and location have the highest relation with our house value.

Prediction errors have approximate normal distribution meaning that our model is in line with statistical assumptions.

Feature Correlation with House Value

Prediction Error Distribution
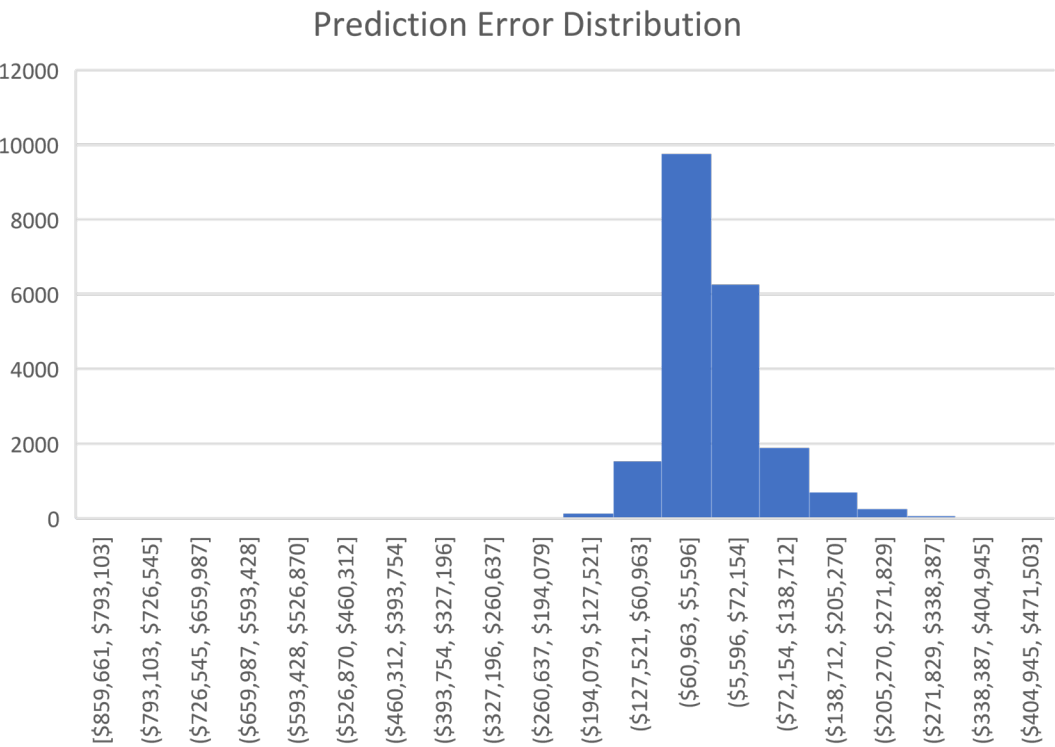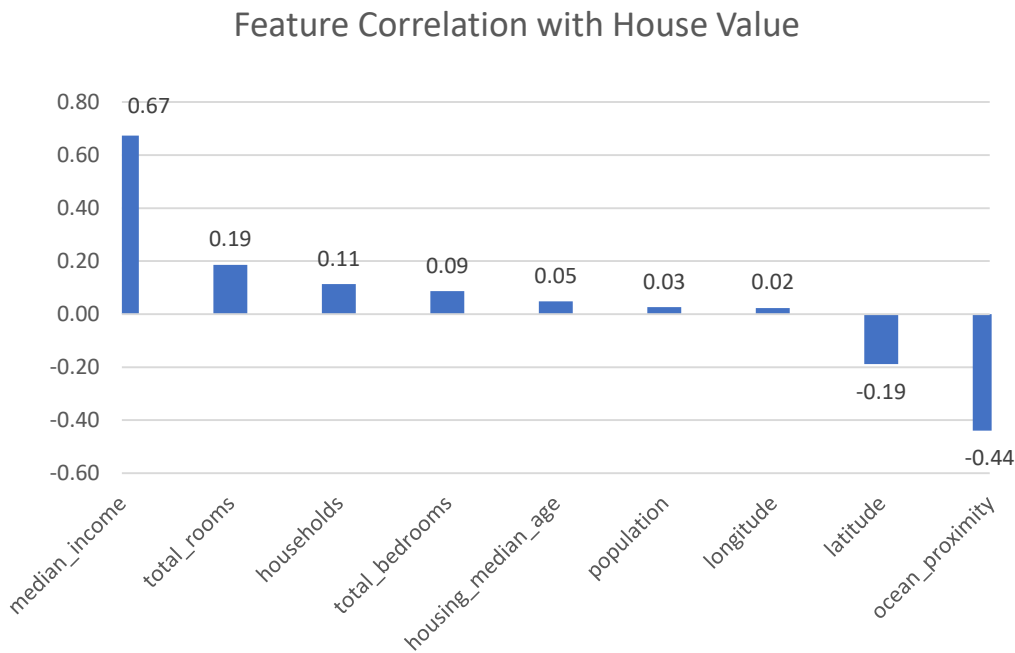
Last but not Least

- As a Result;
  - We successfully predicted California House Value with first version of the model.
  - We detected significant variables that correlate with house value.

- Next Steps for Application Modelling;
  - As we can see from proof of concept analysis, house prices can be modelled. And important factors can be explained.
  - To increase the model efficiency;
    - More accurate location information can be collected
    - Demographics of locations can be added