

IDS ASSIGNMENT

Exploratory Data Analysis

Dataset chosen : House sales in King County USA

<https://www.kaggle.com/harlfoxem/housesalesprediction>

3rd Sem

Section : F

YASH GOEL

PES1201701853

Objectives :

Analysing different features of house with respect to year of construction

Analysing pricing of the houses

Questions:

What's are the most popular house features?

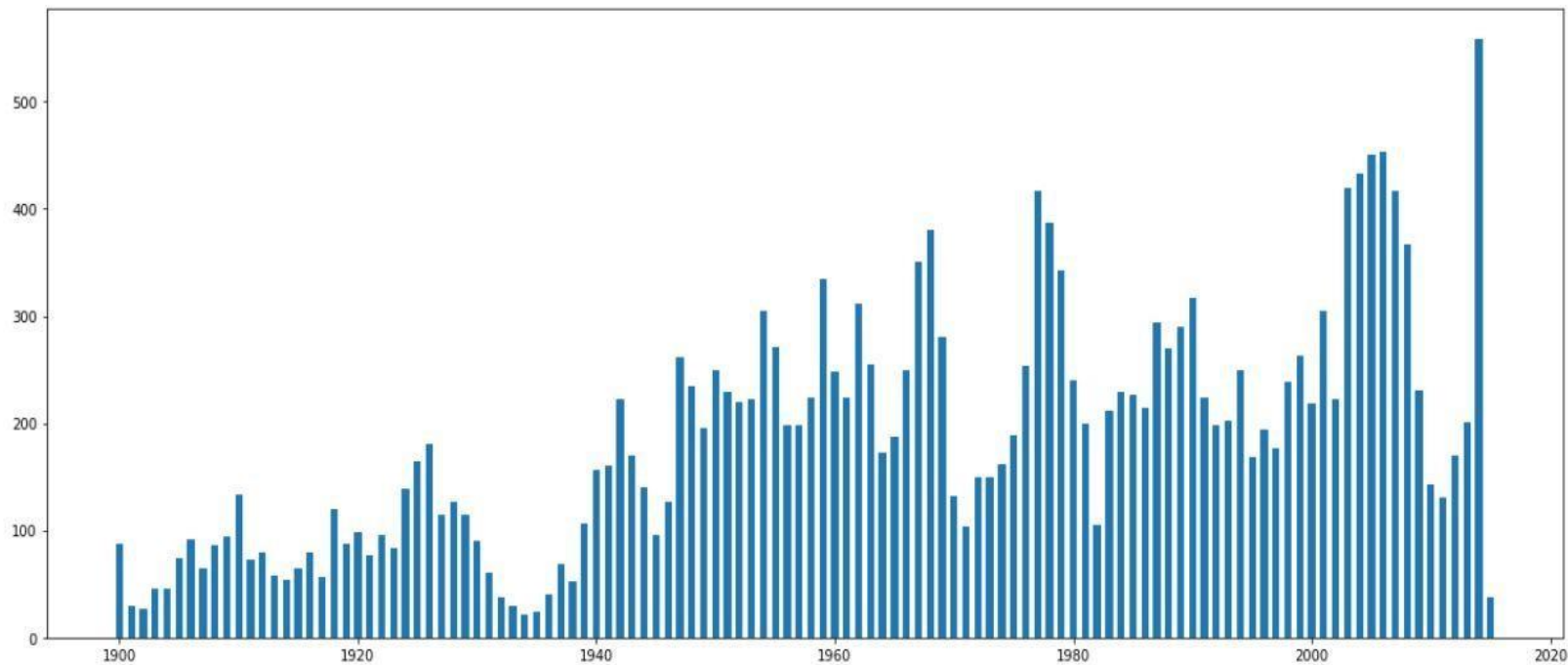
Eg-no. Of bedrooms, no. of floors, etc

Is there a rise in number of houses over the years?

Do people use all the available lot area for constructing house or do they prefer to have gardens?

Is there a correlation between the pricing and features of house like number of rooms, living space, basement, etc?

Number of houses



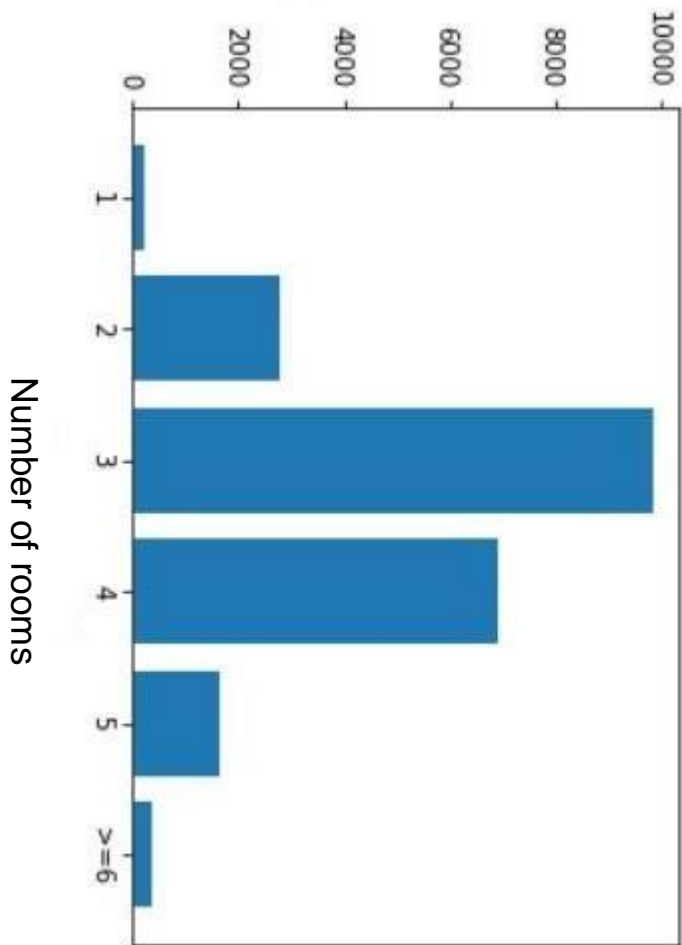
Year constructed

Insights from graph:

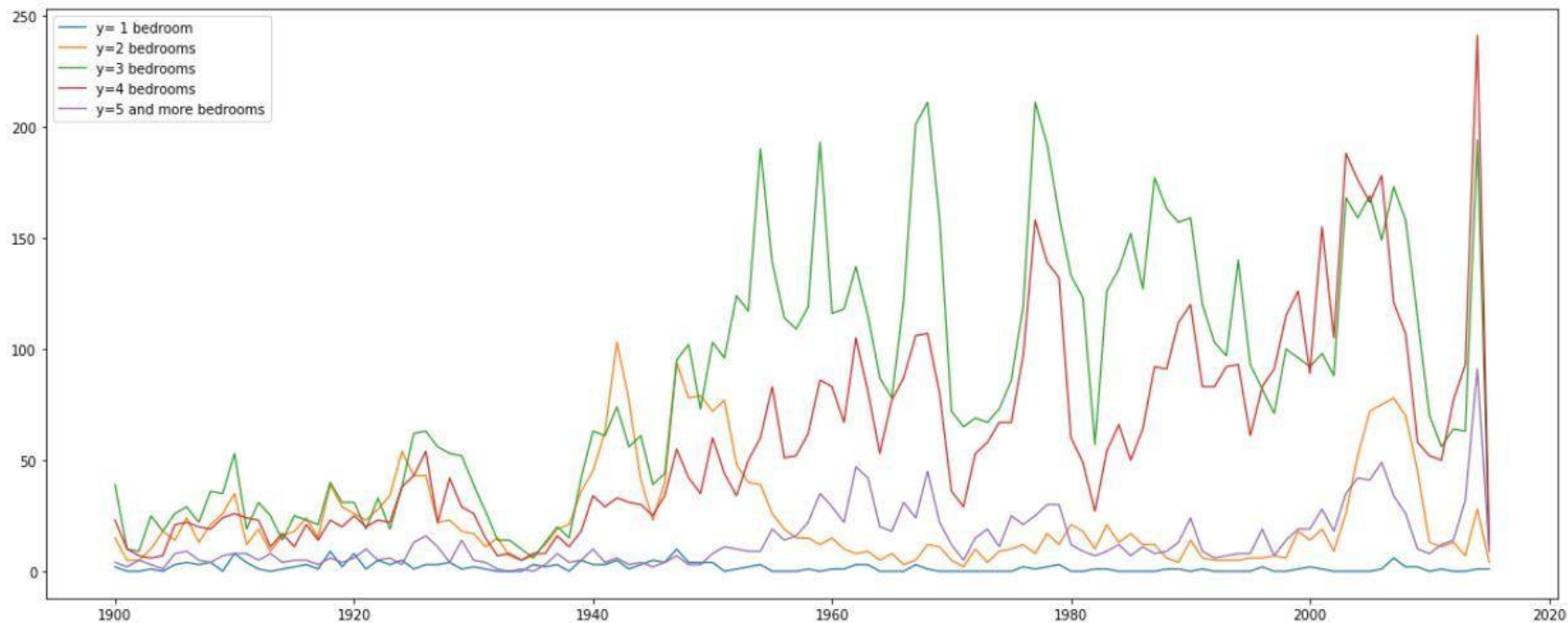
Number of old houses bought are less, houses constructed in 2000s are more sought of.

By this we can infer that demand for recently built houses is slightly more than old houses.

Number of houses



Number of houses



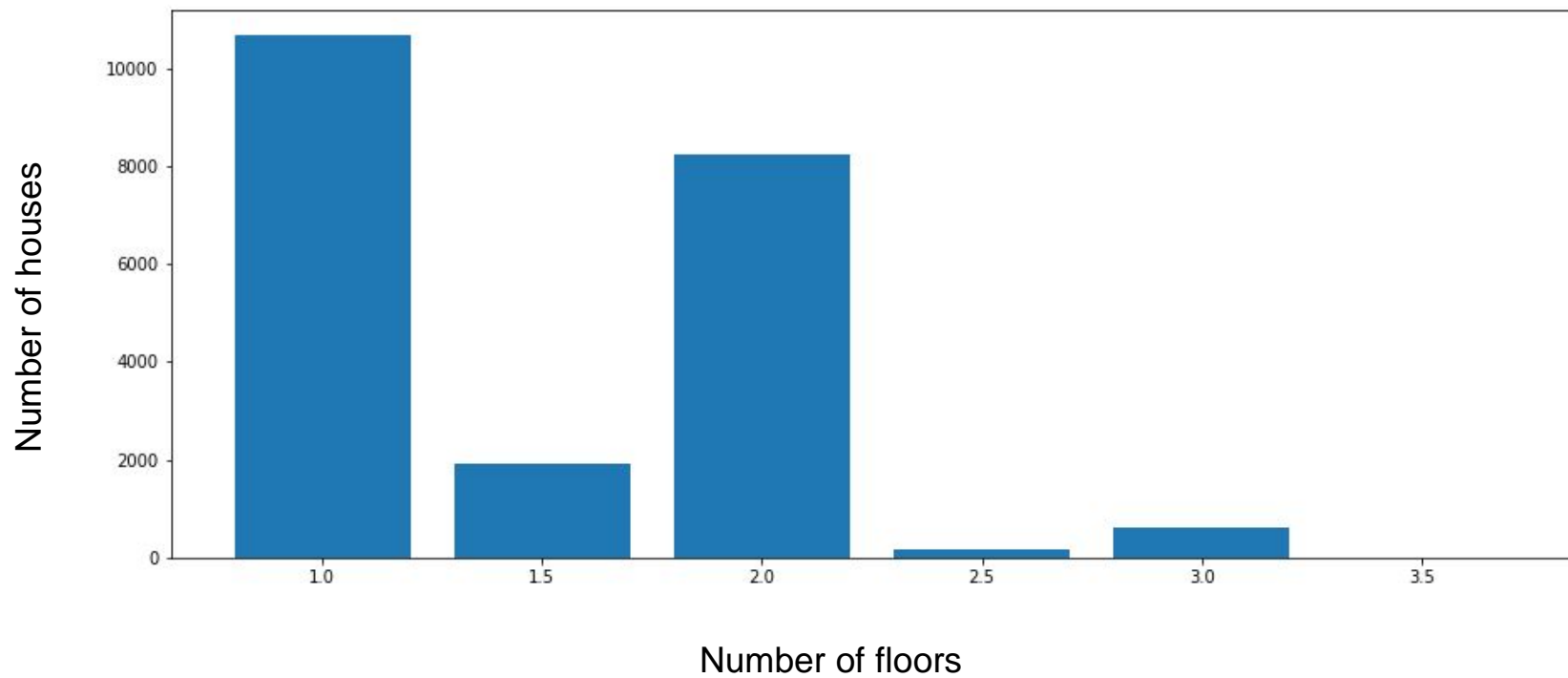
Year constructed

Insights from graph:

It's clear houses with 3 and 4 bedrooms are the most sought houses.

There is also a good demand for 2 and 5 bedroom houses.

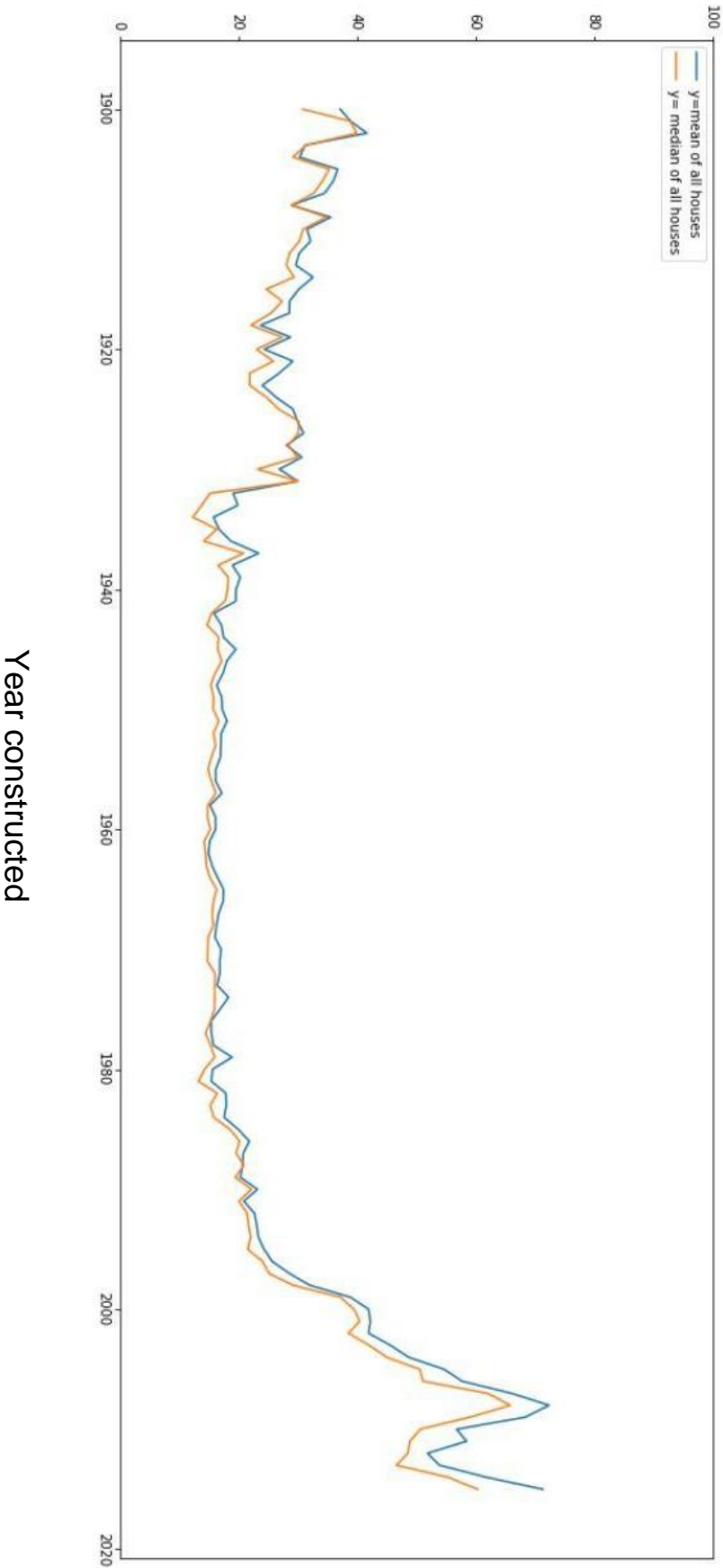
1 bedroom or 6 and more bedroom houses bought are less.



Insights from graph:

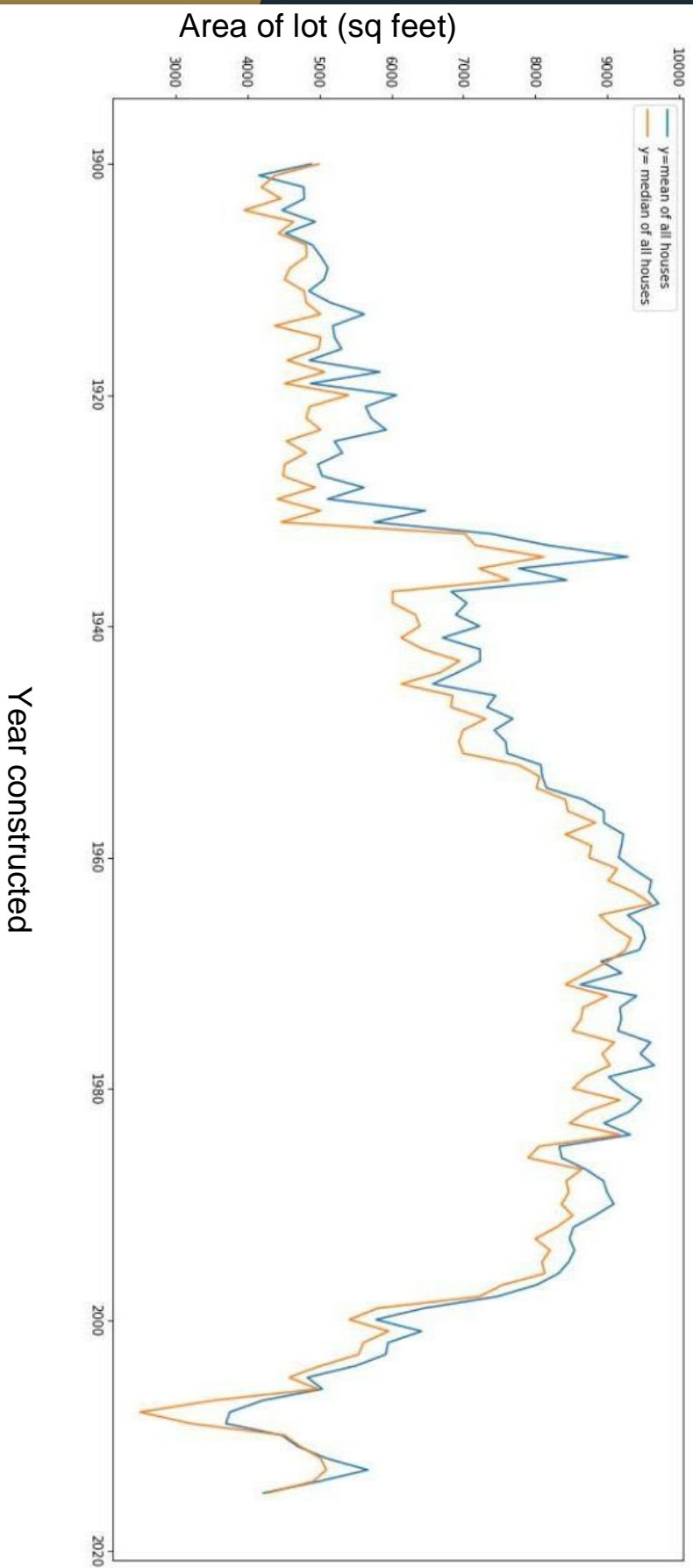
Almost 96% of people opt for houses within 2 floors.

Percentage of living space to that of total lot size

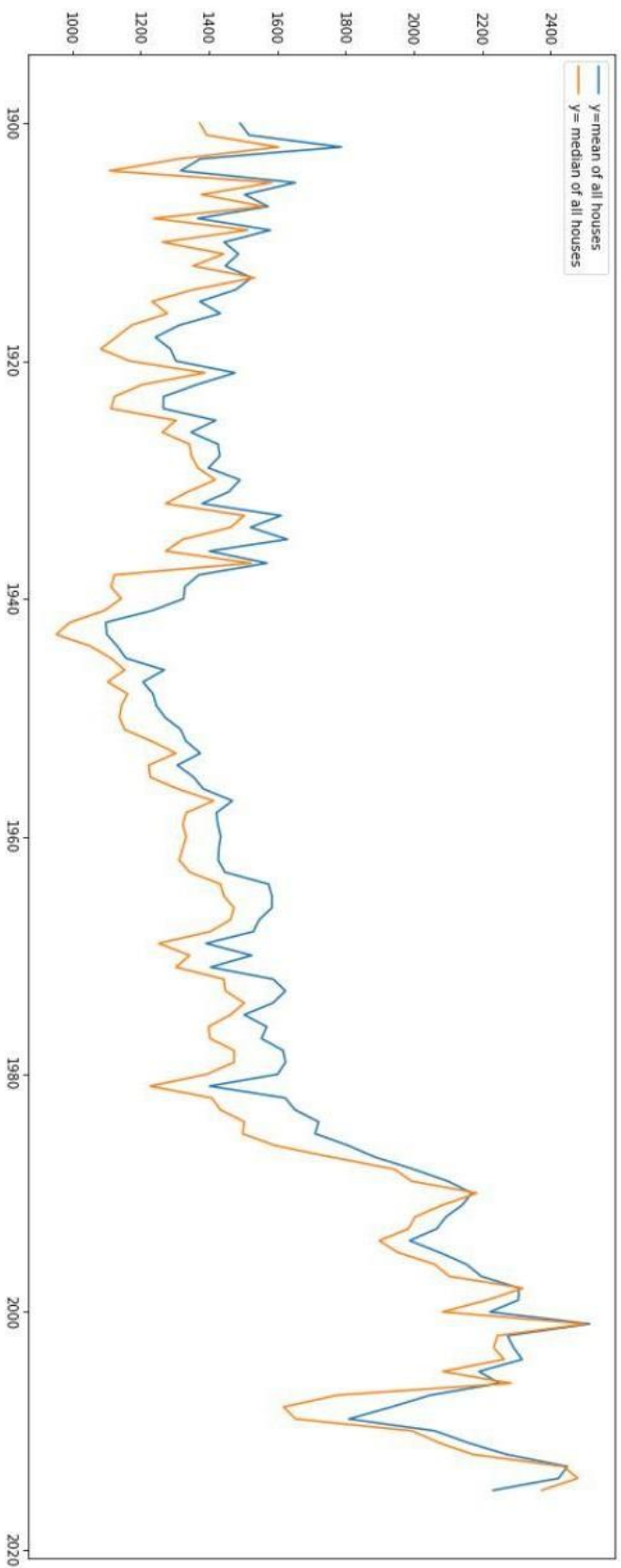


Insights from graph:

Reason for the increase in the percentage cannot be understood directly from the graph, therefore let's take a look on the following graphs and then come to a conclusion.



Area of living space (sq feet)



Year constructed

Insights from graph:

The lot size decreases significantly after the year 2000.

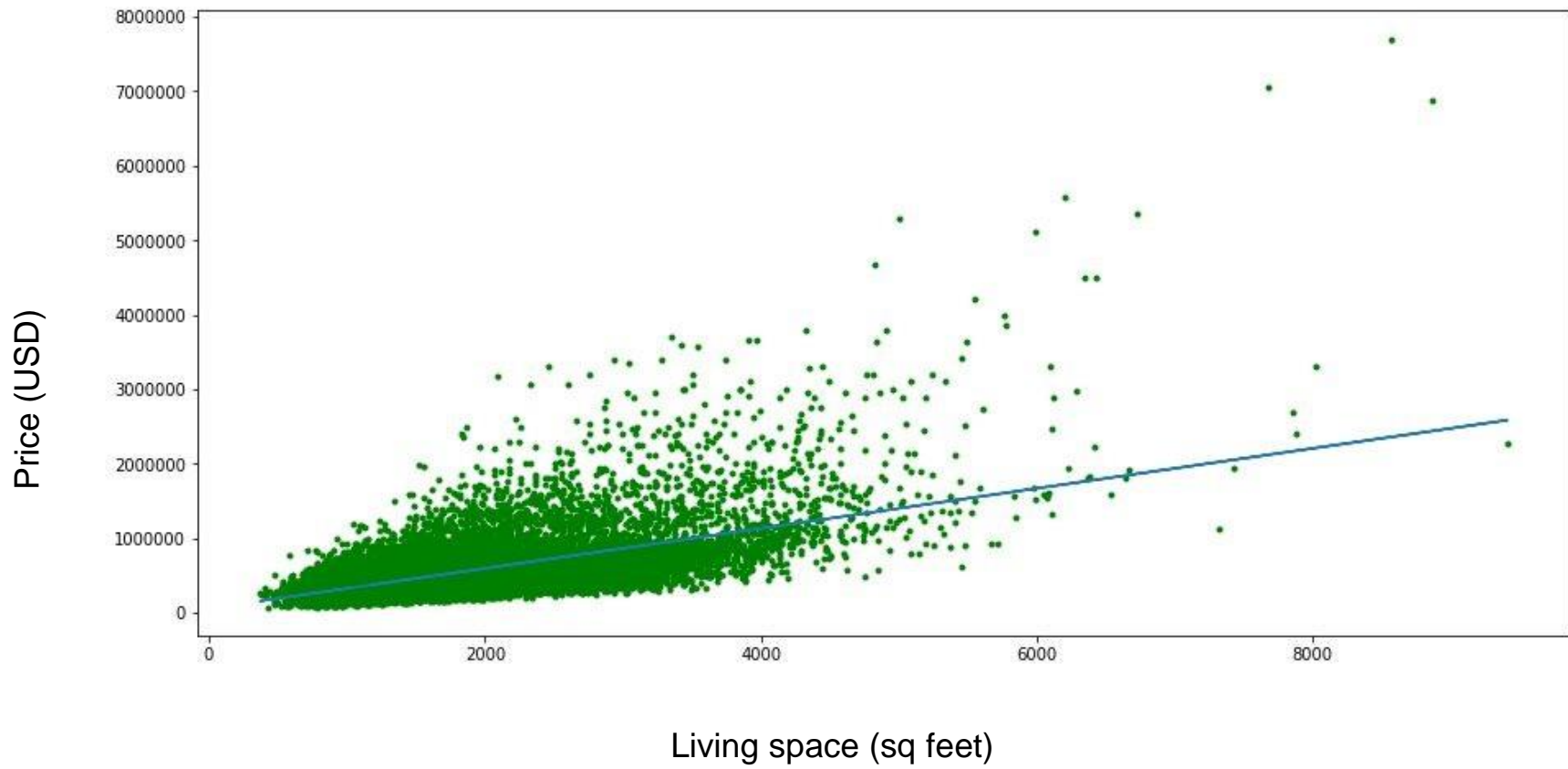
Also there is a little increase in the living space.

Considering both of these factors the reason for increase in the percentage of living space to that of lot size is justified.

Insights from graph:

From the previous graphs, we can infer that the old houses have large gardens compared to the recent houses.

As the percentage of living space to that of lot size is high in recent years, garden space is reduced.

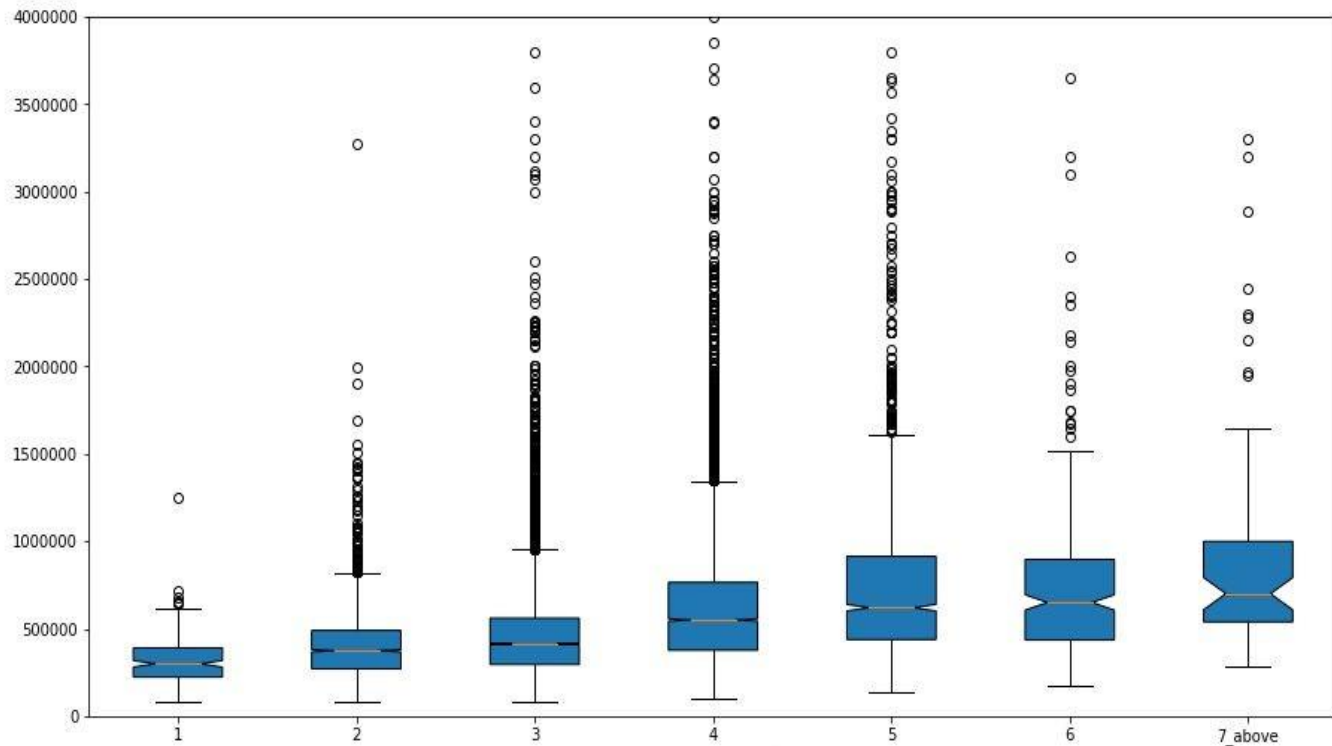


Insights from graph:

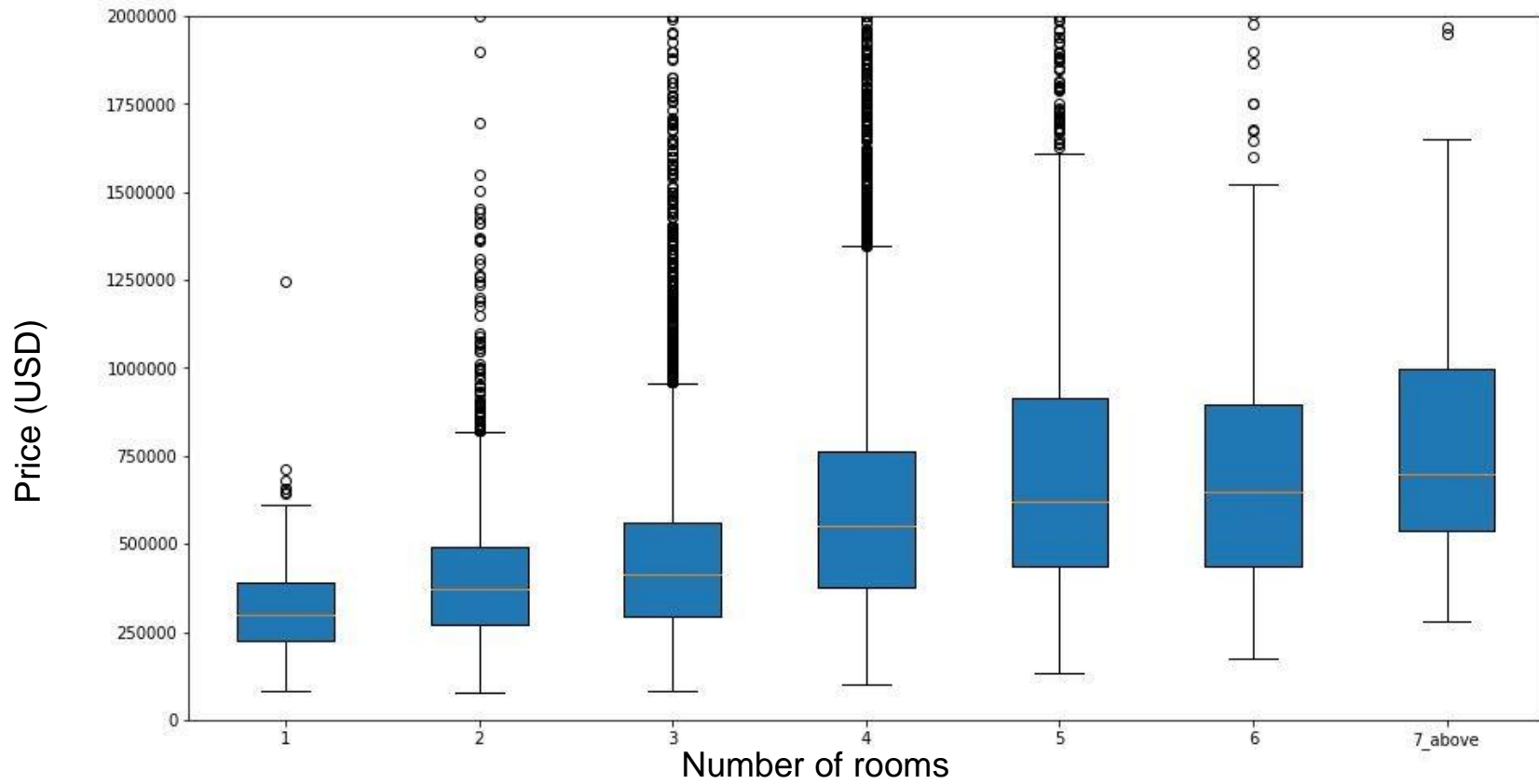
There is a small positive correlation between living space and price.

It's obvious that there is no proper positive correlation, due to fact that price also depends on other factors such as materials used, furnished or unfurnished, etc.

Price (USD)



Number of rooms

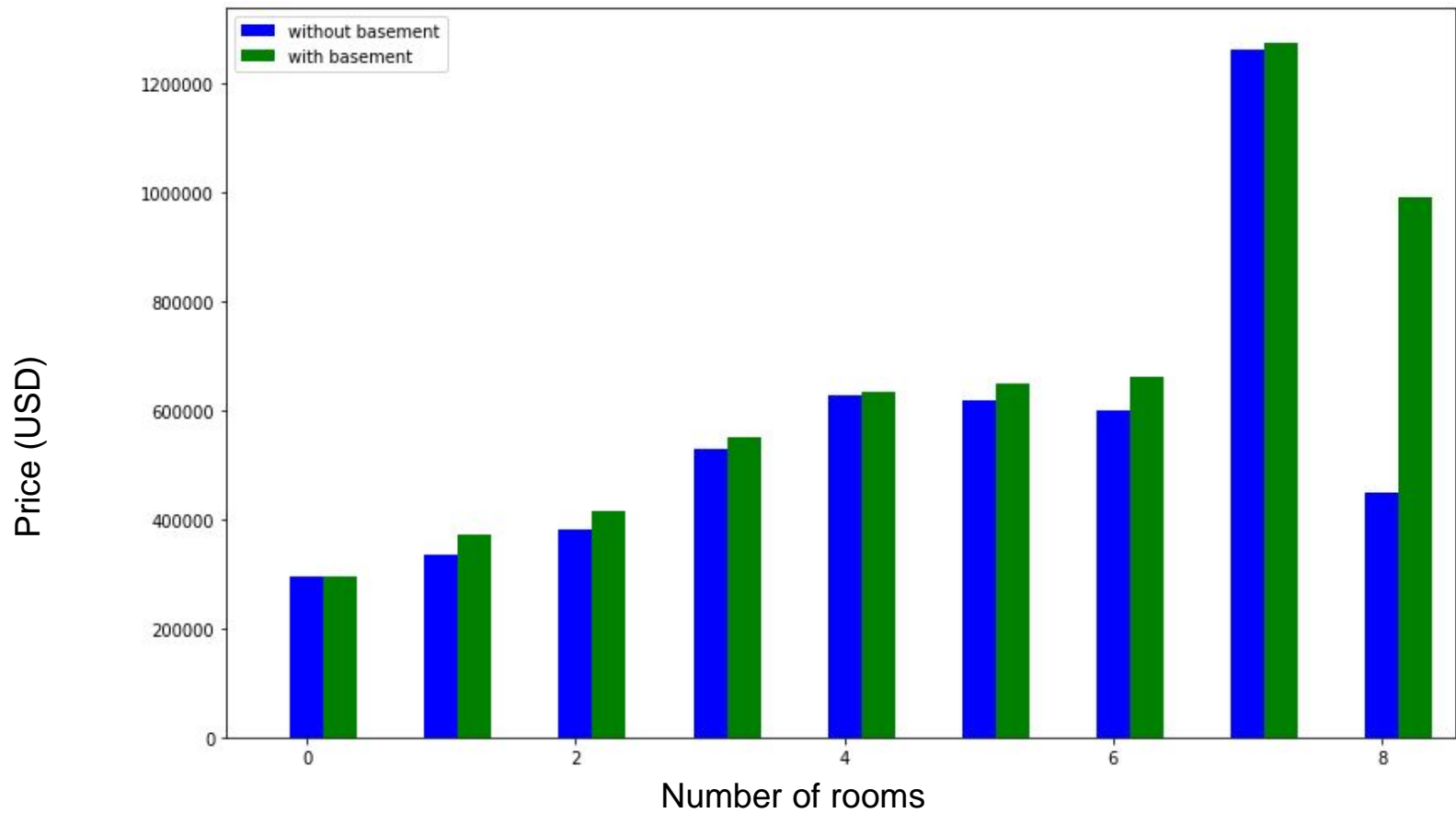


Insights from graph:

The median values of the houses increase with the increase in the number of rooms.

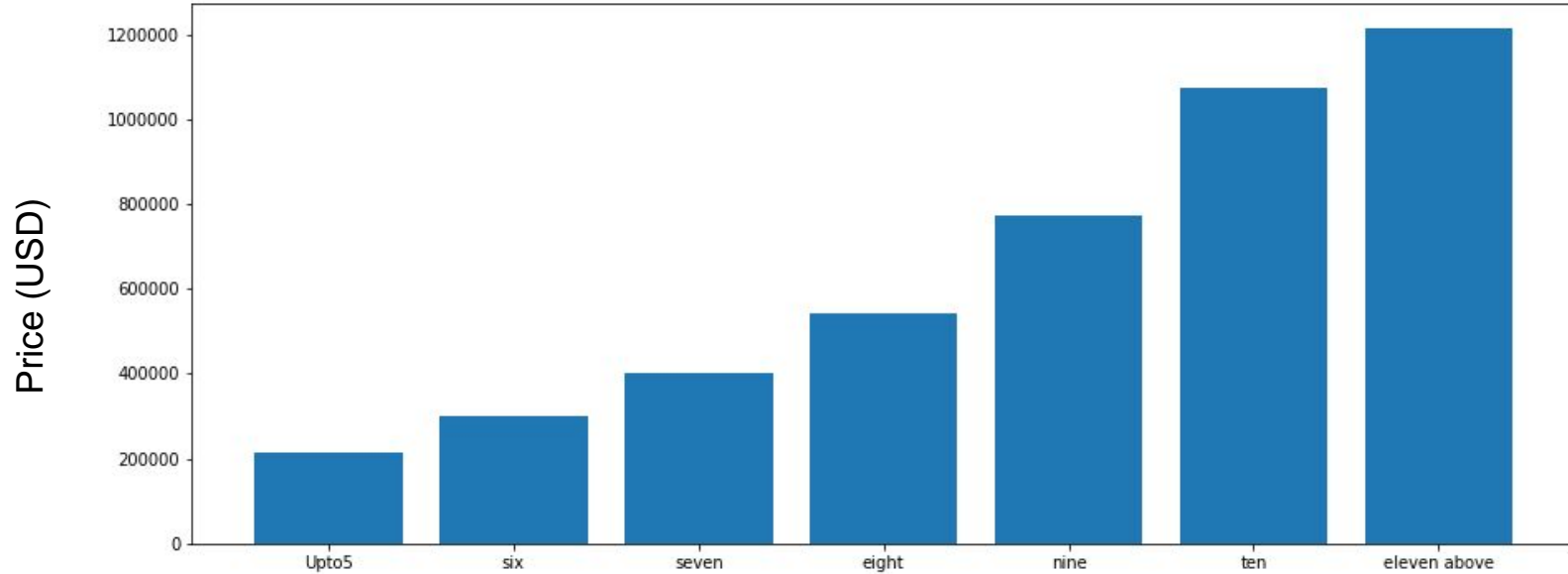
However there is no clear distinction of prices for different bedroom houses as it can due to the reason of different luxuries.

Eg: A 2 bedroom house might have a modular kitchen, luxury lights, etc and a 3 bedroom house might not be having these features and may cost lower



Insights from graph:

It can be seen that except for 8 bedroom houses, there is not much difference in the pricing for houses with and without basement.



Grade (represents the construction quality)

Insights from graph:

Grade is assigned by the King County authorities which represents the construction quality. Grades run from grade 1 to 13.

Its evident from the graph that higher the grade higher the price.

The end