

Seattle Housing Dataset - EDA

by Thomas Roider





Dataset

- Seattle Housing Dataset
- Containing information regarding:
 - prices of sold houses in the last year
 - number of bedrooms / bathrooms
 - size of houses and surrounding houses
 - condition / grade of a house
 - location
 - etc.



Stakeholder: Erin Robinson

- Buyer / Seller
- Socially Responsible
- Goal:
 - Find houses in poor neighborhoods, buy them and sell them with small profit



General Hypotheses

General Insights to the dataset:

- The newer a house, the bigger the house
- If a house has been renovated at least once, it is more likely to have a higher price
- The closer a house is to the city center, the smaller it is



Stakeholder-specific Hypotheses

- Mrs. Robinson wants to invest in poor neighbourhoods
- Definition of a poor neighborhood:
 - houses are graded badly (zipcodes with most houses in lowest 20% of dataset)
 - houses have a lower price (zipcodes with lowest 20% of dataset)
 - houses are small (zipcodes lowest 20% of dataset)
- Zipcodes of poor neighborhoods: 98168, 98148, 98002, 98178
- Goal of the EDA: Find comparatively cheap houses in poor neighborhoods



Stakeholder-specific Hypotheses

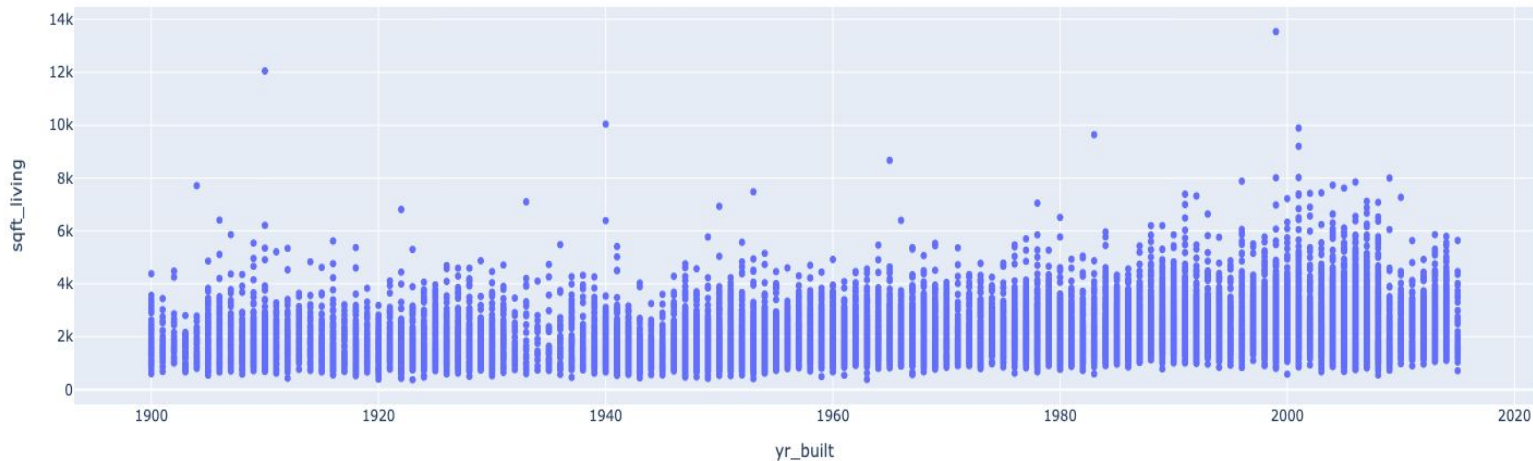
- The smaller the houses around a house, the lower the house price
- If a house is located in an area with small houses, it's more likely to be graded badly
- If a house is in a poor neighborhood, it tends to be smaller and to have a smaller lot



The newer a house, the bigger the house

Correlation: 0.31815229

-> Slight Correlation



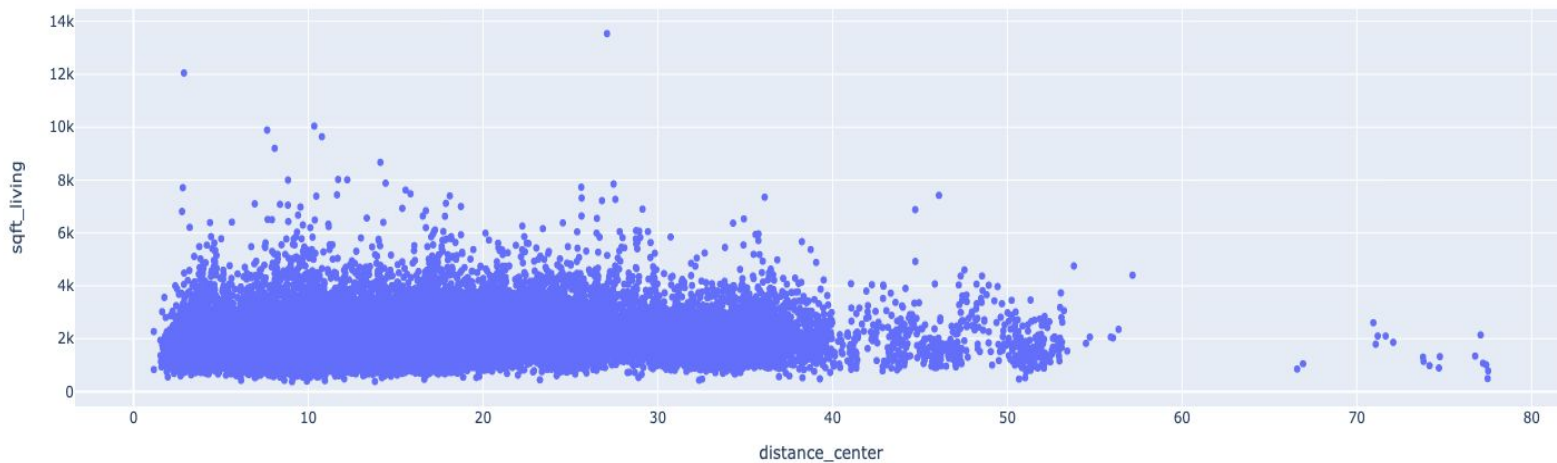


The closer a house is to the city center, the smaller it is

Assumption: The city center in Seattle is "Downtown Seattle" with the zip code "98121"

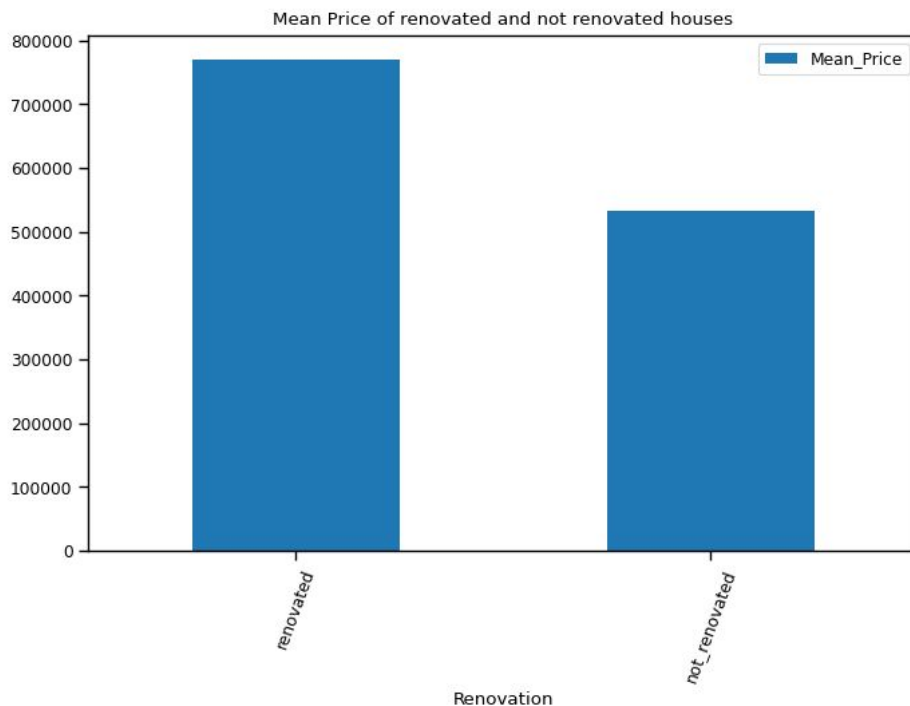
Correlation: 0.07741065

-> No significant Correlation





If a house has been renovated at least once, it tends to have a higher price.

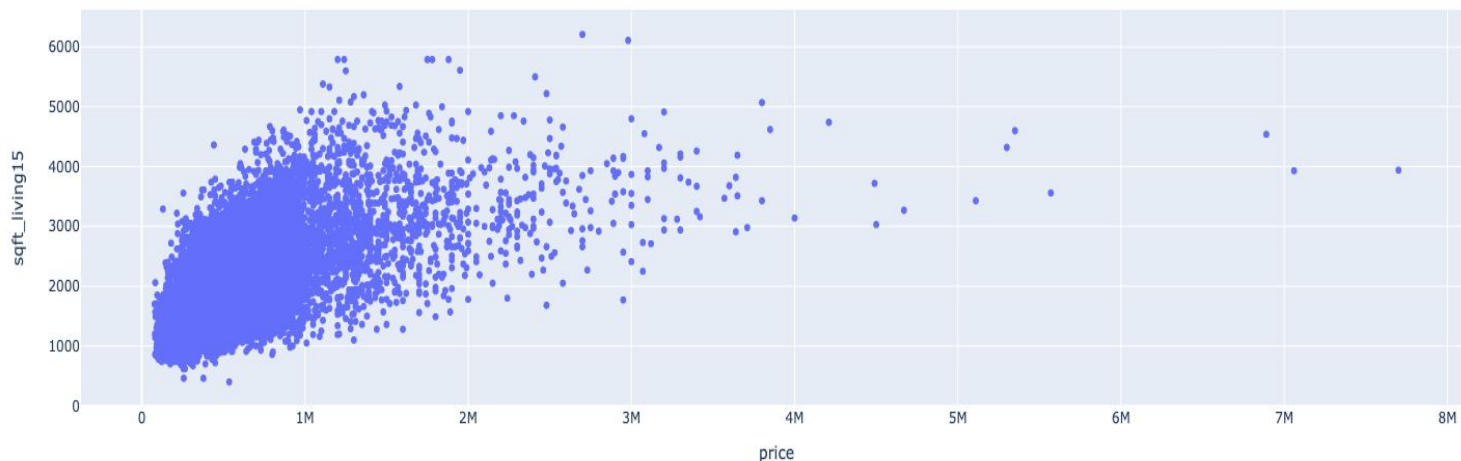




The smaller the houses around a house, the smaller the house price

Correlation: 0.5852412

-> Significant Correlation

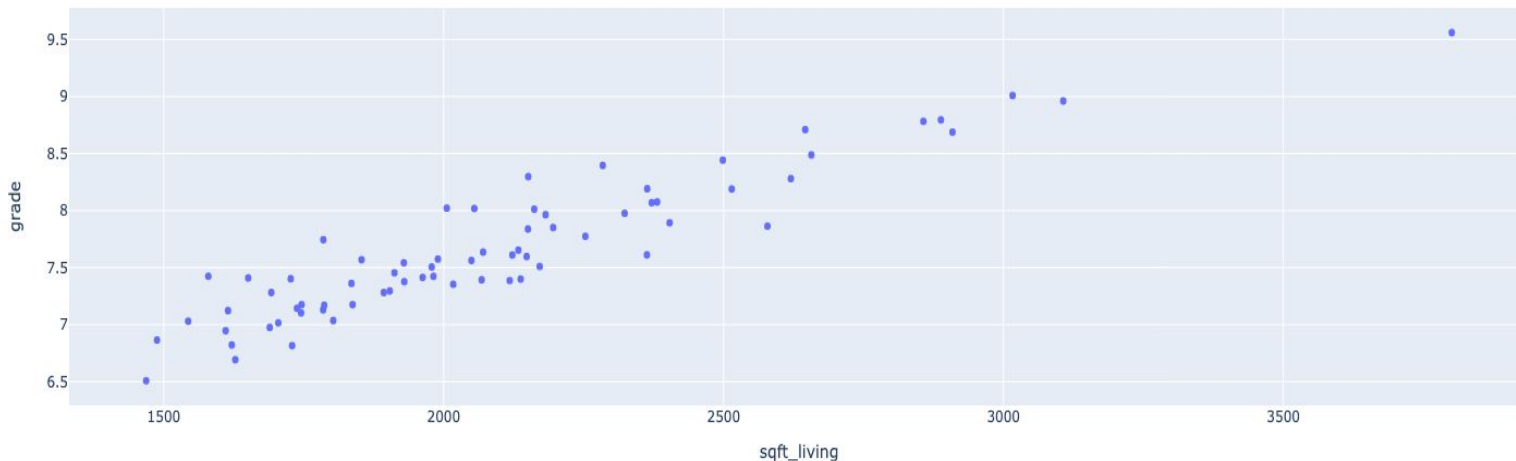




If a house is located in an area with small houses, it's more likely to be graded badly

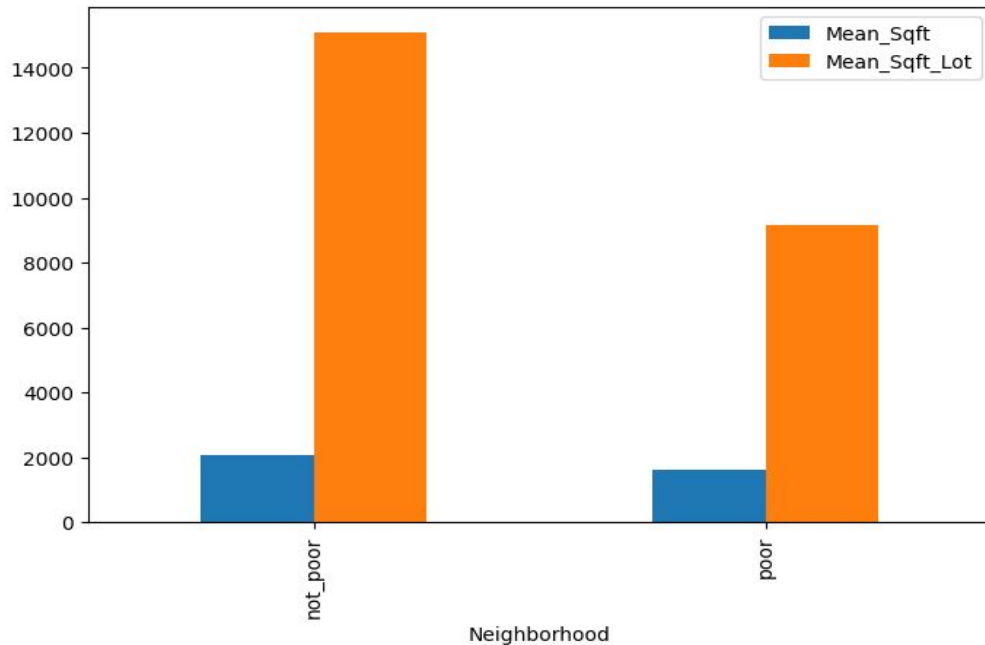
Correlation: 0.91743516

-> Strong correlation





If a house is located in a poor neighborhood, it tends to be smaller and have a smaller lot





Recommendations

Avg. price / sqft_living in poor neighborhoods: 174.61 \$

The following houses could possibly be interesting (IDs): 2891000610, 3800008, 6121800050

- > Houses with the lowest price / sqft

- > They should be looked at more closely

- > A renovation could help to increase the price and the quality of living in the neighborhoods