

# Exploratory Data Analysis

## King County Real Estate Data

Marina Nascimento Souza  
Spiced Academy



01

## Clean Data

- Check for duplicates
- Check missing entries
- Round data
- Fix unreasonable entries

02

## Understand Data

Plot variables that are relevant to my client to establish correlations



03

## Establish and Implement Criteria

Established criteria according to client's needs and implemented it as filter to data



04

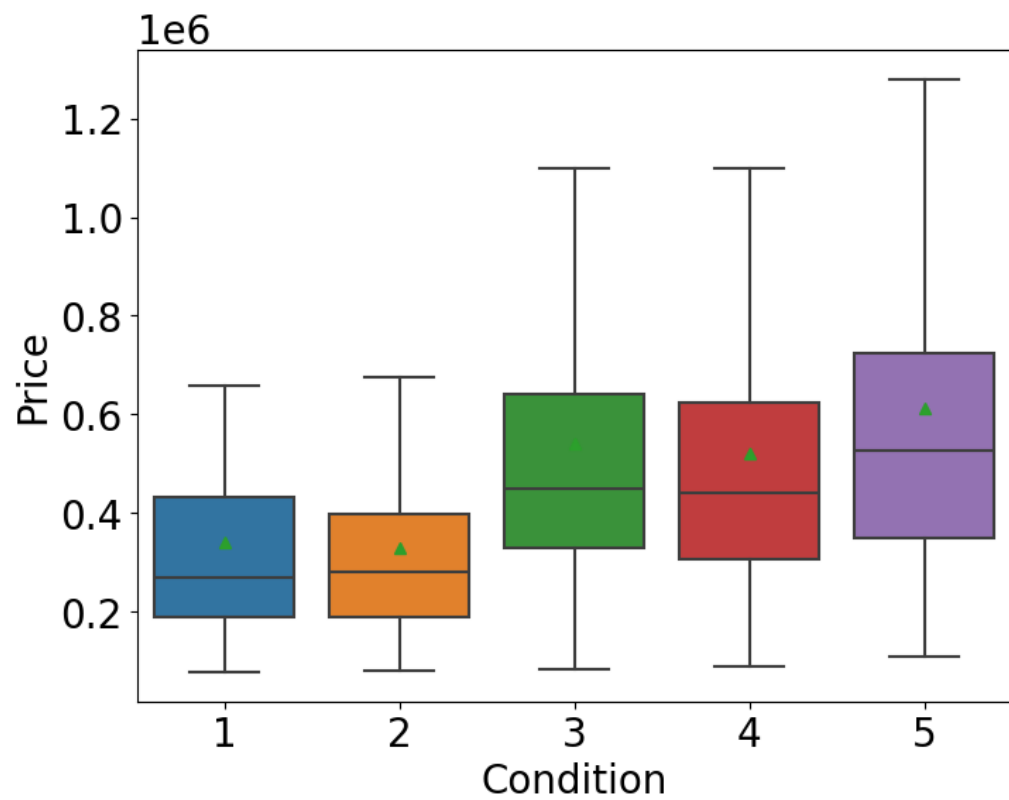
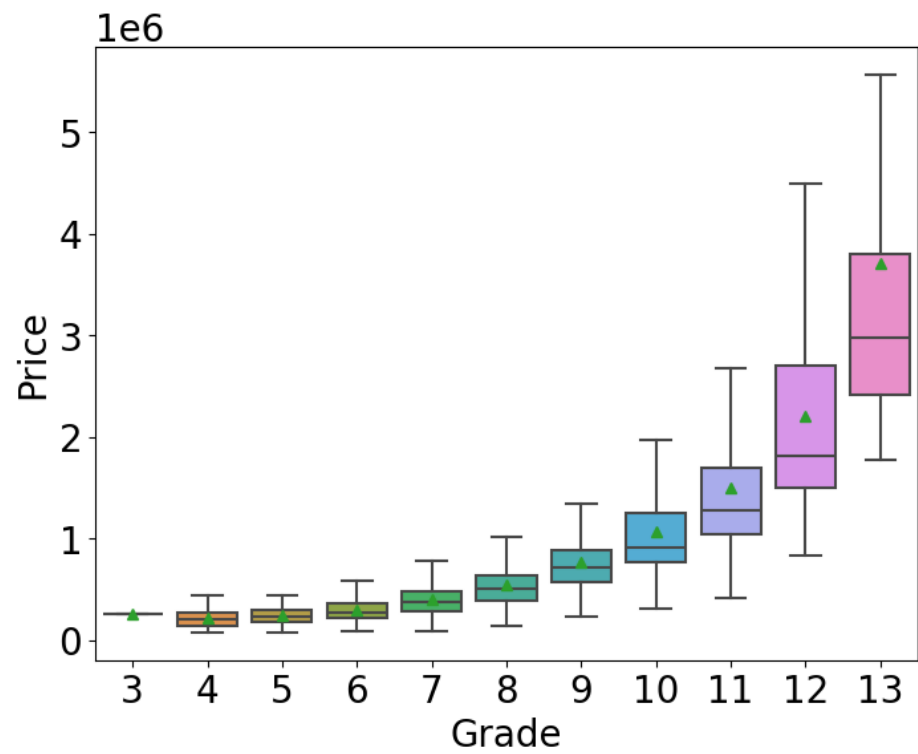
## Reach Conclusion

Ultimate goal: Find properties for my client

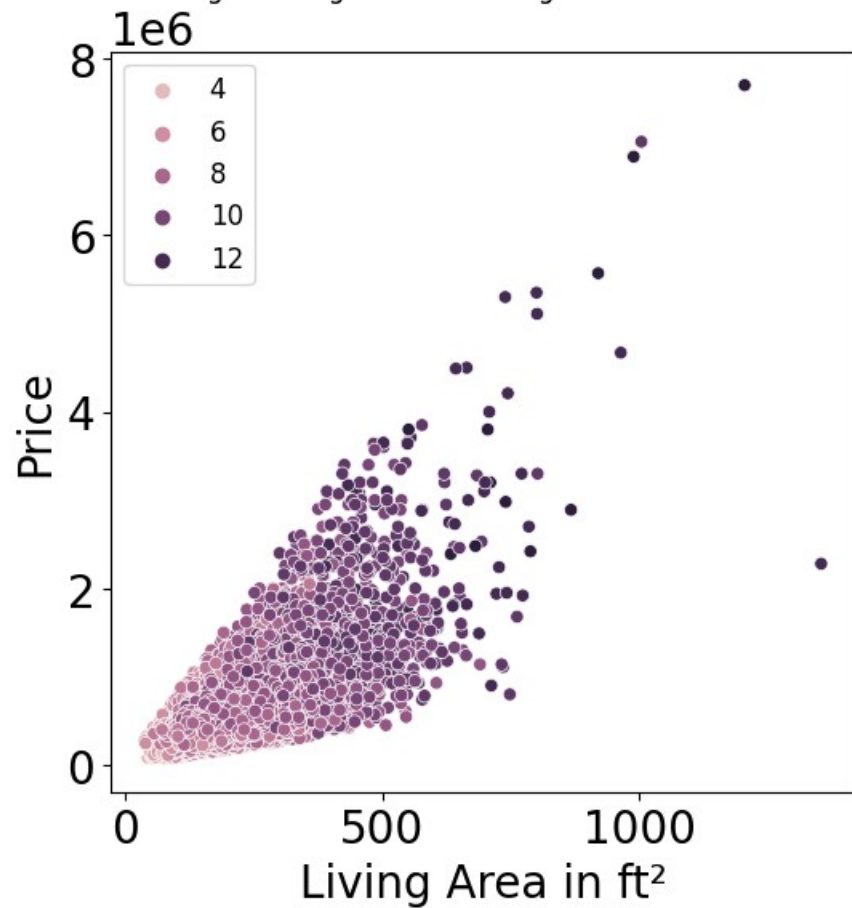


# Understanding Data

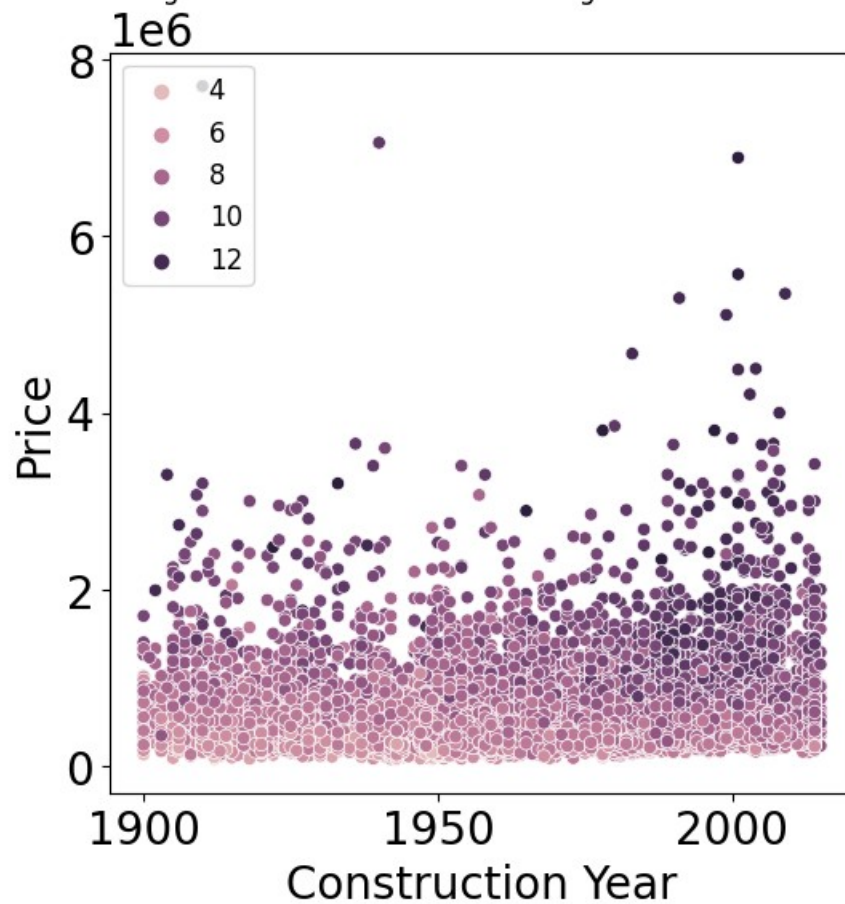
Using Price as Guiding Variable



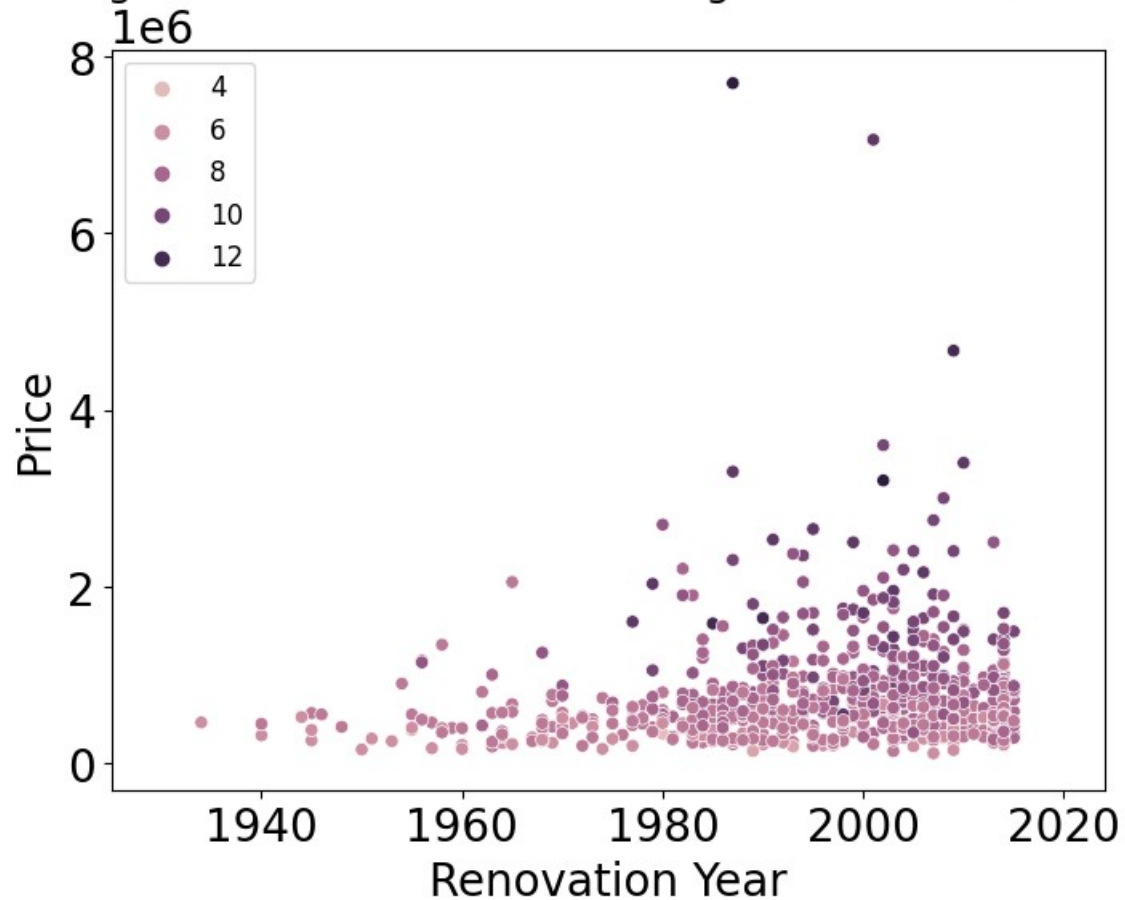
Scattering of Living Area According to Price with Grade Hue

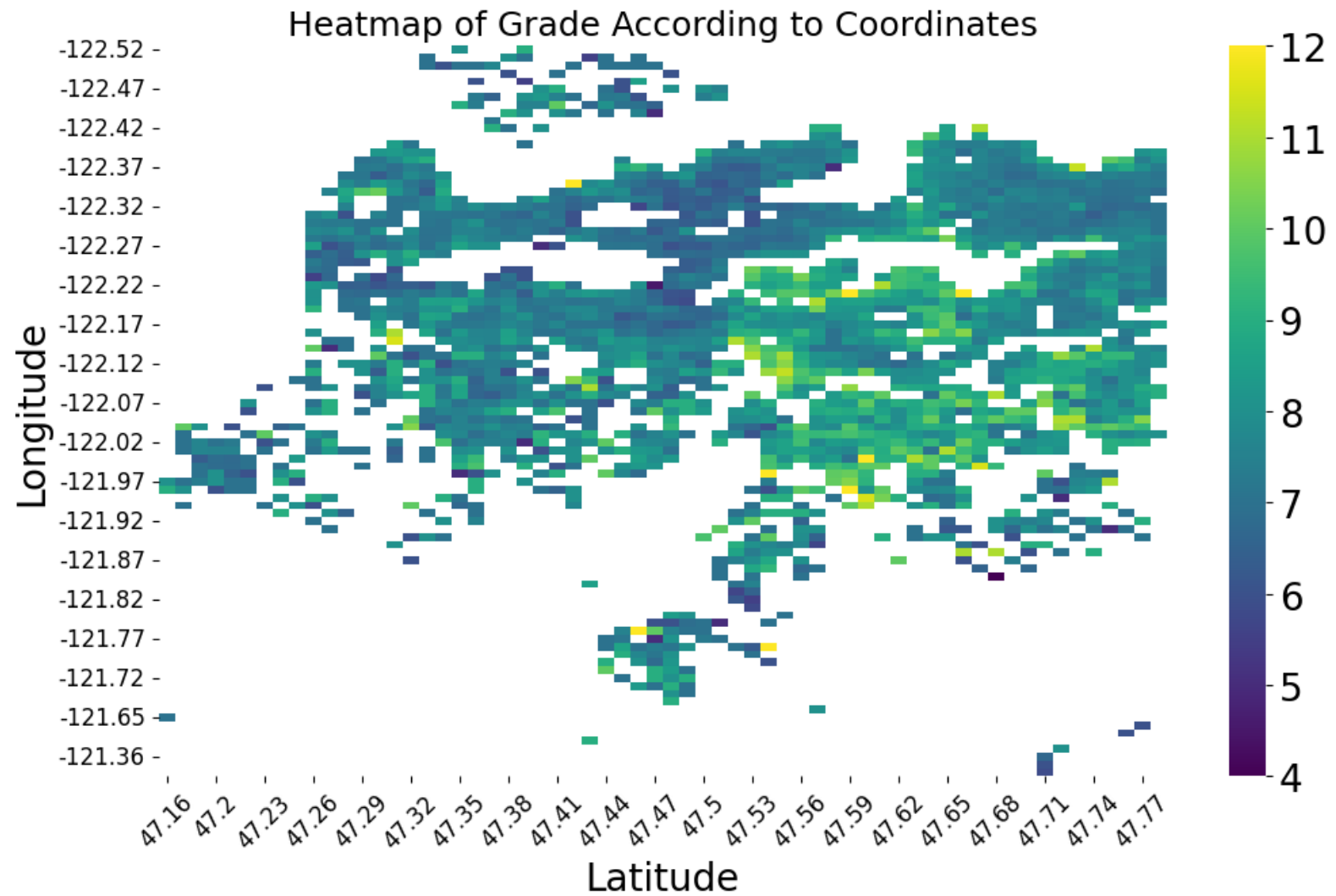


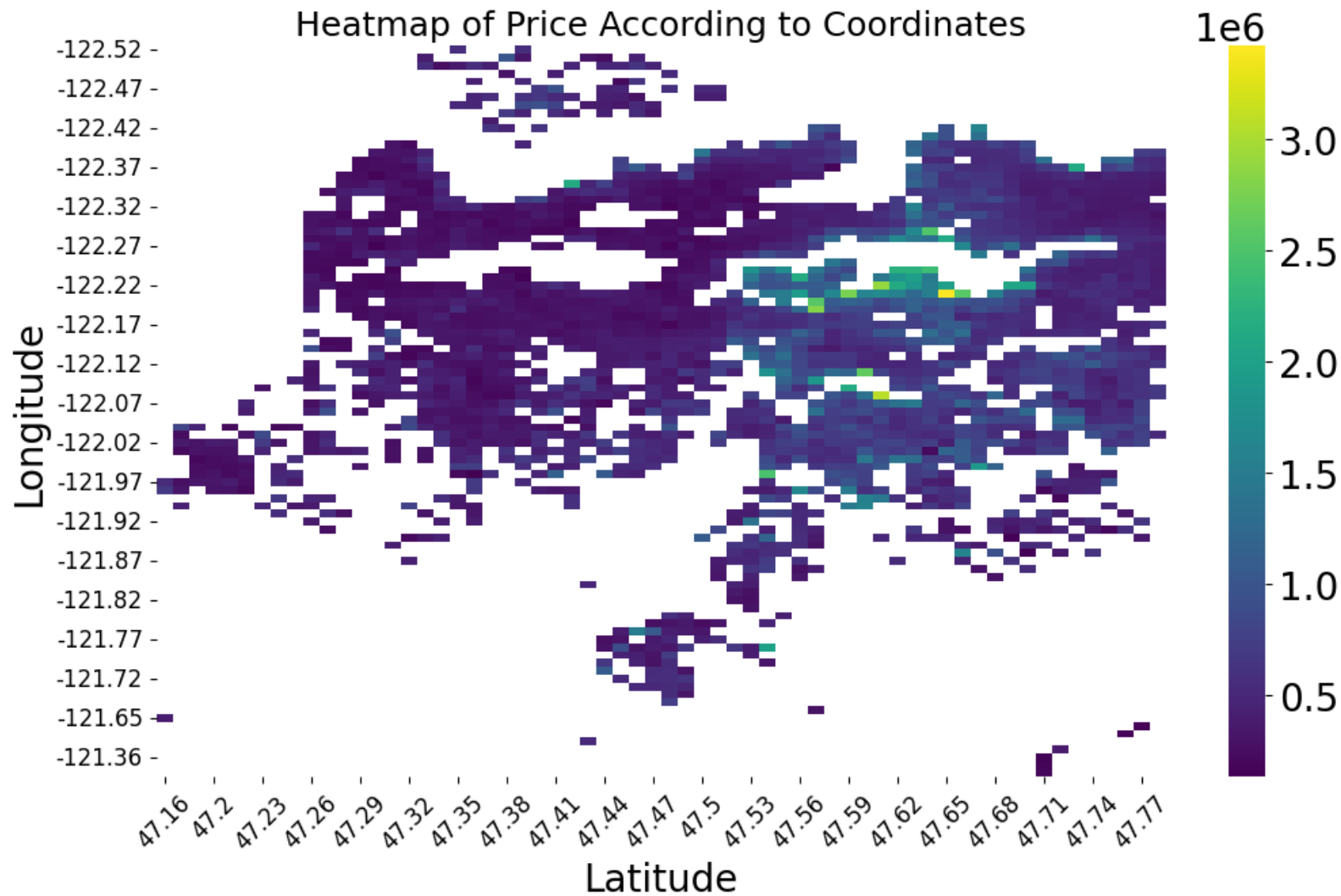
Scattering of Construction Year According to Price with Grade Hue



Scattering of Renovation Year According to Price with Condition Hue









## Client:

5 kids, no money, wants nice (social) neighborhood, Timing?, Location?

## Criteria:

- ✓ Wants house of grade under average, so **grade under 8** as he is handy and willing to renovate himself. He wants **at least 4 bedrooms and 2 bathrooms**, he wants **no waterfront**, as his kids don't know how to swim and it makes the property more expensive. He wants **at least 120m<sup>2</sup> of living area and at least 50m<sup>2</sup> garden** for his kids to play in and his garden does not have to be larger than 100m<sup>2</sup>. He **wants a basement** to storage all his kid's stuff.
  - Garden area is equal to lot area minus living area
- ✓ My client sets the **limit price as average** price for houses within first criteria
- ✓ I define a neighborhood as social, where **houses are of average grade above 8**

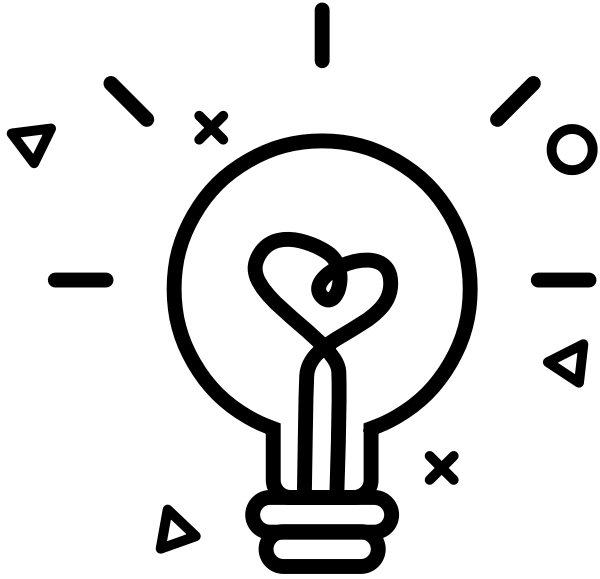


- Dolor Sit

# Results

I filtered out properties that don't fit my criteria and I come up with a list of 10 properties that are a match for my client

house_id	price
3814400125	493000.000
4278900055	599000.000
6373000130	555000.000
6870310010	599950.000
1565930130	429900.000
8812401450	660000.000
710600160	665000.000
3574770100	550000.000
1523059239	475000.000
1776230190	495000.000



THANK YOU

