

# CAPSTONE PROJECT -HOW A SETTLERS WITH THEIR FAMILY TO PICKUP A HOUSE

Applied Data Science Capstone by IBM/Coursera

# Table of contents

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

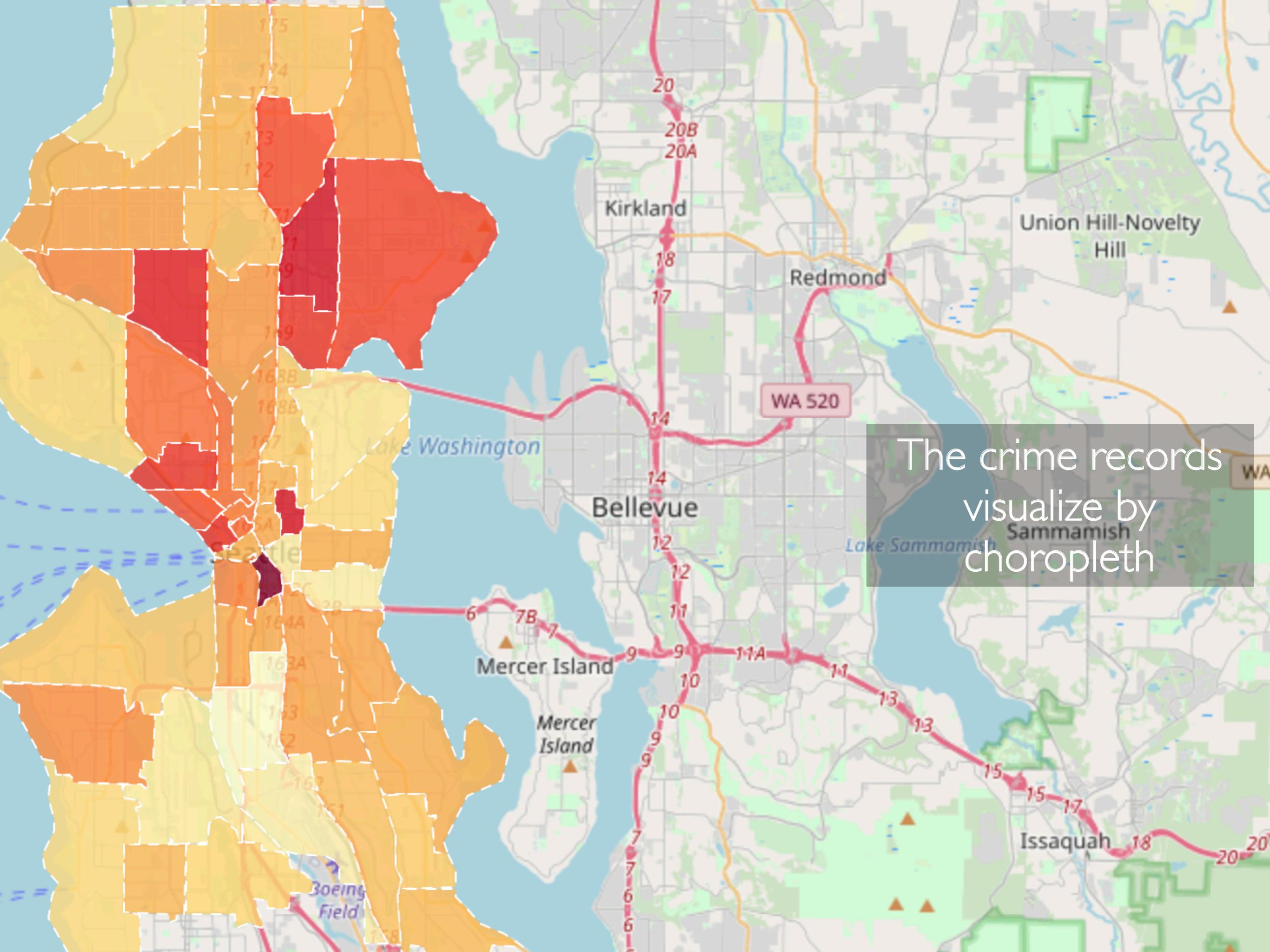
# Introduction: Business Problem

In this project we will try to find an optimal area for a new comer to settle down in Seattle. Specifically, this report will be targeted to the settlers with their family.

Considering the following factors when picking a new place to live you don't end up wasteing your valuable time and money making a move you will end up regretting.

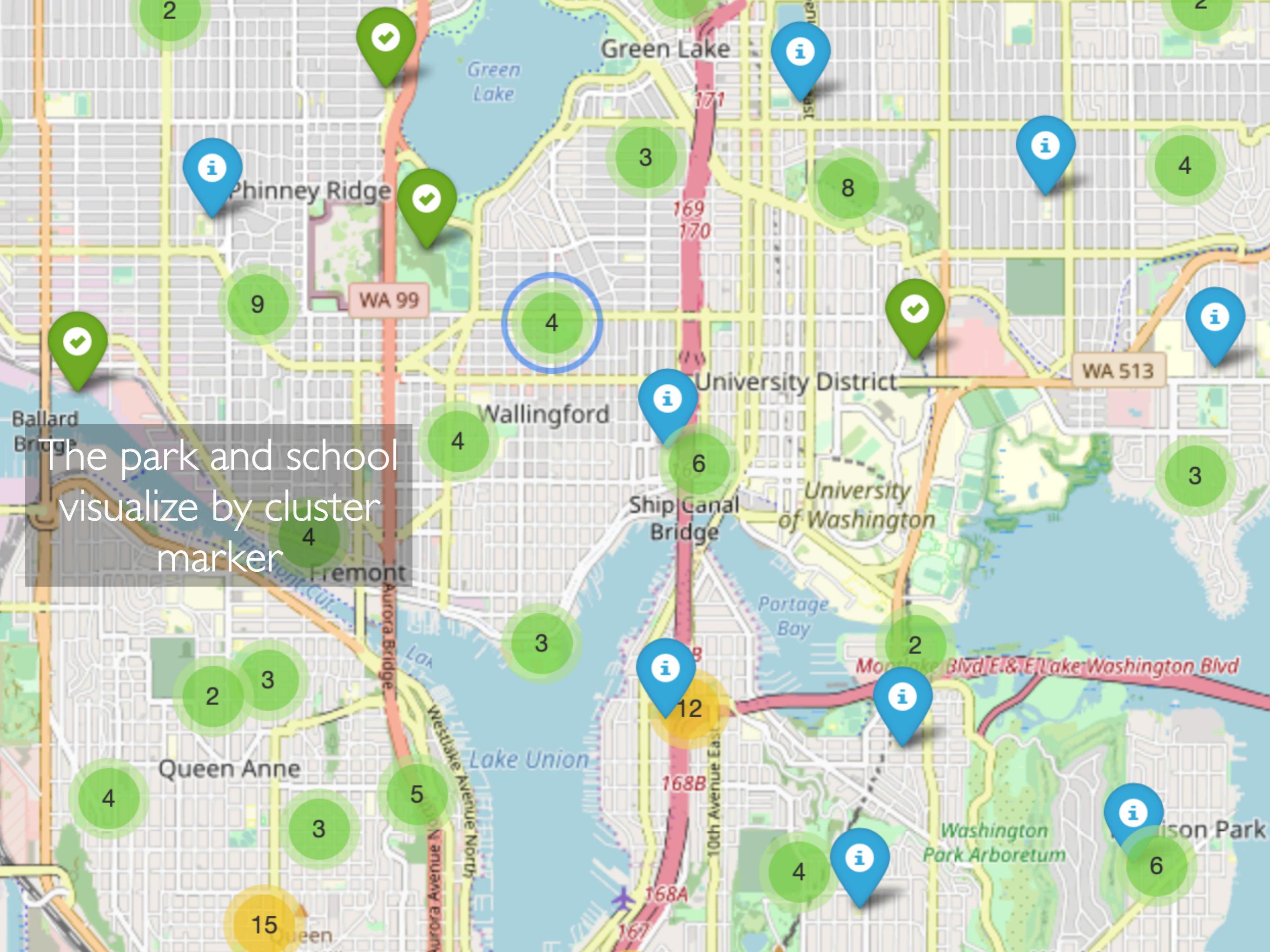
# Data

- number of crime in the area 
- number of public facilities such as park and school also with some living commercial facilities  
- previous the city house selling data 



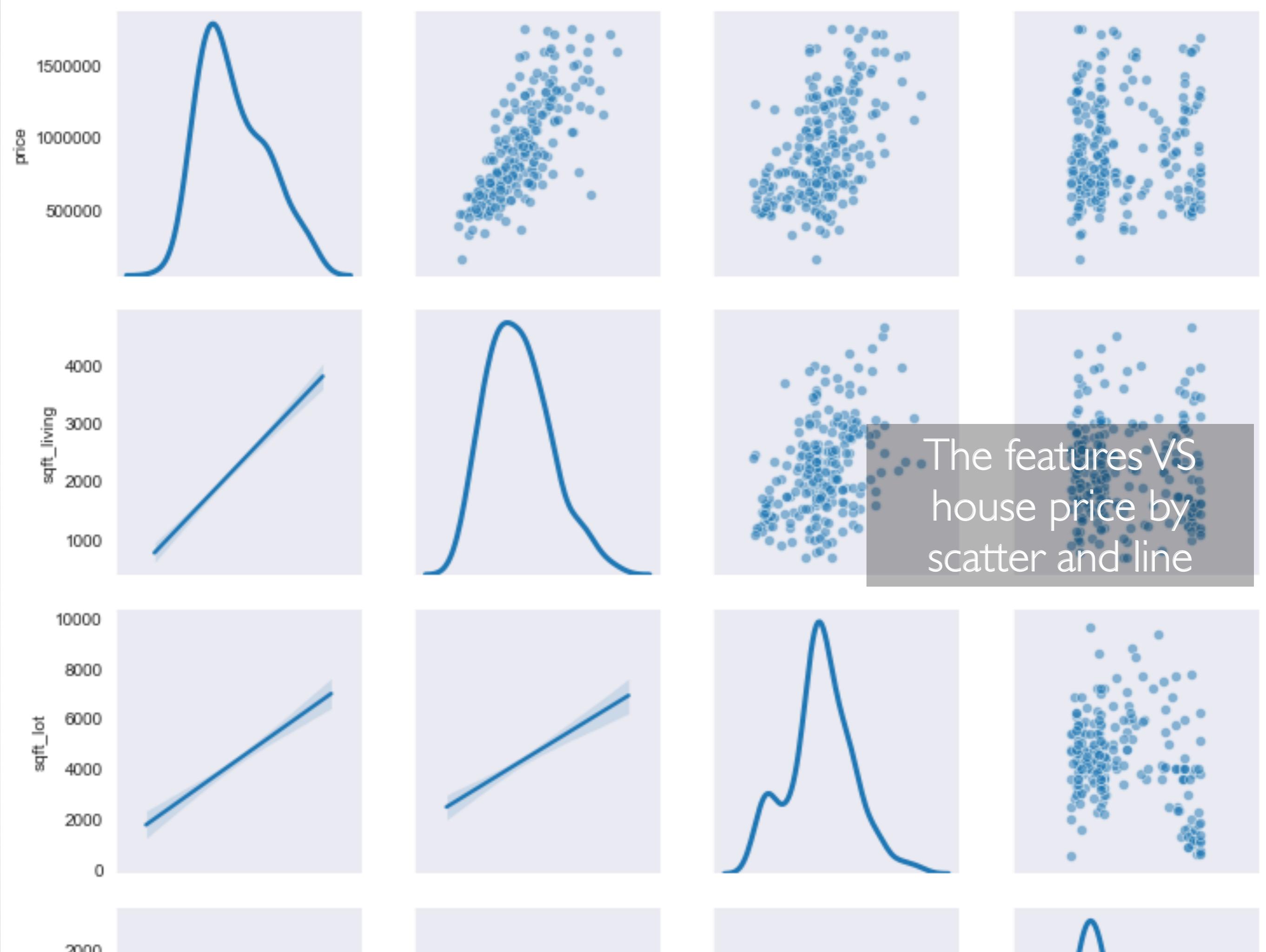
# The crime records visualize by choropleth

# The park and school visualize by cluster marker

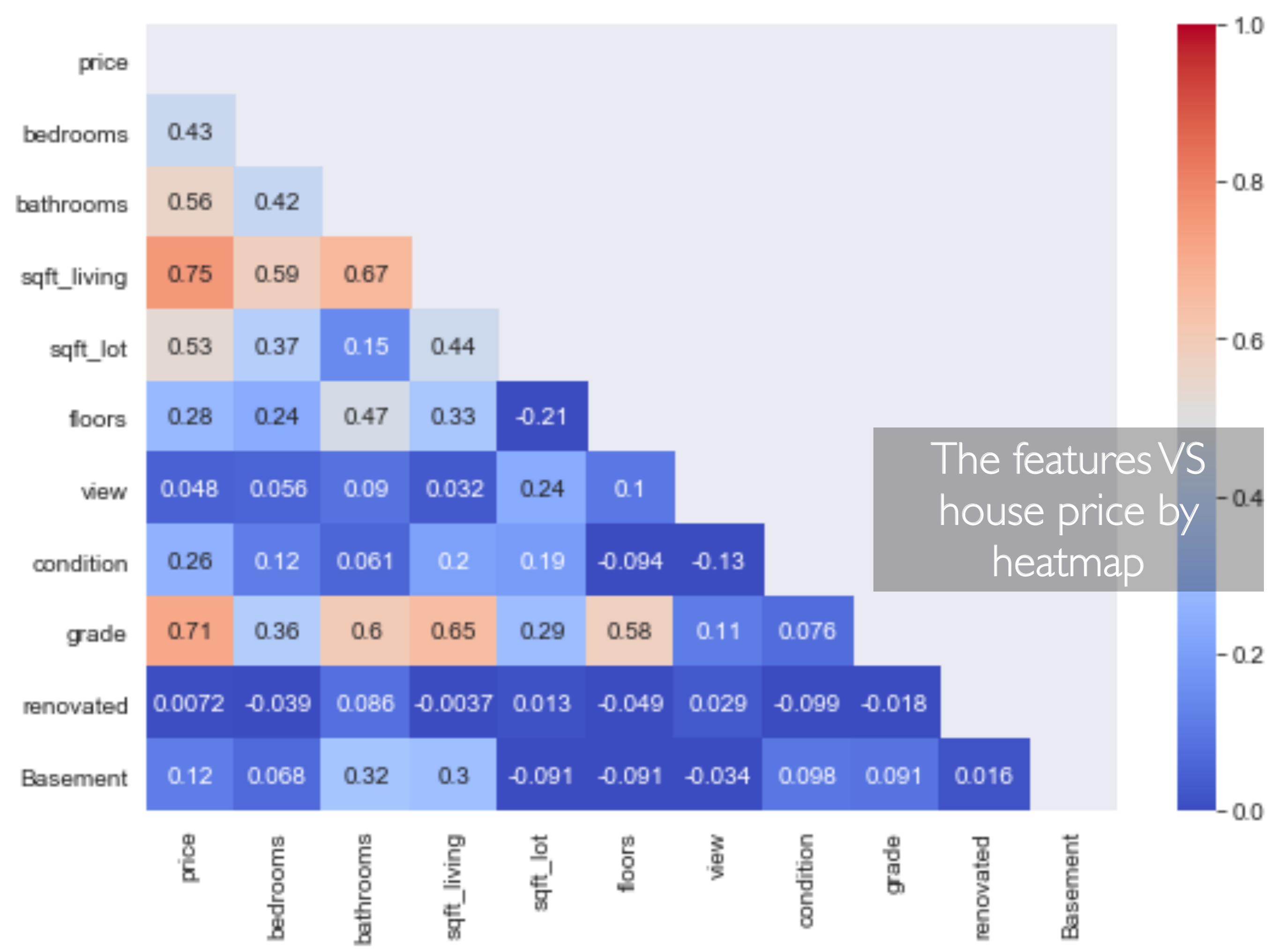


# Methodology

- Filter the crime record before 2000, and highlight the area with lower records
- We`re use the map we generated above, By the linear superposition we can see which area is quite suitable.
- Using the Foursquare API to get the venue data
- Calculate and explore the house price data by heatmap and regression algorithm



The features VS  
house price by  
heatmap





# Results

- Our analysis shows that how a settler to chose a place and he should pay how much for a new house in Seattle . Highest concentration of area was detected Madison ParkWe`re use the map we generated above, By the linear superposition we can see which area is quite suitable.
- Proportionate increase in living quality('bedrooms','bathrooms','floors','grade','sqft') lead to increase in price more linearly. Model seems to be better at estimating prices rather than predicting them.

# Discussion

- With the limited data and skill, not in a proper way to show the first part of selecting the area directly.
- In sales data, specific property types not included, only sales from one year, so estimations do not account for inflation.
- The predict model not too good to estimate with the limited data, should use more algorithm to test the data, finally get the better model.
- During the data clean, I have filtered the outer data as possible in case to avoid the errors.