

# **EDA PROJECT**

## **King County Houses Analyses**

**Selchuk HADZHAAHMED - 11.08.23**

# Goals

- The purpose of this project is to analyze our dataset and suggest suitable house or houses for our client based on their characteristics.
- Obtaining meaningful results from the data set
- At the end of the day, to be able to find a house or houses for our client in the conditions he wants.

# Data Overview

- First, when I examine our data, I can see that it consists of 23 columns (features) and approximately 21k rows.
- Here we only see the first 8 columns.

	id	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view
0	7129300520	3.000	1.000	1180.000	5650.000	1.000	NaN	0.000
1	6414100192	3.000	2.250	2570.000	7242.000	2.000	0.000	0.000
2	5631500400	2.000	1.000	770.000	10000.000	1.000	0.000	0.000
3	2487200875	4.000	3.000	1960.000	5000.000	1.000	0.000	0.000
4	1954400510	3.000	2.000	1680.000	8080.000	1.000	0.000	0.000

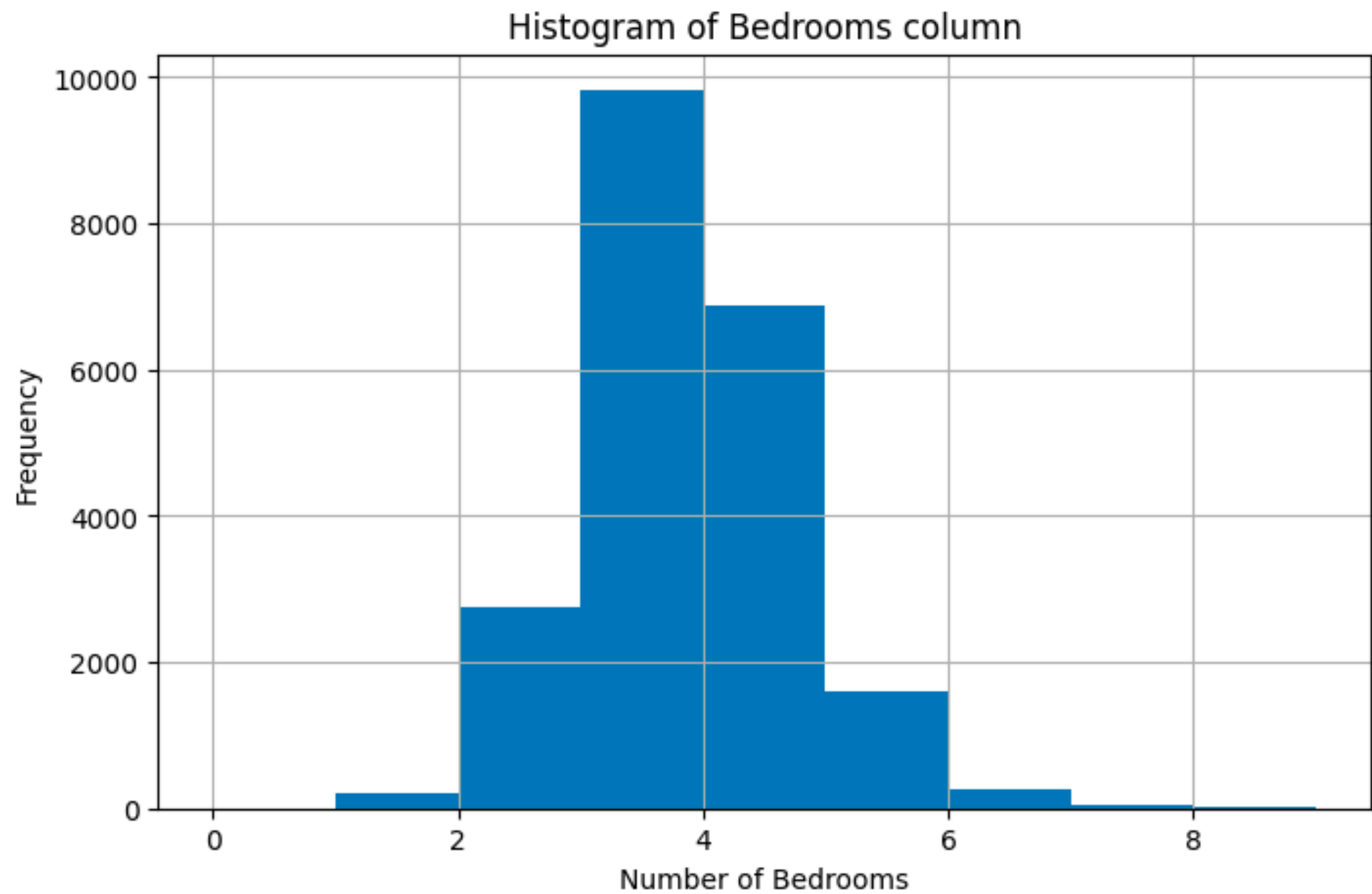
# Data Overview

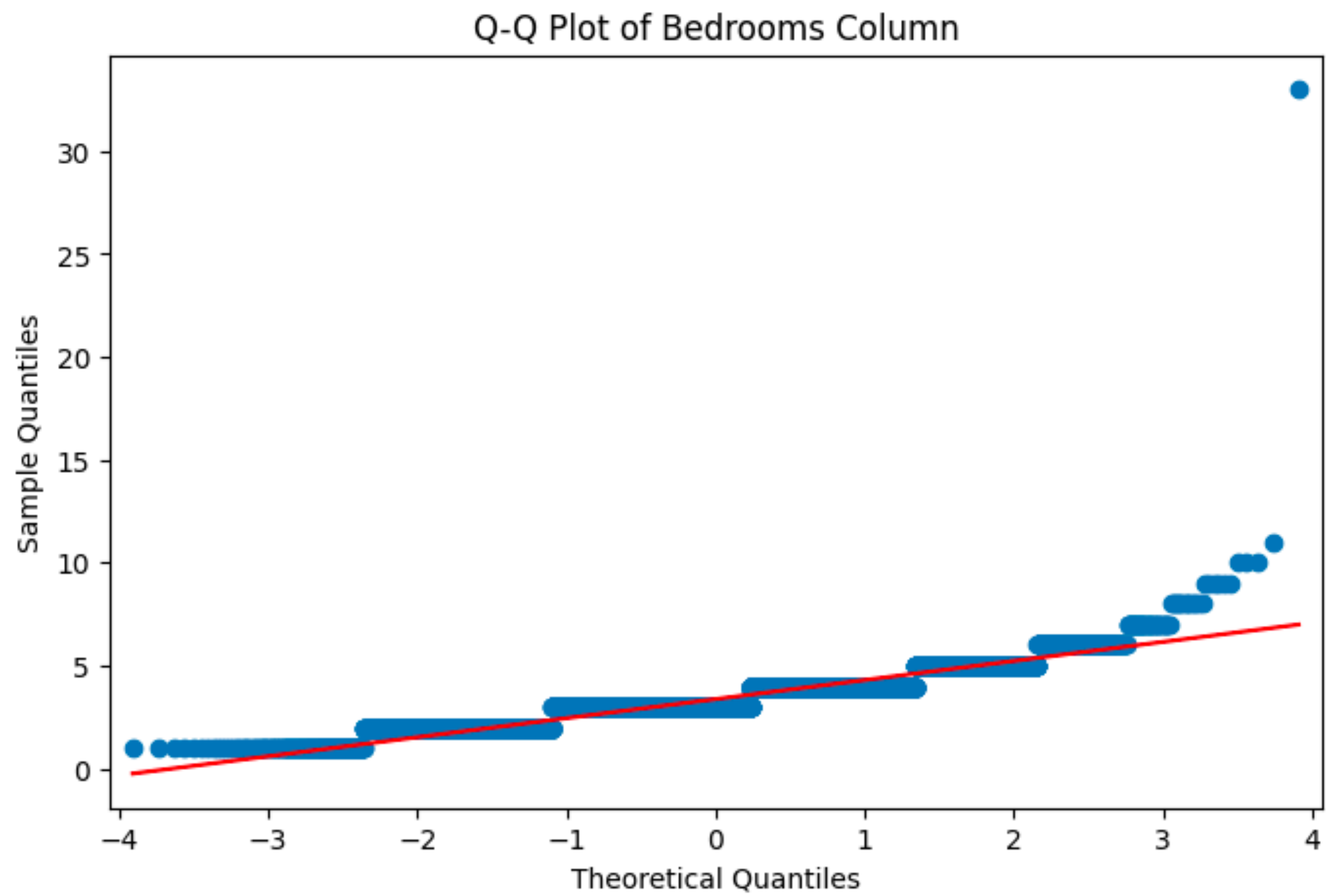
```
RangeIndex: 21597 entries, 0 to 21596
Data columns (total 23 columns):
#   Column              Non-Null Count  Dtype
---  -
0   id                   21597 non-null  int64
1   bedrooms             21597 non-null  float64
2   bathrooms            21597 non-null  float64
3   sqft_living          21597 non-null  float64
4   sqft_lot             21597 non-null  float64
5   floors               21597 non-null  float64
6   waterfront           19206 non-null  float64
7   view                 21534 non-null  float64
8   condition            21597 non-null  int64
9   grade               21597 non-null  int64
10  sqft_above           21597 non-null  float64
11  sqft_basement        21145 non-null  float64
12  yr_built             21597 non-null  int64
13  yr_renovated         17749 non-null  float64
14  zipcode              21597 non-null  int64
15  lat                  21597 non-null  float64
16  long                 21597 non-null  float64
17  sqft_living15        21597 non-null  float64
18  sqft_lot15           21597 non-null  float64
19  date                 21597 non-null  object
...
21  house_id             21597 non-null  int64
22  id.1                 21597 non-null  int64
dtypes: float64(15), int64(7), object(1)
memory usage: 3.8+ MB
```

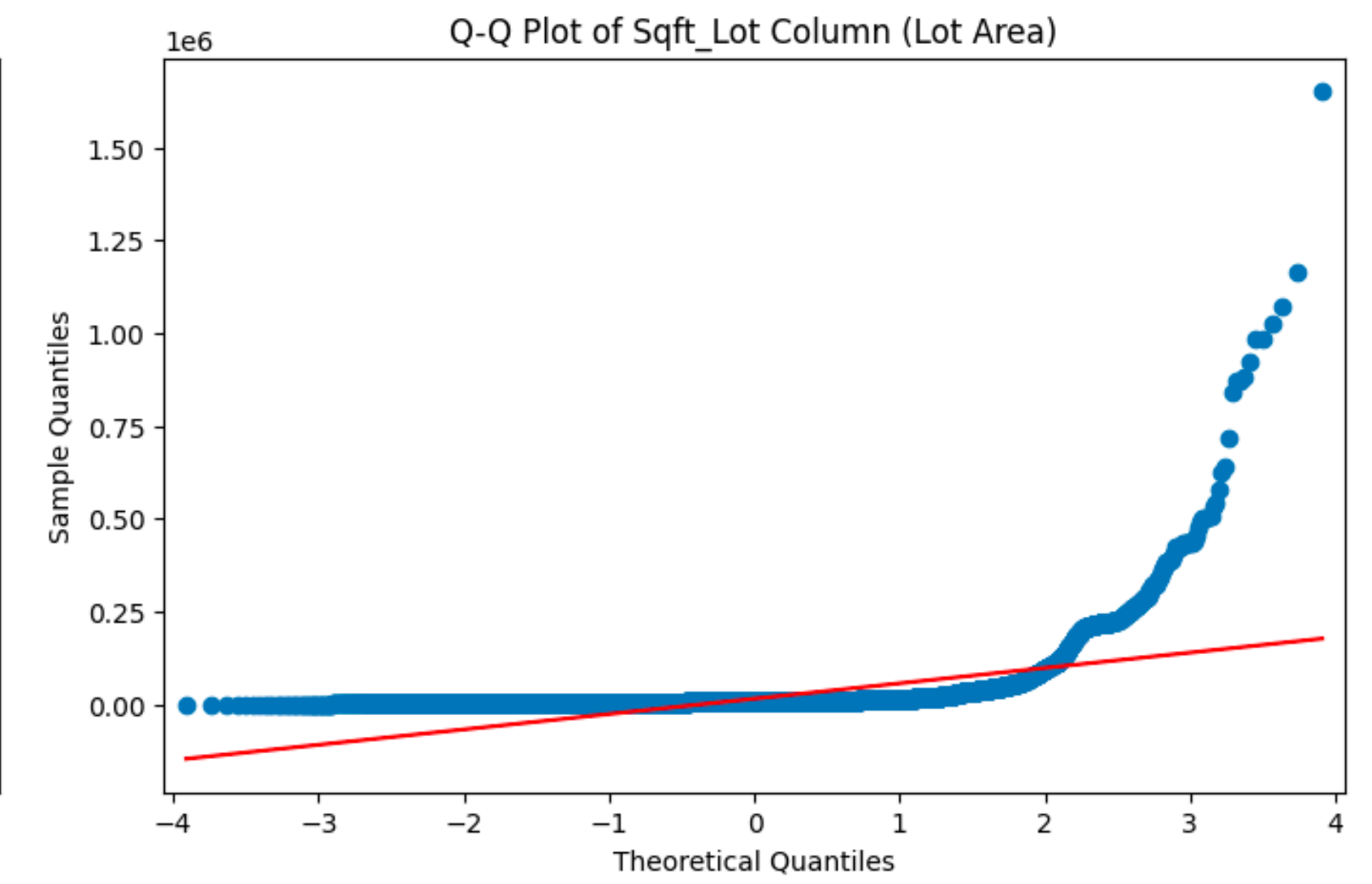
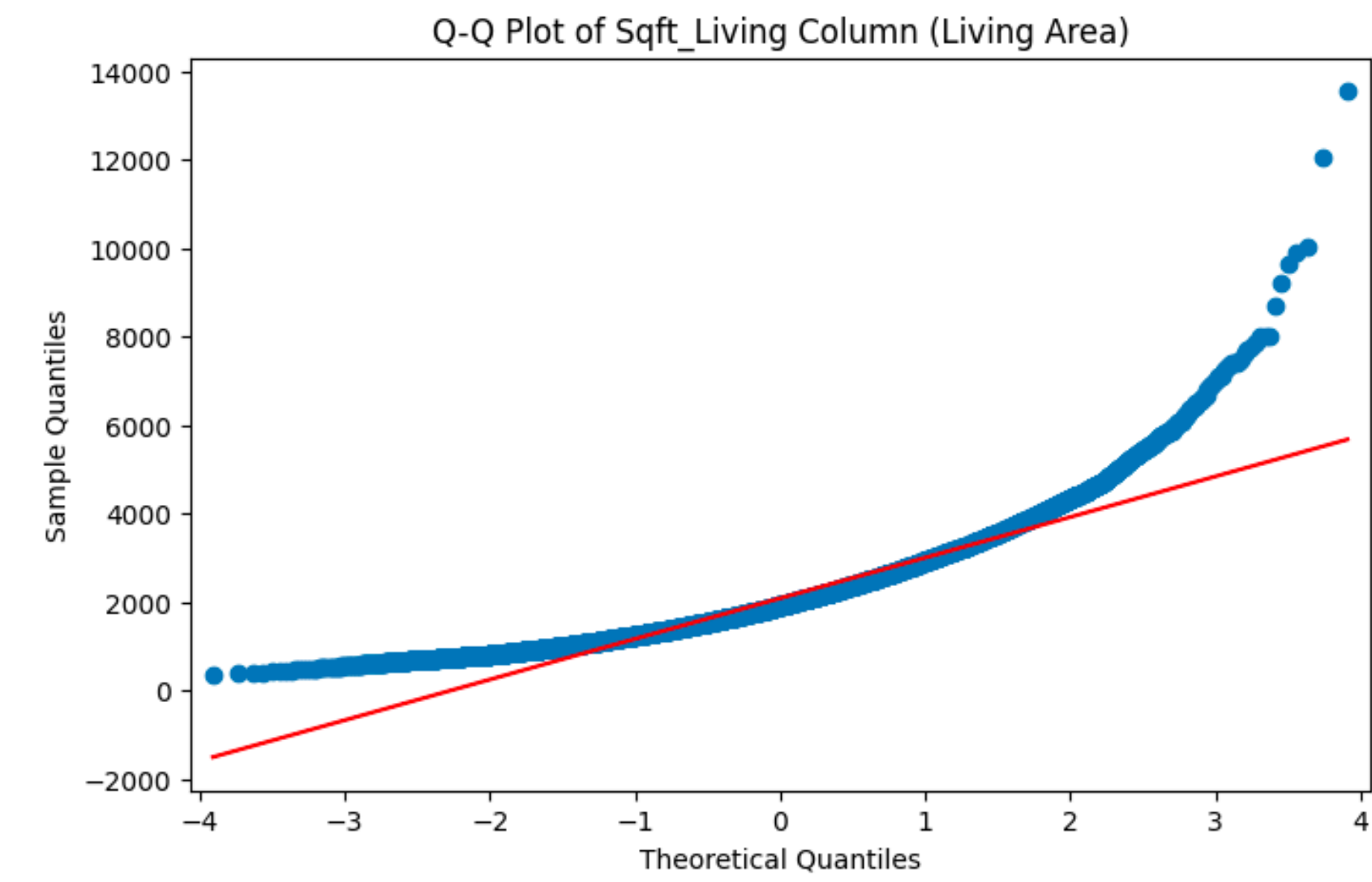
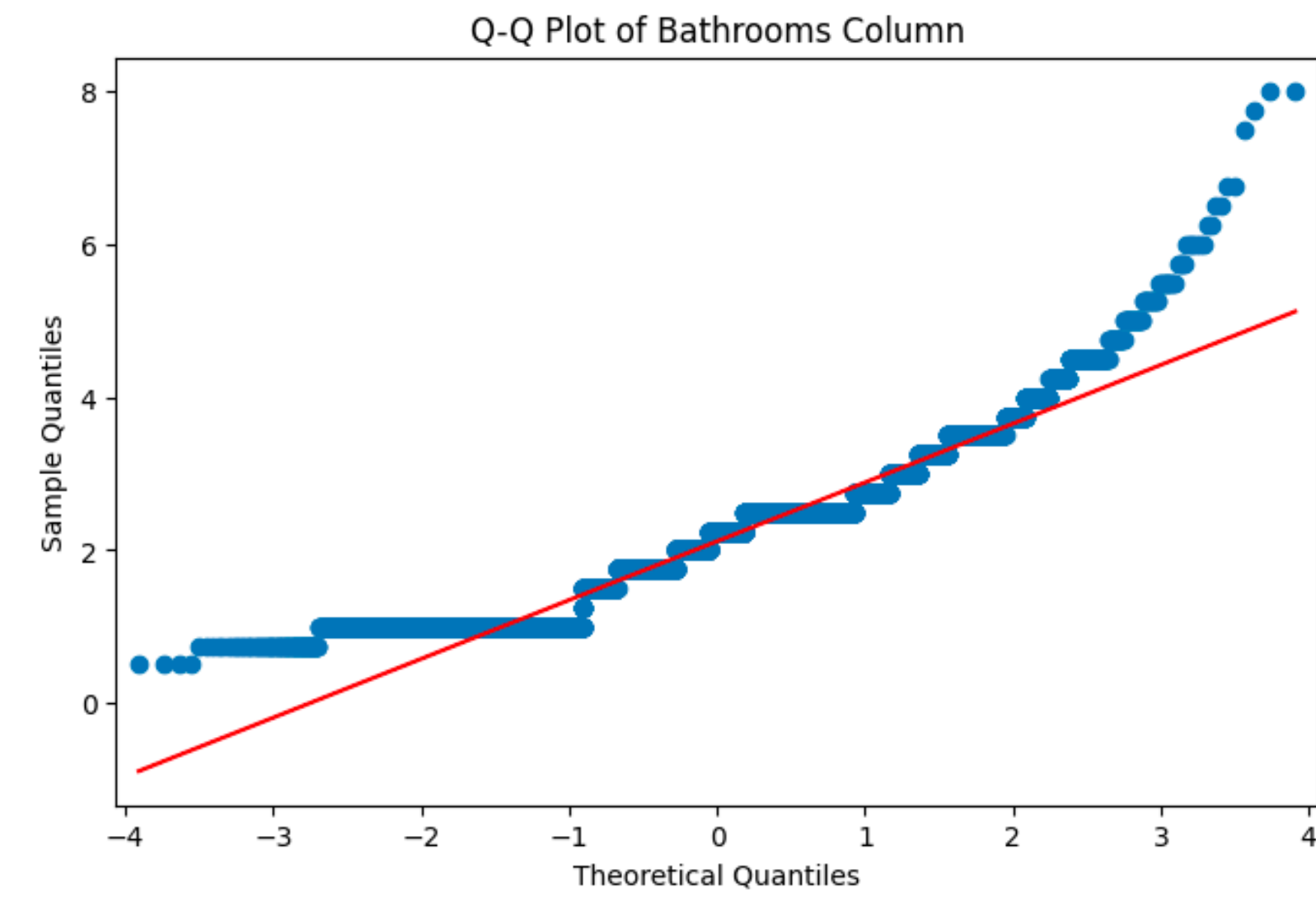
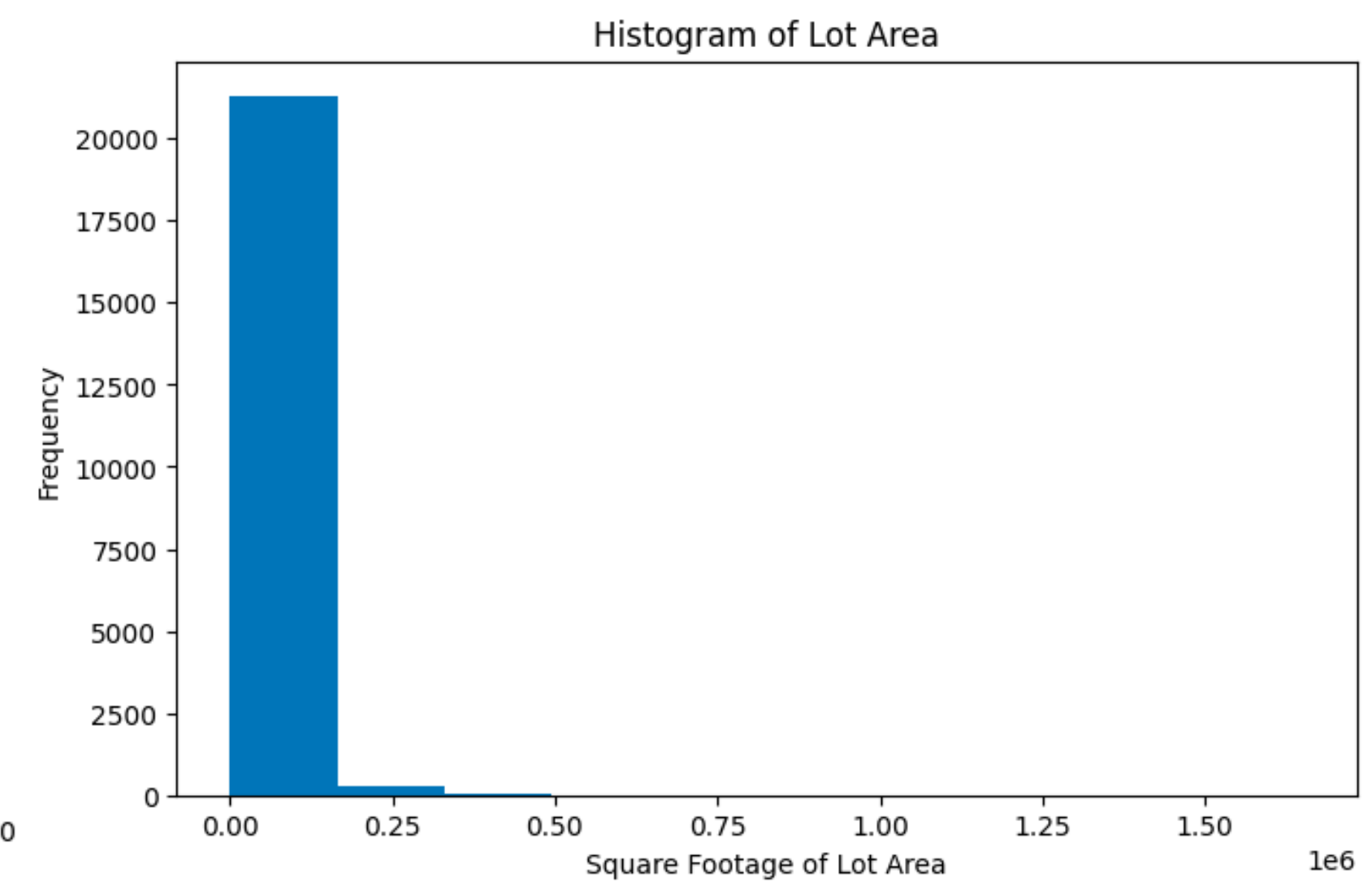
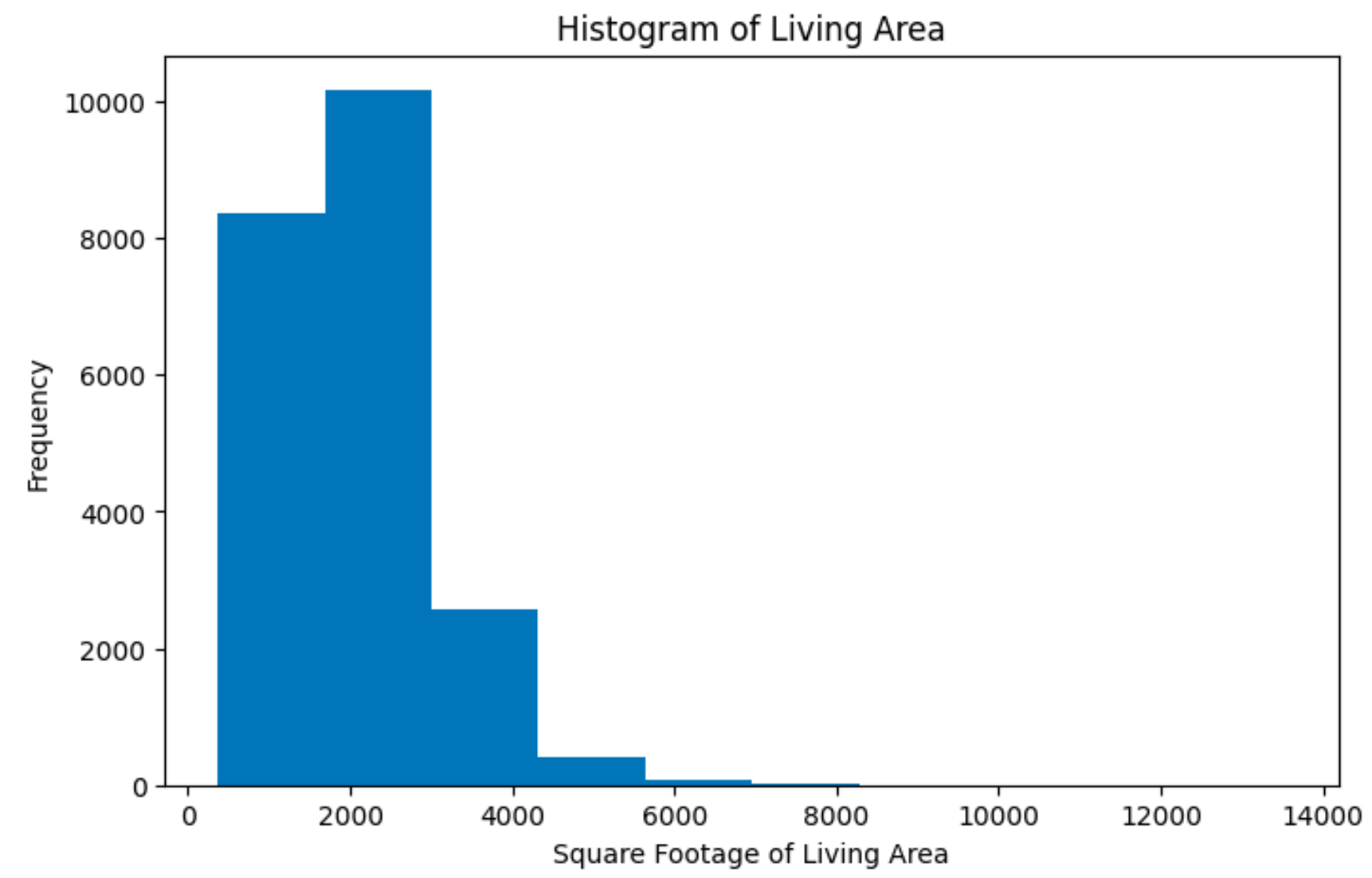
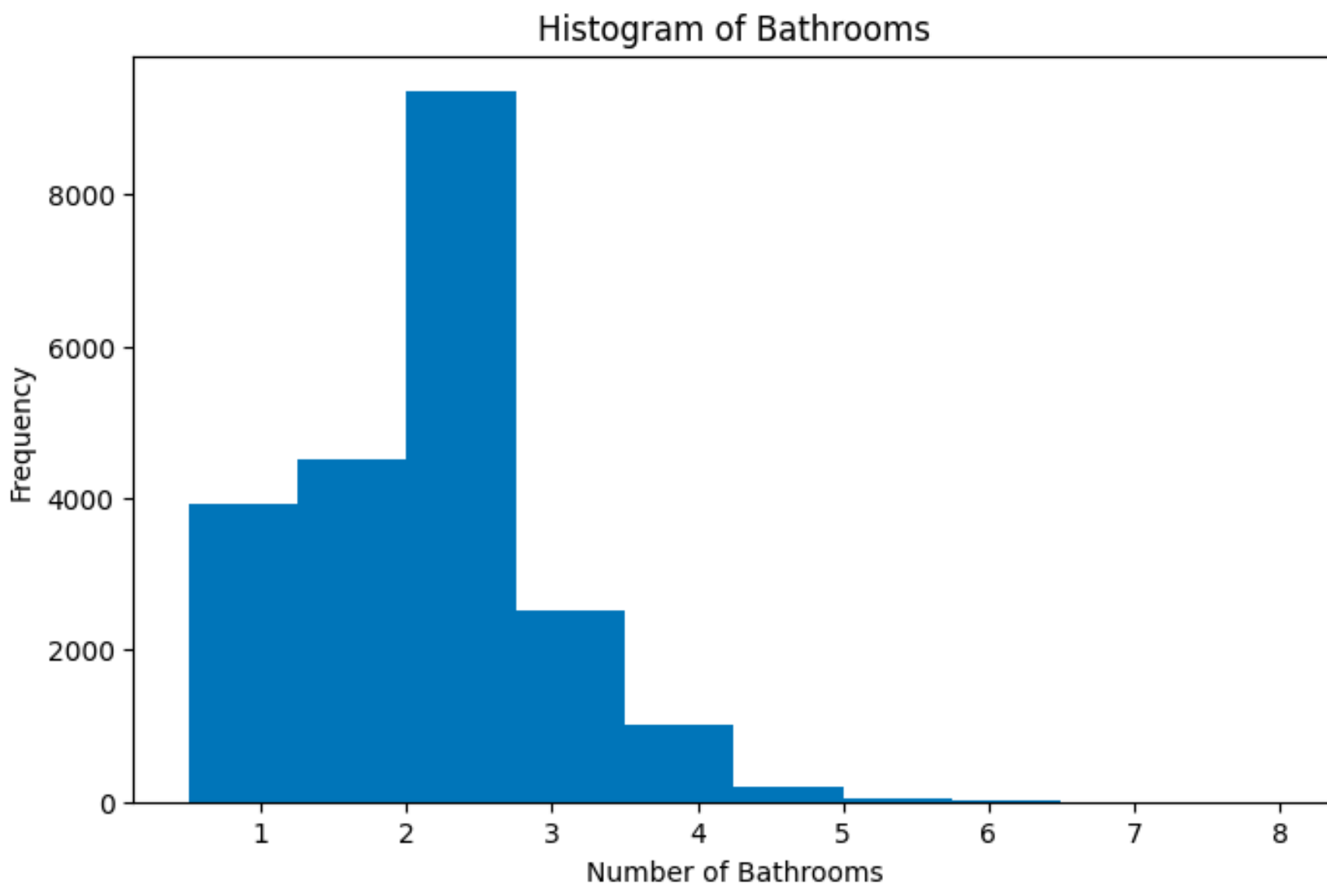
# Technical Side :)

	id	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront
count	21597.000	21597.000	21597.000	21597.000	21597.000	21597.000	19206.000
mean	4580474287.771	3.373	2.116	2080.322	15099.409	1.494	0.008
std	2876735715.748	0.926	0.769	918.106	41412.637	0.540	0.087
min	1000102.000	1.000	0.500	370.000	520.000	1.000	0.000
25%	2123049175.000	3.000	1.750	1430.000	5040.000	1.000	0.000
50%	3904930410.000	3.000	2.250	1910.000	7618.000	1.500	0.000
75%	7308900490.000	4.000	2.500	2550.000	10685.000	2.000	0.000
max	9900000190.000	33.000	8.000	13540.000	1651359.000	3.500	1.000

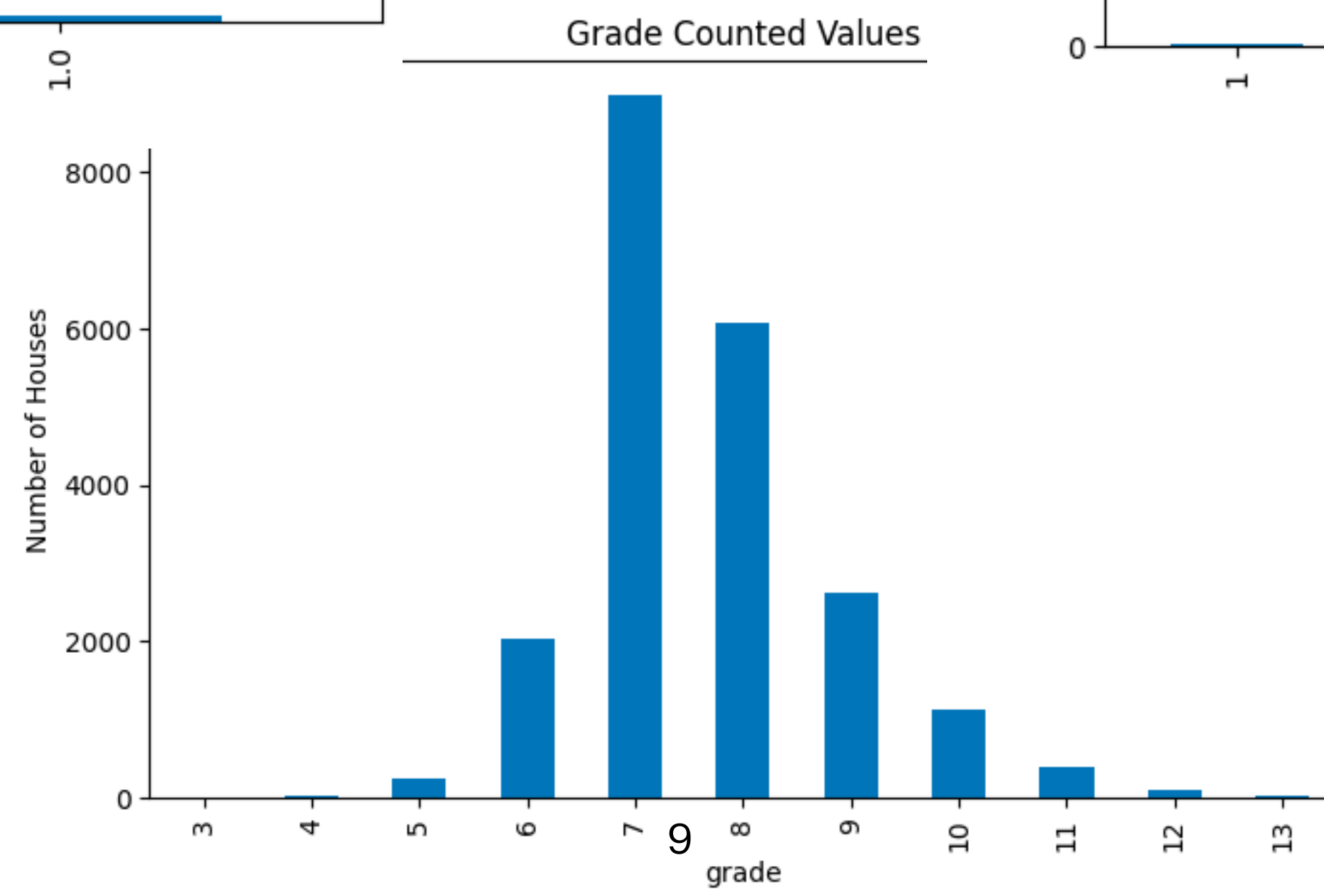
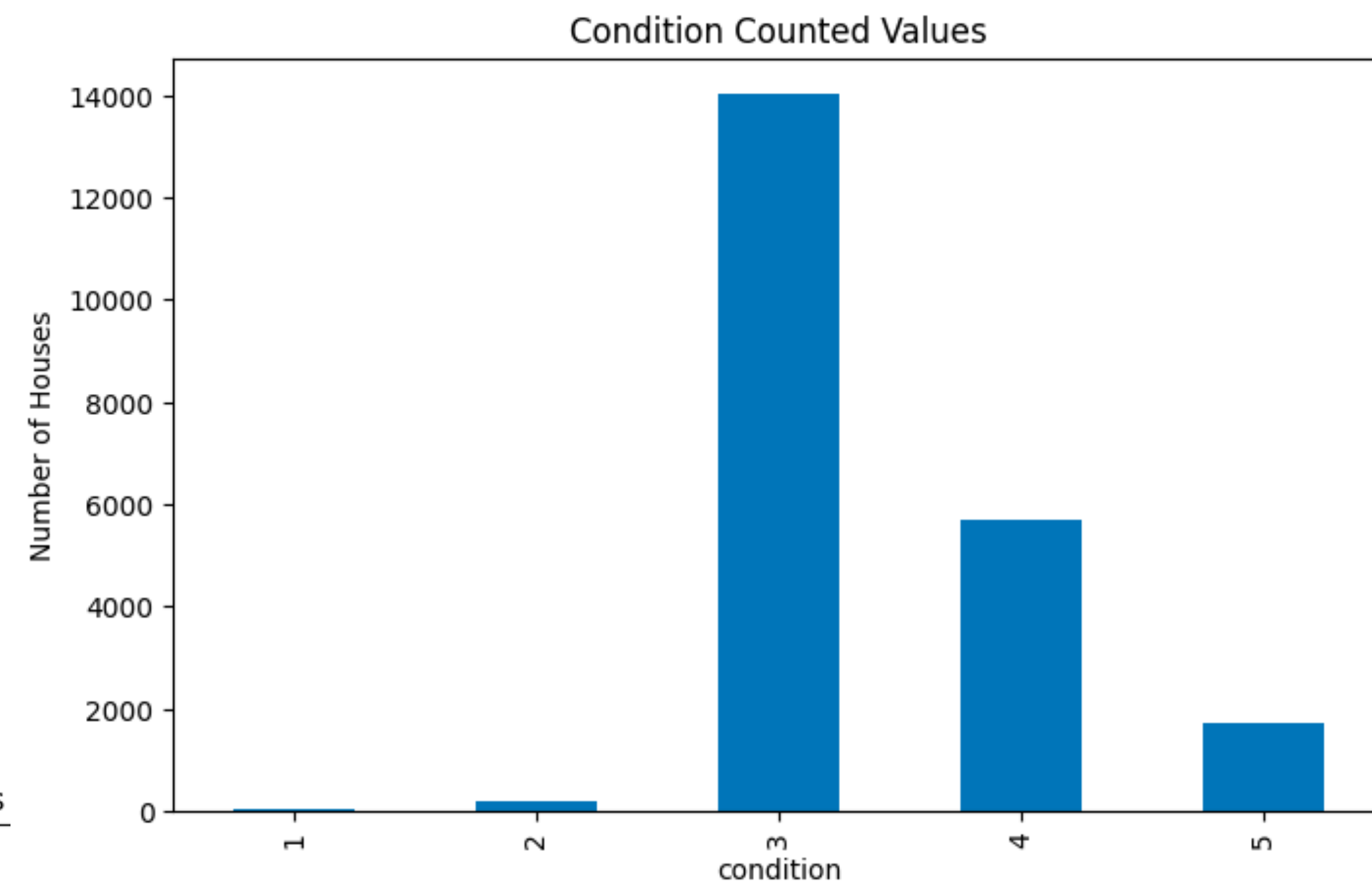
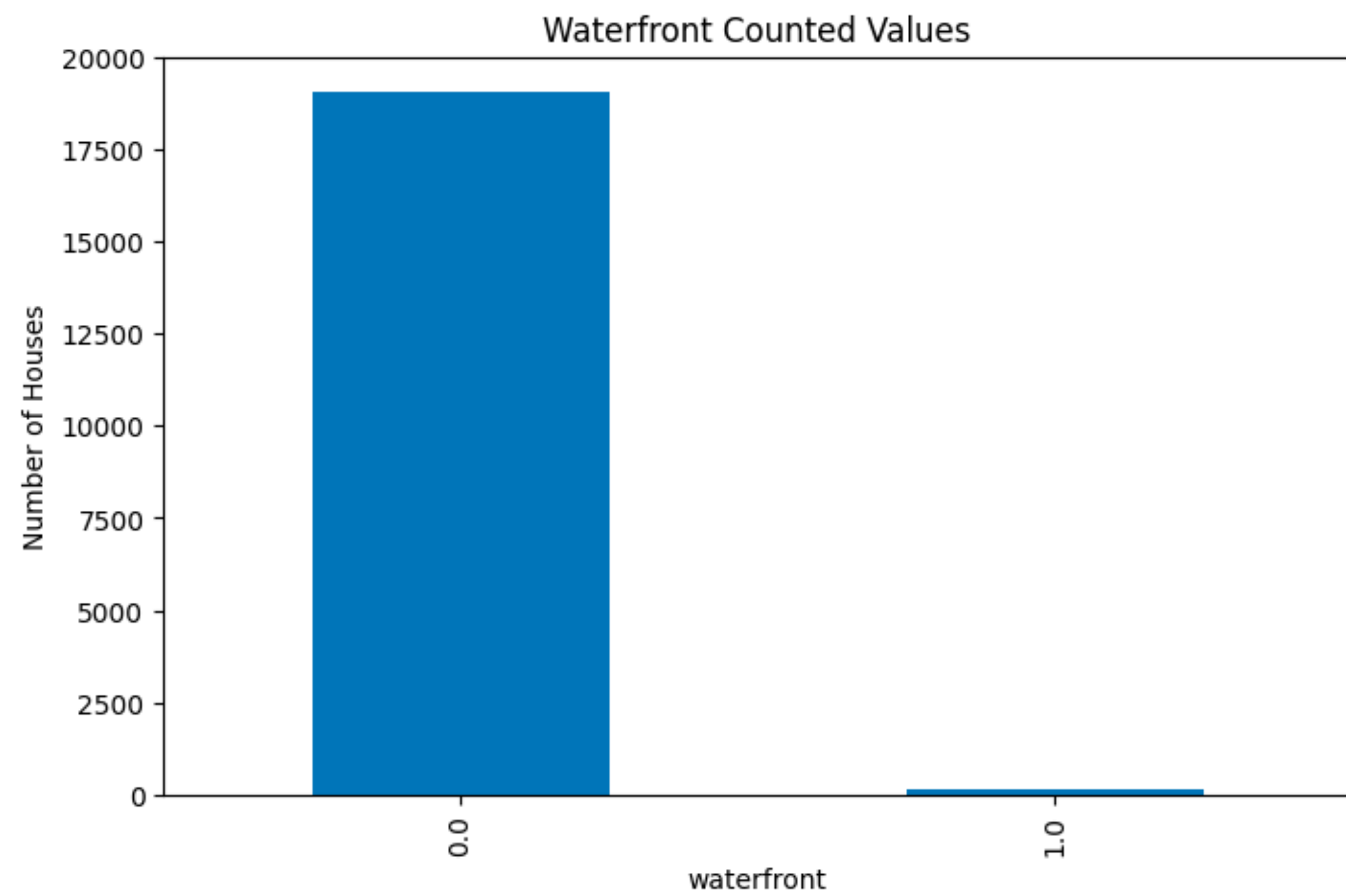
# Histogram Plot







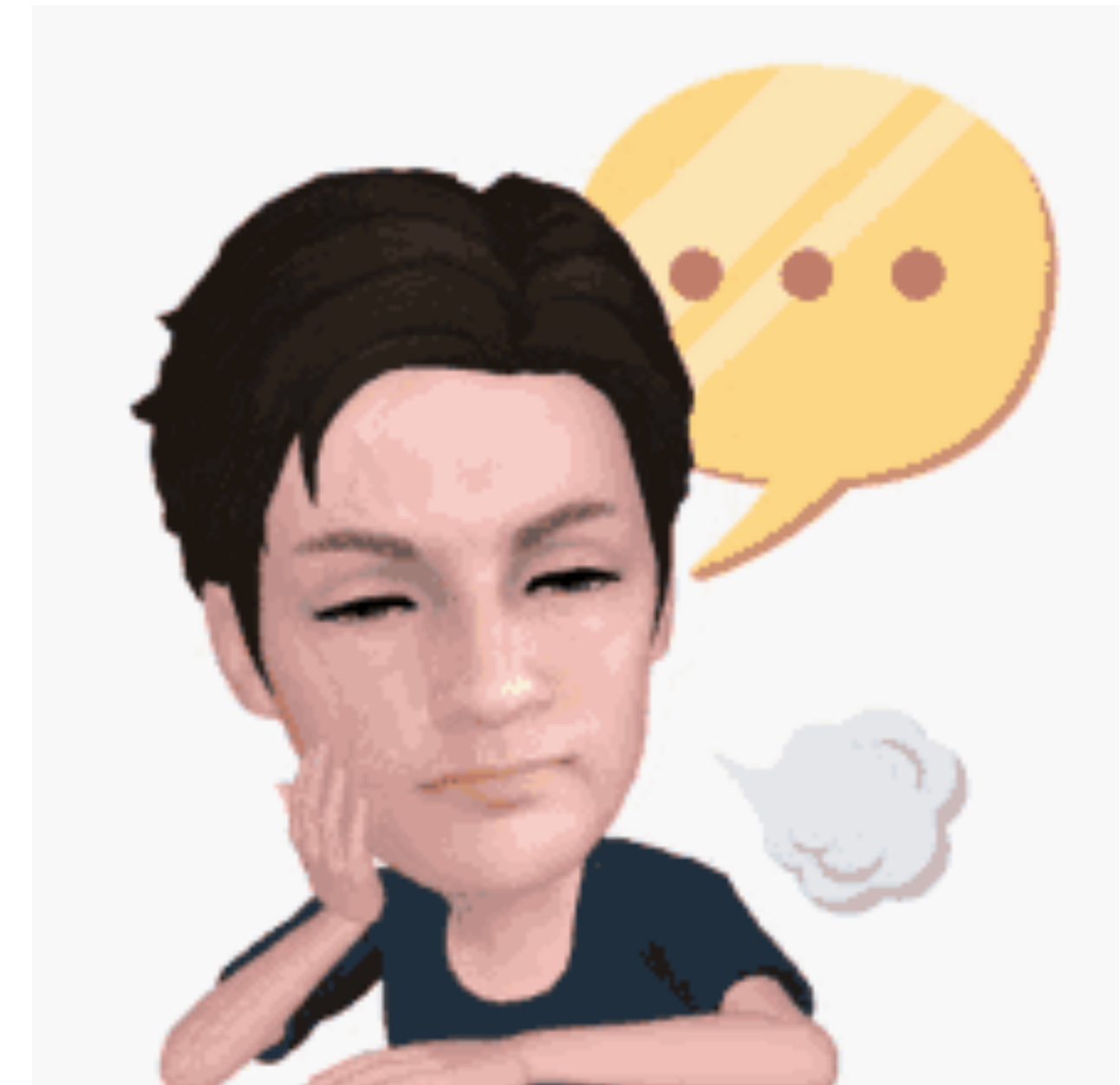




# Client

## Larry Sanders (Buyer)

- Characteristics;
  - Waterfront
  - Limited Budget
  - Nice
  - Isolated
  - Central Neighbourhood without kids
  - But got some of his own, just does not want his kids to play with others - because of germs :)



# Hypotheses

- The price of house with a waterfront is higher than others.
- When the number of bedrooms increases, then the price of the house increases.
- When the grade points increase, then the price of house increases.
- When condition point increases, then the price of house increases.

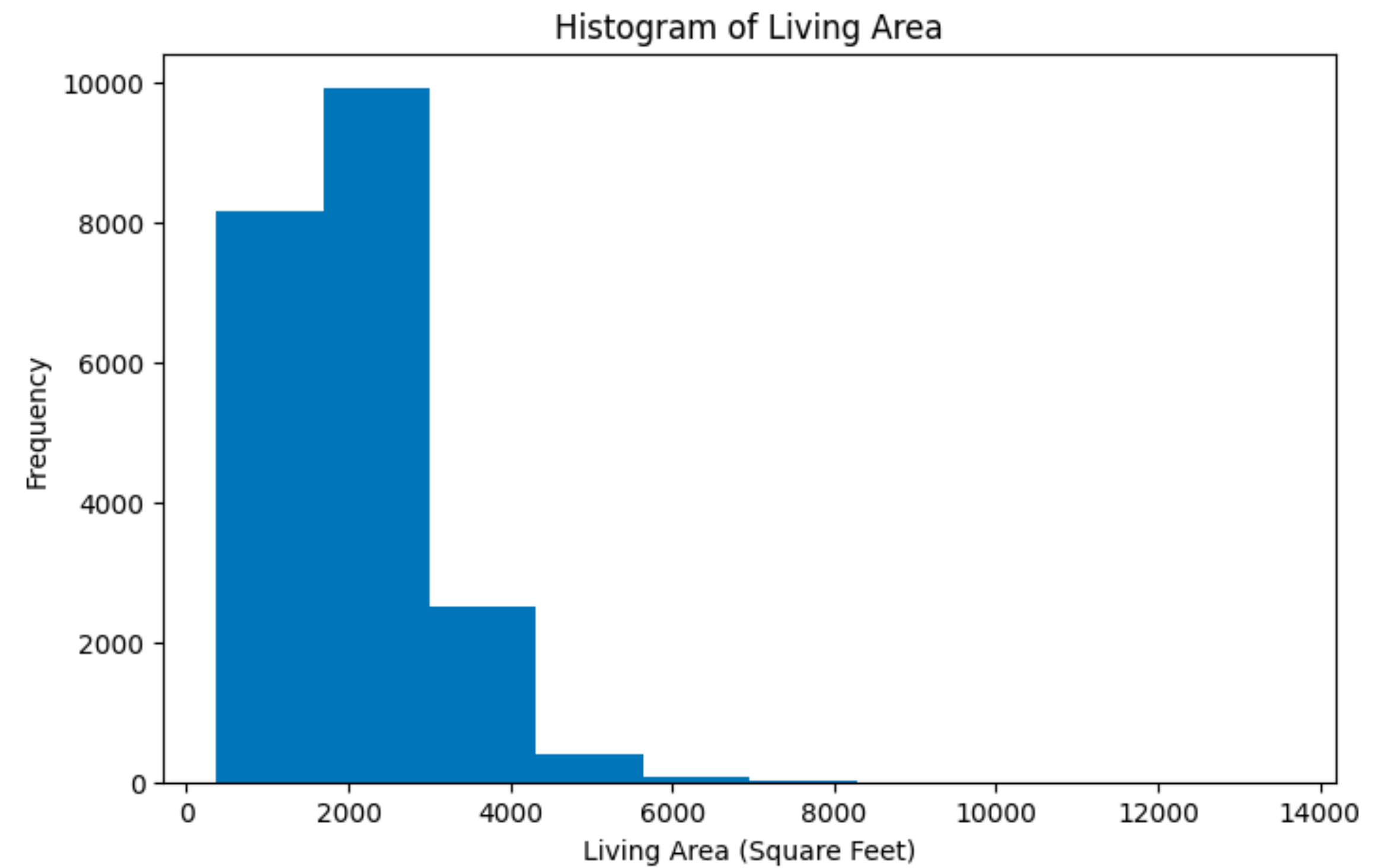
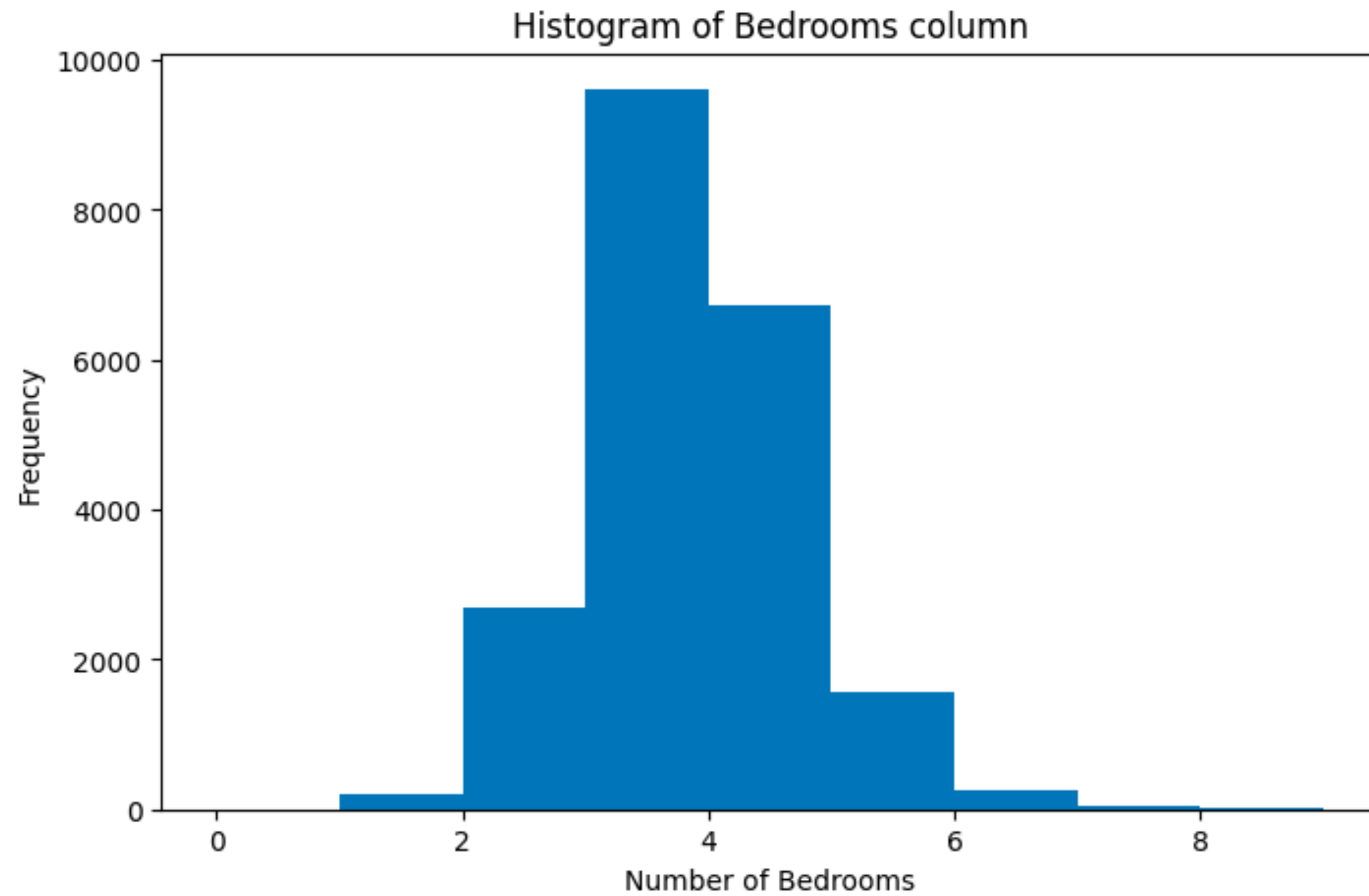
# Questions

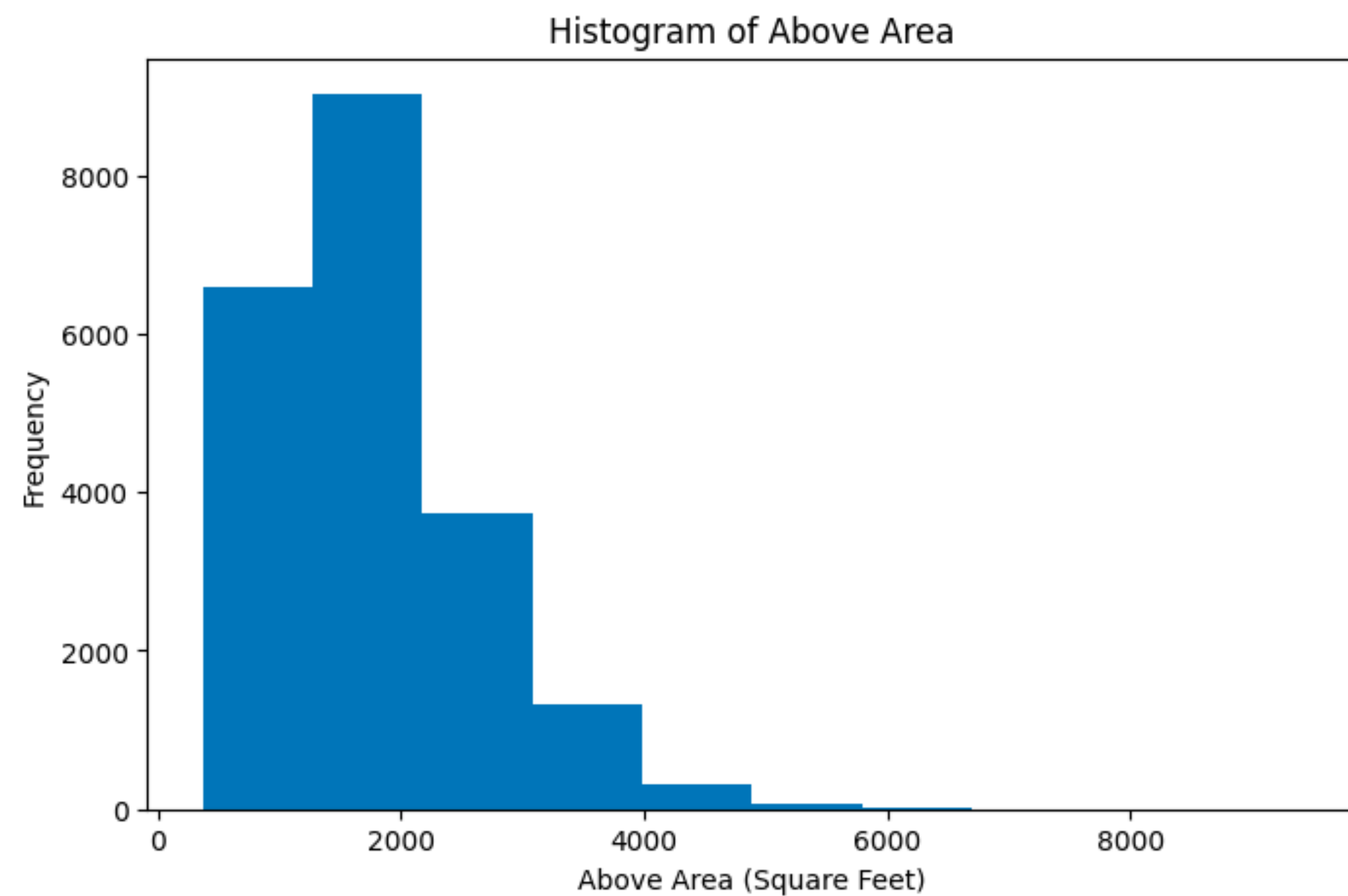
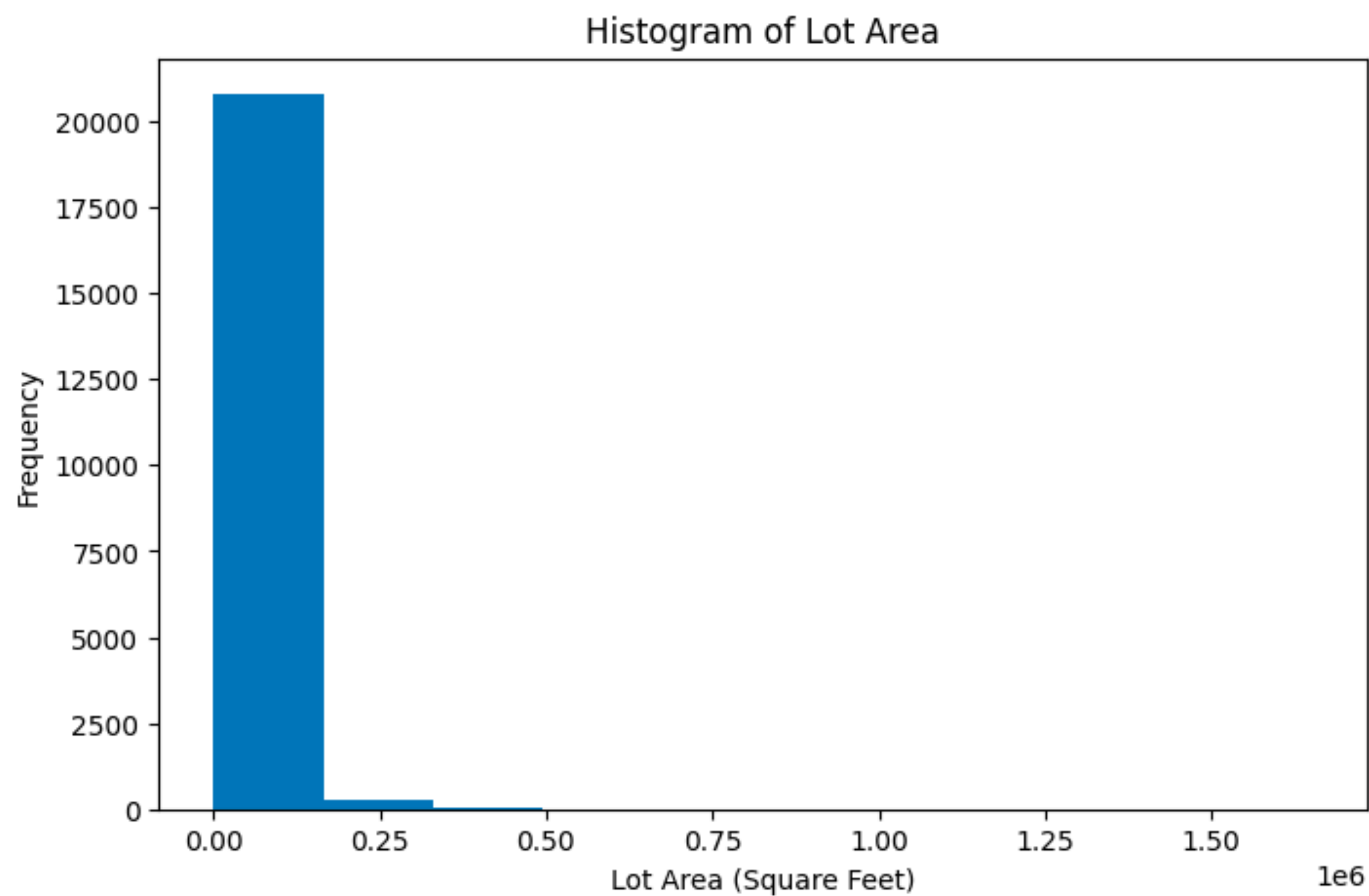
- Do houses with a waterfront cost more than those without?
- Does the price of house increase as the living area increases ?
- Does the price of house increase as the lot area increases ?
- Does the price of house increase as the area of Living15 increase ?
- Does the price of house increase as the area of Lot15 increase ?

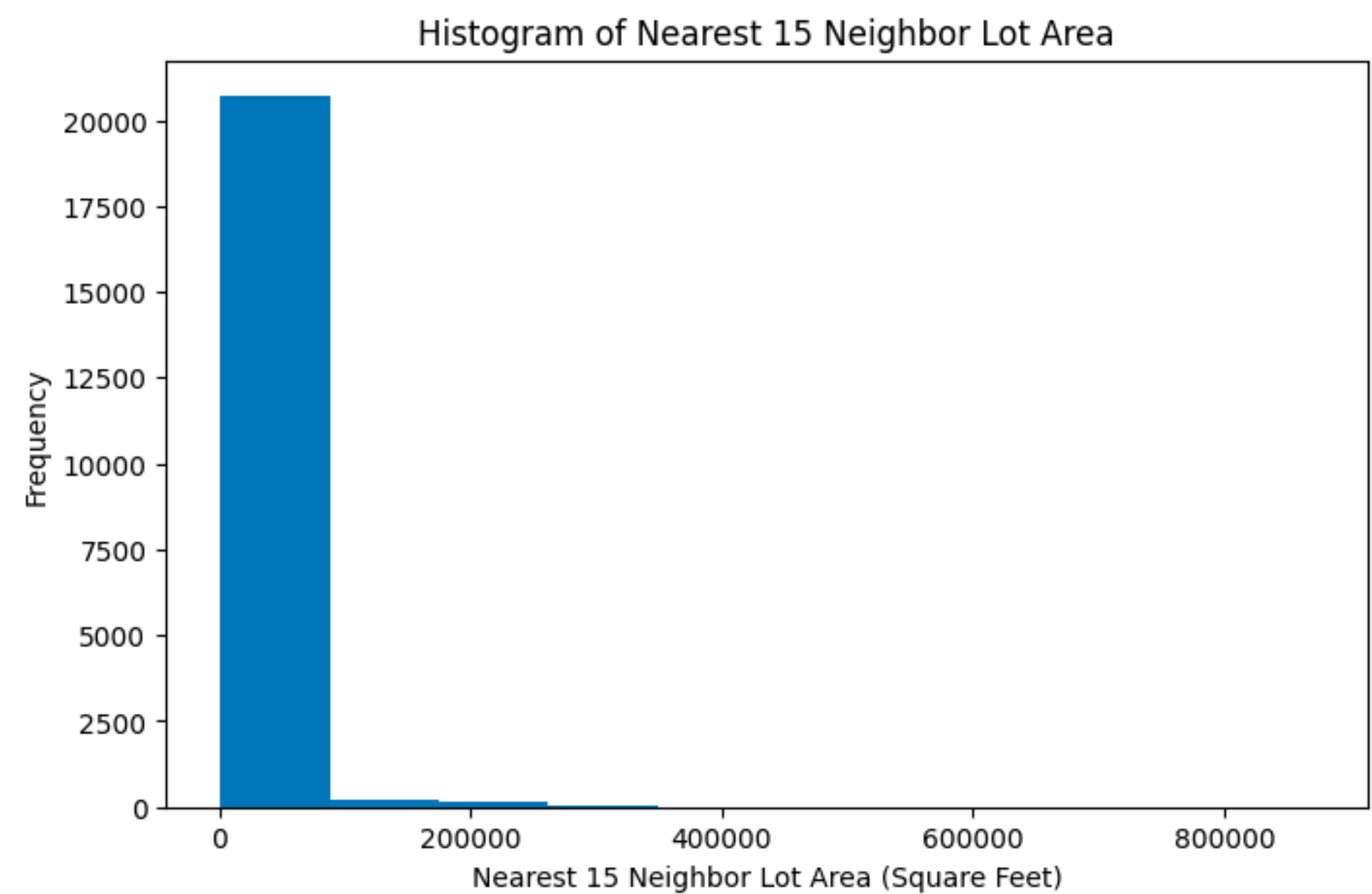
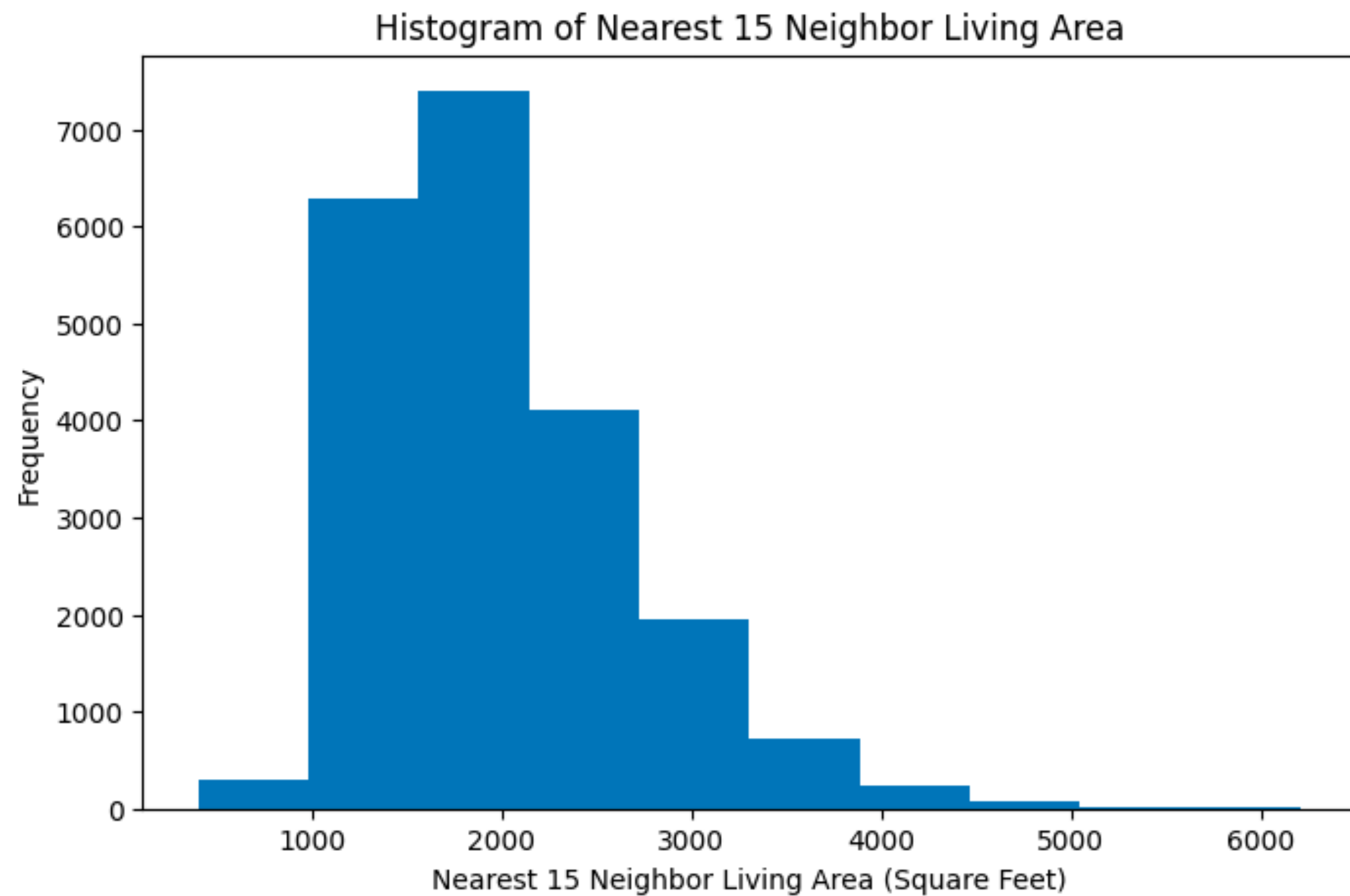
# Features

- **Bedrooms** : Number of bedrooms
- **sqft\_living** : Square Footage of Living area
- **sqft\_lot** : Square Footage of Lot Area
- **Waterfront** : House which has a view to a waterfront
- **Condition** : How good the condition is (Overall point)
- **Sqft\_living15** : The square footage of interior housing living space for the nearest 15 neighbours
- **Sqft\_lot15** : The square footage of the land lots of the nearest 15 neighbors
- **Grade** : overall grade given to the housing unit, based on King County grading system
- **Sqft\_above** : square footage of house apart from basement
- **Zipcode**
- **Lat** :Latitude coordinate
- **Long** : Latitude coordinate
- **Price**

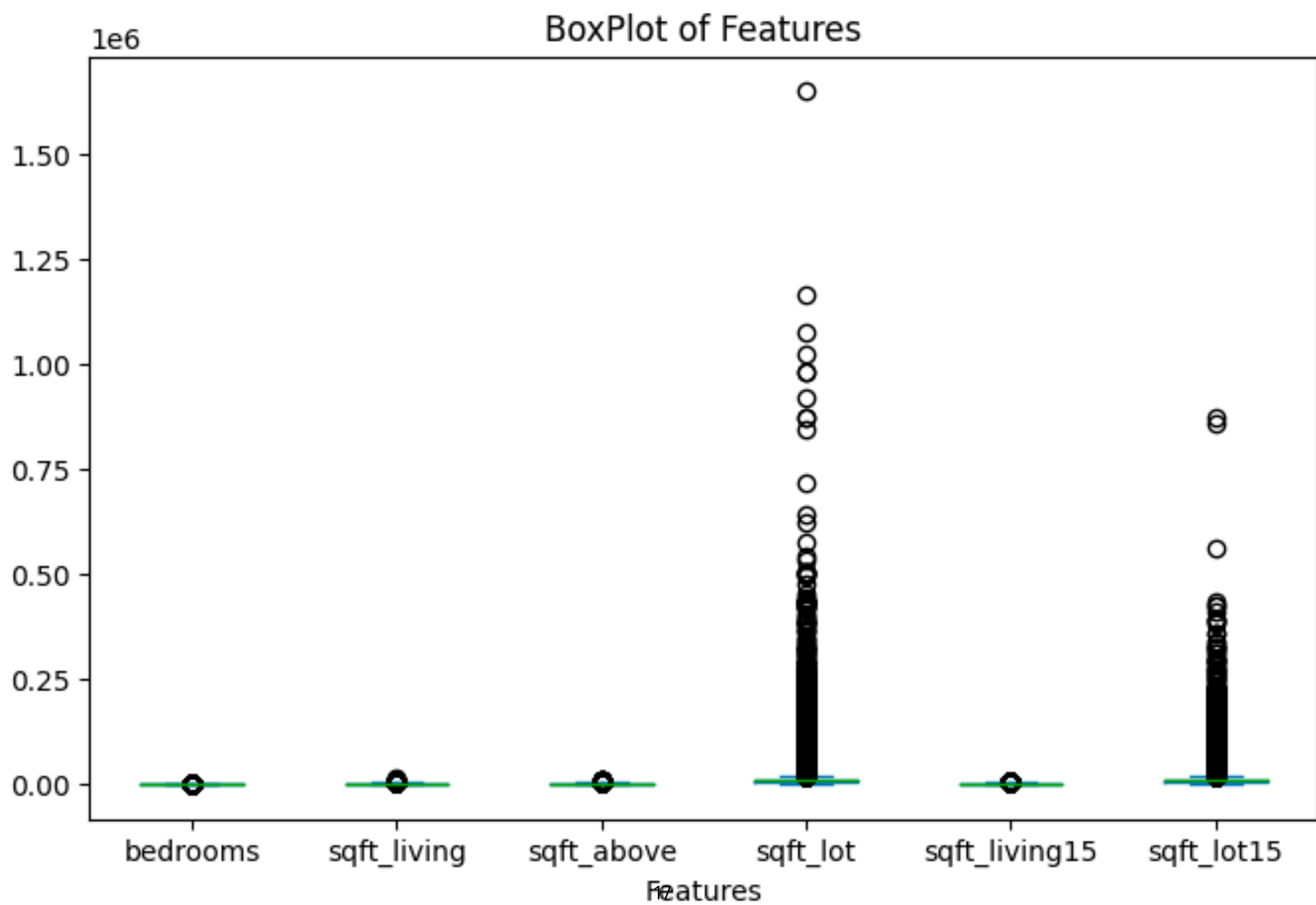
# Visualisation





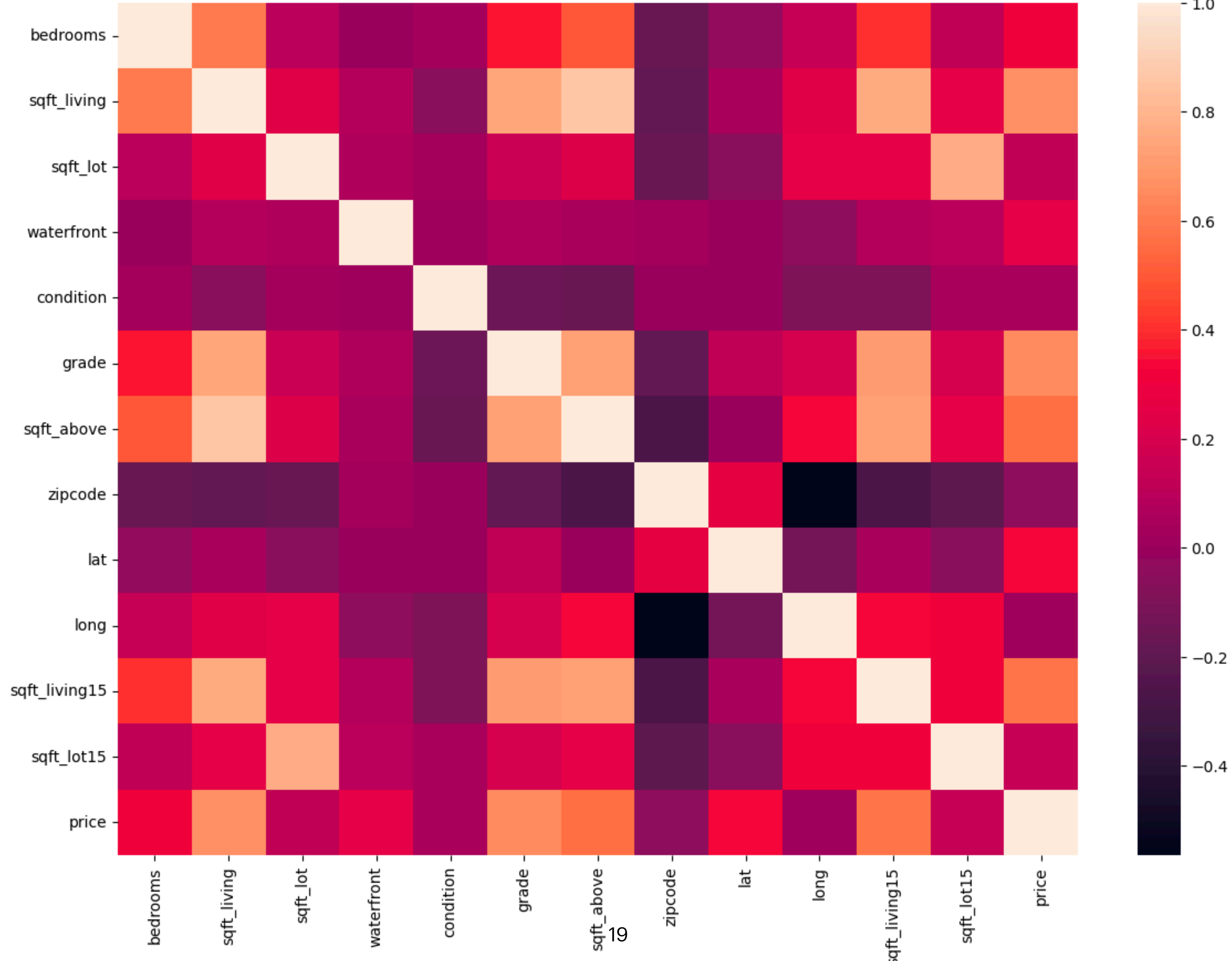


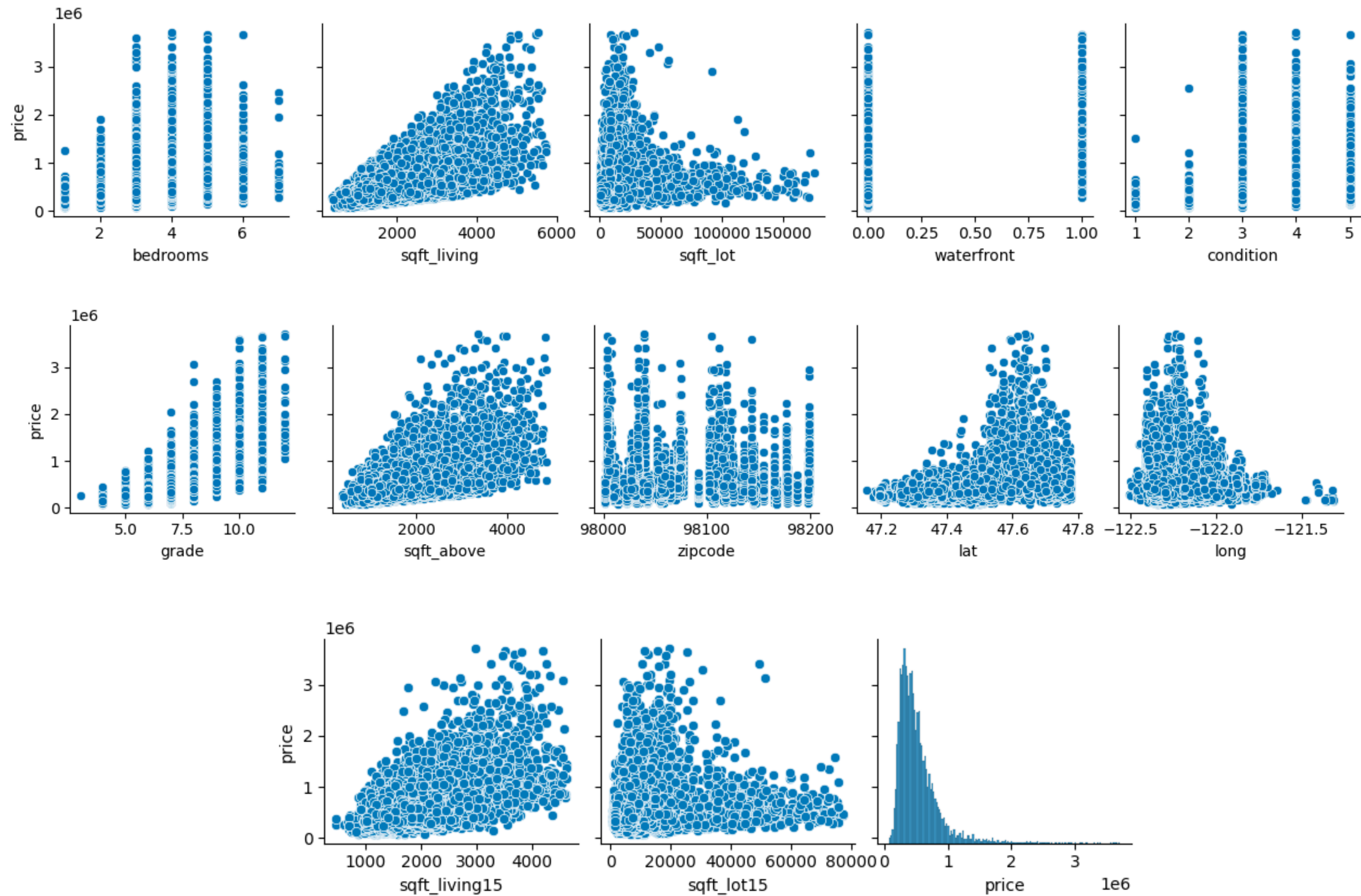




# Relationship in the Data

```
There is 4 strongly correlated values with Price:  
grade            0.645  
sqft_living      0.630  
sqft_living15    0.562  
sqft_above       0.524
```





# Hypotheses

- The price of house with a waterfront is higher than others. **(FALSE)**
- When the number of bedrooms increases, then the price of the house increases. **(FALSE)**
- When the grade points increase, then the price of house increases. **(TRUE)**
- When condition point increases, then the price of house increases. **(FALSE)**



# Questions

- Do houses with a waterfront cost more than those without?
  - No, there is no strong relationship between them. Houses that do not have a waterfront can be expensive.
- Does the price of house increase as the living area increases ?
  - Yes
- Does the price of house increase as the lot area increases ?
  - No, on the contrary, the price decreases as the lot area increases.
- Does the price of house increase as the area of Living15 increase ?
  - Yes
- Does the price of house increase as the area of Lot15 increase ?
  - Just like Lot area, the price decreases.

# Insights and Recommendations

- **Insights**
  - The living areas of the houses affect the price of houses. As living area increases, also the price increases.
  - The grade point given to the house affect the price of houses. Whichever house is given a high grade point , that house is more expensive.
  - As the area of the first floor of the houses increases (sqrt\_above), the price of houses also increase.

- Recommendations
  - Considering the costumer's characteristics;
    - I can suggest, of course, a house with waterfront. :)
    - Besides, I can recommend houses with higher than average grade points because he wants a 'Nice' house.
    - Since my client wants to live in isolation, I can recommend houses with larger lot area but smaller living area in his own house and neighbouring houses.
    - Since he has limited budget, I can recommend the cheapest houses among the houses that meet the above conditions.



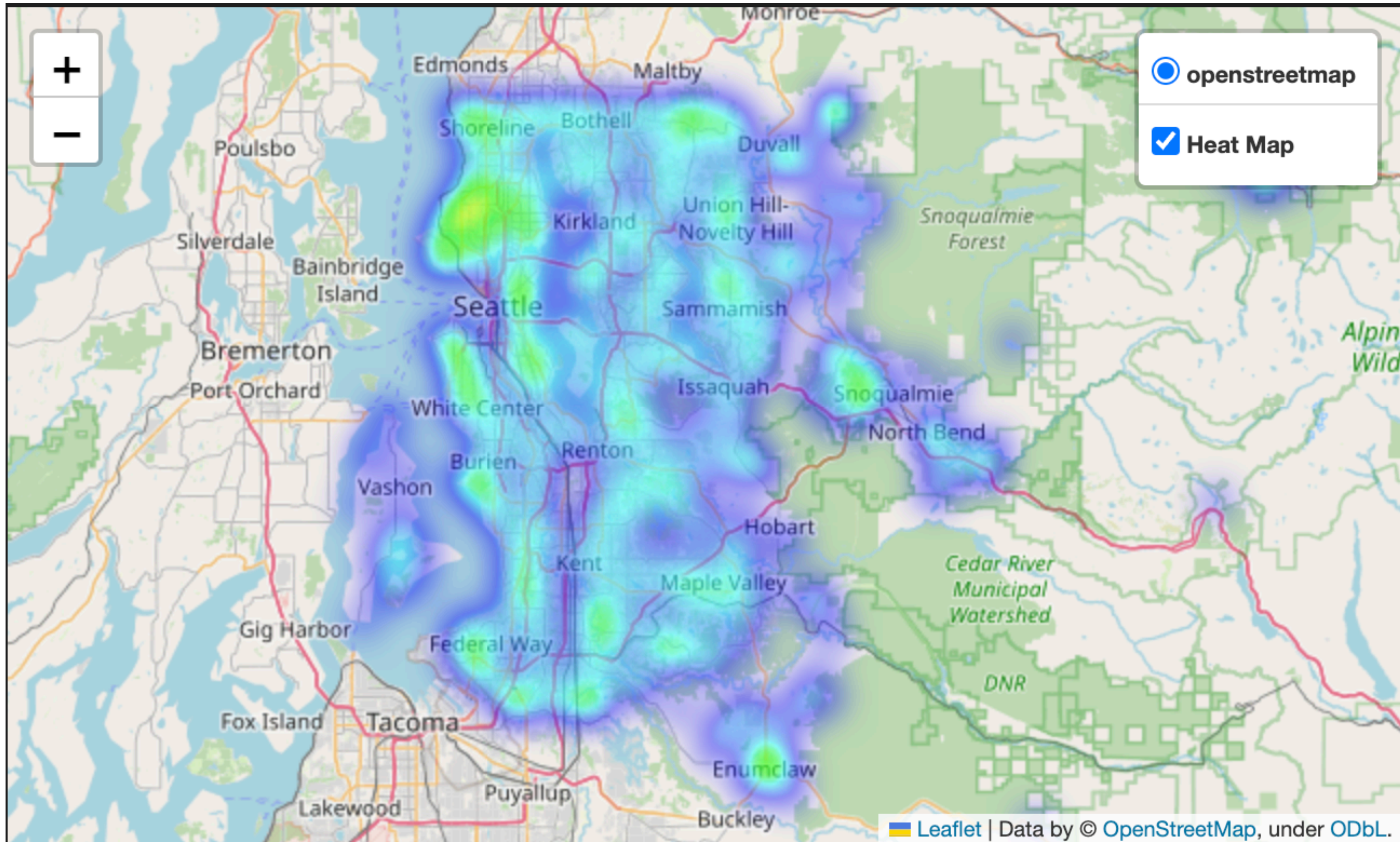
## House 1

bedrooms	3.000
sqft_living	1970.000
sqft_lot	20978.000
waterfront	1.000
condition	4.000
grade	9.000
sqft_above	1770.000
zipcode	98070.000
lat	47.384
long	-122.438
sqft_living15	2280.000
sqft_lot15	75396.000
price	705000.000

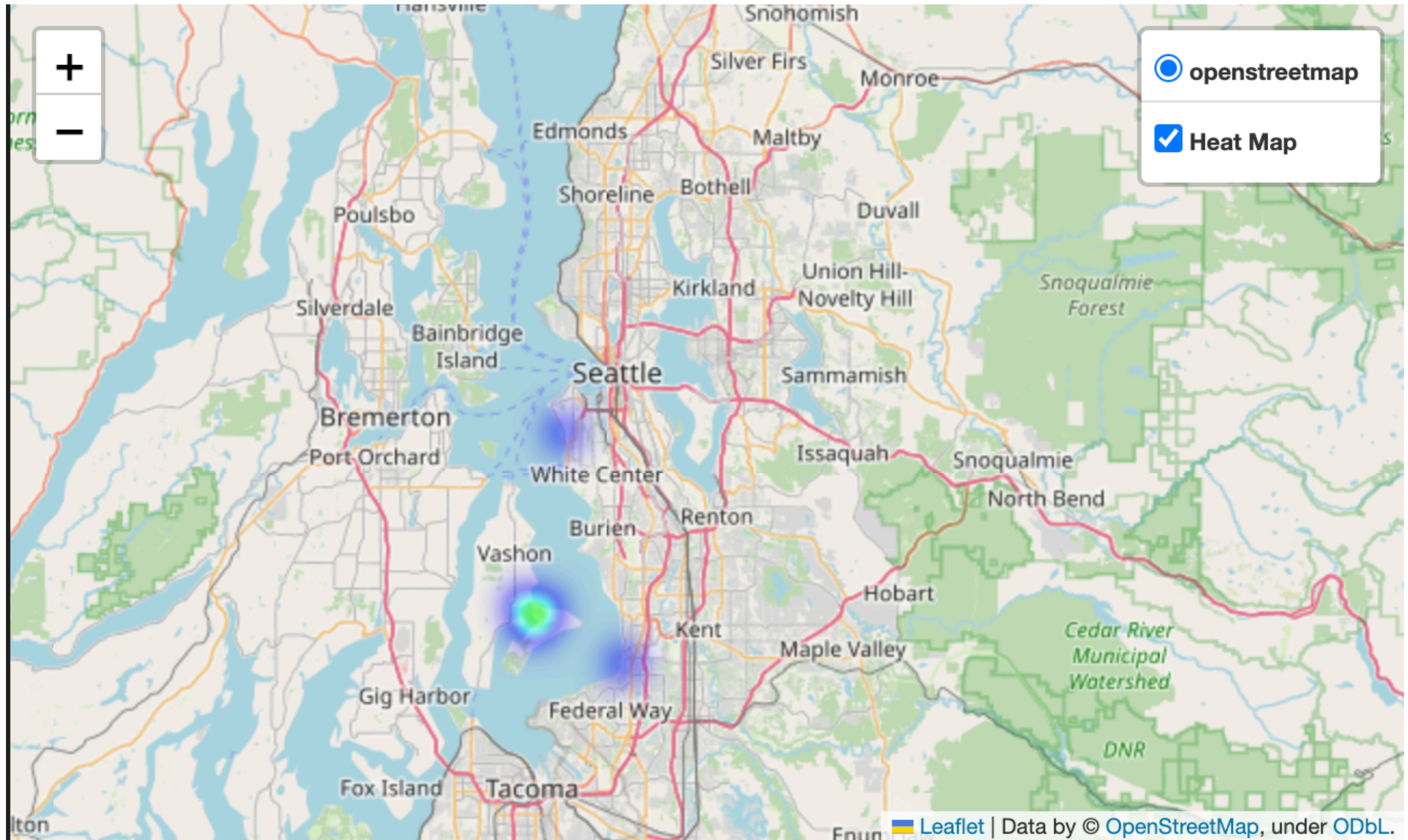
## House 2

bedrooms	3.000
sqft_living	2960.000
sqft_lot	69351.000
waterfront	1.000
condition	4.000
grade	9.000
sqft_above	2960.000
zipcode	98070.000
lat	47.400
long	-122.420
sqft_living15	2350.000
sqft_lot15	41433.000
price	750000.000











**Thank you for your listening!**