**Descriptive Report On Housing Prices**

University Canada West

BUSI 650 (Section- 20)- Business Analytics

Prepared by: Aditya Arte

Student ID: 2305505

Date: February 11, 2024

**Table of Contents**

[**Introduction 3**](#_vz5kf9oi72wr)

[**Background 4**](#_mmjkj3p43t3j)

[**Methodology 5**](#_vuksnbn6cam3)

[**Results and Discussion 6**](#_msdgkcj7zjee)

[Formatting and Missing Data 6](#_g2n02veic2tk)

[Outliers 6](#_bainadmg8kga)

[General Calculations 8](#_wx6veuej527z)

[Correlation and Covariance 10](#_d9o4sdpmhyas)

[**Conclusion 13**](#_b02wkeel7rgc)

[**References 14**](#_692pzjj56ung)

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

# 

# **Introduction**

Data analysis involves processing data from a dataset or sets and converting it to meaningful information (Taherdoost, 2022). In this paper, the author analyzes data on housing prices in the USA for the year 2014 and tries to figure out what influenced the prices. The information retrieved from this analysis helps us understand how different parameters of a house affect its price.

A sequential and detailed report of this analysis will help people from different fields, like real estate, investors, and the general public, looking to buy a house understand the market trend and if the money they spend will get them good returns. The analysis explains how the different parameters influence the fluctuation of the price of the houses. It also explains why houses with seemingly similar features have different costs. This analysis will help people to make an informed decision regarding their purchase.

In the next part of this paper, we will understand more about the topic of the research and analysis, how the author has carried out his analysis, what concepts the author used for analysis, the results he got from it, and what they entail.

# **Background**

Real estate is a type of real property tied to the land in a long-term way, including houses and buildings. Real estate is known to be one of the most critical sectors of the US economy. Real estate contributes to balancing a country's economy by boosting its people's income and providing space for businesses and job opportunities, especially for those in the construction industry (Yang, 2022).

The US housing prices have been increasing continuously in the past. The dataset used by the author for his study consists of data on residential real estate, one of the major categories of real estate. The author has tried to examine whether the increasing house price is a bubble or a continuous trend.

# **Methodology**

The author has started his analysis by formatting the data to make it more presentable and readable. He then checked the data for incorrect values or data which might cause the analysis to misinform. The author then checked for outliers, which again indicated inaccuracy in data, which was done with the help of functions and features of Excel. For the primary analysis, the author has calculated the values of the minimum, maximum, and average prices of houses. The author has also done statistical calculations highlighting the relationship between different parameters. The author has also supported his calculations and work with visualizations covering the chances of confusion caused by using complex statistical Excel functions. In the next part of the paper, the author explains his analysis and shares his visuals along with a descriptive explanation of the analysis result.

# **Results and Discussion**

## **Formatting and Missing Data**

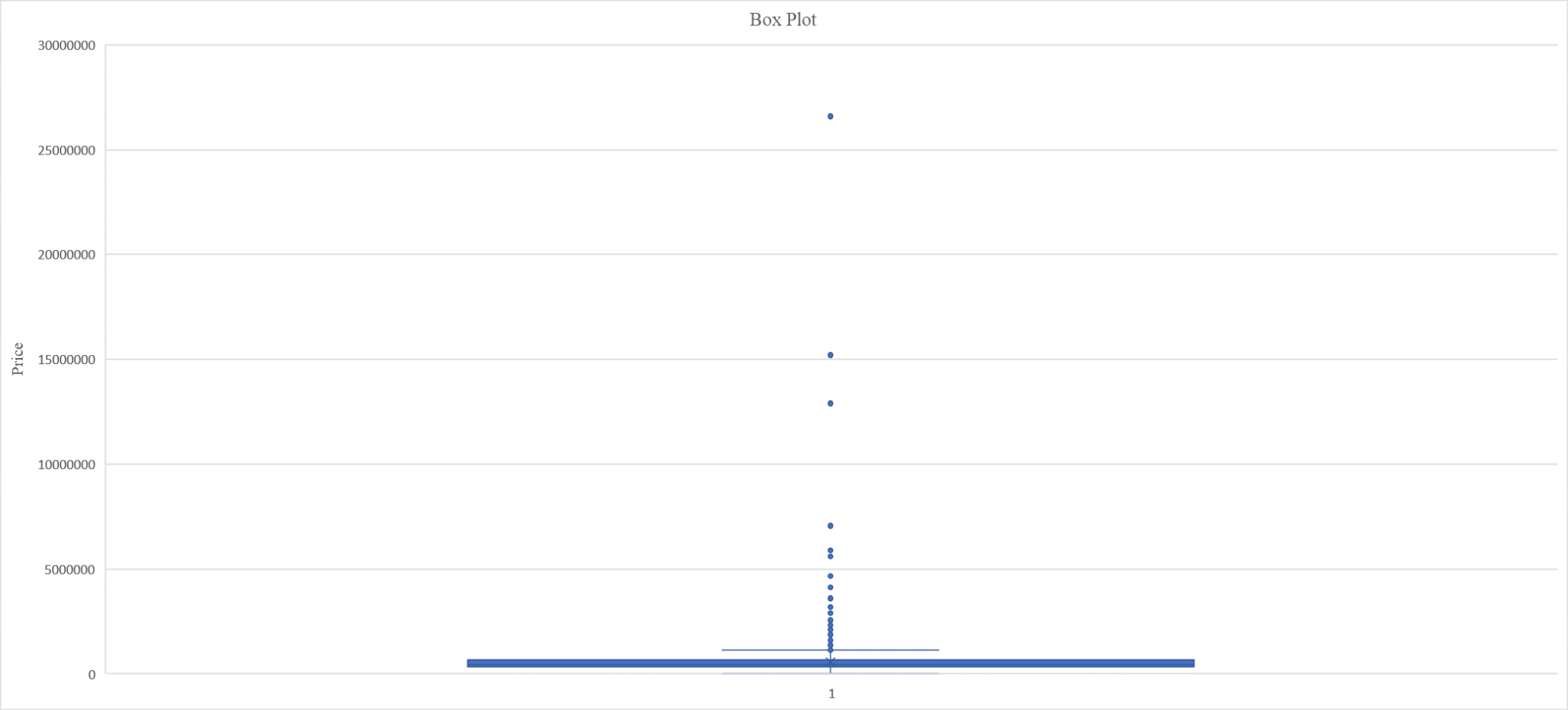
The author enclosed the dataset in a clearly defined table, highlighting the headers and adding filters. He then checked for missing values; in this case, the houses with a price of 0 dollars were filtered out.

## **Outliers**

In the next step, the author has determined the outliers, which are unnatural in form. These were found with the help of a box plot Figure 1. In this case, the outliers mainly were values on the high end. According to the author's findings, the outliers concerning price were the following values - 5600000, 7062500, 12899000, 15208000, 26590000, and 5600000. These outliers indicate that the prices of some houses were unnaturally inflated; this can be attributed to those properties being designer high-end properties or huge skyscrapers or just an entry error. These outlier value records have been marked by the author in the dataset. The same is the case with the lot area of some of the houses.

With the help of the box plot Figure 2, the author discovered some houses with unusually high lot areas. These houses have lots as huge as - 818592504.1 sq ft, which is probably an entry error.

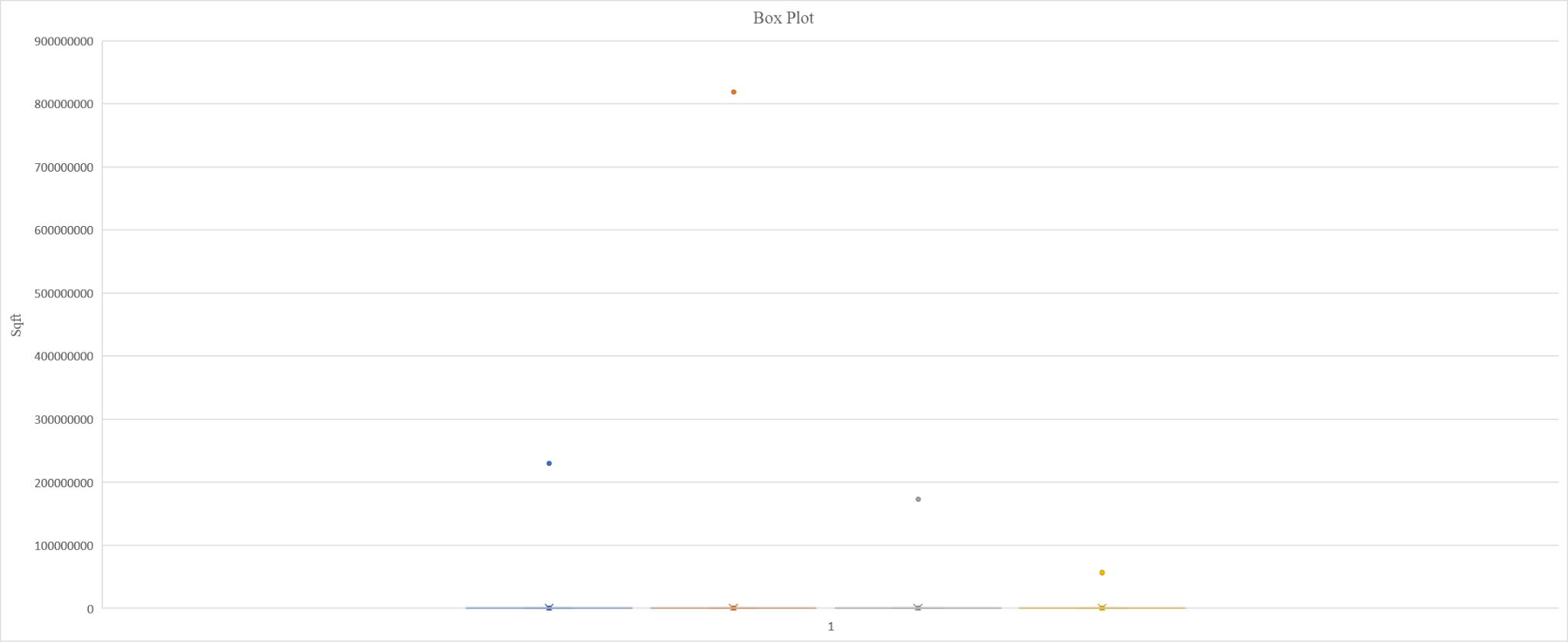
**Figure 1**

Price Box Plot

*Note.* From Aditya\_Arte\_report Excel file

**Figure 2**

Area in square feet box plot



*Note.* From Aditya\_Arte\_report Excel file

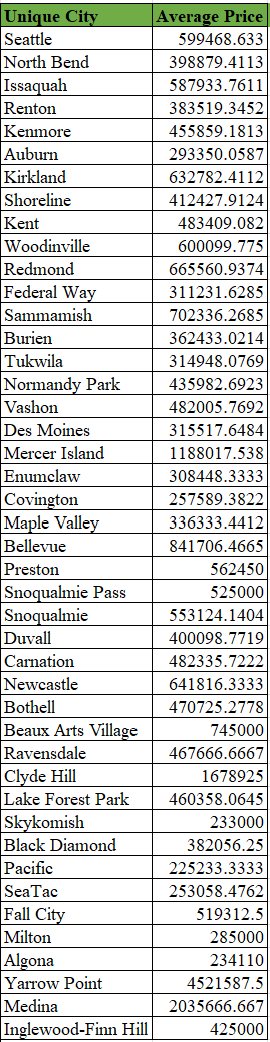
## **General Calculations**

Here, the author has calculated the minimum price of a house, which is 30,375, and the maximum is 26,590000. The author also calculated the average price of a house, which is 562907.6, which means that most houses cost around this value. The author also created a scatter plot of price vs living area; refer to Figure 3. This shows the relationship between the price and the living area. It can be observed that as the size of the living area increases, the price also increases. The author has also created 2 bar charts that show the effect of the house's condition on its price and the relation of the number of bedrooms in a house to its price in Figure 4 and Figure 5, respectively. In both cases, the bar charts indicate that the price also increases as the condition of the house or the number of bedrooms increases.

The author has also calculated the average house price per the different cities (see Table 1). This data can highlight the average cost of the house to the buyers and investors who want to buy in a particular city.

**Table 1**

Average Price of House per City



*Note.* From Aditya\_Arte\_report Excel file

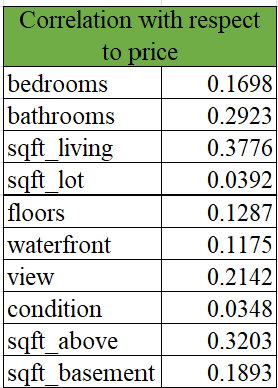
## **Correlation and Covariance**

The author has calculated the correlation between all the different parameters and the price in each case except the year of renovation; as the value of the parameters increases, the house cost also increases. Refer to Table 2 to check the level of correlation. As per the author, if the correlation value is greater, the greater that parameter is, the more it affects the price. This relationship is linearly proportional.

The data in Table 2 explains that if the number of bedrooms, bathrooms, and floors increases, the price also increases; in the author's study, the increase in the number of bathrooms has a more significant effect on price than the number of bedrooms and floors. The increase in the living area will have the most significant effect on the price, and the area of the lot will have the lowest.

**Table 2**

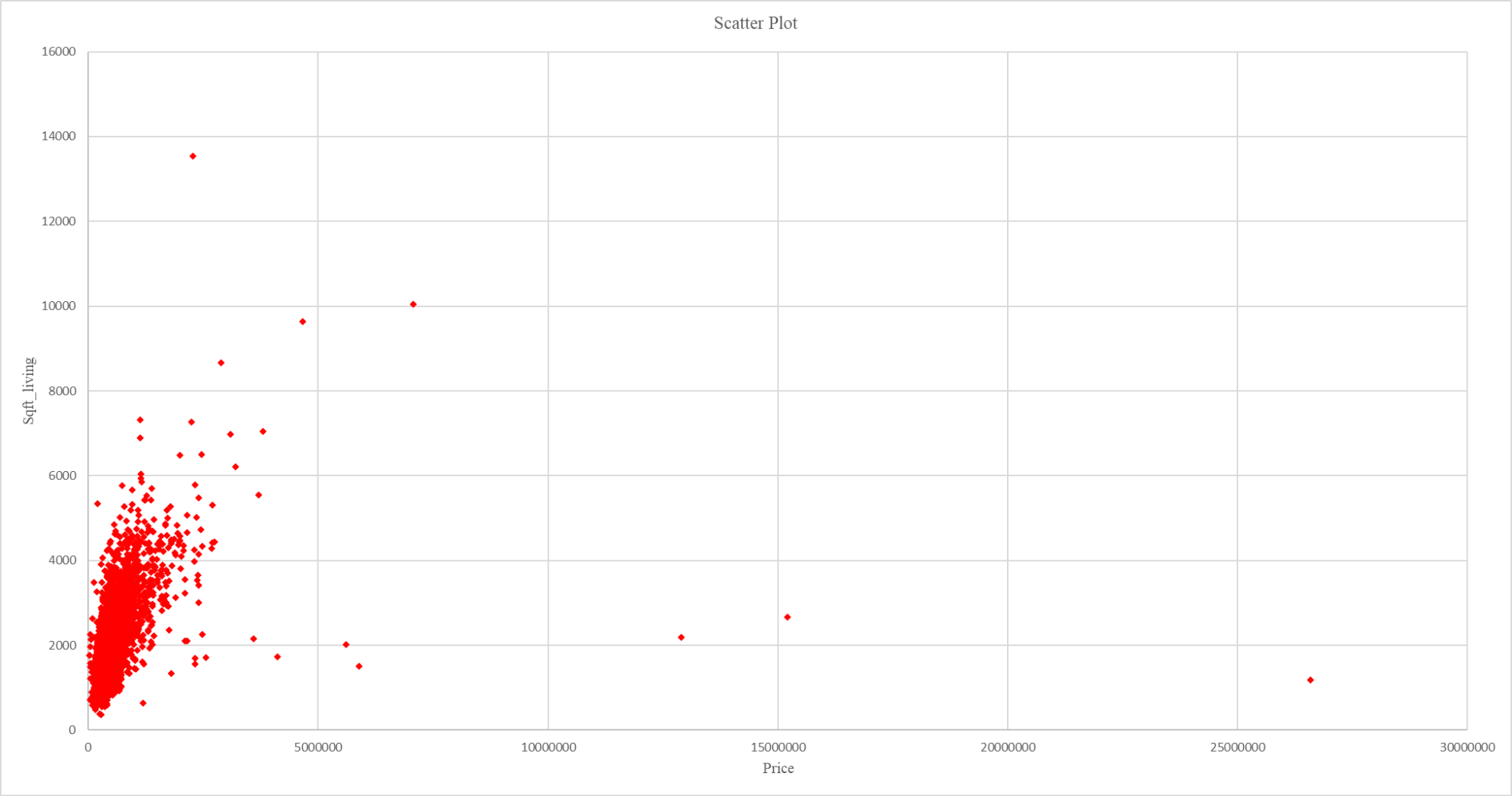
Correlation with respect to Price



*Note.* From Aditya\_Arte\_report Excel file

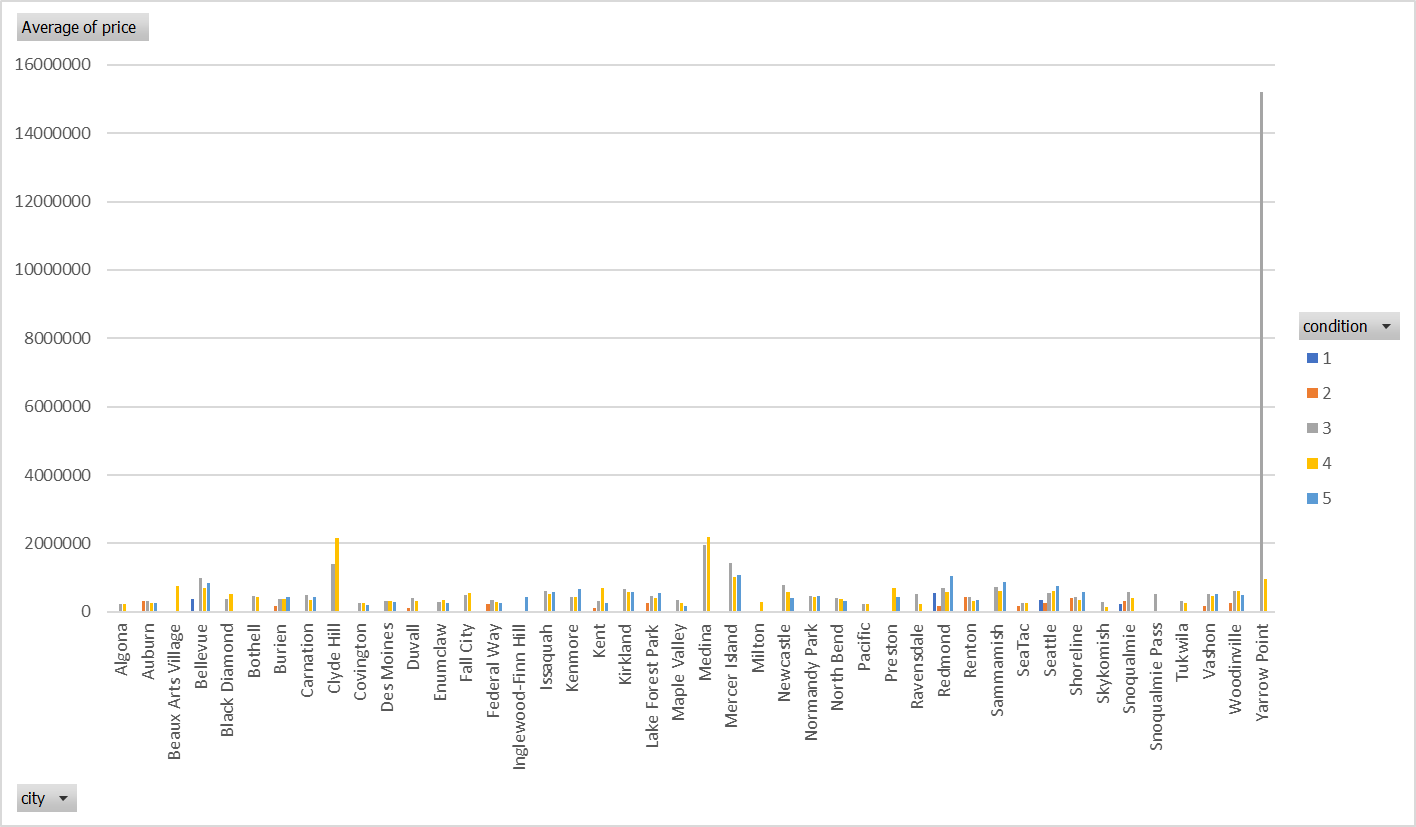
**Figure 3**

Price vs Living area scatter plot



*Note.* From Aditya\_Arte\_report Excel file

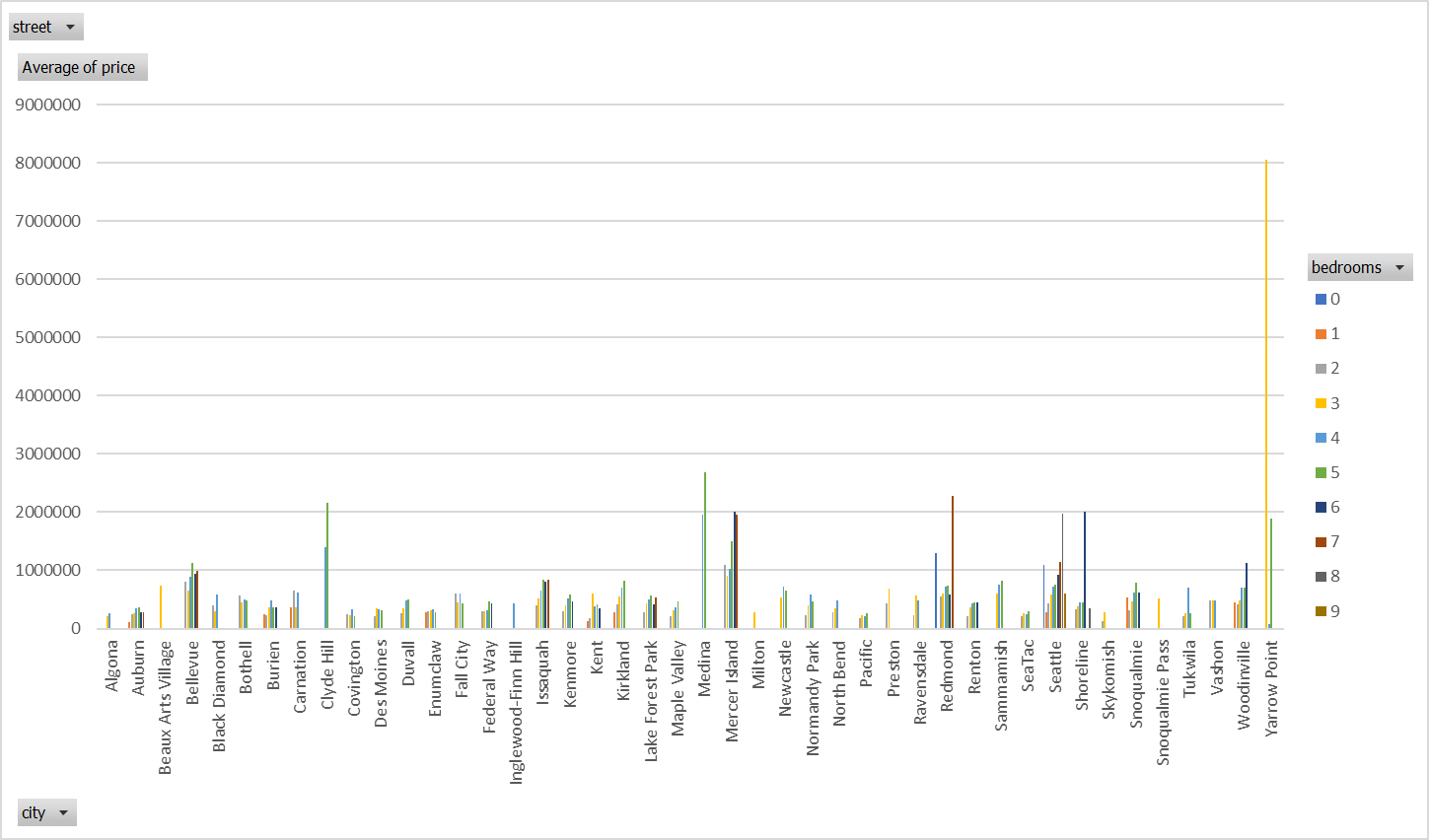
**Figure 4**

Average Price vs Condition of the house

*Note.* From Aditya\_Arte\_report Excel file

**Figure 5**

Average price vs the number of bedrooms in the house



*Note.* From Aditya\_Arte\_report Excel file

# **Conclusion**

Based on the author's analysis, the bigger and better the home, which comes with more bedrooms, bathrooms, a good view, and a waterfront, the more it will cost. It was observed that the prices went up drastically in August, so in the future, the author recommends that buyers time their purchases.

# **References**

Taherdoost, H. (2022, August 1). https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=4178680

Yang, L. (2022). *Research on U.S. Housing Prices and the Real Estate Industry*. Atlantis press. Research on U.S. Housing Prices and the Real Estate Industry