**Introduction:**

The report evaluates the types and sources of data used in the Spring Houses dataset. The dataset provides information about selling prices, number of bathrooms, square footage, and number of bedrooms of houses for sale. The report discusses the measurement scale used for the variables and the lack of information regarding the data collection method. The report then explores other possible information that can be inferred or calculated from the dataset.

**Evaluation of the types and sources of data used:**

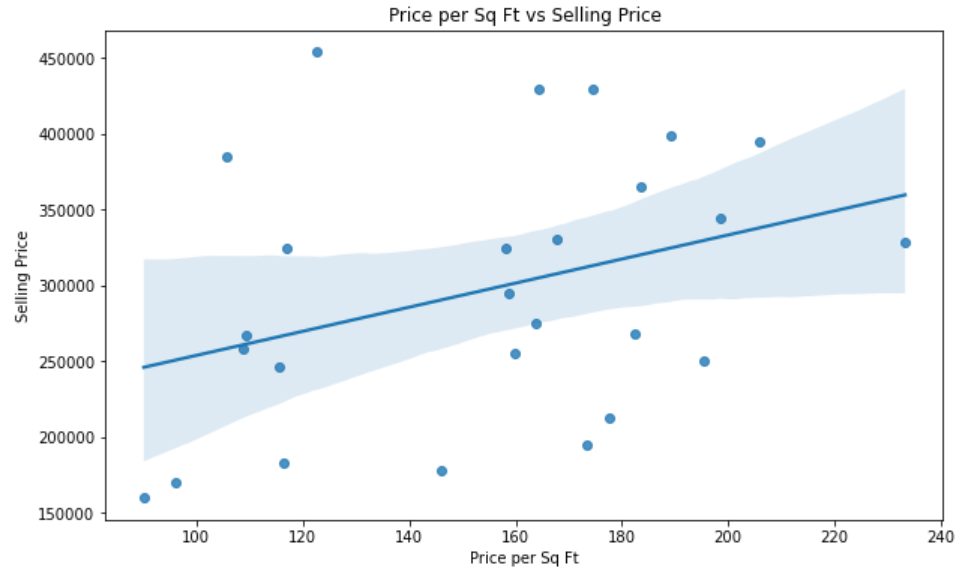
The dataset includes information on selling price, number of bathrooms, square footage, and number of bedrooms of houses for sale. The variables Selling Price, Baths, Sq ft, and Beds are all quantitative variables that are measured on a ratio scale. Ratio scale is a type of measurement scale used in statistics that has a fixed and meaningful zero point. In this scale, the numerical values of a variable represent not only the magnitude of the variable but also the absolute amount of the variable. Variables measured on a ratio scale are very informative and are often used in statistical analysis. In the context of the given dataset, the selling price, number of bathrooms, square footage, and number of bedrooms are all measured on a ratio scale, as they have a fixed and meaningful zero point and their ratios can be calculated and compared. It is unclear from the information provided what sources were used to collect the data.

Without more information on the methodology used to collect and analyze the data, it is difficult to evaluate the quality of the dataset. However, assuming the data was collected and analyzed using appropriate methods like random sampling, this dataset could be useful for predicting selling prices of houses based on their characteristics, as well as for identifying patterns and trends in the housing market.

**Other information can be retrieved from the dataset**

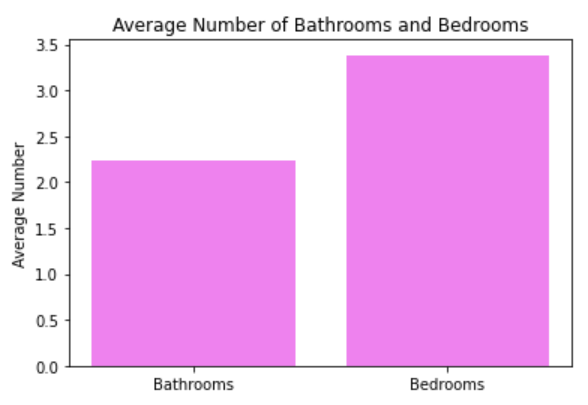
Based on the information provided in the dataset, several other pieces of information can be inferred or calculated above the analysis we did for presentation. For example:

1. Price per square foot: By dividing the selling price by the square footage, we can calculate the price per square foot of each house. This can be useful for comparing the relative values of different houses. we can create a scatter plot to visualize the relationship between the selling price and square footage of a house. This can help us understand how the selling price of a house is affected by its size.



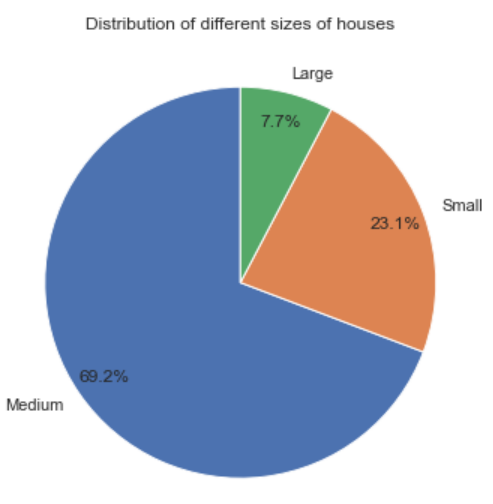
2. Number of Bathrooms and Bedrooms:

We can create a bar plot to visualize the distribution of the number of bathrooms and bedrooms in the dataset. This can help us understand the most common number of bathrooms and bedrooms in the houses for sale.



3. House catagory: Finally, we can create a new column that categorizes the houses based on their Sq Ft size area. For example, we can create three categories: small, medium, and large, based on the square footage of each house based on the below table. This information can be used to analyze the distribution of houses across different size categories and to compare the average selling price of houses in each category.

|  |  |
| --- | --- |
| Sq Ft range | Size category |
| 0 - 1500 | Small |
| 1501 - 3000 | Medium |
| 3001 - 5000 | Large |



4. Predictive modelling: We can build predictive models to predict the selling price of a house based on the other variables in the dataset. For example, we can use linear regression to predict the selling price of a house based on its square footage, number of bedrooms, and number of bathrooms.

**Different limitations of the data set:**

While this dataset provides useful information about houses but there are several limitations that should be considered when using the data:

1. Limited scope: The dataset only includes information on houses sold by certain type of houses. Therefore, the findings based on this dataset may not be generalizable to the entire housing market.

2. Missing data: The dataset is relatively small and may not provide a complete picture of the housing market. Moreover, there may be missing data or errors in the dataset, which can affect the accuracy of any analysis conducted.

3. Lack of context: The dataset provides only basic information about each house, such as the number of bedrooms, bathrooms, and square footage. Additional context, such as the age of the house, the location, and any special features, may be necessary to fully understand the value of the house.

4. No information on sales history: The dataset only provides information on the current selling prices of the houses. It does not provide any information on the sales history of the houses or the reasons why they were sold, which can be crucial in understanding the housing market.

5. Bias: The dataset may suffer from bias, such as selection bias or measurement bias, which can impact the validity of any findings based on the data.

Therefore, it is important to consider these limitations when using the dataset and to interpret the findings based on the context in which the data was collected.

**Conclusion:**

The Spring Houses dataset includes important information on houses for sale that can be useful in predicting selling prices and identifying patterns and trends in the housing market. However, the dataset has limitations such as a limited scope, missing data, and lack of context that must be considered when using the data. Despite these limitations, the report shows that additional information can be retrieved from the dataset using different combinations of variables, including the creation of new variables such as price per square foot and house categories based on square footage. Overall, the Spring Houses dataset is a valuable resource for analysis and prediction of the housing market.