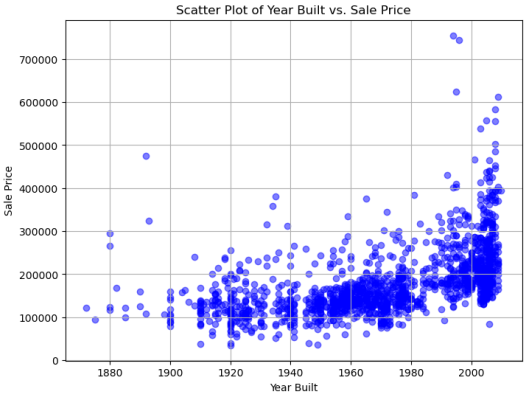
Summary of EDA on House Price Dataset

1. **Determine if the year of built affects the sale price of the house.**

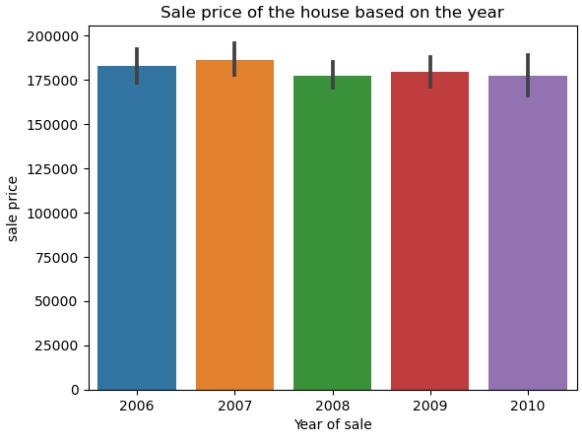



```
print("Degrees of freedom:", dof)
```

```
Chi-square statistic: 411.88818868825433
P-value: 5.358640559461605e-84
Degrees of freedom: 8
```

- We can observe that the houses which were built between the years 1880-1980 have pretty much the same rates with few outliers but for the houses built after 1980 the prices have gone up exponentially.
- The chi-square analysis performed on this hypothesis is done by creating bins for both the variables and we can observe that the chi-square value is very high and the P-value is near zero so this shows that both variables are strongly dependent on each other.

2. **Determine if the sale price has increased based on the year of sale.**

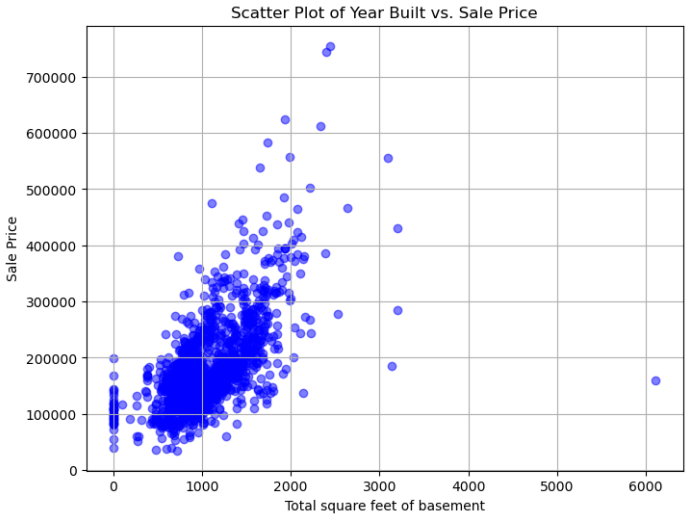


- The above plot shows that the sale price of houses has remained almost the same irrespective of the year.

3. **Determine if the street type affects the price of the house.**

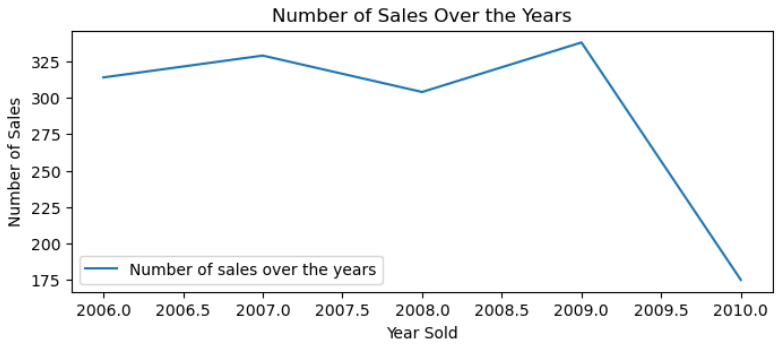
- The above graph shows that sale price is higher for street type “Pave” when compared with “Grvl”.

4. **Determine if there is any relation between the total square feet of basement area and the price of the house**



- The above scatter plot shows a strong increase in price with an increase in total square feet of basement with few outliers.

5. **Determine if the number of sales has varied over the years and which year has the highest sales.**



- There is no significant increase or decrease in the count sales from the year 2006 to 2009 but after 2009 the counts have significantly decreased.
- Year 2009 has the greatest number of sales with 338.