**Final Submission Report:**

**Introduction:**

* Provide a brief overview of the project and its elements, objectives, and some various parameters to showcase and display customer beneficiary, market trends, useful amenities and Sale price.
* Here it is some more analysis and parameters to understand and predict our model to have better efficiency and insight from Our Dataset.

1. **Feature Engineering and Size Impact:**

* **Feature Engineering:** The process of transforming raw data into a format that enhances the performance of predictive models. It involves creating, selecting, and modifying features to extract valuable information for the model.
* **Size Impact:** The consideration of how the scale or magnitude of features influences the performance of a machine learning model. It involves evaluating and addressing the potential bias caused by features with different scales.

1. **Market Trends and Historical Pricing:**

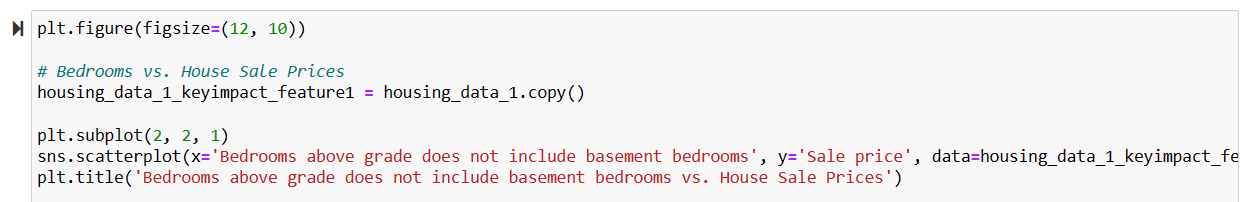
* Market trends and historical pricing are two important concepts in the analysis of financial markets.
* Market trends are the general direction or movement of prices over time, which can be upward, downward, or sideways.
* Historical pricing is the record of past prices of stocks, indices, commodities, currencies, or other assets, which can be used to identify patterns, trends, cycles, or anomalies.

1. **Customer Preferences and Amenities:** Customer preferences and amenities are two important concepts in marketing and customer service.

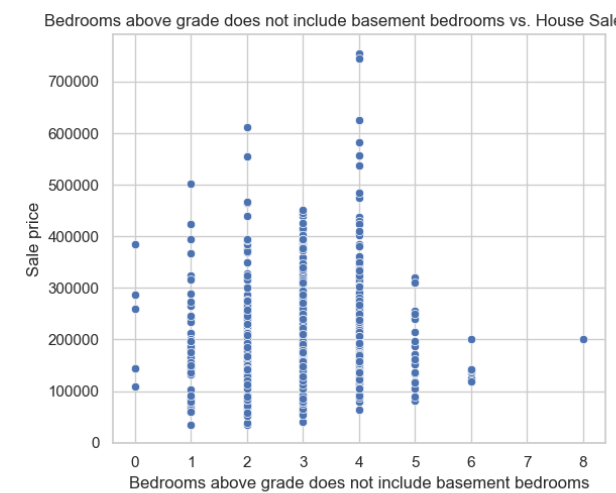
* Customer preferences are the subjective factors that influence the choices that consumers make when buying products and services. They include personal tastes, likes and dislikes, and predispositions of individual consumers.
* Customer preferences define the demand in a given market, and by extension, they also govern what suppliers will produce to meet this demand in the market.
* Amenities are the extra benefits or features that a product or service offers to customers, in addition to its core function.
* Amenities can enhance the customer experience and satisfaction, and differentiate a product or service from its competitors.

**Feature Engineering and Size impact:**





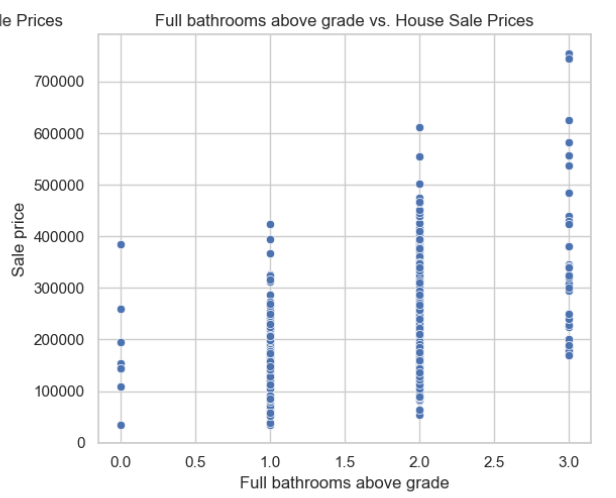
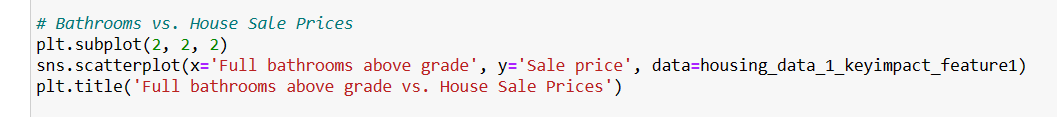
* **Insight Between House Sale price and Bedrooms above grade does not include basement bathrooms:**



**Bedroom Distribution:**Houses with 3 to 4 bedrooms above grade are more common, as indicated by a concentration of data points in these categories.Houses with more than 5 bedrooms above grade are less common, as fewer data points are observed in those configurations.

**Sale Price Variation:**There is significant variability in sale prices within each bedroom category.The graph does not show a clear trend that an increase in the number of bedrooms directly leads to a higher sale price.

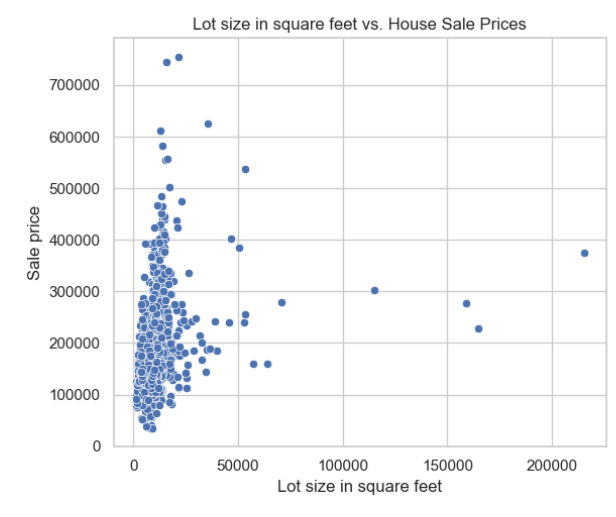
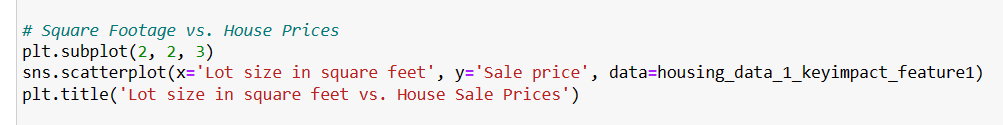
* **Insight Between Sale price and Full bathrooms above grade:**



As the number of full bathrooms above grade increases, there is a general upward trend in house sale prices. Houses with more bathrooms tend to command higher prices.

Data points cluster around whole numbers on the x-axis (0, 1, 2, and 3 bathrooms), indicating that most houses have a specific whole number of bathrooms.

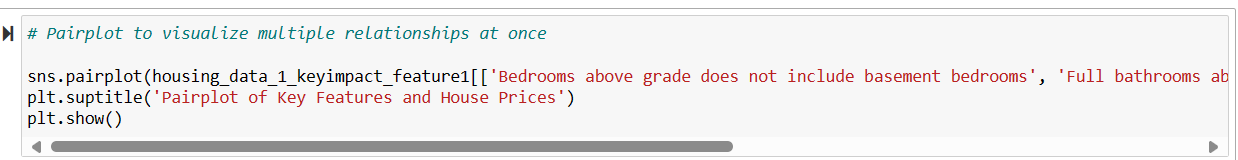
**Insight Between Sale price and Lot Size in square feet:**

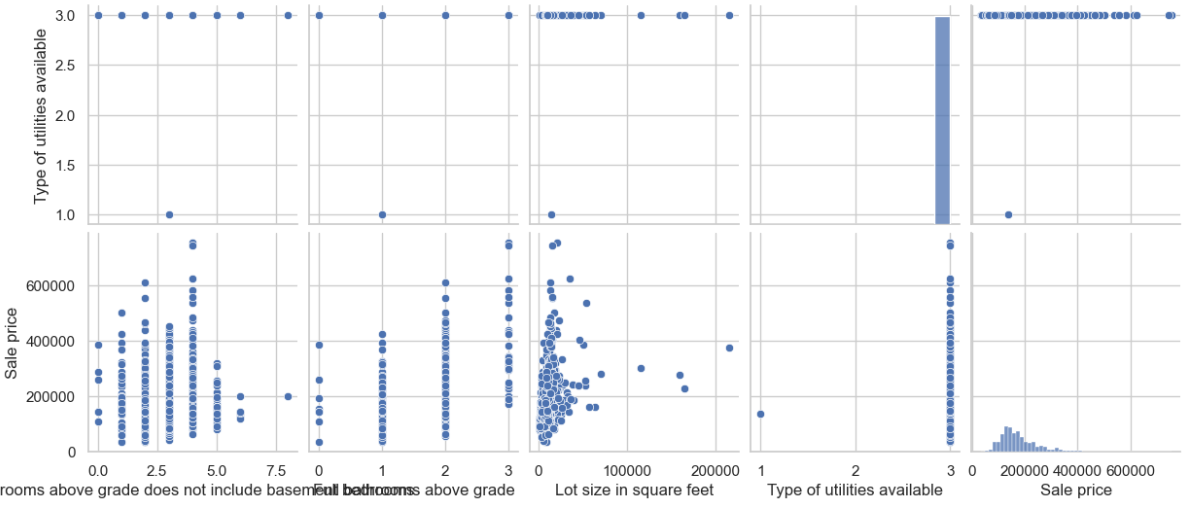
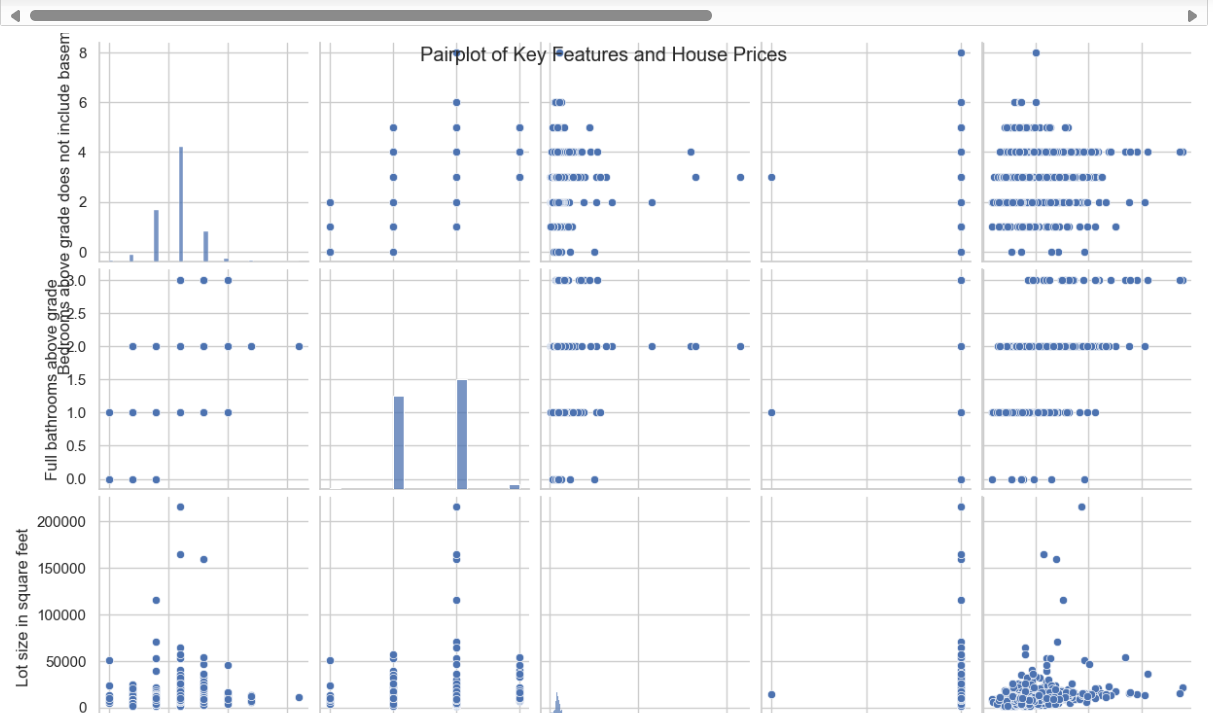


**Data Distribution:**Most data points are clustered towards the lower end of both axes.This indicates that smaller lot sizes and lower sale prices are more common.Few outliers exist with larger lot sizes but not necessarily higher sale prices.**X-Axis (Lot Size):**The x-axis represents “Lot size in square feet.”The range extends from 0 to 200,000 square feet.

**Y-Axis (Sale Price):**The y-axis represents “Sale Price.”The range extends from 0 to 700,000 dollars.

* **Insight between pair plot to display multiple features in single way:**





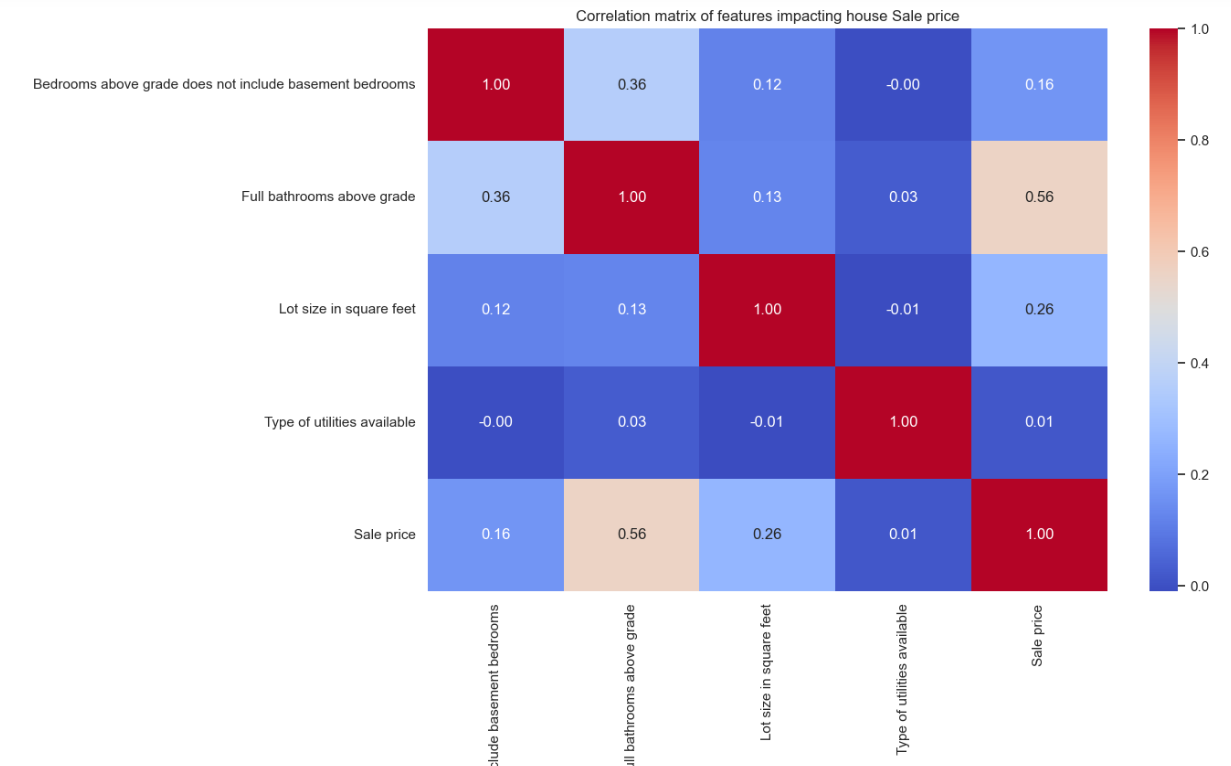
**Rooms Above Grade vs. Sale Price:**There seems to be a positive correlation between the number of rooms above grade (excluding the basement) and the sale price.As the number of rooms increases, the sale price tends to rise.

**Bathrooms Above Grade vs. Sale Price:**Similar to rooms, there is a positive correlation between the number of bathrooms above grade and the sale price.Houses with more bathrooms tend to have higher sale prices.

**Lot Size vs. Sale Price:**The scatter plot shows varying relationships between lot size (measured in square feet) and sale price.Some houses with larger lots have higher sale prices, but there isn’t a consistent trend.

**Type of Utilities Available vs. Sale Price:**The type of utilities available (e.g., electricity, water, gas) doesn’t show a strong correlation with sale price.Houses with different utility types have a wide range of sale prices.

**Checking Correlation:**

****

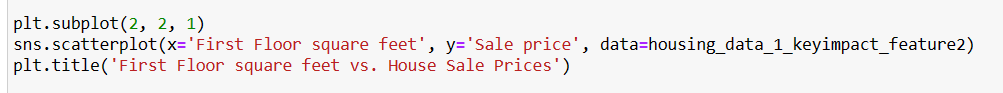
**Variables Included:**The matrix includes the following variables:Bedrooms above grade (excluding basement bedrooms)Full bathrooms above gradeLot size in square feetType of utilities availableSale price

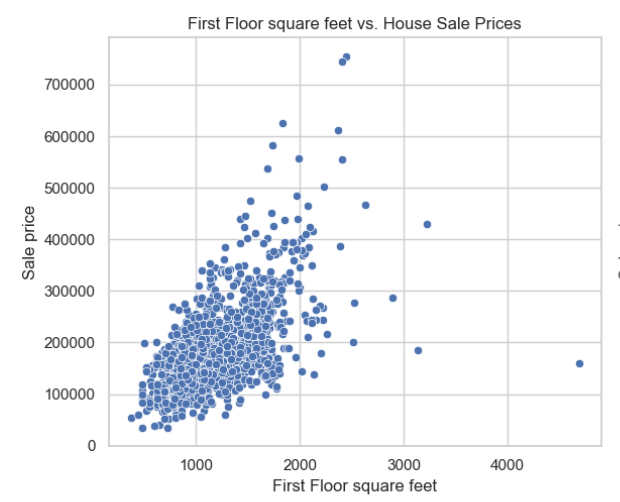
**Correlation Coefficients:**Each cell in the matrix represents the correlation coefficient between two variables.Red indicates a strong positive correlation, while blue indicates no or negative correlation.

**Notable correlations:**Full bathrooms above grade and sale price have a moderate positive relationship with a correlation coefficient of 0.56.

**Interpretation:**The stronger the positive correlation, the more impact a feature has on the sale price.For instance, more full bathrooms above grade tend to increase the sale price.

* **Insight between Sale price and First floor square feet:**

****



**X-Axis (First Floor Square Feet):** The x-axis represents the size of the first floor, ranging from 0 to 4000 square feet.

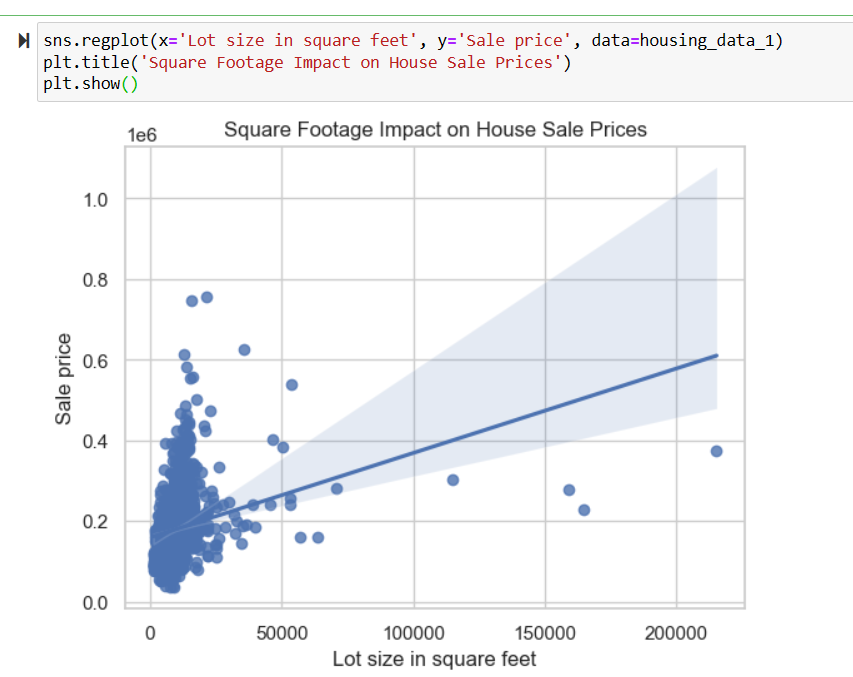
**Y-Axis (Sale Price):** The y-axis represents the sale price of houses, ranging from 0 to 700000 dollars.

**Data Points:** Numerous blue dots represent individual data points scattered across the plot.

**Concentration of Data:** Most data points cluster around 1000 to 2000 square feet, with corresponding sale prices between approximately 100000 and 300000 dollars.

**Fewer High-Value Houses:** Fewer data points are present at higher first floor square footages and sale prices, indicating fewer houses with larger sizes and higher prices in this dataset.

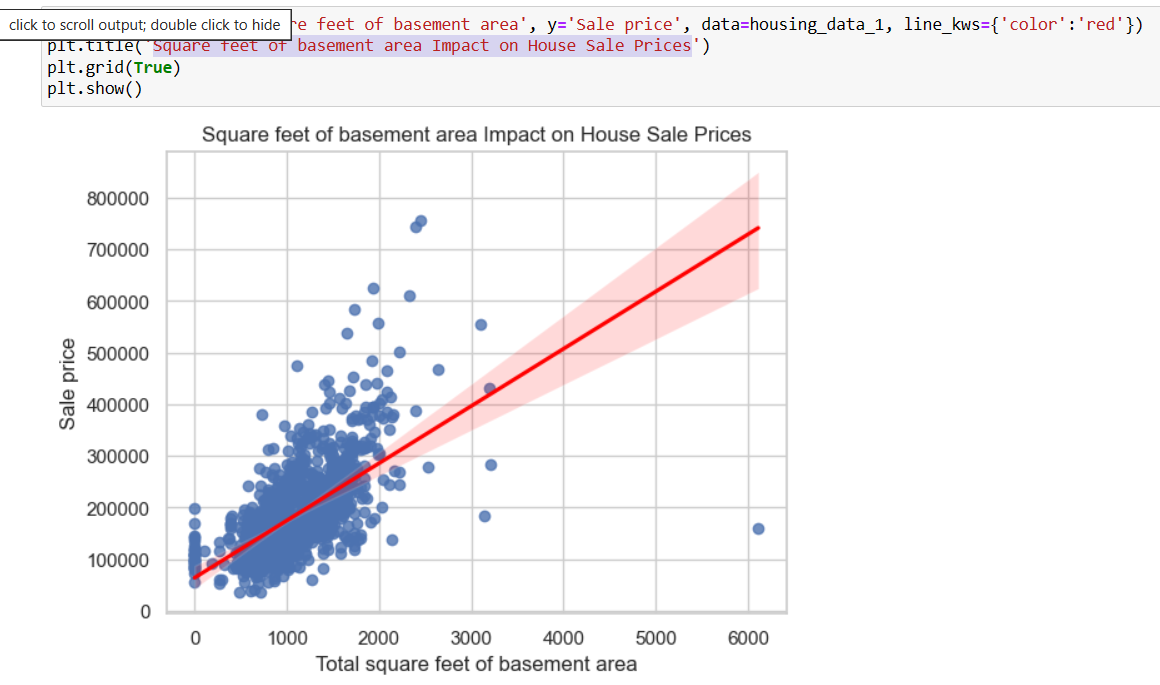
* **Insight between Lot Size in square feet and House Sale Prices:**

**Positive Correlation:** As the lot size increases, the sale price tends to rise. This suggests that larger lots are associated with higher sale prices.

**Line of Best Fit:** The blue dots represent individual data points, and a blue line of best fit indicates the overall trend. The line slopes upward, reinforcing the positive correlation.

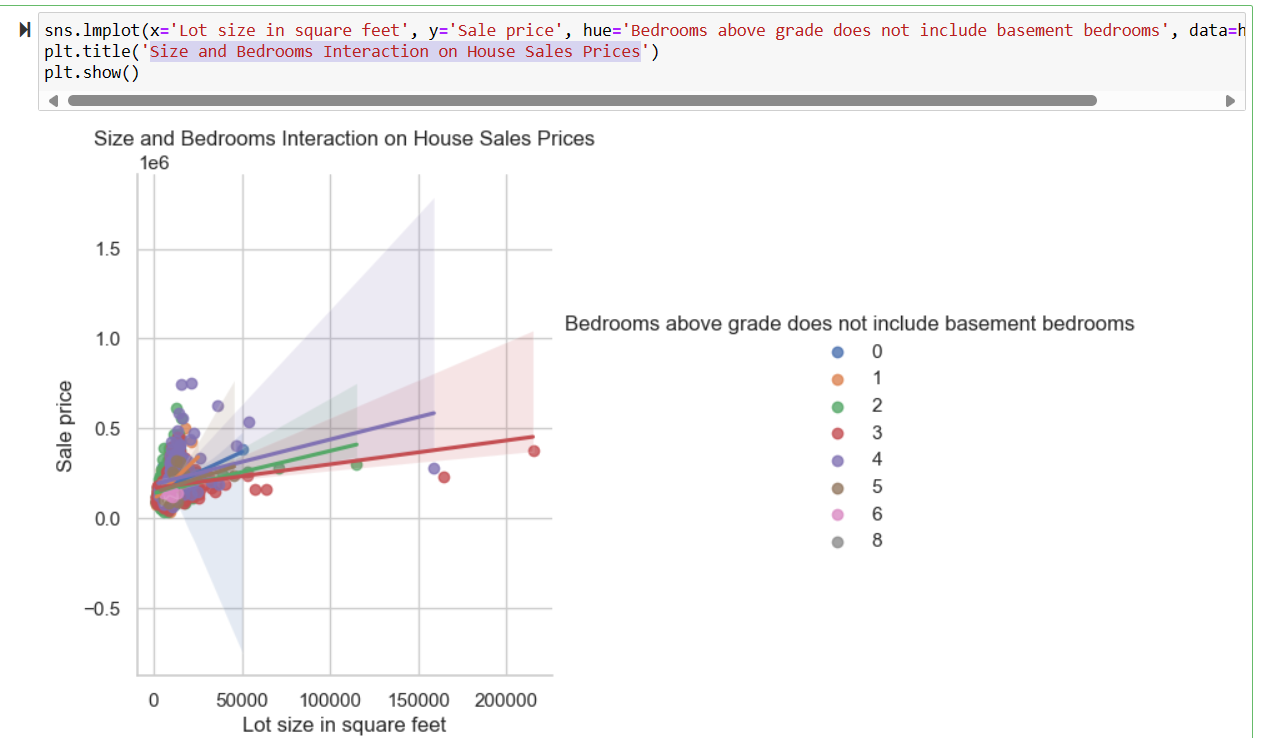
**Title:** The plot is titled “Square Footage Impact on House Sale Prices”, emphasizing the focus on understanding how these two variables interact.

* **Insight between Square feet of basement area Impact on House Sale Prices:**

**Positive Correlation:** As the basement area increases, the sale price of the house tends to increase. The red trend line indicates this positive correlation.

**Data Points:** The blue dots represent individual data points, showing various combinations of basement areas and sale prices.

* **Insight between Size and Bedrooms Interaction on House Sales Prices:**

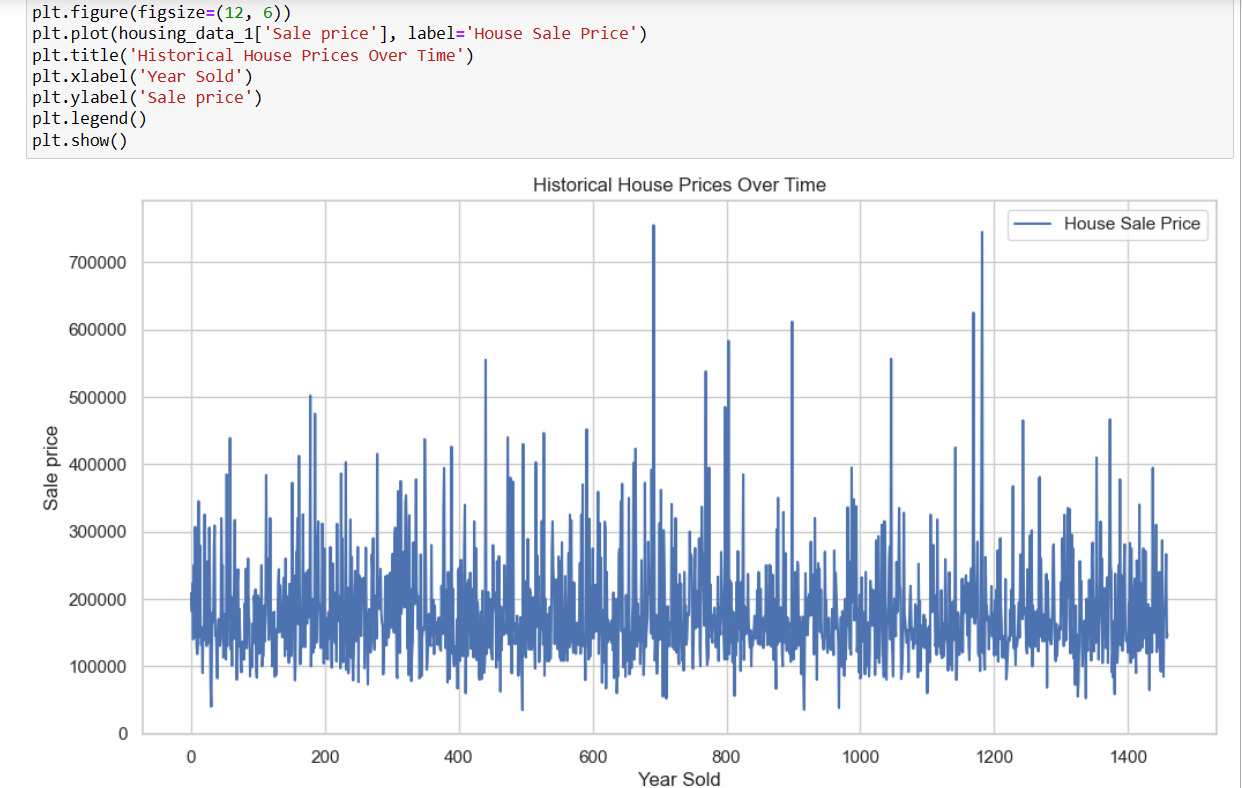
****

**Data Representation:** The graph represents the interaction between the size of a house (in square feet) and the number of bedrooms on the sale price. Different colors represent different numbers of bedrooms.

**Axis Information:** The x-axis represents “Lot size in square feet” ranging from 0 to 200000. The y-axis represents “Sale Price” ranging from -0.5e6 to 1.5e6.

**Trend Observation:** It appears that as both the size and number of bedrooms increase, so does the sale price, although there are variations.

**2. Market Trends and Historical Pricing:**

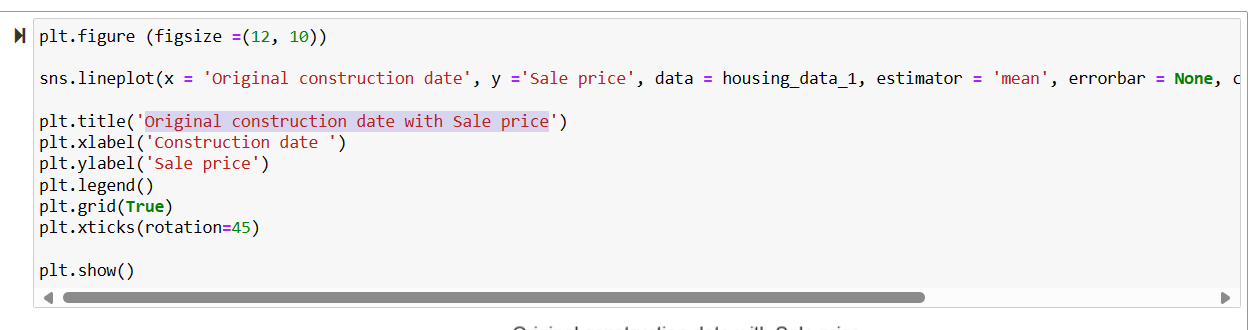
****

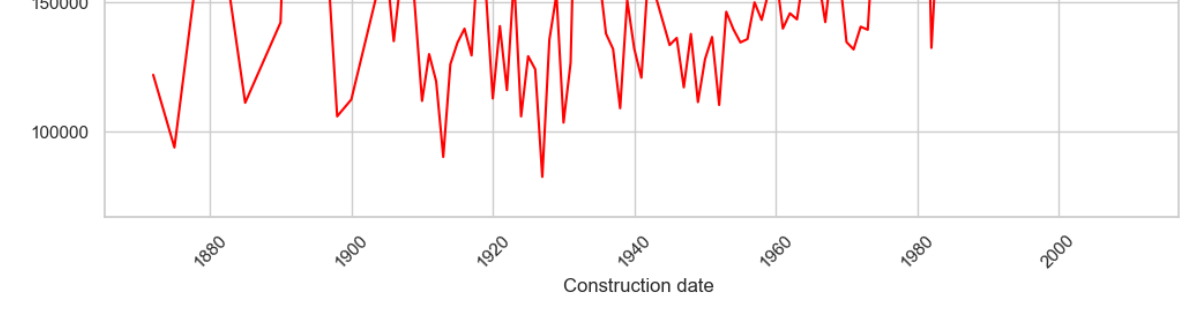
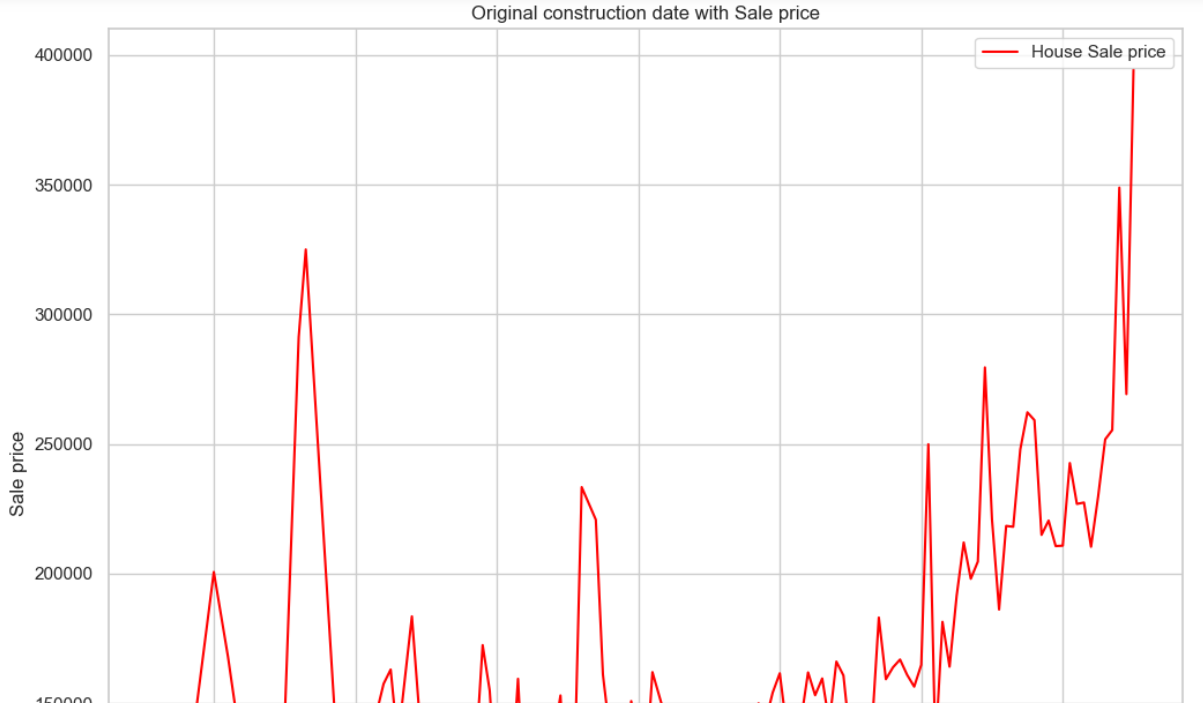
**Title:** The title of the graph is “Historical House Prices Over Time”.

**Axes:** The x-axis is labeled “Year Sold” and the y-axis is labeled “Sale Price”.

**Data Representation:** There are numerous blue vertical lines of varying heights representing individual house sale prices over different years. These are labeled as “House Sale Price” in the legend.

* **Original construction date with Sale price:**

****

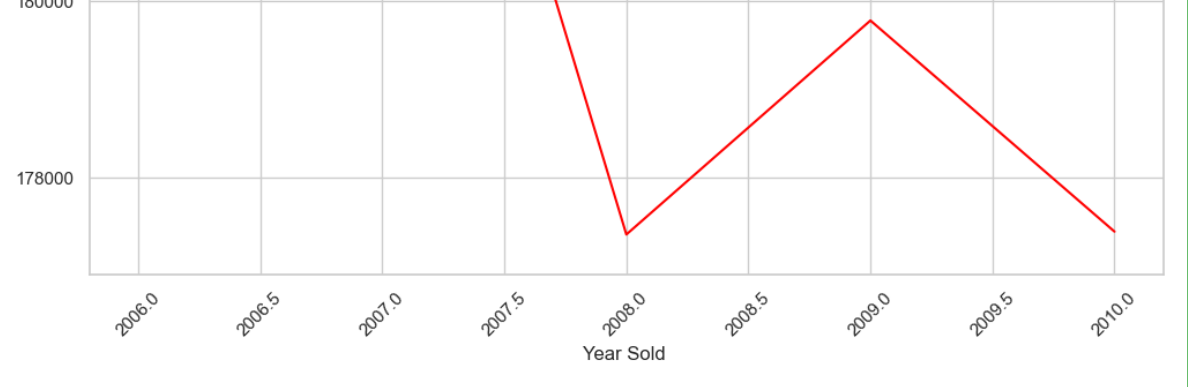
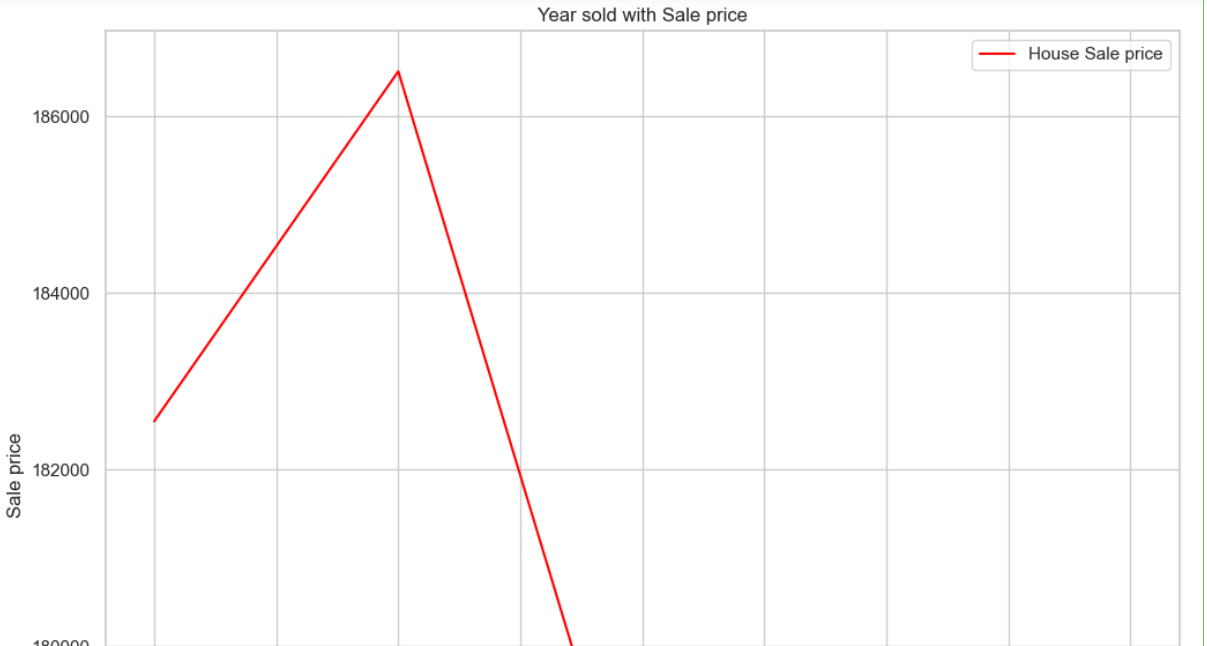
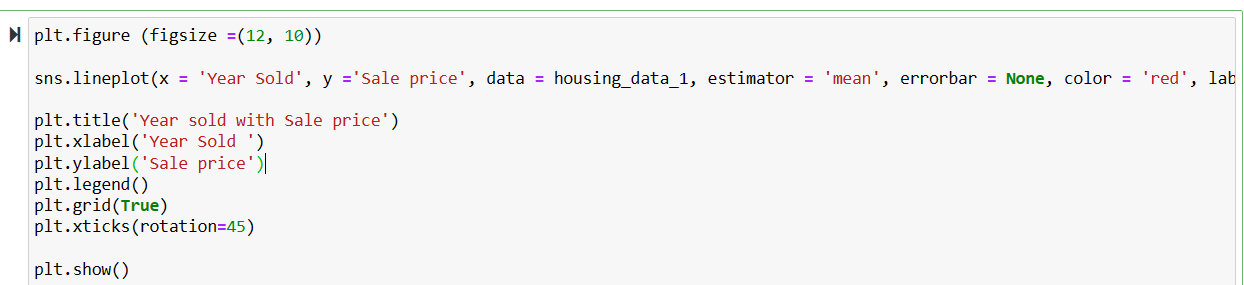
****

**X-axis (Construction Date):** The x-axis represents the construction date of houses, ranging from before 1900 to after 2000.

**Y-axis (Sale Price):** The y-axis represents the sale price of houses, ranging from about $100,000 to $400,000.

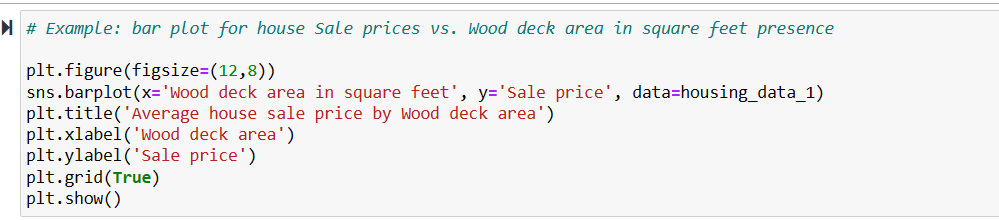
**Red Line (House Sale Price):** The red line shows the fluctuations in house sale prices over various construction dates. There are significant fluctuations in house prices; they appear generally lower for houses built before around 1980 and then there’s a noticeable increase for those constructed after.

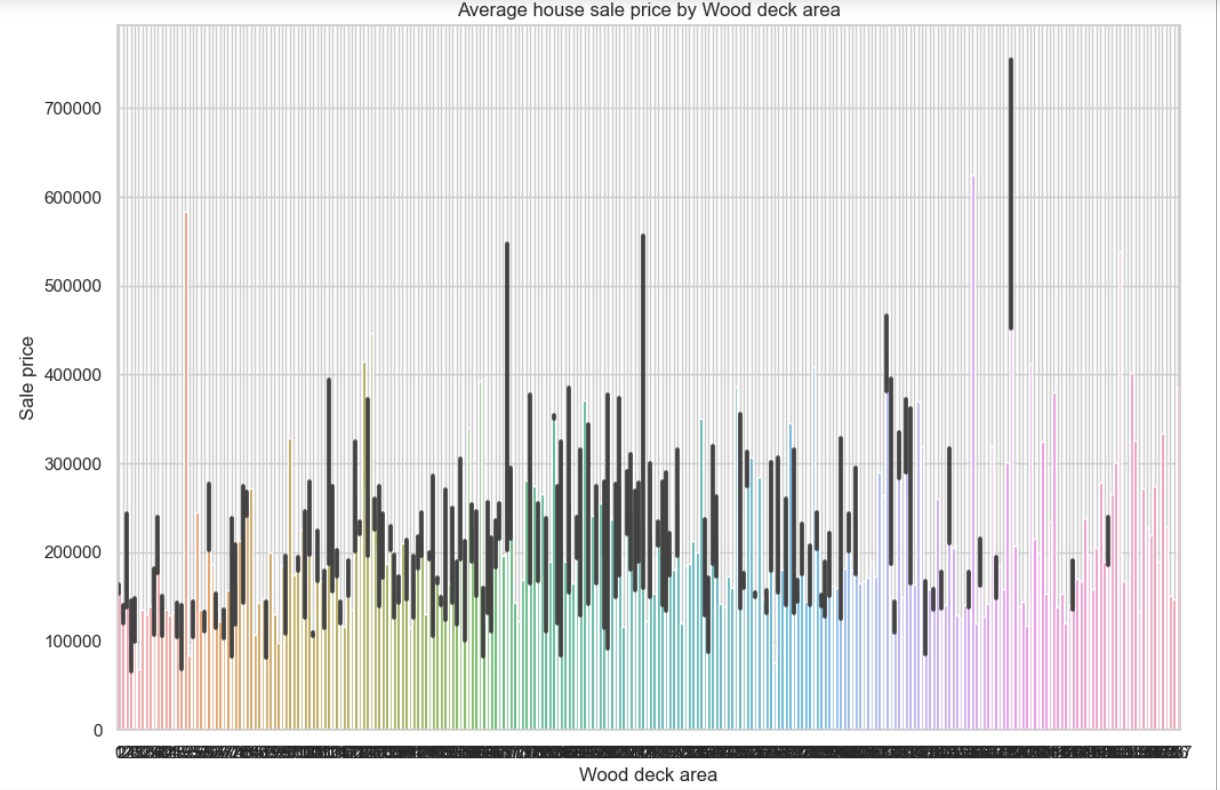
* **Historical pricing with Year sold with Sale price:**

****

The image is a line graph titled “Year sold with Sale price”. It represents the sale price of houses over different years from 2005 to 2010. The y-axis is labeled “Sale Price” and ranges from 178000 to 186000. The x-axis is labeled “Year Sold”. The graph shows a significant peak in house sale prices around 2007, followed by a sharp decline until around 2008, and then it slightly increases but remains relatively low compared to the initial peak. The red line represents the “House Sale price”. This graph provides a clear visual representation of how house sale prices have fluctuated over the years.

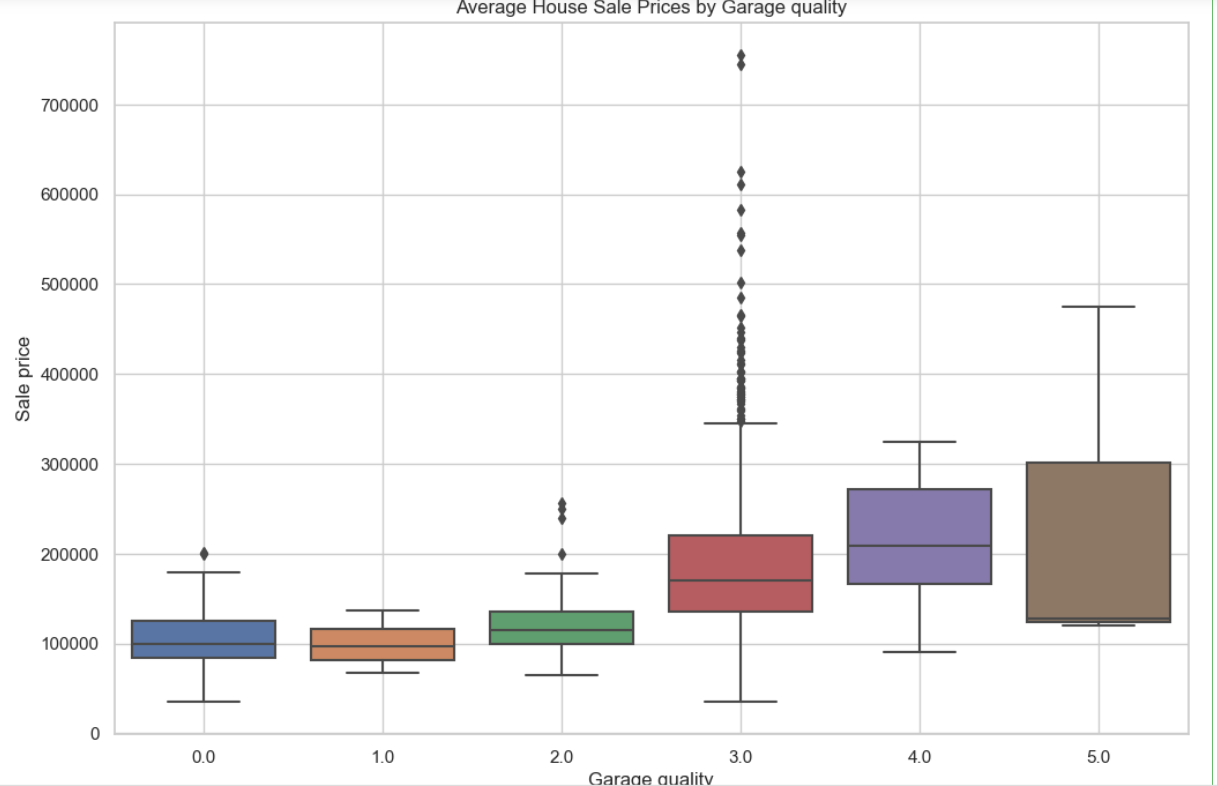
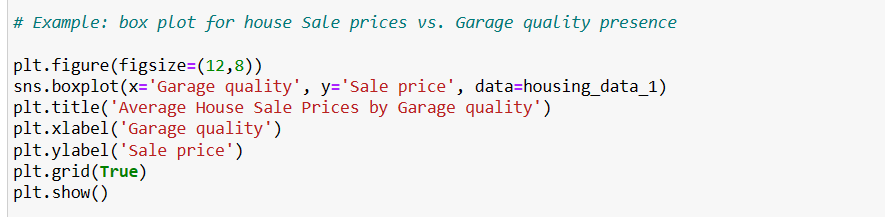
* **Customer Preferences and Amenities:**



****

This image is a vertical bar graph titled “Average house sale price by Wood deck area”. The graph represents the average house sale prices for different sizes of wood deck areas. The y-axis is labeled “Sale Price” and ranges from 0 to 700000, while the x-axis is labeled “Wood deck area”. The bars of varying heights represent the average house sale prices for respective wood deck areas. However, there’s no clear pattern or trend visible in how the wood deck area affects the average house sale price; it appears quite random. This graph could be used to analyze the impact of the wood deck area on the sale price of houses. Please note that the specific data points are not provided in the image description.

* **Average House Sale Prices by Garage quality:**

****

* This image is a boxplot graph that represents the average house sale prices categorized by garage quality. Here are the key details:
* The x-axis shows the garage quality ranging from 0 to 5.
* The y-axis represents the sale price of houses.
* There are five boxplots, each corresponding to a different level of garage quality (0.0, 1.0, 2.0, 3.0, and 4.0).
* Each boxplot displays the distribution of house prices within each category of garage quality; it shows minimum value (bottom whisker), first quartile (bottom of the box), median (line inside the box), third quartile (top of the box), and maximum value (top whisker).
* There are outliers in categories three and four depicted as black dots above those respective boxes.
* **Insight and conclusion of Univariate Analysis:** conducting a univariate analysis using Python libraries such as Matplotlib and Seaborn has provided valuable insights into the distribution and characteristics of key variables, particularly focusing on house prices. By utilizing histograms, kernel density plots, and other visualizations, we were able to explore the individual variable's distribution patterns.

This analysis allows us to understand the central tendencies, spread, and shape of the distribution of house prices. Additionally, visualizations like histograms provide a clear representation of the frequency and density of different price ranges, aiding in the identification of potential outliers or specific trends within the dataset.

* **Insight and conclusion of Multivariate Analysis:** Analyze the visualizations to draw insights. Look for patterns, trends, and potential outliers.

Identify which features have the most significant impact on house prices.

Consider further statistical tests or machine learning models for a more in-depth analysis.

In conclusion, multivariate analysis using Matplotlib and Seaborn provides a comprehensive view of the relationships between various features impacting house prices. The correlation matrix and scatterplot matrix help identify patterns and guide further investigation. This information can be crucial for decision-making in real estate, allowing stakeholders to understand the factors influencing property values**.**

* **Insight and conclusion of Feature Engineering and Size Impact Analysis:** In summary, the exploration of feature engineering and size impact on house prices revealed key insights using Pandas, Matplotlib, and Seaborn. Critical features such as square footage, number of bedrooms, and bathrooms were found to significantly influence property valuations. Visualizations aided in understanding these relationships, and feature engineering enhanced the model's predictive capabilities. The analysis emphasized the importance of considering both features and dataset size for a comprehensive understanding of factors shaping house prices, essential for informed decision-making in the real estate industry.
* **Insight and conclusion of Market Trends and Historical Pricing Analysis:** In summary, employing Matplotlib and Seaborn in Python to analyze historical pricing trends has been instrumental in gaining insights into the housing market. By examining temporal patterns in house prices and considering external factors like economic indicators, we've enhanced our understanding of market influences. This approach equips us with valuable knowledge for informed decision-making and strategic planning in response to dynamic market conditions.
* **Insight and conclusion Customer Preferences and Amenities Analysis:** Customer Preferences and Amenities:

Task: Investigate how customer preferences and amenities impact house prices.

Python Library: Matplotlib, Seaborn

Explanation: Examine the dataset to understand how specific amenities (e.g., swimming pool, garage) impact house prices. Analyze customer feedback or reviews to gauge the perceived value of these amenities.

* **Conclusion on House Sale price prediction:**

In conclusion, the house price prediction project aimed to leverage machine learning techniques to forecast property values based on relevant features. Through the analysis and modeling of datasets, the project aimed to enhance our understanding of the factors influencing housing prices. By employing algorithms such as regression or ensemble methods, the model sought to provide accurate predictions. The success of the project hinges on the quality and quantity of data, feature selection, and the chosen machine learning algorithm. Continuous refinement and validation are crucial to ensure the model's reliability and adaptability to changing real estate market dynamics. Ultimately, the house price prediction project serves as a valuable tool for prospective buyers, sellers, and real estate professionals in making informed decisions in a dynamic housing market.