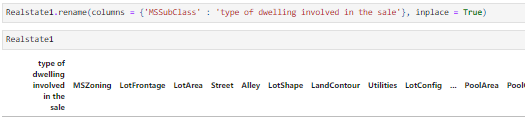
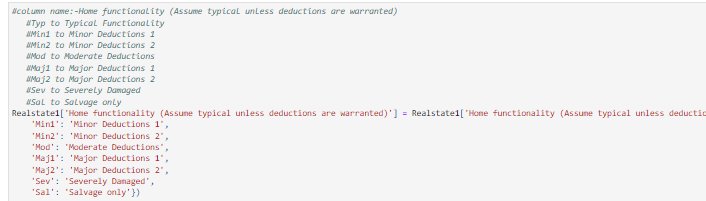
1. **Introduction**: An introduction that outlines the purpose of the interim submission report, which is to provide a snapshot of progress and insights gained during the initial stages of the Real Estate Pricing project.

At first my aim is to make the raw data look good to read and visualise (here we changed the columns name and element names of the columns as per dictionary given to understand the project area) so that afterwards with the visual analysis I can understand the pattern of the project.



1. **Project Overview**: Providing a overview of the Real Estate Pricing project, including its objectives, scope, and the dataset being analyzed.

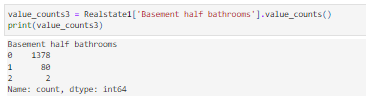
After renaming the column and column element name, I figured out that which column is necessary to take and which not with the help of missing values and values count and as well as finding out the duplicate in the columns.

After understanding the pattern of the project I can apply the logic and do feature engineering and to take the good key finding of the project with the best possible outcome features such as square footage, location, amenities, and property type.

Checking out duplicates



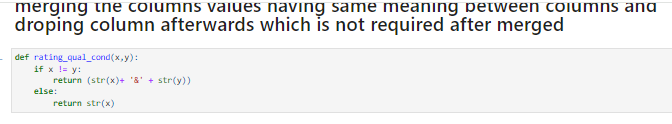
Value counts

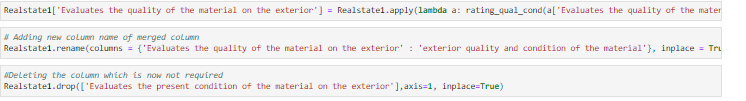


Dropping columns



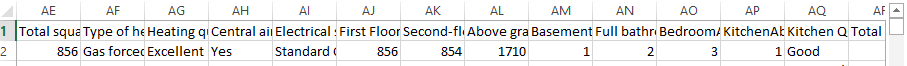
Merging columns name





Printing out the final cleaned dataset for further findings such outliers, univariate and much more.





1. **Key Findings**: Key findings discovered during the preliminary stages of the project. This may include initial trends observed in the real estate pricing data, such as price distributions or property characteristics influencing pricing.

Now we have dataset cleaned with important features that are useful for the project such as square footage, location, amenities, and property type and with these feature it’s time to understand the impact pattern when sale price of houses as per time fluctuates with various techniques.

Price Distributions:

The distribution of property prices exhibits a right-skewed pattern, with a majority of homes falling within a certain price range.

Outliers at the higher end of the price spectrum may indicate luxury properties or unique features.

Property Characteristics Influencing Pricing:

Square footage (size of the property) strongly correlates with price.

The number of bedrooms, bathrooms, and garage spaces also plays a crucial role.

Features like fireplaces, and outdoor spaces contribute to property value.

Next Steps:

Further exploration of feature interactions.

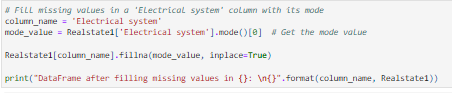
Deeper analysis of neighborhood-specific trends.

Model development and evaluation.

These initial insights provide a foundation for our project. As I delve deeper, we’ll refine our understanding and build robust pricing models.

1. **Challenges Encountered**: Identifying any challenges or obstacles faced during the analysis process. This could involve data quality issues, missing values, outliers, or difficulties in interpreting certain variables.

There were some missing values in important features and I tried to fill the missing value with the help of mode so that I do not lose that feature



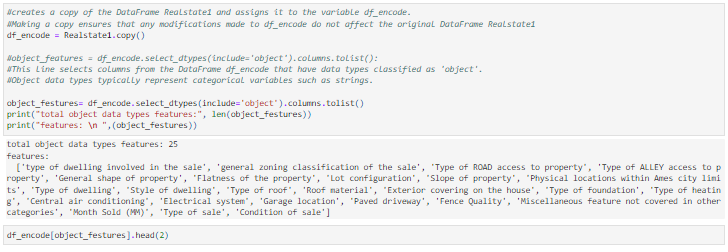
After all correction it’s time to find out the outliers which are already in numeric form and for me but it is a good challenge and difficult to convert the categorical data into numeric form so that I can take out the outliers of the columns

And with the help of **ordinal encoding and one-hot encoding I could able to change the** categorical data into numeric form

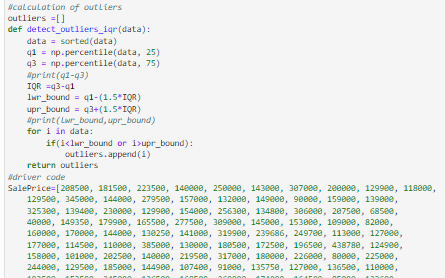
Ordinal encoding

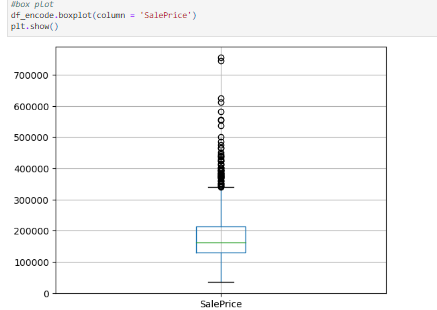


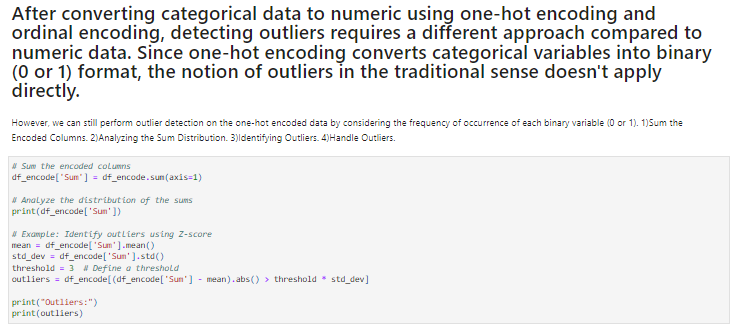
One-hot encoding



Outliers

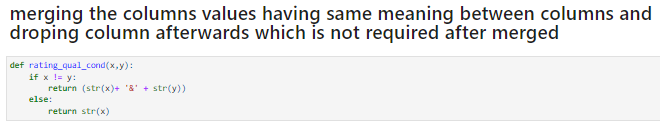


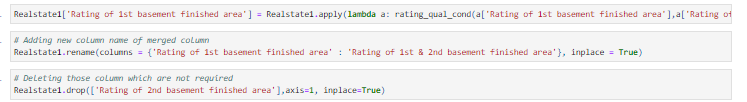


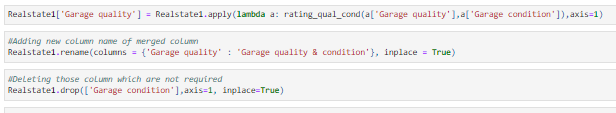


1. **Adjustments to the Original Plan**:

Adjustments made to the original project plan in response to the challenges encountered. This involves changes in the analysis approach, data preprocessing techniques, or exploration of additional variables.

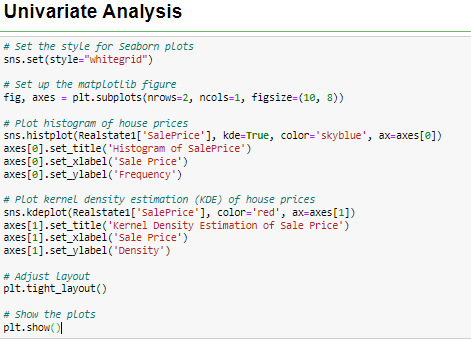


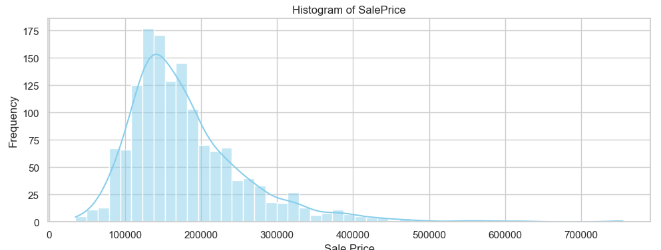


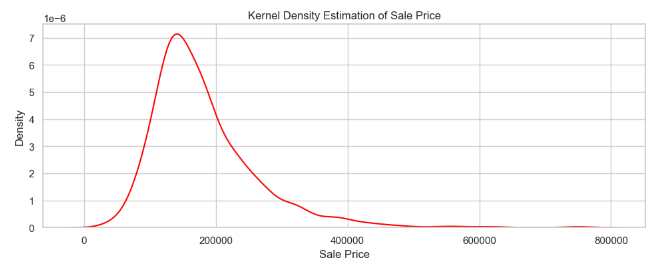
And 

1. **Insights from Univariate Analysis**: Present insights gained from univariate analysis, focusing on individual variables and their distributions. Highlight any notable patterns, outliers, or anomalies identified during this stage.

**INSIGHT IN TERM OF SALEPRICE**







1. **Histogram of Sale Price:**

The top plot is a histogram that displays the distribution of sale prices.

The x-axis represents the Sale Price, and the y-axis represents the frequency (how often each price occurs).

Most of the sale prices seem to be clustered around a specific range (approximately 100,000 to 200,000).

1. **Kernel Density Estimation (KDE) of Sale Price:**

The second plot is a Kernel Density Estimation (KDE), which provides a smoothed representation of the distribution.

It shows the density of sale prices across the same range.

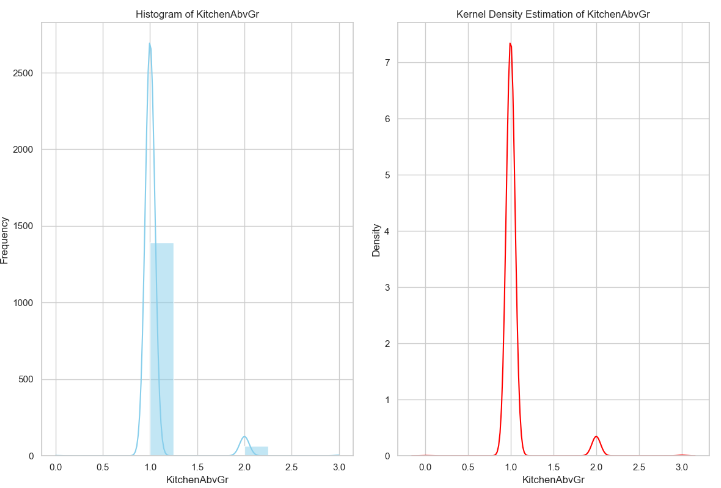
Again, the concentration appears to be around 100,000 to 200,000.

These insights help in understanding the distribution and concentration of sale prices within the real estate data. The histogram provides a visual representation of the frequency of different sale price ranges, while the KDE plot offers a smoothed estimation of the density of sale prices across the same range. This information can be valuable for further analysis and decision-making in the real estate domain.

**INSIGHT IN TERM OF KITCHEN ABOVE GROUND, BEDROOM ABOVE GROUND AND ALSO WITH FULL BATHROOM ABOVE GROUND**



KITCHEN ABOVE GROUND



1. **Histogram of KitchenAbvGr:**

The histogram displays the distribution of the variable "KitchenAbvGr."

On the x-axis, values of "KitchenAbvGr" ranging from 0 to 3 are represented.

The y-axis represents the frequency, indicating the number of occurrences for each value.

The most common value for "KitchenAbvGr" appears to be 1, as evidenced by the peak in the histogram.

1. **Kernel Density Estimation of KitchenAbvGr:**

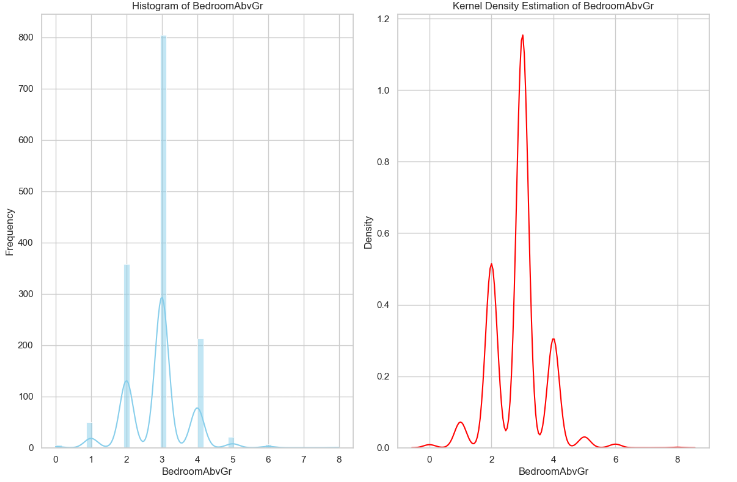
The kernel density estimation plot provides a smooth estimate of the density of "KitchenAbvGr."

Similar to the histogram, there's a prominent peak around the value of 1.

The y-axis represents the density, which ranges from 0 to approximately 7.

In summary, the variable "KitchenAbvGr" is most frequently observed at a value of 1 in the dataset. Both the histogram and kernel density estimation confirm this pattern, indicating that a single kitchen above ground is the most common configuration among the properties in the dataset. This insight helps in understanding the distribution and prevalence of kitchen configurations within the dataset.

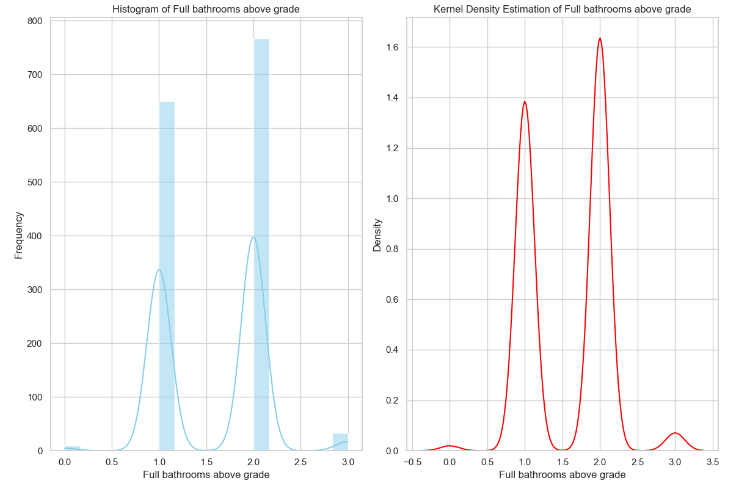
**BEDROOM ABOVE GROUND**



1. **Histogram of BedroomAbvGr**:
   * The histogram showcases the distribution of the variable “BedroomAbvGr.”
   * On the x-axis, we see the values of “BedroomAbvGr,” ranging from 0 to 8.
   * The y-axis represents the frequency (number of occurrences).
   * Notably, there are prominent bars at the values of 2, 3, and 4 on the x-axis, indicating that these are common numbers of bedrooms above grade in the dataset.
2. **Kernel Density Estimation (KDE) of BedroomAbvGr**:
   * The KDE plot provides a smooth estimate of the density of “BedroomAbvGr.”
   * Similar to the histogram, there’s a significant peak around the value of 3 on the x-axis.
   * The y-axis represents the density, which ranges from 0 to approximately 1.2.

In summary, most homes in the dataset have either 2 or 3 bedrooms above grade, with 3 being the most common configuration. Both the histogram and KDE reinforce this observation.

**FULL BATHROOM ABOVE GROUND**

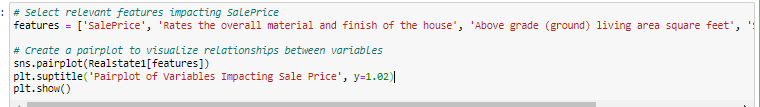


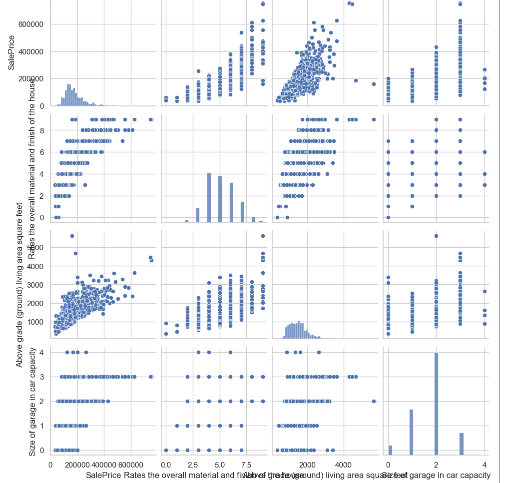
1. **Histogram of Full Bathrooms above Grade**:
   * The histogram showcases the distribution of the variable “Full Bathrooms above Grade.”
   * On the x-axis, we see the values of “Full Bathrooms above Grade,” ranging from 0 to 3.
   * The y-axis represents the frequency (number of occurrences).
   * Notably, there are two prominent bars at the values of 1 and 2 full bathrooms, indicating that these are common numbers of bathrooms to have above grade.
2. **Kernel Density Estimation (KDE) of Full Bathrooms above Grade**:
   * The KDE plot provides a smooth estimate of the density of “Full Bathrooms above Grade.”
   * Similar to the histogram, there’s a significant peak around the value of 1 on the x-axis.
   * The y-axis represents the density, which ranges from 0 to approximately 1.6.

In summary, most homes in the dataset have either 1 or 2 full bathrooms above grade, with 2 being the most common configuration. Both the histogram and KDE reinforce this observation.

1. **Preliminary Observations from Exploratory Data Analysis (EDA)**: Summarize preliminary observations from EDA, including bivariate and multivariate analysis techniques. Discuss relationships between variables, correlation patterns, based on the data exploration process.

**Insight from Multivariate analysis**





**Selecting features**:

Selecting specific features believed to influence the sale price:

‘Rates the overall material and finish of the house’

‘Above grade (ground) living area square feet’

‘Size of garage in car capacity’

These features are crucial for understanding pricing dynamics.

**Pairplot Visualization**:

The pairplot displays scatter plots and histograms for the selected features.

On the diagonal, histograms show the distribution of each feature.

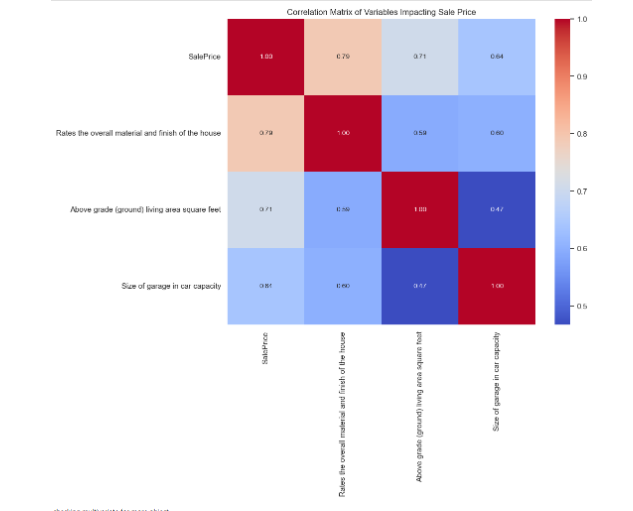
Off-diagonal elements show scatter plots, revealing bivariate relationships between pairs of features.

Notably, there seems to be a positive correlation between Sale Price and Above grade (ground) living area square feet.

In summary, this visualization provides an initial glimpse into how these features relate to the sale price of real estate properties

**Correlation heat matrix**





**Strong Positive Correlations**:

The variable “Rates the overall material and finish of the house” exhibits a strong positive correlation with the sale price (0.79).

This suggests that higher-rated material and finish tend to lead to higher sale prices.

**Moderate Positive Correlations:**

The variable “Above ground (ground) living area square foot” (square footage) also shows a positive correlation with sale price (0.71).

Larger living areas tend to be associated with higher prices.

**Negative Correlation**:

Surprisingly, there’s a negative correlation (-0.64) between the “Size of garage in car capacity” and sale price.

This could imply that very large garages may not significantly impact the price.

In summary, material quality, living area size, and garage capacity play crucial roles in determining real estate prices.