# BUSINESS CASE REPORT: HOUSE PRICES EDA (EXPLORATORY DATA ANALYSIS)

This business case report focuses on conducting an exploratory data analysis (EDA) of house prices, with an emphasis on the Sale Price variable. The objective is to gain insights into the distribution of sale prices, their relationship with various numerical and categorical variables, and their impact on the housing market.

1. Analysis of Sale Price Distribution
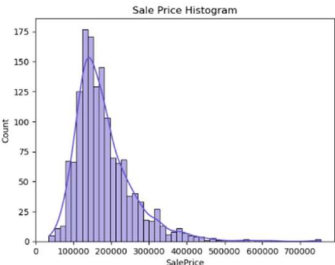
The histogram of Sale Prices reveals that the distribution is right-skewed. The skewness is calculated to be 1.88, indicating a rightward tail in the distribution. Additionally, the kurtosis is 6.54, signifying heavy tails and the presence of outliers.

The right-skewed distribution suggests that most houses have lower sale prices, while fewer have significantly higher prices. The kurtosis value highlights that there are potential outliers in the data, which may require further investigation.

2. Sale Price Trends Over Time

The analysis of Sale Price against the year of sale indicates a decreasing trend in median house prices. Specifically, between 2006 and 2007, there was an increasing pattern in prices, with the highest values observed in 2007. However, from 2007 onward, there has been a continuous decline in house prices, reaching their lowest point in 2010.

3. Correlation Analysis of Numerical Variables
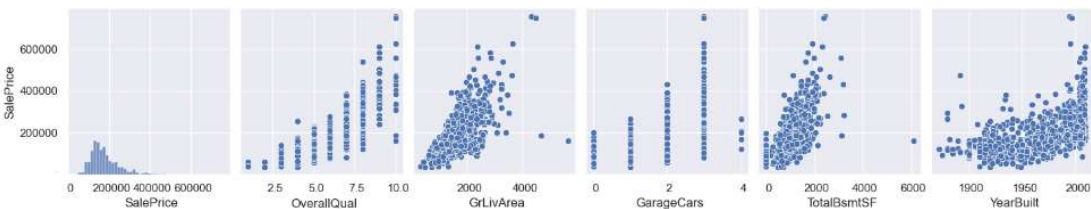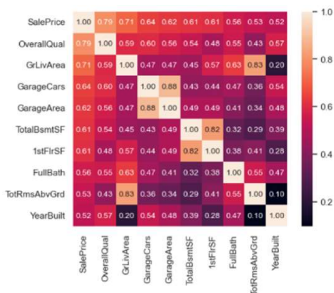
Important Numerical Variables (Correlation with Sale Price): Overall Quality, Living Area (ft2), Garage Space (Car #), Basement Area (ft2), and Year Built. Correlation Findings:

- Sale Price is highly correlated with Overall Quality, with a correlation coefficient of 0.79. Higher quality materials and finishes result in higher prices.
- There is a positive correlation between Living Area (ft2) and Sale Price, with a coefficient of 0.71. Larger living areas are associated with higher prices.
- Garage Car Space is correlated with Sale Price (coefficient of 0.64), with the highest prices linked to houses having 3 car spaces. Prices decrease for houses with 4 car spaces.
- There is a slight positive correlation between Year Built and Sale Price, with a coefficient of 0.52, indicating slightly higher prices for more recently built houses.
- Basement Area (ft2) is positively correlated with Sale Price, with a coefficient of 0.61. Larger basements command higher prices.

4. Relationship with Categorical Variables

Several categorical variables significantly influence sale prices. These include Zoning Classification, Road Access Type, Utilities Type, Neighborhoods, Dwelling Type, Roof Material, Exterior Covering, Foundation Type, Basement Height and Rating, Heating Type and Quality, Central Air Conditioning, Electrical System, Kitchen Quality, Garage Location and Quality.

Additionally, certain variables like Type of Sale and Sale Condition can also impact sale prices.