NEXTHIKES IT SOLUTION

Exploratory Data Analysis for Real Estate Pricing

Project Type: Exploratory Data Analysis (EDA)

Tools Used: Python, Pandas, Numpy, Matplotlib, Seaborn, Jupyter Notebook

Name: Goutam Soni

Role: Data Science Intern

Date: 31-01-2026

INTRODUCTION

The main aim of this project is to understand how different factors affect house prices.

Real estate pricing depends on many variables like size, quality, location, and amenities.

Through EDA, I tried to find patterns, trends, and relationships in the dataset.

PROBLEM STATEMENT

House prices vary due to many reasons. The objective is to:

- Identify important factors that influence house price
- Understand market trends over time
- Analyze how features like quality, size, and location impact valuation

DATASET UNDERSTANDING

The dataset contains 1460 records and 80+ features related to houses.

It includes information about:

- Living area, basement, rooms
- Quality and condition
- Neighborhood (location)
- Amenities like garage, fireplace, porch
- Sale price and sale date

DATA CLEANING

During data cleaning:

- Handled missing values using median and mode
- Removed outliers using IQR method
- Checked for duplicate entries

This step ensured data consistency before analysis.

KEY PRICING DRIVERS IDENTIFIED

The strongest factors affecting house prices were:

- Overall Quality of house
- Above ground living area (GrLivArea)
- Neighborhood (Location)
- Basement area
- Garage capacity

IMPACT OF LIVING AREA AND QUALITY

A strong positive relationship was observed between living area and price.

Overall quality showed even stronger impact than size.

Better material and finishing increased the house valuation significantly.

LOCATION IMPACT (NEIGHBORHOOD)

Different neighborhoods showed huge variation in house prices.

Premium areas had much higher median prices than other areas.

This confirms that location is a major pricing factor.

FEATURE ENGINEERING

Created new features:

- Price per square foot for standard comparison

- Property age to understand effect of construction year

Older properties showed lower prices compared to newer ones.

LAYOUT IMPACT (ROOMS AND BATHROOMS)

Bathrooms showed stronger influence on price than bedrooms.

Total rooms increased price more consistently than bedroom count.

MARKET TRENDS OVER TIME

Year-wise analysis showed price drop during 2008 housing crisis.

Month-wise analysis showed seasonal buying patterns.

AMENITIES IMPACT

Garage capacity and fireplaces added premium value to houses.

Deck and porch areas showed moderate positive impact on price.

BUSINESS INSIGHTS

Buyers prioritize quality and location over just size.

Newer houses and better amenities attract higher prices.

Market trends and seasonality also affect valuation.

CONCLUSION

EDA helped in understanding important pricing drivers in real estate.

The analysis can help buyers, sellers, and real estate agents in better decision making.

THANK YOU

This project helped me understand real-world data analysis and business insights in real estate pricing.

Thank you for reviewing my project.