Project Documentation

Bengaluru Housing Price Analysis — Exploratory Data Analysis (EDA)

---

1. Introduction

Exploratory Data Analysis (EDA) is an essential step in the data analysis process used to understand datasets before building predictive models. This project focuses on analyzing the Bengaluru housing dataset to identify trends, patterns, and relationships influencing property prices.

The analysis aims to transform raw housing data into meaningful insights using statistical summaries and visualization techniques.

---

2. Project Objective

The main objectives of this project are:

- To understand housing price distribution.
- To clean and preprocess raw data.
- To analyze relationships between housing features.
- To identify factors affecting property prices.
- To visualize trends using graphical analysis.
- To generate insights useful for real estate decision-making.

---

3. Dataset Description

The dataset contains information related to residential properties in Bengaluru.

Key Attributes:

- Location of property
- Size and total square feet area
- Number of bedrooms (BHK)
- Number of bathrooms
- Price of the property

The dataset required preprocessing due to missing values and inconsistent entries.

---

4. Tools and Technologies Used

- **Python** – Data analysis and processing
- **Pandas** – Data cleaning and manipulation
- **NumPy** – Numerical computations
- **Matplotlib** – Data visualization
- **Jupyter Notebook** – Analysis environment

---

5. Methodology

Data Loading

The dataset was imported using Pandas and inspected to understand its structure.

Data Understanding

Statistical summaries and dataset information were analyzed to identify numerical and categorical variables.

Data Cleaning

- Missing values were handled.
- Duplicate records were removed.
- Inconsistent data entries were corrected.
- Data types were standardized.

Exploratory Data Analysis

EDA was performed to examine:

- Price distribution
- Location-wise price variations
- Relationship between area and price
- Impact of bedrooms and bathrooms on price

Data Visualization

Various charts were created including:

- Histograms
- Box plots
- Scatter plots
- Bar charts

These visualizations helped detect patterns and outliers.

---

6. Key Findings

- Property prices vary significantly across locations.
- Larger properties generally have higher prices.
- Some locations show extreme price outliers.
- Data cleaning significantly improved analysis accuracy.
- Visualizations revealed meaningful pricing trends.

---

7. Challenges Faced

- Handling missing and inconsistent values.
- Cleaning irregular area measurements.
- Managing outliers affecting price distribution.

---

8. Results and Insights

The exploratory analysis successfully identified major factors influencing housing prices. Location, size, and number of rooms were found to be strong contributors to price variation. The insights derived from this analysis can support better real estate decision-making.

---

9. Future Improvements

- Apply machine learning models for price prediction.
- Perform feature engineering.
- Create interactive dashboards using Power BI.
- Deploy analysis using Streamlit.

---

10. Conclusion

This project demonstrates how Exploratory Data Analysis helps transform raw housing datasets into meaningful insights. Through systematic data cleaning, visualization, and analysis, important pricing patterns were identified.

EDA plays a critical role in understanding data before advanced modeling and provides a strong foundation for data-driven decision making in real-world applications.

Author

Puttoor Yuvaratna Reddy
Data Analyst Enthusiast