**Bangalore House Prediction Report**

**Objective:** To develop a predictive model that accurately estimates the prices of residential properties in Bangalore based on various features such as location, size, amenities, and market trends.

**Libraries:**

* Pandas : for data manipulation
* Numpy: For numerical calculations
* Matplotlib: for data visualization
* Scikit learn: for training model and evaluation

**Data Cleaning**

**Handling Missing Values**

Missing data can skew analysis and model predictions. Here's how missing values are addressed:

* **area\_type**: No missing values detected.
* **availability**: No missing values detected.
* **location**: A few entries missing; such records are removed to maintain data integrity.
* **size**: Standardized to a consistent format (e.g., converting "2 BHK" to numerical values).
* **society**: Missing values handled by imputing with the mode or marking as "Unknown".
* **total\_sqft**: Analyzed for consistency; outliers are treated by capping or removal.
* **bath & balcony**: No significant missing values; ensured data types are correct.
* **price**: Essential target variable; ensured no missing values are present.

**Data Transformation**

Ensuring data is in a suitable format for analysis and modeling:

* **Standardizing Units**: Ensuring all measurements are in consistent units (e.g., square feet).
* **Encoding Categorical Variables**: Converting categorical variables like area\_type and location into numerical formats using techniques pandas dummy variable.
* **Feature Engineering**: Creating new features that might enhance model performance, such as:
  + **Price per Square Foot**: price\_per\_sqft = price / total\_sqft

**Data Visualization**

Visual representations aid in uncovering patterns and insights that might not be evident through raw data alone.

**Price Distribution**

* **Histogram of Prices**: Shows the frequency distribution of house prices, highlighting common price ranges and skewness.

**Feature Selection**

Identifying the most influential features for predicting house prices ensures model efficiency and accuracy.

**Selected Features**

Based on the analysis, the following features are selected for modeling:

* total\_sqft
* bath
* balcony
* location
* area\_type
* size
* price\_per\_sqft

**Model Building and Evaluation:A screenshot of a computer

Description automatically generated**