Practice Set

Mastering Data Cleaning and Data Preparation

PW SKILLS

Dataset Link : Real_Estate

Dataset Description

- **Property_ID:** Unique ID for property
- **Location:** City where property is located
- **Property_Type:** Type (Apartment, Villa, Plot, Studio)
- **Price:** Property price (in lakhs, may have missing values)
- **Size_sqft:** Size of property in square feet (may have missing values)
- **Bedrooms:** Number of bedrooms (may have missing values)

Tools and Libraries

- **Python 3** → Programming language
- **NumPy** → For numerical operations (mean, median, std, correlation)
- **Pandas** → For data manipulation, cleaning, and grouping

Questions :

1. **Load Dataset**
   - Import the CSV file into a pandas DataFrame and display the first 5 rows.
2. **Check Missing Values**
   - Find out how many missing values are there in each column.
3. **Handle Missing Values**
   - Fill missing values in the Bedrooms column with the mode.
   - Fill missing values in the Size_sqft column with the median.
4. **Handle Missing Price**
   - Drop rows where Price is missing. How many rows remain?
5. **Remove Duplicate Records**
   - Check for duplicate rows and remove them. How many duplicates were found?
6. **Statistical Summary**
   - Generate summary statistics (mean, median, min, max) for Price and Size_sqft.
7. **DataFrame from Series**
   - **Create a new DataFrame with two Series:**
     - City: ["Delhi", "Mumbai", "Bangalore"]
     - Avg_Price: [80, 120, 95] (in lakhs)
8. **Filter Data**
   - Select all properties located in Mumbai with more than 2 bedrooms.
9. **Conditional Filtering**
   - Select properties with Price greater than 1 crore (10,000,000).
10. **Merge DataFrames**
    - Merge the main DataFrame with the new DataFrame (from Q7) on Location = City.
    - Compare actual Price with Avg_Price.
11. **Group Analysis**
    - Find the average property Price for each Property_Type.
12. **Sorting**

- Sort the DataFrame by Price in descending order.