Experiment No. 10

Aim : Data Wrangling.

Problem Statement: Data Wrangling on Real Estate Market, and perform following tasks – 1. Import the dataset and explore. 2. Handle missing values. 3. Perform data merging. 4. Filter and subset the data. 5. Handle categorical variables. 6. Aggregate the data, and 7. Identify and handle outliers.

Dataset: "RealEstate_Prices.csv"

Software Requirements : Python and Jupyter Notebook.

Hardware Requirements : 8GB RAM, Storge and Processor.

Objectives : Clean, Integrate, and format data to remove errors, handle missing values, standardize format, and prepare for analysis, ensuring accuracy and reliability.

Theory : Data Wrangling

Data wrangling, also known as data munging or data preprocessing, is the process of cleaning, transforming, and organizing raw data into a structured format that is suitable for analysis or modeling. It's a critical step in the data analysis pipeline, as it ensures that the data is accurate, consistent, and usable.

key aspects of data wragling -

1.Data Collection and Acquisition

- **Source Identification**: Determine the sources of data, which may include databases, spreadsheets, APIs, or web scraping.
- **Data Extraction**: Extract data from various sources while ensuring data quality and completeness.

2. Data Cleaning

- **Handling Missing Values**: Address missing data through imputation (e.g., filling in missing values with the mean or median) or by removing incomplete records if necessary.
- **Removing Duplicates**: Identify and eliminate duplicate entries to avoid redundancy and ensure the accuracy of the dataset.

- **Error Correction**: Fix errors such as typos, incorrect data types, or out-of-range values to maintain data integrity.

3. Data Transformation

- **Normalization and Standardization**: Scale numerical features to a consistent range or standardize them to have zero mean and unit variance.
- **Encoding Categorical Variables**: Convert categorical data into numerical formats using techniques like one-hot encoding or label encoding.
- **Feature Engineering**: Create new features from existing ones to capture additional insights, such as interaction terms or polynomial features.

4. Data Integration

- **Merging Datasets**: Combine data from different sources or tables using join operations, ensuring that the merging process maintains data integrity and consistency.
- **Consolidation**: Aggregate data to a higher level of granularity or summarize it for easier analysis (e.g., aggregating sales data by month).

5. Data Reshaping and Aggregation

- **Pivoting**: Transform data from long to wide format or vice versa using pivot tables to facilitate analysis.
- **Aggregation**: Compute summary statistics such as totals, averages, or counts to consolidate data into meaningful metrics.

6. Outlier Detection and Handling

- **Detection Methods**: Identify outliers using statistical methods (e.g., z-scores, IQR) or visual techniques (e.g., box plots).
- **Handling Outliers**: Decide on the appropriate action for outliers, which may involve removing, adjusting, or retaining them based on their impact on the analysis.

7. Data Validation

- **Consistency Checks**: Verify that data is consistent across different records and within expected ranges (e.g., ensuring dates are in the correct format).

- **Accuracy Verification**: Cross-check data against external sources or validation rules to ensure correctness.

8. Data Enrichment

- **Incorporating External Data**: Enhance your dataset by adding information from external sources, such as demographic or geographical data.
- **Creating New Insights**: Derive new features or insights from enriched data to improve the quality of analysis.

9. Data Formatting

- **Standardizing Formats**: Ensure uniform data formats across the dataset, such as consistent date formats or numerical representations.
- **Cleaning Text Data**: Process text data by removing unnecessary characters, normalizing text (e.g., lowercasing), and handling text encoding issues.

10. Documentation and Metadata

- **Documenting Changes**: Keep detailed records of the data wrangling process, including transformations, cleaning steps, and feature engineering tasks.
- **Metadata Management**: Track metadata such as data source, definitions of features, and modifications to facilitate transparency and reproducibility.

Implementation

Step No.1 - Import the Dataset and Clean the Columns name.

```
import pandas as pd

import numpy as np

df = pd.read_csv(r"C:\Users\saira\Downloads\Mumbai_Property.csv")

df.columns

Index(['Property_Name', 'Location', 'Region', 'Property_Age', 'Availability',
```

```
      'Area_Tpye', 'Area_SqFt', 'Rate_SqFt', 'Floor_No', 'Bedroom',
      'Bathroom', 'Price_Lakh'],
     dtype='object')
df.columns = df.columns.str.replace(' ', '_').str.replace('(', '').str.replace(')', '').str.lower()
df.columns
Index(['property_name', 'location', 'region', 'property_age', 'availability',
      'area_tpye', 'area_sqft', 'rate_sqft', 'floor_no', 'bedroom',
      'bathroom', 'price_lakh'],
     dtype='object')
df.head()
```

	property_name	location	region	property_age	availability	area_tpye	area_sqft	rate_sqft	floor_no	bedroom	bathroom	price_lakh
0	Omkar Alta Monte	W E Highway Malad East Mumbai	Malad Mumbai	0 to 1 Year	Ready To Move	Super Built Up Area	2900.0	17241	14	3	4	500.0
1	T Bhimjyani Neelkanth Woods	Manpada Thane Mumbai	Manpada Thane	1 to 5 Year	Ready To Move	Super Built Up Area	1900.0	12631	8	3	3	240.0
2	Legend 1 Pramila Nagar	Dahisar West Mumbai	Dahisar Mumbai	10+ Year	Ready To Move	Super Built Up Area	595.0	15966	3	1	2	95.0
3	Unnamed Property	Vidyavihar West Central Mumbai	Vidyavihar West Central Mumbai...	5 to 10 Year	Ready To Move	Built Up Area	1450.0	25862	1	3	3	375.0

4 Unnamed Property 176 Cst Road Kalina Mumbai 400098 Santacruz Ea... Santacruz Mumbai 5 to 10 Year Ready To Move Carpet Area 876.0 39954 5 2 2 350.0

df.tail()

property_name location region property_age availability area_tpye area_sqft rate_sqft floor_no bedroom bathroom price_lakh

2575 Shagun White Woods Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1180.0 10338 2 2 2 122.0

2576 Guru Anant Sector 2 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 0 to 1 Year Ready To Move Built Up Area 1090.0 8073 11 2 2 88.0

2577 Balaji Mayuresh Delta Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1295.0 10579 6 2 2 137.0

2578 Balaji Mayuresh Delta Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1850.0 9243 6 3 3 171.0

2579 Gurukrupa Tulsi Heights Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 0 to 1 Year Ready To Move Built Up Area 1100.0 8636 4 2 2 95.0

Step No.2 - Handle Missing Values.

df.isnull().sum()

property_name 0

location 0

region 0

property_age 0

availability 0

area_tpye 0

area_sqft 0

rate_sqft 0

floor_no 0

bedroom 0

bathroom 0

price_lakh 0

dtype: int64

Step No.3 - Perform Data Merging.

```
additional_df = pd.read_csv(r"C:\Users\saira\Downloads\Cleaned_Mumbai_RealEstate_Data.csv")
additional_df.columns
Index(['Property_Name', 'Location', 'Region', 'Availability', 'Area_Type',
       'Area_SqFt', 'Rate_SqFt', 'Floor_No', 'Bedroom', 'Bathroom',
       'Price_Lakh', 'Property_Age_0 to 1 Year', 'Property_Age_1 to 5 Year',
       'Property_Age_10+ Year', 'Property_Age_5 to 10 Year',
       'Property_Age_Under Construction'],
      dtype='object')
additional_df.columns = additional_df.columns.str.replace(' ', '_').str.replace('(', '').str.replace(')', '').str.lower()
additional_df.columns
```

```
Index(['property_name', 'location', 'region', 'availability', 'area_type',
       'area_sqft', 'rate_sqft', 'floor_no', 'bedroom', 'bathroom',
       'price_lakh', 'property_age_0_to_1_year', 'property_age_1_to_5_year',
       'property_age_10+_year', 'property_age_5_to_10_year',
       'property_age_under_construction'],
      dtype='object')
df_merged = pd.merge(df, additional_df, on='property_name', how='left')
df_merged
```

property_name	location_x	region_x	property_age	availability_x	area_tpye	area_sqft_x	rate_sqft_x	floor_no_x	bedroom_x	...	rate_sqft_y	floor_no_y	bedroom_y	bathroom_y	price_lakh_y	property_age_0_to_1_year	property_age_1_to_5_year	property_age_10+_year	property_age_5_to_10_year	property_age_under_construction
0	Omkar Alta Monte	W E Highway Malad East Mumbai	Malad Mumbai	0 to 1 Year	Ready To Move	Super Built Up Area	2900.0	17241	14	3	...	17241.0	14.0	3.0	4.0	500.0	True	False	False	False	False
1	Omkar Alta Monte	W E Highway Malad East Mumbai	Malad Mumbai	0 to 1 Year	Ready To Move	Super Built Up Area	2900.0	17241	14	3	...	20238.0	7.0	5.0	5.0	850.0	True	False	False	False	False
2	Omkar Alta Monte	W E Highway Malad East Mumbai	Malad Mumbai	0 to 1 Year	Ready To Move	Super Built Up Area	2900.0	17241	14	3	...	16220.0	4.0	2.0	2.0	212.0	False	True	False	False	False
3	Omkar Alta Monte	W E Highway Malad East Mumbai	Malad Mumbai	0 to 1 Year	Ready To Move	Super Built Up Area	2900.0	17241	14	3	...	16466.0	6.0	3.0	3.0	316.0	False	True	False	False	False

4 Omkar Alta Monte W E Highway Malad East Mumbai Malad Mumbai 0 to 1 Year Ready To Move Super Built Up Area 2900.0 17241 14 3 ... 19404.0 25.0 3.0 3.0 326.0 False True False False False

... ...

763428 Balaji Mayuresh Delta Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1850.0 9243 6 3 ... 10240.0 10.0 2.0 2.0 127.0 False True False False False

763429 Balaji Mayuresh Delta Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1850.0 9243 6 3 ... 9569.0 8.0 3.0 3.0 178.0 False True False False False

763430 Balaji Mayuresh Delta Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1850.0 9243 6 3 ... 10579.0 6.0 2.0 2.0 137.0 False True False False False

763431 Balaji Mayuresh Delta Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1850.0 9243 6 3 ... 9243.0 6.0 3.0 3.0 171.0 False True False False False

763432 Gurukrupa Tulsi Heights Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 0 to 1 Year Ready To Move Built Up Area 1100.0 8636 4 2 ... 8636.0 4.0 2.0 2.0 95.0 True False False False False

763433 rows × 27 columns

Step No.4 - Filter and Subset the Data.

Filter the data based on location

df_filtered_location = df[df['location'] == 'Sector 23 Ulwe Navi Mumbai Mumbai']

df_filtered_location.head()

property_name location region property_age availability area_tpye area_sqft rate_sqft floor_no bedroom bathroom price_lakh

1154 Titanium One Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 640.0 7500 4 1 2 48.0

1184 Unnamed Property Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 0 to 1 Year Ready To Move Carpet Area 680.0 9558 2 2 2 65.0

1539 Unnamed Property Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Carpet Area 680.0 8382 1 2 2 57.0

1584 Platinum Palacio Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Super Built Up Area 665.0 8307 5 1 1 54.0

2529 Shagun White Woods Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Super Built Up Area 1160.0 10775 2 2 2 125.0

df_filtered_location.tail()

property_name location region property_age availability area_tpye area_sqft rate_sqft floor_no bedroom bathroom price_lakh

1184 Unnamed Property Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 0 to 1 Year Ready To Move Carpet Area 680.0 9558 2 2 2 65.0

1539 Unnamed Property Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Carpet Area 680.0 8382 1 2 2 57.0

1584 Platinum Palacio Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Super Built Up Area 665.0 8307 5 1 1 54.0

2529 Shagun White Woods Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Super Built Up Area 1160.0 10775 2 2 2 125.0

2575 Shagun White Woods Sector 23 Ulwe Navi Mumbai Mumbai Ulwe Navi-Mumbai 1 to 5 Year Ready To Move Built Up Area 1180.0 10338 2 2 2 122.0

df_filtered_property_type = df[(df['area_tpye'] == 'Plot Area') & (df['bedroom'] >2)]

df_filtered_property_type.head()

property_name location region property_age availability area_tpye area_sqft rate_sqft floor_no bedroom bathroom price_lakh

97 Unnamed Property Ramdev Park Ramdev Park Mira Road And Beyond M... Mira Road 10+ Year Ready To Move Plot Area 3000.0 18666 4 6 7 560.0

104 Unnamed Property New Panvel Navi Mumbai Mumbai Panvel Navi-Mumbai 0 to 1 Year Ready To Move Plot Area 210.0 247619 4 5 4 520.0

183 Unnamed Property Wada Mumbai Beyond Thane Mumbai Wada Mumbai 5 to 10 Year Ready To Move Plot Area 2400.0 2291 1 3 2 55.0

237 Unnamed Property O 13 Sector 9 Belapur Navi Mumbai Mumbai Belapur Navi-Mumbai 10+ Year Ready To Move Plot Area 2000.0 8750 2 3 2 175.0

281 Unnamed Property Sector 2 Airoli Navi Mumbai Mumbai Airoli Navi-Mumbai 10+ Year Ready To Move Plot Area 1200.0 13750 3 3 2 165.0

df_filtered_property_type.tail()

property_name location region property_age availability area_tpye area_sqft rate_sqft floor_no bedroom bathroom price_lakh

345 Unnamed Property Sector 9 Belapur Navi Mumbai Mumbai Belapur Navi-Mumbai 10+ Year Ready To Move Plot Area 922.0 27223 2 4 3 251.0

521 Unnamed Property Sector 1 Koparkhairane Sector 1 Koparkhairane ... Koparkhairane Navi-Mumbai 5 to 10 Year Ready To Move Plot Area 200.0 34500 3 3 3 69.0

567 Unnamed Property 101 Manpada Thane Mumbai Manpada Thane 5 to 10 Year Ready To Move Plot Area 2700.0 27777 2 4 4 750.0

1201 Ravi Gaurav Greens Mira Road East Mira Road And Beyond Mumbai Mira Road 5 to 10 Year Ready To Move Plot Area 100000.0 290 2 4 4 290.0

1984 Unnamed Property Khardi Mumbai Beyond Thane Mumbai Mumbai Thane 10+ Year Ready To Move Plot Area 8500.0 1752 2 3 3 149.0

Step No.5 - Handle Categorical Variables.

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 2580 entries, 0 to 2579

Data columns (total 12 columns):

```
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   property_name  2580 non-null   object
 1   location       2580 non-null   object
 2   region         2580 non-null   object
 3   property_age   2580 non-null   object
 4   availability   2580 non-null   object
 5   area_tpye      2580 non-null   object
 6   area_sqft      2580 non-null   float64
 7   rate_sqft      2580 non-null   int64
 8   floor_no       2580 non-null   int64
 9   bedroom        2580 non-null   int64
 10  bathroom       2580 non-null   int64
 11  price_lakh     2580 non-null   float64
dtypes: float64(2), int64(4), object(6)
memory usage: 242.0+ KB
```

```
df['location'].unique()
```

```
array(['W E Highway Malad East Mumbai', 'Manpada Thane Mumbai',
       'Dahisar West Mumbai', ..., '501 Sector 5 Ulwe Navi Mumbai Mumbai',
       '1503 Mira Road East Mira Road And Beyond Mumbai',
```

```
      'Sector 22 Kamothe Navi Mumbai Mumbai'], dtype=object)

df['area_tpye'].unique()

array(['Super Built Up Area', 'Built Up Area', 'Carpet Area', 'Plot Area'],

     dtype=object)

# Apply One-Hot Encoding to categorical columns

df_encoded = pd.get_dummies(df, columns=['location', 'area_tpye'])

df_encoded.head()
```

property_name region property_age availability area_sqft rate_sqft floor_no bedroom bathroom price_lakh ... location_Yagna Nagar Mumbai location_kavesar Thane Mumbai location_kurla west Central Mumbai suburbs Mumbai location_secter 7 koparkhairne Navi Mumbai Mumbai location_thakurli Mumbai Beyond Thane Mumbai location_y K nagar Nx virar west Mira Road And Beyond Mumbai area_tpye_Built Up Area area_tpye_Carpet Area area_tpye_Plot Area area_tpye_Super Built Up Area

0 Omkar Alta Monte Malad Mumbai 0 to 1 Year Ready To Move 2900.0 17241 14 3 4 500.0 ... False False False False False False False False False True

1 T Bhimjyani Neelkanth Woods Manpada Thane 1 to 5 Year Ready To Move 1900.0 12631 8 3 3 240.0 ... False False False False False False False False False True

2 Legend 1 Pramila Nagar Dahisar Mumbai 10+ Year Ready To Move 595.0 15966 3 1 2 95.0 ... False False False False False False False False False True

3 Unnamed Property Central Mumbai 5 to 10 Year Ready To Move 1450.0 25862 1 3 3 375.0 ... False False False False False False True False False False

4 Unnamed Property Santacruz Mumbai 5 to 10 Year Ready To Move 876.0 39954 5 2 2 350.0 ... False False False False False False False True False False

5 rows × 1322 columns

```
from sklearn.preprocessing import LabelEncoder

# Initialize LabelEncoder

label_encoder = LabelEncoder()

# Apply Label Encoding to categorical columns

df['Location_encoded'] = label_encoder.fit_transform(df['location'])

df['Area_Tpye_encoded'] = label_encoder.fit_transform(df['area_tpye'])

df[['location', 'Location_encoded', 'area_tpye', 'Area_Tpye_encoded']].head()
```

	location	Location_encoded	area_tpye	Area_Tpye_encoded
0	W E Highway Malad East Mumbai	1276	Super Built Up Area	3
1	Manpada Thane Mumbai	886	Super Built Up Area	3
2	Dahisar West Mumbai	683	Super Built Up Area	3
3	Vidyavihar West Vidyavihar West Central Mumbai...	1263	Built Up Area	0
4	176 Cst Road Kalina Mumbai 400098 Santacruz Ea...	246	Carpet Area	1

Step No.6 - Aggregate the Data.

```
# Group by 'Location' and calculate the average sale price

average_price_by_neighborhood = df.groupby('location')['price_lakh'].mean().reset_index()

average_price_by_neighborhood
```

location price_lakh

0 000 4 Bunglows Mumbai 700.0000

1 000 Anand Nagar Thane Mumbai 69.0000

2 000 Andheri West Mumbai 450.0000

3 000 Balkum Thane Mumbai 210.0000

4 000 Borivali West Mumbai 102.0000

... ...

1303 kavesar Thane Mumbai 120.3750

1304 kurla west Central Mumbai suburbs Mumbai156.6000

1305 secter 7 koparkhairne Navi Mumbai Mumbai 110.0000

1306 thakurli Mumbai Beyond Thane Mumbai 61.5125

1307 y K nagar Nx virar west Mira Road And Beyond M... 38.0000

1308 rows × 2 columns

Group by 'area_Tpye' and calculate the average sale price

average_price_by_area_type = df.groupby('area_tpye')['price_lakh'].mean().reset_index()

average_price_by_area_type

area_tpye price_lakh

0 Built Up Area 138.621000

1 Carpet Area 184.005299

2 Plot Area 215.687500

3 Super Built Up Area 177.751004

Step No.7 - Identify and Handle Outliers.

```
from scipy import stats

# Calculate Z-scores

z_scores = stats.zscore(df['price_lakh'])

df['z_score'] = z_scores

z_scores
```

0 0.881426

1 0.177607

2 -0.214908

3 0.543052

4 0.475377

...

2575 -0.141819

2576 -0.233857

2577 -0.101214

2578 -0.009176

2579 -0.214908

Name: price_lakh, Length: 2580, dtype: float64

```
# Identify outliers (e.g., Z-score > 3 or < -3)
```

```
outliers_z_score = df[(df['z_score'] > 3) | (df['z_score'] < -3)]

# Display outliers

outliers_z_score
```

	property_name	location	region	property_age	availability	area_tpye	area_sqft	rate_sqft	floor_no	bedroom	bathroom	price_lakh	Location_encoded	Area_Tpye_encoded	z_score
39	Swan Lake Apartment1	101 Khar West Mumbai	South West Mumbai	South Mumbai	10+ Year	Ready To Move	Carpet Area	2715.0	66298	0	4	4	1800.0	69	1	4.400525
203	Sagar Mahal	Opposite Gopi Birla School And Sheetal Baug Wa...	Walkeshwar Mumbai	10+ Year	Ready To Move	Built Up Area	2450.0	67350	9	4	5	1650.0	991	0	3.994475
329	Jolly Maker Apartment	Cuffe Parade South Mumbai	Mumbai South Mumbai	10+ Year	Ready To Move	Built Up Area	2135.0	74941	20	5	4	1600.0	673	0	3.859125
605	Hiranandani Gardens Richmond Tower	Hiranandani Gardens Powai Hiranandani Gardens ...	Central Mumbai	1 to 5 Year	Ready To Move	Super Built Up Area	5000.0	33000	6	5	6	1650.0	766	3	3.994475
634	Kalpataru Solitaire	Juhu Mumbai South West Mumbai	Juhu Mumbai	1 to 5 Year	Ready To Move	Super Built Up Area	3000.0	46000	6	3	5	1380.0	781	3	3.263586
635	Kalpataru Solitaire	Juhu Mumbai South West Mumbai	Juhu Mumbai	1 to 5 Year	Ready To Move	Super Built Up Area	2800.0	46428	5	3	3	1300.0	781	3	3.047026
1064	Unnamed Property	Juhu Mumbai South West Mumbai	Juhu Mumbai	10+ Year	Ready To Move	Built Up Area	4363.0	36672	2	4	4	1600.0	781	0	3.859125

1067 Unnamed Property Juhu Mumbai South West Mumbai Juhu Mumbai 10+ Year Ready To Move Super Built Up Area 4200.0 45238 1 4 4 1900.0 781 3 4.671225

1416 Unnamed Property Juhu Mumbai South West Mumbai Juhu Mumbai 10+ Year Ready To Move Super Built Up Area 5700.0 42105 1 4 4 2400.0 781 3 6.024724

1675 Piramal Aranya Byculla East Byculla East Mumbai Harbour Mumbai Mumbai Harbour 0 to 1 Year Ready To Move Carpet Area 2800.0 49107 21 4 4 1375.0 637 1 3.250051

2065 White City 005 Kandivali East Mumbai Kandivali Mumbai 0 to 1 Year Ready To Move Super Built Up Area 1000.0 1650000 21 2 2 16500.0 60 3 44.193409

Remove outliers based on Z-scores

df_cleaned = df[(df['z_score'] <= 3) & (df['z_score'] >= -3)]

Drop the Z-score column if no longer needed

df_cleaned = df_cleaned.drop(columns=['z_score'])

df_cleaned.head()

property_name location region property_age availability area_tpye area_sqft rate_sqft floor_no bedroom bathroom price_lakh Location_encoded Area_Tpye_encoded

0 Omkar Alta Monte W E Highway Malad East Mumbai Malad Mumbai 0 to 1 Year Ready To Move Super Built Up Area 2900.0 17241 14 3 4 500.0 1276 3

1 T Bhimjyani Neelkanth Woods Manpada Thane Mumbai Manpada Thane 1 to 5 Year Ready To Move Super Built Up Area 1900.0 12631 8 3 3 240.0 886 3

2 Legend 1 Pramila Nagar Dahisar West Mumbai Dahisar Mumbai 10+ Year Ready To Move Super Built Up Area 595.0 15966 3 1 2 95.0 683 3

3 Unnamed Property Vidyavihar West Vidyavihar West Central Mumbai... Central Mumbai 5 to 10 Year Ready To Move Built Up Area 1450.0 25862 1 3 3 375.0 1263 0

4 Unnamed Property 176 Cst Road Kalina Mumbai 400098 Santacruz Ea... Santacruz Mumbai 5 to 10 Year Ready To Move Carpet Area 876.0 39954 5 2 2 350.0 246 1

Conclusion : We can successfully done effective data wrangling ensures clean, integrated on a “RealEstate_Price.csv” dataset, and also enhancing insights and decisions making and also addressing errors and standardizing formats it ensures accuracy and reliability.