Assignment 2

Name – Kajal

Reg no – 21BAC10039

Campus – Vit Bhopal

Perform the Below Tasks to complete the assignment:- Tasks:-

1. Download the dataset: Dataset

2. Load the dataset.

3. Perform the Below Visualizations.

* Univariate Analysis
* Bi - Variate Analysis
* Multivariate Analysis

4. Perform descriptive statistics on the dataset.

5. Handle the Missing values

Code:

1.

import pandas as pd

dataset = pd.read\_csv('House Price India.csv')

import matplotlib.pyplot as plt

import seaborn as sns

# Example: Histogram of the 'Price' column

plt.figure(figsize=(8, 6))

sns.histplot(data=dataset, x='Price', bins=30, kde=True)

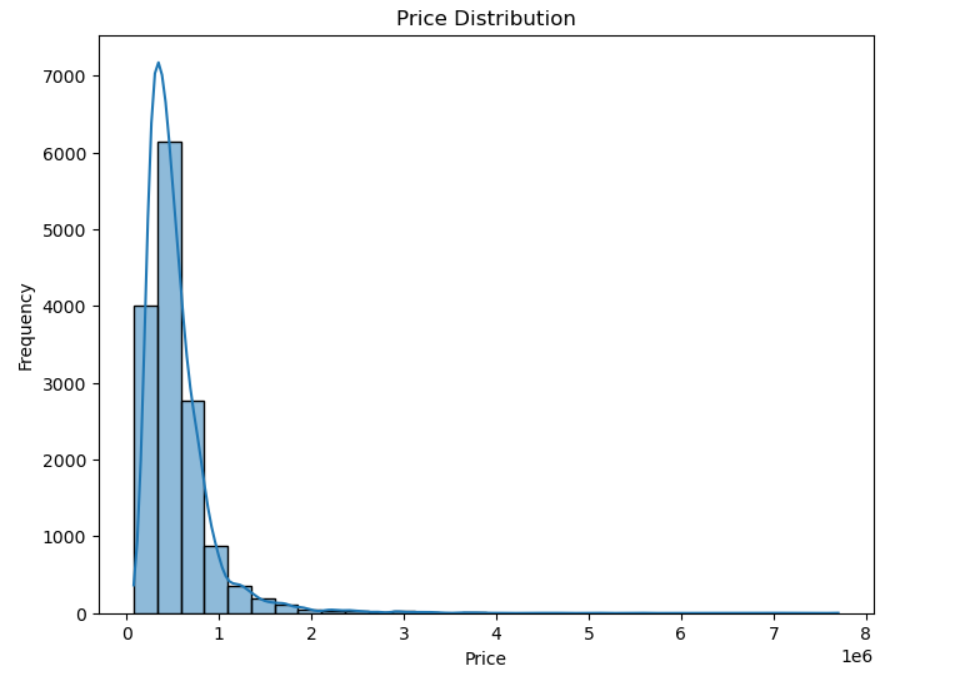
plt.title('Price Distribution')

plt.xlabel('Price')

plt.ylabel('Frequency')

plt.show()

Output:



2.

# Example: Scatter plot of 'Living Area' vs. 'Price'

plt.figure(figsize=(8, 6))

sns.scatterplot(data=dataset, x='living area', y='Price')

plt.title('Living Area vs. Price')

plt.xlabel('Living Area')

plt.ylabel('Price')

plt.show()

Output:



3.

# Calculate correlations between numeric columns

correlation\_matrix = dataset.corr()

# Create a heatmap

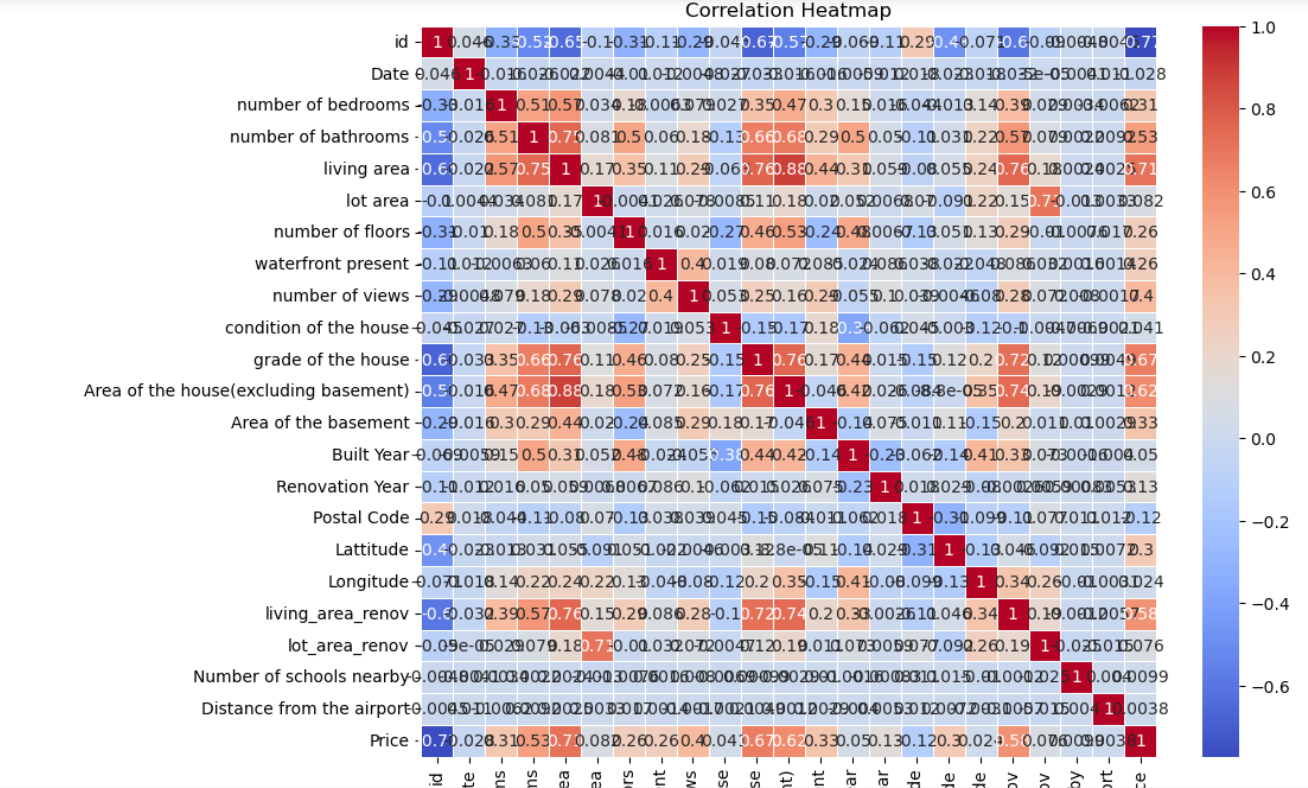
plt.figure(figsize=(10, 8))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', linewidths=0.5)

plt.title('Correlation Heatmap')

plt.show()

Output:



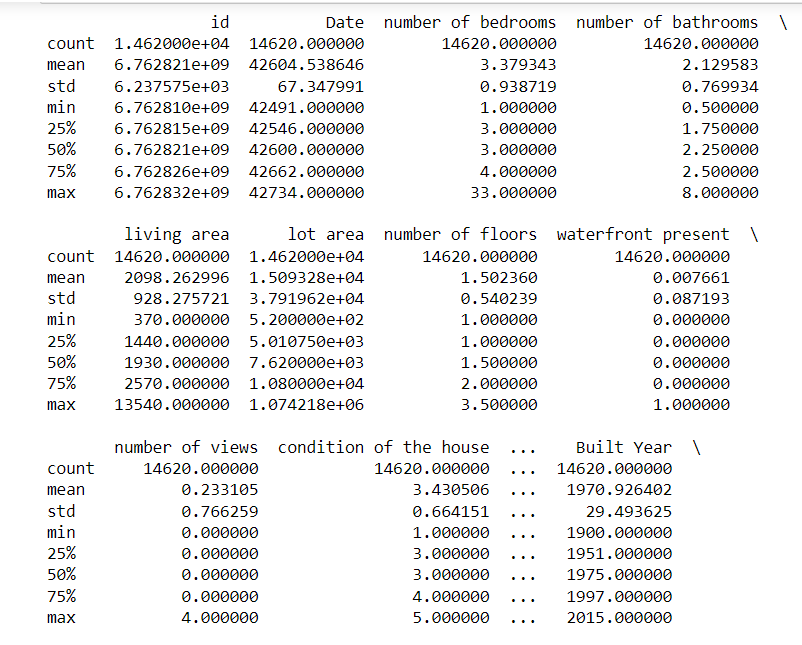
4.

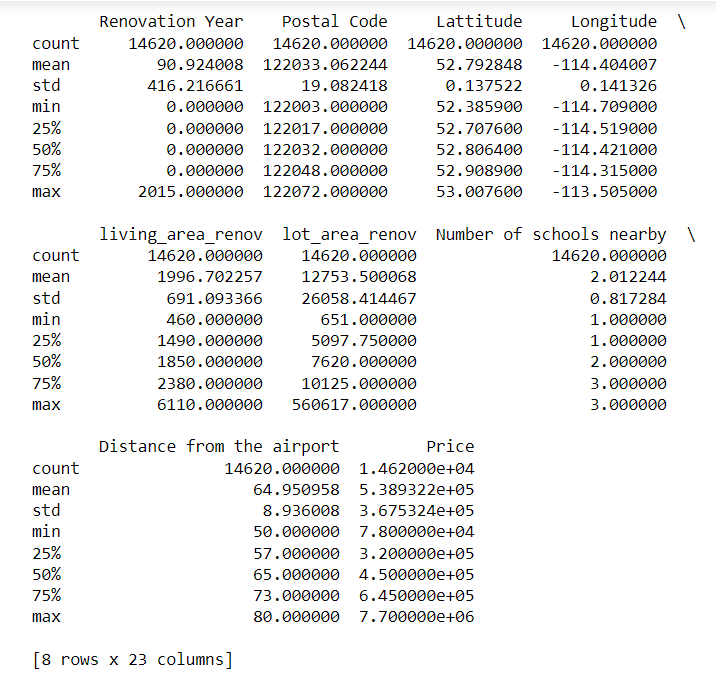
# Summary statistics

summary\_stats = dataset.describe()

print(summary\_stats)

Output:





5.

# Check for missing values before imputation

print(dataset.isnull().sum())

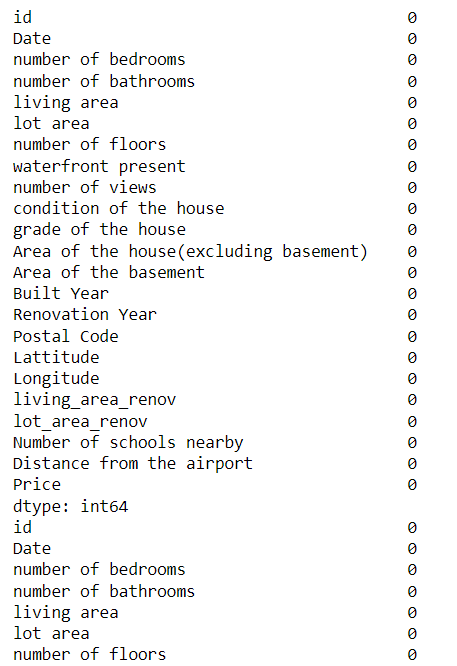
# Impute missing values with mean

dataset.fillna(dataset.mean(), inplace=True)

# Check for missing values after imputation

print(dataset.isnull().sum())

Output:



A screenshot of a computer code

Description automatically generated