Spyder

File Edit Search Source Run Debug Consoles Projects Tools View Help

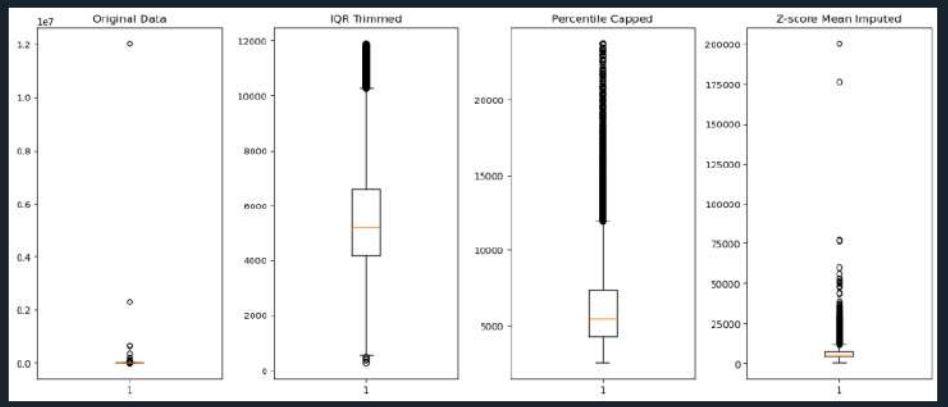
C:\Users\sayan

C:\Users\sayan\.spyder-py3\main_program.py

main_program.py

```
70  plt.boxplot(df_imputed['price_per_sqft'])
71  plt.title("Z-score Mean Imputed")
72
73  plt.tight_layout()
74  plt.show()
75
76  # Normality check with histogram
77  plt.figure(figsize=(14, 6))
78  plt.hist(df_trimmed['price_per_sqft'], bins=50, edgecolor='black', alpha=0.7)
79  plt.title("Histogram of Price per Sqft (After IQR Trimming)")
80  plt.xlabel("Price per Sqft")
81  plt.ylabel("Frequency")
82  plt.show()
83
84  # Skewness and Kurtosis before and after log transformation
85  skewness_before = skew(df_trimmed['price_per_sqft'])
86  kurtosis_before = kurtosis(df_trimmed['price_per_sqft'])
87  df_trimmed['price_per_sqft_log'] = np.log1p(df_trimmed['price_per_sqft'])
88  skewness_after = skew(df_trimmed['price_per_sqft_log'])
89  kurtosis_after = kurtosis(df_trimmed['price_per_sqft_log'])
90
91  print("Skewness before transformation:", skewness_before)
92  print("Kurtosis before transformation:", kurtosis_before)
93  print("Skewness after transformation:", skewness_after)
94  print("Kurtosis after transformation:", kurtosis_after)
95
96  # Compute correlation matrix for numerical columns only
97  numerical_df = df_trimmed.select_dtypes(include=[np.number])
98  corr_matrix = numerical_df.corr()
99  print("Correlation Matrix:\n", corr_matrix)
100
101 # Scatter plots
102 plt.figure(figsize=(14, 6))
103 plt.subplot(1, 2, 1)
104 plt.scatter(df_trimmed['total_sqft'], df_trimmed['price'], alpha=0.5)
105 plt.title("Total Sqft vs Price")
106 plt.xlabel("Total Sqft")
107 plt.ylabel("Price")
108
109 plt.subplot(1, 2, 2)
```

57 %



Help Debugger Plots Files

Console 1/A

```
Kurtosis before transformation: 0.6784399584501744
Skewness after transformation: -0.12988210833124453
Kurtosis after transformation: 1.3369230684904956
Correlation Matrix:
                  total_sqft      bath  ...   z_score  price_per_sqft_log
total_sqft          1.000000  0.455416  ...  0.136414            0.057160
bath                0.455416  1.000000  ...  0.348669            0.320142
price               0.660443  0.571890  ...  0.612102            0.573693
bhk                 0.426239  0.881677  ...  0.303220            0.270817
price_per_sqft      0.136414  0.348669  ...  1.000000            0.970182
z_score             0.136414  0.348669  ...  1.000000            0.970182
price_per_sqft_log  0.057160  0.320142  ...  0.970182            1.000000

[7 rows x 7 columns]
```

IPython Console History

Inline Conda: spyder-runtime (Python 3.11.11) LSP: Python Line 90, Col 1 UTF-8 CRLF RW Mem 91%

Spyder

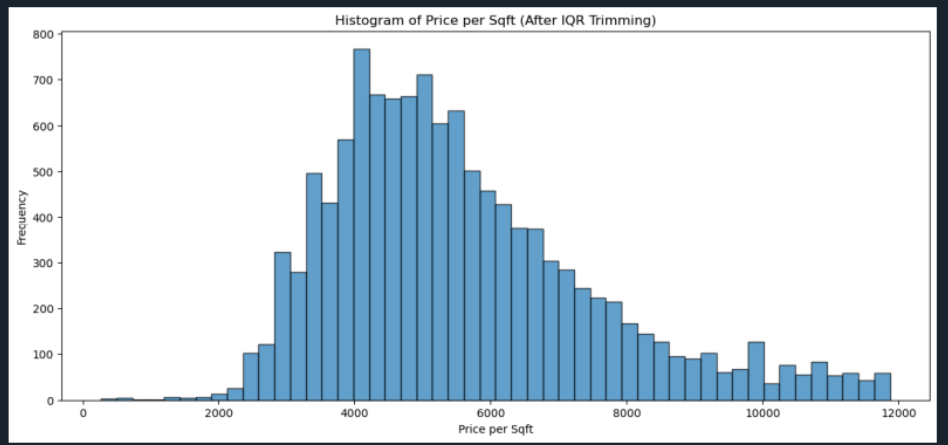
File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\sayan

C:\Users\sayan\.spyder-py3\main_program.py

main_program.py

```
70  plt.boxplot(df_imputed['price_per_sqft'])
71  plt.title("Z-score Mean Imputed")
72
73  plt.tight_layout()
74  plt.show()
75
76  # Normality check with histogram
77  plt.figure(figsize=(14, 6))
78  plt.hist(df_trimmed['price_per_sqft'], bins=50, edgecolor='black', alpha=0.7)
79  plt.title("Histogram of Price per Sqft (After IQR Trimming)")
80  plt.xlabel("Price per Sqft")
81  plt.ylabel("Frequency")
82  plt.show()
83
84  # Skewness and Kurtosis before and after log transformation
85  skewness_before = skew(df_trimmed['price_per_sqft'])
86  kurtosis_before = kurtosis(df_trimmed['price_per_sqft'])
87  df_trimmed['price_per_sqft_log'] = np.log1p(df_trimmed['price_per_sqft'])
88  skewness_after = skew(df_trimmed['price_per_sqft_log'])
89  kurtosis_after = kurtosis(df_trimmed['price_per_sqft_log'])
90
91  print("Skewness before transformation:", skewness_before)
92  print("Kurtosis before transformation:", kurtosis_before)
93  print("Skewness after transformation:", skewness_after)
94  print("Kurtosis after transformation:", kurtosis_after)
95
96  # Compute correlation matrix for numerical columns only
97  numerical_df = df_trimmed.select_dtypes(include=[np.number])
98  corr_matrix = numerical_df.corr()
99  print("Correlation Matrix:\n", corr_matrix)
100
101 # Scatter plots
102 plt.figure(figsize=(14, 6))
103 plt.subplot(1, 2, 1)
104 plt.scatter(df_trimmed['total_sqft'], df_trimmed['price'], alpha=0.5)
105 plt.title("Total Sqft vs Price")
106 plt.xlabel("Total Sqft")
107 plt.ylabel("Price")
108
109 plt.subplot(1, 2, 2)
```



68 %

Help Debugger Plots Files

Console 1/A

```
Kurtosis before transformation: 0.6784399584501744
Skewness after transformation: -0.12988210833124453
Kurtosis after transformation: 1.3369230684904956
Correlation Matrix:
                  total_sqft      bath  ...   z_score  price_per_sqft_log
total_sqft          1.000000  0.455416  ...  0.136414            0.057160
bath                0.455416  1.000000  ...  0.348669            0.320142
price               0.660443  0.571890  ...  0.612102            0.573693
bhk                 0.426239  0.881677  ...  0.303220            0.270817
price_per_sqft      0.136414  0.348669  ...  1.000000            0.970182
z_score             0.136414  0.348669  ...  1.000000            0.970182
price_per_sqft_log  0.057160  0.320142  ...  0.970182            1.000000

[7 rows x 7 columns]
```

IPython Console History

Inline Conda: spyder-runtime (Python 3.11.11) LSP: Python Line 90, Col 1 UTF-8 CRLF RW Mem 88%

Spyder

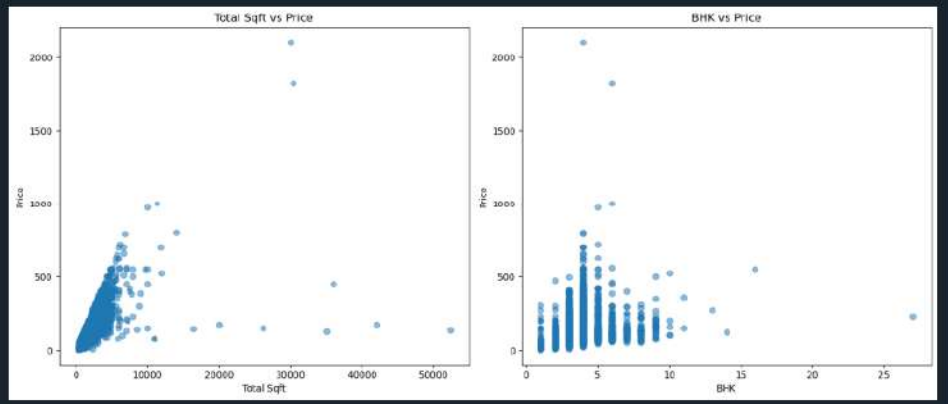
File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\sayan

C:\Users\sayan\.spyder-py3\main_program.py

main_program.py

```
79  plt.title("Histogram of Price per Sqft (After IQR Trimming)")
80  plt.xlabel("Price per Sqft")
81  plt.ylabel("Frequency")
82  plt.show()
83
84  # Skewness and Kurtosis before and after log transformation
85  skewness_before = skew(df_trimmed['price_per_sqft'])
86  kurtosis_before = kurtosis(df_trimmed['price_per_sqft'])
87  df_trimmed['price_per_sqft_log'] = np.log1p(df_trimmed['price_per_sqft'])
88  skewness_after = skew(df_trimmed['price_per_sqft_log'])
89  kurtosis_after = kurtosis(df_trimmed['price_per_sqft_log'])
90
91  print("Skewness before transformation:", skewness_before)
92  print("Kurtosis before transformation:", kurtosis_before)
93  print("Skewness after transformation:", skewness_after)
94  print("Kurtosis after transformation:", kurtosis_after)
95
96  # Compute correlation matrix for numerical columns only
97  numerical_df = df_trimmed.select_dtypes(include=[np.number])
98  corr_matrix = numerical_df.corr()
99  print("Correlation Matrix:\n", corr_matrix)
100
101 # Scatter plots
102 plt.figure(figsize=(14, 6))
103 plt.subplot(1, 2, 1)
104 plt.scatter(df_trimmed['total_sqft'], df_trimmed['price'], alpha=0.5)
105 plt.title("Total Sqft vs Price")
106 plt.xlabel("Total Sqft")
107 plt.ylabel("Price")
108
109 plt.subplot(1, 2, 2)
110 plt.scatter(df_trimmed['bhk'], df_trimmed['price'], alpha=0.5)
111 plt.title("BHK vs Price")
112 plt.xlabel("BHK")
113 plt.ylabel("Price")
114
115 plt.tight_layout()
116 plt.show()
117
```



57 %

Help Debugger Plots Files

Console 1/A

```
Kurtosis before transformation: 0.6784399584501744
Skewness after transformation: -0.12988210833124453
Kurtosis after transformation: 1.3369230684904956
Correlation Matrix:
                 total_sqft      bath  ...   z_score  price_per_sqft_log
total_sqft         1.000000  0.455416  ...  0.136414            0.057160
bath               0.455416  1.000000  ...  0.348669            0.320142
price              0.660443  0.571890  ...  0.612102            0.573693
bhk                0.426239  0.881677  ...  0.303220            0.270817
price_per_sqft     0.136414  0.348669  ...  1.000000            0.970182
z_score            0.136414  0.348669  ...  1.000000            0.970182
price_per_sqft_log 0.057160  0.320142  ...  0.970182            1.000000

[7 rows x 7 columns]
```

IPython Console History

Inline Conda: spyder-runtime (Python 3.11.11) LSP: Python Line 90, Col 1 UTF-8 CRLF RW Mem 90%