```
 1 C:\Users\kavan\PyCharmMiscProject\.venv\Scripts\python.exe "C:\Users\kavan\
   PyCharmMiscProject\DATA ANALYST INTERNSHIP\task 10.py"
 2 Shape of dataset: (20, 7)
 3
 4 Dataset Info:
 5 <class 'pandas.core.frame.DataFrame'>
 6 RangeIndex: 20 entries, 0 to 19
 7 Data columns (total 7 columns):
 8  #   Column        Non-Null Count  Dtype
 9 ---  ------        --------------  -----
10  0   Id            20 non-null     int64
11  1   LotArea       20 non-null     int64
12  2   LotFrontage   20 non-null     float64
13  3   OverallQual   20 non-null     int64
14  4   YearBuilt     20 non-null     int64
15  5   MasVnrType    20 non-null     object
16  6   SalePrice     20 non-null     int64
17 dtypes: float64(1), int64(5), object(1)
18 memory usage: 1.2+ KB
19 None
20
21 First 5 rows:
22    Id  LotArea  LotFrontage  OverallQual  YearBuilt MasVnrType  SalePrice
23 0   1     8450         65.0            7       2003    BrkFace     208500
24 1   2     9600         80.0            6       1976    BrkFace     181500
25 2   3    11250         72.0            7       2001    BrkFace     223500
26 3   4     9550         70.0            7       1915    BrkFace     140000
```

```
27 4    5    14260    72.0    2000    Stone    250000
28
29 Descriptive Statistics:
30              Id       LotArea  ...    YearBuilt      SalePrice
31 count  20.000000     20.000000  ...   20.000000      20.000000
32 mean   10.50000   10415.100000  ... 1978.650000  202670.000000
33 std     5.91608    2404.765398  ...   28.084974   95420.311751
34 min     1.00000    6120.000000  ... 1915.000000  118000.000000
35 25%     5.75000    9275.000000  ... 1961.750000  138000.000000
36 50%    10.50000   10721.500000  ... 1984.500000  169250.000000
37 75%    15.25000   11418.500000  ... 2003.250000  230125.000000
38 max    20.00000   14260.000000  ... 2009.000000  500000.000000
39
40 [8 rows x 6 columns]
41
42 Missing Value Percentage:
43        Column  Missing_Percentage
44 0          Id                 0.0
45 1     LotArea                 0.0
46 2  LotFrontage                0.0
47 3  OverallQual                0.0
48 4    YearBuilt                0.0
49 5    MasVnrType               0.0
50 6    SalePrice                0.0
51
52 Number of rows with outliers: 5
53
```

```
54 Outliers capped using IQR method.
55
56 Correlation Matrix:
57                   Id      LotArea  LotFrontage  OverallQual  YearBuilt  SalePrice
58 Id          1.000000  -0.022926     0.063007    -0.084254   0.041338  -0.006718
59 LotArea    -0.022926   1.000000     0.711083     0.026375   0.512230   0.301421
60 LotFrontage 0.063007   0.711083     1.000000    -0.050843   0.541134   0.352427
61 OverallQual -0.084254  0.026375    -0.050843     1.000000   0.316234   0.620939
62 YearBuilt   0.041338   0.512230     0.541134     0.316234   1.000000   0.709158
63 SalePrice  -0.006718   0.301421     0.352427     0.620939   0.709158   1.000000
64
65 Cleaned dataset saved as cleaned_dataset.csv
66
67 Process finished with exit code 0
68
```