Multiple linear regression

Accessing the data

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

churn = pd.read_csv("chennai_house_price_prediction.csv")

churn
```

```
      PRT_ID        AREA  INT_SQFT  DIST_MAINROAD  N_BEDROOM  
N_BATHROOM  \
0     P03210   Karapakkam      1004            131        1.0
1.0
1     P09411   Anna Nagar      1986             26        2.0
1.0
2     P01812        Adyar       909             70        1.0
1.0
3     P05346    Velachery      1855             14        3.0
2.0
4     P06210   Karapakkam      1226             84        1.0
1.0
...      ...          ...       ...            ...        ...          .
..
7104  P03834   Karapakkam       598             51        1.0
1.0
7105  P10000    Velachery      1897             52        3.0
2.0
7106  P09594    Velachery      1614            152        2.0
1.0
7107  P06508   Karapakkam       787             40        1.0
1.0
7108  P09794    Velachery      1896            156        3.0
2.0

      N_ROOM    SALE_COND  PARK_FACIL   BUILDTYPE UTILITY_AVAIL 
STREET  \
0          3     AbNormal           1  Commercial        AllPub
Paved
1          5     AbNormal           0  Commercial        AllPub
Gravel
2          3     AbNormal           1  Commercial           ELO
Gravel
3          5       Family           0      Others       NoSewr 
Paved
4          3     AbNormal           1      Others        AllPub
Gravel
...      ...          ...         ...         ...           ...
```

```
...
7104       2      AdjLand           0      Others           ELO  No 
Access  
7105       5       Family           1      Others        NoSeWa  No 
Access  
7106       4  Normal Sale           0       House        NoSeWa     
Gravel  
7107       2      Partial           1  Commercial           ELO     
Paved  
7108       5      Partial           1      Others           ELO     
Paved  

     MZZONE  QS_ROOMS  QS_BATHROOM  QS_BEDROOM  QS_OVERALL  COMMIS  \
0         A       4.0          3.9         4.9       4.330  144400   
1        RH       4.9          4.2         2.5       3.765  304049   
2        RL       4.1          3.8         2.2       3.090   92114   
3         I       4.7          3.9         3.6       4.010   77042   
4         C       3.0          2.5         4.1       3.290   74063   
...     ...       ...          ...         ...         ...     ...   
7104     RM       3.0          2.2         2.4       2.520  107060   
7105     RH       3.6          4.5         3.3       3.920  205551   
7106      I       4.3          4.2         2.9       3.840  167028   
7107     RL       4.6          3.8         4.1       4.160  119098   
7108      I       3.1          3.5         4.3       3.640   79812   

      SALES_PRICE  
0         7600000  
1        21717770  
2        13159200  
3         9630290  
4         7406250  
...           ...  
7104      5353000  
7105     10818480  
7106      8351410  
7107      8507000  
7108      9976480  

[7109 rows x 19 columns]
```

```
churn.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7109 entries, 0 to 7108
Data columns (total 19 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   PRT_ID         7109 non-null   object 
 1   AREA           7109 non-null   object 
 2   INT_SQFT       7109 non-null   int64  
```

```
 3   DIST_MAINROAD  7109 non-null   int64
 4   N_BEDROOM      7108 non-null   float64
 5   N_BATHROOM     7104 non-null   float64
 6   N_ROOM         7109 non-null   int64
 7   SALE_COND      7109 non-null   object
 8   PARK_FACIL     7109 non-null   int64
 9   BUILDTYPE      7109 non-null   object
 10  UTILITY_AVAIL  7109 non-null   object
 11  STREET         7109 non-null   object
 12  MZZONE         7109 non-null   object
 13  QS_ROOMS       7109 non-null   float64
 14  QS_BATHROOM    7109 non-null   float64
 15  QS_BEDROOM     7109 non-null   float64
 16  QS_OVERALL     7061 non-null   float64
 17  COMMIS         7109 non-null   int64
 18  SALES_PRICE    7109 non-null   int64
dtypes: float64(6), int64(6), object(7)
memory usage: 1.0+ MB
```

data preprocessing

```
churn.iloc[0:10,0:15]
```

```
   PRT_ID        AREA  INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM  
N_ROOM  \
0  P03210  Karapakkam      1004            131        1.0         1.0  
3   
1  P09411  Anna Nagar      1986             26        2.0         1.0  
5   
2  P01812       Adyar       909             70        1.0         1.0  
3   
3  P05346   Velachery      1855             14        3.0         2.0  
5   
4  P06210  Karapakkam      1226             84        1.0         1.0  
3   
5  P00219    Chrompet      1220             36        2.0         1.0  
4   
6  P09105    Chrompet      1167            137        1.0         1.0  
3   
7  P09679   Velachery      1847            176        3.0         2.0  
5   
8  P03377    Chrompet       771            175        1.0         1.0  
2   
9  P09623   Velachery      1635             74        2.0         1.0  
4   

  SALE_COND  PARK_FACIL   BUILDTYPE UTILITY_AVAIL     STREET MZZONE  
QS_ROOMS  \
0  AbNormal           1  Commercial        AllPub      Paved      A  
4.0   
1  AbNormal           0  Commercial        AllPub     Gravel     RH  
```

```
4.9
2  AbNormal           1  Commercial          ELO     Gravel     RL
4.1
3    Family           0      Others       NoSewr       Paved      I
4.7
4  AbNormal           1      Others       AllPub     Gravel      C
3.0
5   Partial           0  Commercial       NoSeWa  No Access     RH
4.5
6   Partial           0       Other       AllPub  No Access     RL
3.6
7    Family           0  Commercial       AllPub     Gravel     RM
2.4
8   AdjLand           0      Others       NoSewr       Paved     RM
2.9
9  AbNormal           0      Others          ELO  No Access      I
3.1

   QS_BATHROOM
0          3.9
1          4.2
2          3.8
3          3.9
4          2.5
5          2.6
6          2.1
7          4.5
8          3.7
9          3.1
```

```
churn.iloc[0:10,15:]
```

```
   QS_BEDROOM  QS_OVERALL  COMMIS  SALES_PRICE
0         4.9       4.330  144400      7600000
1         2.5       3.765  304049     21717770
2         2.2       3.090   92114     13159200
3         3.6       4.010   77042      9630290
4         4.1       3.290   74063      7406250
5         3.1       3.320  198316     12394750
6         2.5       2.670   33955      8488790
7         2.1       3.260  235204     16800250
8         4.0       3.550   33236      8308970
9         3.3       3.160  121255      8083650
```

```
# target
y=churn['SALES_PRICE']
y
```

```
0        7600000
1       21717770
2       13159200
```

```
3        9630290
4        7406250
          ...
7104     5353000
7105    10818480
7106     8351410
7107     8507000
7108     9976480
Name: SALES_PRICE, Length: 7109, dtype: int64
```

```python
## features

X= churn.drop(['SALES_PRICE'], axis=1)
X
```

```
      PRT_ID        AREA  INT_SQFT  DIST_MAINROAD  N_BEDROOM
N_BATHROOM  \
0     P03210  Karapakkam      1004            131        1.0
1.0
1     P09411  Anna Nagar      1986             26        2.0
1.0
2     P01812       Adyar       909             70        1.0
1.0
3     P05346   Velachery      1855             14        3.0
2.0
4     P06210  Karapakkam      1226             84        1.0
1.0
...      ...         ...       ...            ...        ...         .
..
7104  P03834  Karapakkam       598             51        1.0
1.0
7105  P10000   Velachery      1897             52        3.0
2.0
7106  P09594   Velachery      1614            152        2.0
1.0
7107  P06508  Karapakkam       787             40        1.0
1.0
7108  P09794   Velachery      1896            156        3.0
2.0

      N_ROOM    SALE_COND  PARK_FACIL   BUILDTYPE UTILITY_AVAIL
STREET  \
0          3     AbNormal           1  Commercial        AllPub
Paved
1          5     AbNormal           0  Commercial        AllPub
Gravel
2          3     AbNormal           1  Commercial           ELO
Gravel
3          5       Family           0      Others       NoSewr
Paved
```

```
4           3     AbNormal           1      Others        AllPub
Gravel
...       ...          ...         ...         ...           ...
...
7104        2      AdjLand           0      Others           ELO   No
Access
7105        5       Family           1      Others        NoSeWa   No
Access
7106        4  Normal Sale           0       House        NoSeWa
Gravel
7107        2      Partial           1  Commercial           ELO
Paved
7108        5      Partial           1      Others           ELO
Paved

     MZZONE  QS_ROOMS  QS_BATHROOM  QS_BEDROOM  QS_OVERALL  COMMIS
0         A       4.0          3.9         4.9       4.330  144400
1        RH       4.9          4.2         2.5       3.765  304049
2        RL       4.1          3.8         2.2       3.090   92114
3         I       4.7          3.9         3.6       4.010   77042
4         C       3.0          2.5         4.1       3.290   74063
...     ...       ...          ...         ...         ...     ...
7104     RM       3.0          2.2         2.4       2.520  107060
7105     RH       3.6          4.5         3.3       3.920  205551
7106      I       4.3          4.2         2.9       3.840  167028
7107     RL       4.6          3.8         4.1       4.160  119098
7108      I       3.1          3.5         4.3       3.640   79812

[7109 rows x 18 columns]
```

```
# Drop irrelevant features

X_1 =X.drop(['PRT_ID','SALE_COND','UTILITY_AVAIL'],axis=1)

X_1
```

```
            AREA  INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM
N_ROOM  \
0     Karapakkam      1004            131        1.0         1.0
3
1     Anna Nagar      1986             26        2.0         1.0
5
2          Adyar       909             70        1.0         1.0
3
3      Velachery      1855             14        3.0         2.0
5
4     Karapakkam      1226             84        1.0         1.0
3
...          ...       ...            ...        ...         ...      .
..
```

```
7104  Karapakkam       598             51        1.0         1.0
2
7105   Velachery      1897             52        3.0         2.0
5
7106   Velachery      1614            152        2.0         1.0
4
7107  Karapakkam       787             40        1.0         1.0
2
7108   Velachery      1896            156        3.0         2.0
5

      PARK_FACIL   BUILDTYPE     STREET MZZONE  QS_ROOMS  QS_BATHROOM
\
0              1  Commercial      Paved      A       4.0          3.9

1              0  Commercial     Gravel     RH       4.9          4.2

2              1  Commercial     Gravel     RL       4.1          3.8

3              0      Others      Paved      I       4.7          3.9

4              1      Others     Gravel      C       3.0          2.5

...          ...         ...        ...    ...       ...          ...

7104           0      Others  No Access     RM       3.0          2.2

7105           1      Others  No Access     RH       3.6          4.5

7106           0       House     Gravel      I       4.3          4.2

7107           1  Commercial      Paved     RL       4.6          3.8

7108           1      Others      Paved      I       3.1          3.5


      QS_BEDROOM  QS_OVERALL  COMMIS
0            4.9       4.330  144400
1            2.5       3.765  304049
2            2.2       3.090   92114
3            3.6       4.010   77042
4            4.1       3.290   74063
...          ...         ...     ...
7104         2.4       2.520  107060
7105         3.3       3.920  205551
7106         2.9       3.840  167028
7107         4.1       4.160  119098
7108         4.3       3.640   79812
```

```
[7109 rows x 15 columns]
```

```python
X_1['AREA'].unique()
```

```
array(['Karapakkam', 'Anna Nagar', 'Adyar', 'Velachery', 'Chrompet',
       'KK Nagar', 'TNagar', 'T Nagar', 'Chrompt', 'Chrmpet',
'Karapakam',
       'Ana Nagar', 'Chormpet', 'Adyr', 'Velchery', 'Ann Nagar',
       'KKNagar'], dtype=object)
```

```python
X_1['BUILDTYPE'].unique()
```

```
array(['Commercial', 'Others', 'Other', 'House', 'Comercial'],
      dtype=object)
```

```python
X_1['STREET'].unique()
```

```
array(['Paved', 'Gravel', 'No Access', 'Pavd', 'NoAccess'],
dtype=object)
```

```python
X_1['MZZONE'].unique()
```

```
array(['A', 'RH', 'RL', 'I', 'C', 'RM'], dtype=object)
```

```python
# Conver categorical to numeric using one hot encoding

X_1 =
pd.get_dummies(X_1,columns=['AREA','BUILDTYPE','STREET','MZZONE'])

X_1
```

```
      INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM  N_ROOM
PARK_FACIL  \
0         1004            131        1.0         1.0       3
1
1         1986             26        2.0         1.0       5
0
2          909             70        1.0         1.0       3
1
3         1855             14        3.0         2.0       5
0
4         1226             84        1.0         1.0       3
1
...        ...            ...        ...         ...     ...          .
..
7104       598             51        1.0         1.0       2
0
7105      1897             52        3.0         2.0       5
1
7106      1614            152        2.0         1.0       4
0
```

```
7107       787            40         1.0          1.0       2
1
7108      1896           156         3.0          2.0       5
1

      QS_ROOMS  QS_BATHROOM  QS_BEDROOM  QS_OVERALL  ...  STREET_No
Access  \
0          4.0          3.9         4.9       4.330  ...
0
1          4.9          4.2         2.5       3.765  ...
0
2          4.1          3.8         2.2       3.090  ...
0
3          4.7          3.9         3.6       4.010  ...
0
4          3.0          2.5         4.1       3.290  ...
0
...        ...          ...         ...         ...  ...
...
7104       3.0          2.2         2.4       2.520  ...
1
7105       3.6          4.5         3.3       3.920  ...
1
7106       4.3          4.2         2.9       3.840  ...
0
7107       4.6          3.8         4.1       4.160  ...
0
7108       3.1          3.5         4.3       3.640  ...
0

      STREET_NoAccess  STREET_Pavd  STREET_Paved  MZZONE_A
MZZONE_C  \
0                   0            0             1         1         0

1                   0            0             0         0         0

2                   0            0             0         0         0

3                   0            0             1         0         0

4                   0            0             0         0         1

...               ...          ...           ...       ...       ...

7104                0            0             0         0         0

7105                0            0             0         0         0

7106                0            0             0         0         0
```

```
7107                0            0              1         0         0

7108                0            0              1         0         0


      MZZONE_I  MZZONE_RH  MZZONE_RL  MZZONE_RM
0            0          0          0          0
1            0          1          0          0
2            0          0          1          0
3            1          0          0          0
4            0          0          0          0
...        ...        ...        ...        ...
7104         0          0          0          1
7105         0          1          0          0
7106         1          0          0          0
7107         0          0          1          0
7108         1          0          0          0

[7109 rows x 44 columns]
```

```
X_1.iloc[0:10,0:15]
```

```
   INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM  N_ROOM  PARK_FACIL
\
0      1004            131        1.0         1.0       3           1

1      1986             26        2.0         1.0       5           0

2       909             70        1.0         1.0       3           1

3      1855             14        3.0         2.0       5           0

4      1226             84        1.0         1.0       3           1

5      1220             36        2.0         1.0       4           0

6      1167            137        1.0         1.0       3           0

7      1847            176        3.0         2.0       5           0

8       771            175        1.0         1.0       2           0

9      1635             74        2.0         1.0       4           0


   QS_ROOMS  QS_BATHROOM  QS_BEDROOM  QS_OVERALL  COMMIS
AREA_Adyar  \
0       4.0          3.9         4.9       4.330  144400           0
```

```
1       4.9          4.2          2.5        3.765  304049           0
2       4.1          3.8          2.2        3.090   92114           1
3       4.7          3.9          3.6        4.010   77042           0
4       3.0          2.5          4.1        3.290   74063           0
5       4.5          2.6          3.1        3.320  198316           0
6       3.6          2.1          2.5        2.670   33955           0
7       2.4          4.5          2.1        3.260  235204           0
8       2.9          3.7          4.0        3.550   33236           0
9       3.1          3.1          3.3        3.160  121255           0

   AREA_Adyr  AREA_Ana Nagar  AREA_Ann Nagar
0          0               0               0
1          0               0               0
2          0               0               0
3          0               0               0
4          0               0               0
5          0               0               0
6          0               0               0
7          0               0               0
8          0               0               0
9          0               0               0
```

```
X_1.iloc[0:10,15:]
```

```
   AREA_Anna Nagar  AREA_Chormpet  AREA_Chrmpet  AREA_Chrompet
AREA_Chrompt  \
0                0              0             0              0
0
1                1              0             0              0
0
2                0              0             0              0
0
3                0              0             0              0
0
4                0              0             0              0
0
5                0              0             0              1
0
6                0              0             0              1
```

```
0
7                 0              0              0              0
0
8                 0              0              0              1
0
9                 0              0              0              0
0

   AREA_KK Nagar  AREA_KKNagar  AREA_Karapakam  AREA_Karapakkam
AREA_T Nagar  \
0               0              0                0                1
0
1               0              0                0                0
0
2               0              0                0                0
0
3               0              0                0                0
0
4               0              0                0                1
0
5               0              0                0                0
0
6               0              0                0                0
0
7               0              0                0                0
0
8               0              0                0                0
0
9               0              0                0                0
0

   ...  STREET_No Access  STREET_NoAccess  STREET_Pavd
STREET_Paved  \
0  ...                 0                0            0             1

1  ...                 0                0            0             0

2  ...                 0                0            0             0

3  ...                 0                0            0             1

4  ...                 0                0            0             0

5  ...                 1                0            0             0

6  ...                 1                0            0             0

7  ...                 0                0            0             0
```

```
8  ...                0                0            0             1

9  ...                1                0            0             0


   MZZONE_A  MZZONE_C  MZZONE_I  MZZONE_RH  MZZONE_RL  MZZONE_RM
0         1         0         0          0          0          0
1         0         0         0          1          0          0
2         0         0         0          0          1          0
3         0         0         1          0          0          0
4         0         1         0          0          0          0
5         0         0         0          1          0          0
6         0         0         0          0          1          0
7         0         0         0          0          0          1
8         0         0         0          0          0          1
9         0         0         1          0          0          0

[10 rows x 29 columns]
```

```
y
```

```
0        7600000
1       21717770
2       13159200
3        9630290
4        7406250
          ...
7104     5353000
7105    10818480
7106     8351410
7107     8507000
7108     9976480
Name: SALES_PRICE, Length: 7109, dtype: int64
```

```
import statsmodels.api as sm
X_1 = sm.add_constant(X_1)
X_1
```

```
      const  INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM  N_ROOM  \
0       1.0      1004            131        1.0         1.0       3
1       1.0      1986             26        2.0         1.0       5
2       1.0       909             70        1.0         1.0       3
3       1.0      1855             14        3.0         2.0       5
4       1.0      1226             84        1.0         1.0       3
...     ...       ...            ...        ...         ...     ...
7104    1.0       598             51        1.0         1.0       2
7105    1.0      1897             52        3.0         2.0       5
7106    1.0      1614            152        2.0         1.0       4
7107    1.0       787             40        1.0         1.0       2
7108    1.0      1896            156        3.0         2.0       5
```

```
      PARK_FACIL  QS_ROOMS  QS_BATHROOM  QS_BEDROOM  ...  STREET_No
Access  \
0              1       4.0          3.9         4.9  ...
0
1              0       4.9          4.2         2.5  ...
0
2              1       4.1          3.8         2.2  ...
0
3              0       4.7          3.9         3.6  ...
0
4              1       3.0          2.5         4.1  ...
0
...          ...       ...          ...         ...  ...
...
7104           0       3.0          2.2         2.4  ...
1
7105           1       3.6          4.5         3.3  ...
1
7106           0       4.3          4.2         2.9  ...
0
7107           1       4.6          3.8         4.1  ...
0
7108           1       3.1          3.5         4.3  ...
0

      STREET_NoAccess  STREET_Pavd  STREET_Paved  MZZONE_A
MZZONE_C  \
0                   0            0             1         1         0

1                   0            0             0         0         0

2                   0            0             0         0         0

3                   0            0             1         0         0

4                   0            0             0         0         1

...               ...          ...           ...       ...       ...

7104                0            0             0         0         0

7105                0            0             0         0         0

7106                0            0             0         0         0

7107                0            0             1         0         0

7108                0            0             1         0         0
```

```
      MZZONE_I  MZZONE_RH  MZZONE_RL  MZZONE_RM
0            0          0          0          0
1            0          1          0          0
2            0          0          1          0
3            1          0          0          0
4            0          0          0          0
...        ...        ...        ...        ...
7104         0          0          0          1
7105         0          1          0          0
7106         1          0          0          0
7107         0          0          1          0
7108         1          0          0          0

[7109 rows x 45 columns]
```

```
X_1.shape
```

```
(7109, 45)
```

```
X_1['QS_OVERALL'].fillna(value=5, inplace=True)
```

```
X_1['N_BEDROOM'].fillna(value=1, inplace=True)
```

```
X_1['N_BATHROOM'].fillna(value=1, inplace=True)
```

Splitting data into train and test

```
from sklearn.model_selection import train_test_split
```

```
X_train_1, X_test_1, y_train_1, y_test_1
=train_test_split(X_1,y ,test_size =0.2, random_state=10)
```

```
X_train_1.shape, X_test_1.shape, y_train_1.shape, y_test_1.shape
```

```
((5687, 45), (1422, 45), (5687,), (1422,))
```

```
X_train_1
```

```
      const  INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM  N_ROOM  \
591     1.0       764             36        1.0         1.0       2
2651    1.0      1848             66        2.0         1.0       5
5737    1.0      1592             95        2.0         2.0       4
5601    1.0      1303             74        1.0         1.0       3
1114    1.0       790            200        1.0         1.0       2
...     ...       ...            ...        ...         ...     ...
1180    1.0      1042             82        1.0         1.0       3
3441    1.0      1091             49        1.0         1.0       3
1344    1.0      1864             46        3.0         2.0       5
4623    1.0       917            139        1.0         1.0       3
1289    1.0       897             79        1.0         1.0       3

      PARK_FACIL  QS_ROOMS  QS_BATHROOM  QS_BEDROOM  ...  STREET_No
```

```
Access  \
591            0       3.9          4.8          4.2  ...
0
2651           0       4.0          2.7          4.7  ...
0
5737           1       4.4          4.3          5.0  ...
1
5601           1       3.4          4.6          4.2  ...
0
1114           0       2.7          4.3          2.4  ...
0
...          ...       ...          ...          ...  ...
...
1180           1       3.0          3.4          2.7  ...
0
3441           1       2.7          4.8          2.4  ...
0
1344           1       2.7          2.0          3.1  ...
0
4623           1       2.1          3.7          3.4  ...
1
1289           1       2.2          2.2          5.0  ...
1

      STREET_NoAccess  STREET_Pavd  STREET_Paved  MZZONE_A
MZZONE_C  \
591                 0            0              0         0         0

2651                0            0              1         0         0

5737                0            0              0         1         0

5601                0            0              0         0         0

1114                0            0              0         0         0

...               ...          ...            ...       ...       ...

1180                0            0              1         0         0

3441                0            0              1         0         0

1344                0            0              0         0         0

4623                0            0              0         0         0

1289                0            0              0         0         0
```

```
      MZZONE_I  MZZONE_RH  MZZONE_RL  MZZONE_RM
591          0          0          0          1
2651         0          0          0          1
5737         0          0          0          0
5601         0          1          0          0
1114         1          0          0          0
...        ...        ...        ...        ...
1180         1          0          0          0
3441         1          0          0          0
1344         0          1          0          0
4623         0          0          1          0
1289         0          0          0          1

[5687 rows x 45 columns]
```

X_test_1

```
      const  INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM  N_ROOM  \
461     1.0      1563             57        2.0         1.0       4
3358    1.0      1580             35        1.0         1.0       4
3751    1.0      1807             83        2.0         1.0       5
2386    1.0      2495            191        4.0         2.0       6
1125    1.0      1420             30        2.0         2.0       4
...     ...       ...            ...        ...         ...     ...
6010    1.0      1738             19        1.0         1.0       4
4903    1.0      1690             59        2.0         1.0       4
6806    1.0      1297              0        1.0         1.0       3
3832    1.0      1934            183        2.0         1.0       5
364     1.0       802            161        1.0         1.0       3

      PARK_FACIL  QS_ROOMS  QS_BATHROOM  QS_BEDROOM  ...  STREET_No
Access  \
461            1       4.1          4.7         2.7  ...
0
3358           0       2.9          3.9         5.0  ...
0
3751           1       3.1          2.4         5.0  ...
0
2386           0       3.0          3.0         2.7  ...
0
1125           0       5.0          3.7         2.2  ...
0
...          ...       ...          ...         ...  ...
...
6010           0       3.2          3.4         3.2  ...
0
4903           1       2.7          3.4         4.3  ...
0
6806           1       3.1          3.2         3.0  ...
1
```

```
3832           0       4.6          3.2         4.6  ...
0
364            0       3.4          4.2         3.6  ...
1

      STREET_NoAccess  STREET_Pavd  STREET_Paved  MZZONE_A
MZZONE_C  \
461                 0            0             1         0         0

3358                0            0             1         0         0

3751                0            0             1         0         0

2386                0            0             0         0         0

1125                0            0             0         0         0

...               ...          ...           ...       ...       ...

6010                0            0             1         0         0

4903                0            0             0         0         1

6806                0            0             0         0         1

3832                0            0             0         0         0

364                 0            0             0         0         0


      MZZONE_I  MZZONE_RH  MZZONE_RL  MZZONE_RM
461          0          0          0          1
3358         0          0          0          1
3751         0          0          0          1
2386         0          0          0          1
1125         0          0          1          0
...        ...        ...        ...        ...
6010         0          0          0          1
4903         0          0          0          0
6806         0          0          0          0
3832         0          0          0          1
364          0          0          0          1

[1422 rows x 45 columns]
```

```
X_1.isnull().sum()
```

```
const                    0
INT_SQFT                 0
```

```
DIST_MAINROAD           0
N_BEDROOM               0
N_BATHROOM              0
N_ROOM                  0
PARK_FACIL              0
QS_ROOMS                0
QS_BATHROOM             0
QS_BEDROOM              0
QS_OVERALL              0
COMMIS                  0
AREA_Adyar              0
AREA_Adyr               0
AREA_Ana Nagar          0
AREA_Ann Nagar          0
AREA_Anna Nagar         0
AREA_Chormpet           0
AREA_Chrmpet            0
AREA_Chrompet           0
AREA_Chrompt            0
AREA_KK Nagar           0
AREA_KKNagar            0
AREA_Karapakam          0
AREA_Karapakkam         0
AREA_T Nagar            0
AREA_TNagar             0
AREA_Velachery          0
AREA_Velchery           0
BUILDTYPE_Comercial     0
BUILDTYPE_Commercial    0
BUILDTYPE_House         0
BUILDTYPE_Other         0
BUILDTYPE_Others        0
STREET_Gravel           0
STREET_No Access        0
STREET_NoAccess         0
STREET_Pavd             0
STREET_Paved            0
MZZONE_A                0
MZZONE_C                0
MZZONE_I                0
MZZONE_RH               0
MZZONE_RL               0
MZZONE_RM               0
dtype: int64
```

```
X_1['QS_OVERALL'].fillna(value=5, inplace=True)
```

```
X_1.isnull().sum()
```

```
const                   0
INT_SQFT                0
```

```
DIST_MAINROAD           0
N_BEDROOM               0
N_BATHROOM              0
N_ROOM                  0
PARK_FACIL              0
QS_ROOMS                0
QS_BATHROOM             0
QS_BEDROOM              0
QS_OVERALL              0
COMMIS                  0
AREA_Adyar              0
AREA_Adyr               0
AREA_Ana Nagar          0
AREA_Ann Nagar          0
AREA_Anna Nagar         0
AREA_Chormpet           0
AREA_Chrmpet            0
AREA_Chrompet           0
AREA_Chrompt            0
AREA_KK Nagar           0
AREA_KKNagar            0
AREA_Karapakam          0
AREA_Karapakkam         0
AREA_T Nagar            0
AREA_TNagar             0
AREA_Velachery          0
AREA_Velchery           0
BUILDTYPE_Comercial     0
BUILDTYPE_Commercial    0
BUILDTYPE_House         0
BUILDTYPE_Other         0
BUILDTYPE_Others        0
STREET_Gravel           0
STREET_No Access        0
STREET_NoAccess         0
STREET_Pavd             0
STREET_Paved            0
MZZONE_A                0
MZZONE_C                0
MZZONE_I                0
MZZONE_RH               0
MZZONE_RL               0
MZZONE_RM               0
dtype: int64
```

```python
X_1['N_BEDROOM'].fillna(value=1, inplace=True)
```

```python
X_1['N_BATHROOM'].fillna(value=1, inplace=True)
```

```python
X_1.isnull().sum()
```

```
const                    0
INT_SQFT                 0
DIST_MAINROAD            0
N_BEDROOM                0
N_BATHROOM               0
N_ROOM                   0
PARK_FACIL               0
QS_ROOMS                 0
QS_BATHROOM              0
QS_BEDROOM               0
QS_OVERALL               0
COMMIS                   0
AREA_Adyar               0
AREA_Adyr                0
AREA_Ana Nagar           0
AREA_Ann Nagar           0
AREA_Anna Nagar          0
AREA_Chormpet            0
AREA_Chrmpet             0
AREA_Chrompet            0
AREA_Chrompt             0
AREA_KK Nagar            0
AREA_KKNagar             0
AREA_Karapakam           0
AREA_Karapakkam          0
AREA_T Nagar             0
AREA_TNagar              0
AREA_Velachery           0
AREA_Velchery            0
BUILDTYPE_Comercial      0
BUILDTYPE_Commercial     0
BUILDTYPE_House          0
BUILDTYPE_Other          0
BUILDTYPE_Others         0
STREET_Gravel            0
STREET_No Access         0
STREET_NoAccess          0
STREET_Pavd              0
STREET_Paved             0
MZZONE_A                 0
MZZONE_C                 0
MZZONE_I                 0
MZZONE_RH                0
MZZONE_RL                0
MZZONE_RM                0
dtype: int64
```

Building the model

```
mlr_1 = sm.OLS(y_train_1, X_train_1)
```

```
mlr_1 = mlr_1.fit()
```

```
mlr_1.params
```

```
const                    2.700263e+06
INT_SQFT                 3.782721e+03
DIST_MAINROAD           -6.888493e+01
N_BEDROOM                2.823435e+05
N_BATHROOM              -4.031941e+05
N_ROOM                   1.206859e+05
PARK_FACIL               9.914752e+05
QS_ROOMS                -1.695507e+04
QS_BATHROOM             -1.560937e+04
QS_BEDROOM               2.789208e+03
QS_OVERALL               8.513923e+04
COMMIS                   2.840802e+00
AREA_Adyar               4.579575e+05
AREA_Adyr                1.168757e+06
AREA_Ana Nagar           4.596758e+06
AREA_Ann Nagar           1.213446e+06
AREA_Anna Nagar          1.912341e+06
AREA_Chormpet           -9.981541e+03
AREA_Chrmpet             5.050556e+05
AREA_Chrompet           -3.400180e+03
AREA_Chrompt             1.940305e+05
AREA_KK Nagar           -1.692252e+06
AREA_KKNagar            -2.066746e+06
AREA_Karapakam          -2.502585e+06
AREA_Karapakkam         -2.033962e+06
AREA_T Nagar             2.051998e+06
AREA_TNagar              2.221258e+06
AREA_Velachery          -1.474939e+06
AREA_Velchery           -1.837474e+06
BUILDTYPE_Comercial      2.026595e+06
BUILDTYPE_Commercial     3.164816e+06
BUILDTYPE_House         -1.349369e+06
BUILDTYPE_Other         -4.606319e+05
BUILDTYPE_Others        -6.811467e+05
STREET_Gravel            1.116480e+06
STREET_No Access         4.600433e+03
STREET_NoAccess          4.874171e+05
STREET_Pavd              4.796557e+05
STREET_Paved             6.121090e+05
MZZONE_A                -9.312163e+05
MZZONE_C                -4.019670e+05
MZZONE_I                 1.401905e+05
MZZONE_RH                7.031580e+05
MZZONE_RL                1.286094e+06
MZZONE_RM                1.904003e+06
dtype: float64
```

Diagnosing the model

```
mlr_1.summary2()

<class 'statsmodels.iolib.summary2.Summary'>
"""
                          Results: Ordinary least squares
======================================================================
====================
Model:                       OLS                     Adj. R-squared:
0.952
Dependent Variable:          SALES_PRICE             AIC:
171092.7797
Date:                        2022-11-10 01:46        BIC:
171365.2632
No. Observations:            5687                    Log-Likelihood:
-85505.
Df Model:                    40                      F-statistic:
2827.
Df Residuals:                5646                    Prob (F-
statistic):        0.00
R-squared:                   0.952                   Scale:
6.7625e+11
----------------------------------------------------------------------
--------------------
                         Coef.       Std.Err.     t     P>|t|
[0.025        0.975]
----------------------------------------------------------------------
--------------------
const                 2700262.6253 133901.4227  20.1660 0.0000
2437764.3863  2962760.8642
INT_SQFT                 3782.7207    101.9670  37.0975 0.0000
3582.8262     3982.6152
DIST_MAINROAD             -68.8849    190.3813  -0.3618 0.7175     -
442.1055      304.3356
N_BEDROOM              282343.4565  50202.0172   5.6241 0.0000
183928.2131   380758.6998
N_BATHROOM            -403194.0971  50465.7131  -7.9895 0.0000  -
502126.2859  -304261.9084
N_ROOM                 120685.8893  49649.9672   2.4307 0.0151
23352.8760   218018.9026
PARK_FACIL             991475.2304  22074.5239  44.9149 0.0000
948200.6817  1034749.7791
QS_ROOMS               -16955.0748  24316.7456  -0.6973 0.4857   -
64625.2397    30715.0901
QS_BATHROOM            -15609.3741  26057.3976  -0.5990 0.5492   -
66691.8856    35473.1375
QS_BEDROOM               2789.2083  29137.5059   0.0957 0.9237   -
54331.4990    59909.9156
QS_OVERALL              85139.2250  70710.7591   1.2040 0.2286   -
53481.0328   223759.4829
```

```
COMMIS                        2.8408      0.2011  14.1231 0.0000
2.4465        3.2351
AREA_Adyar              457957.5015 119915.1918   3.8190 0.0001
222877.6492   693037.3538
AREA_Adyr              1168756.9164 784444.8204   1.4899 0.1363  -
369056.3478  2706570.1806
AREA_Ana Nagar         4596757.6997 784668.4589   5.8582 0.0000
3058506.0179  6135009.3814
AREA_Ann Nagar         1213446.1788 560789.6282   2.1638 0.0305
114083.0288  2312809.3289
AREA_Anna Nagar        1912341.4195 122213.1100  15.6476 0.0000
1672756.7645  2151926.0745
AREA_Chormpet            -9981.5407 464024.4963  -0.0215 0.9828  -
919647.8510   899684.7697
AREA_Chrmpet            505055.5965 337033.3664   1.4985 0.1341  -
155659.3039  1165770.4970
AREA_Chrompet            -3400.1804 115578.4249  -0.0294 0.9765  -
229978.3033   223177.9425
AREA_Chrompt            194030.5498 296549.9201   0.6543 0.5129  -
387321.2404   775382.3400
AREA_KK Nagar         -1692251.6215 119992.7799 -14.1029 0.0000 -
1927483.5764 -1457019.6667
AREA_KKNagar          -2066746.4053 783928.1856  -2.6364 0.0084 -
3603546.8668  -529945.9439
AREA_Karapakam        -2502585.1522 463351.4545  -5.4011 0.0000 -
3410932.0419 -1594238.2625
AREA_Karapakkam       -2033961.5584 117218.1490 -17.3519 0.0000 -
2263754.1706 -1804168.9462
AREA_T Nagar           2051997.7054 124485.9407  16.4838 0.0000
1807957.4290  2296037.9818
AREA_TNagar            2221257.9227 405559.9557   5.4770 0.0000
1426204.5765  3016311.2689
AREA_Velachery        -1474938.7710 118305.6171 -12.4672 0.0000 -
1706863.2385 -1243014.3035
AREA_Velchery         -1837473.6355 784870.6218  -2.3411 0.0193 -
3376121.6342  -298825.6369
BUILDTYPE_Comercial    2026594.7061 481654.6730   4.2076 0.0000
1082366.4752  2970822.9370
BUILDTYPE_Commercial   3164815.8045 111404.3121  28.4084 0.0000
2946420.5467  3383211.0624
BUILDTYPE_House       -1349369.2327 110654.0102 -12.1945 0.0000 -
1566293.6105 -1132444.8549
BUILDTYPE_Other        -460631.9271 181581.7847  -2.5368 0.0112  -
816601.9965  -104661.8578
BUILDTYPE_Others       -681146.7255 110814.9170  -6.1467 0.0000  -
898386.5426  -463906.9085
STREET_Gravel          1116480.4732  78695.8205  14.1873 0.0000
962206.4269  1270754.5196
STREET_No Access          4600.4335  79117.7594   0.0581 0.9536  -
150500.7753   159701.6422
```

```
STREET_NoAccess        487417.0973 282303.8921   1.7266 0.0843   -
66007.0040  1040841.1987
STREET_Pavd            479655.6547 214706.0734   2.2340 0.0255
58749.2519   900562.0575
STREET_Paved           612108.9664  78715.8453   7.7762 0.0000
457795.6638   766422.2690
MZZONE_A              -931216.3144  43206.4938 -21.5527 0.0000 -
1015917.6440  -846514.9848
MZZONE_C              -401967.0306  42368.1625  -9.4875 0.0000  -
485024.9087  -318909.1525
MZZONE_I               140190.5487  43827.7971   3.1987 0.0014
54271.2260   226109.8714
MZZONE_RH              703158.0169  32263.5037  21.7942 0.0000
639909.1527   766406.8811
MZZONE_RL             1286094.4877  32511.6068  39.5580 0.0000
1222359.2461  1349829.7294
MZZONE_RM             1904002.9169  32809.7085  58.0317 0.0000
1839683.2814  1968322.5524
----------------------------------------------------------------------
--------------------
Omnibus:                   132.790           Durbin-Watson:
2.043
Prob(Omnibus):             0.000             Jarque-Bera (JB):
192.697
Skew:                      0.255             Prob(JB):
0.000
Kurtosis:                  3.743             Condition No.:
11448429555476168
======================================================================
====================
* The condition number is large (1e+16). This might indicate
strong
multicollinearity or other numerical problems.
"""
```

Note

only ODI_WKTS and BASE PRICE are relevant features.

Multicollinearity

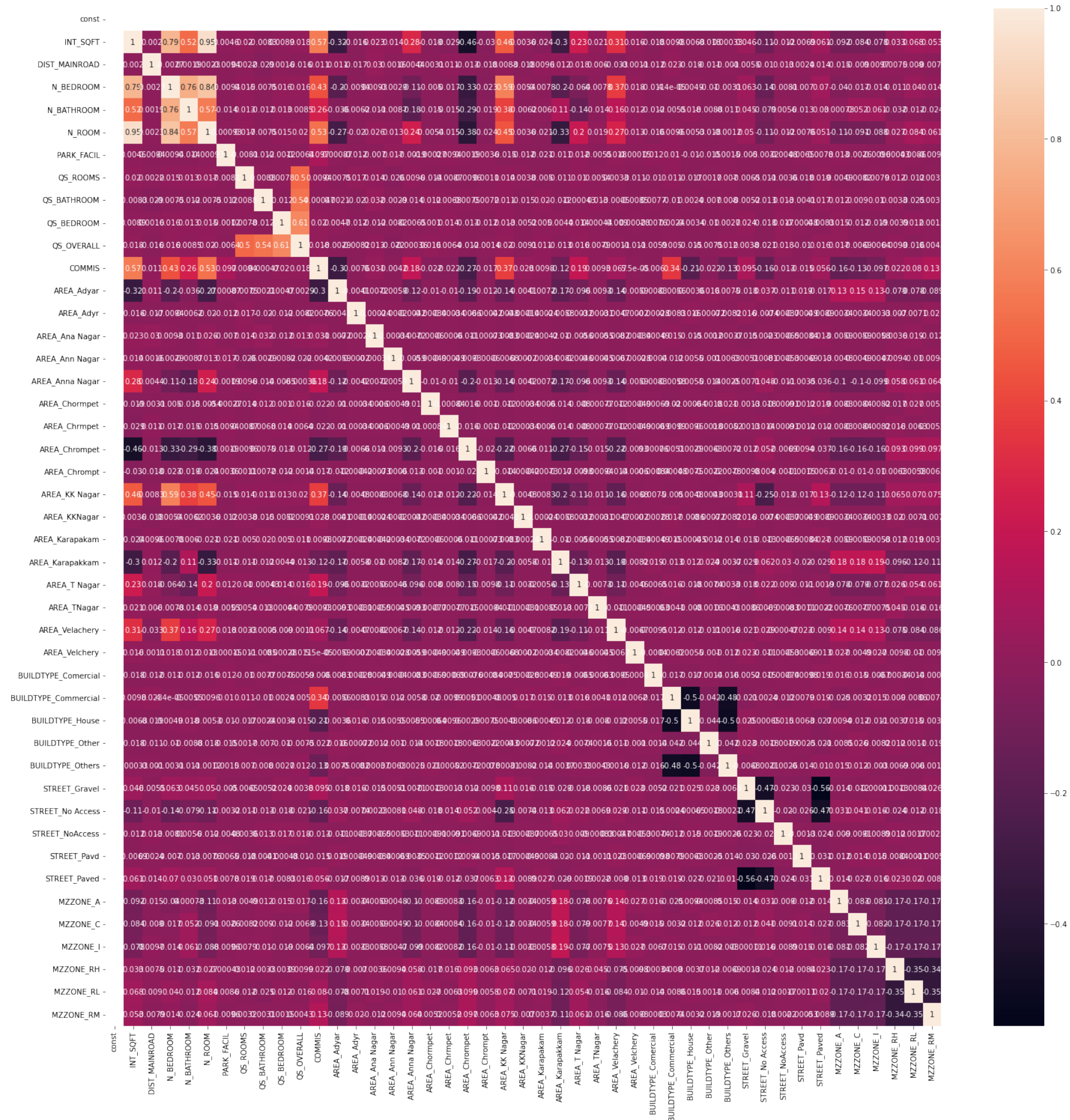
```
from statsmodels.stats.outliers_influence import
variance_inflation_factor

def var_inf_factor(data):
    vif=pd.DataFrame()
    vif['Feature']=data.columns
    vif['VIF_Value']= [variance_inflation_factor(data.values,i) for i
in range(data.shape[1])]
    print(vif)

var_inf_factor(X_1)
```

```
C:\Users\hp\anaconda3\lib\site-packages\statsmodels\regression\
linear_model.py:1736: RuntimeWarning: divide by zero encountered in
double_scalars
  return 1 - self.ssr/self.centered_tss
C:\Users\hp\anaconda3\lib\site-packages\statsmodels\stats\
outliers_influence.py:195: RuntimeWarning: divide by zero encountered
in double_scalars
  vif = 1. / (1. - r_squared_i)
```

	Feature	VIF_Value
0	const	0.000000
1	INT_SQFT	18.539409
2	DIST_MAINROAD	1.006672
3	N_BEDROOM	13.567914
4	N_BATHROOM	3.577235
5	N_ROOM	21.475609
6	PARK_FACIL	1.023800
7	QS_ROOMS	4.012710
8	QS_BATHROOM	4.620034
9	QS_BEDROOM	5.649446
10	QS_OVERALL	12.293831
11	COMMIS	2.110629
12	AREA_Adyar	inf
13	AREA_Adyr	inf
14	AREA_Ana Nagar	inf
15	AREA_Ann Nagar	inf
16	AREA_Anna Nagar	inf
17	AREA_Chormpet	inf
18	AREA_Chrmpet	inf
19	AREA_Chrompet	inf
20	AREA_Chrompt	inf
21	AREA_KK Nagar	inf
22	AREA_KKNagar	inf
23	AREA_Karapakam	inf
24	AREA_Karapakkam	inf
25	AREA_T Nagar	inf
26	AREA_TNagar	inf
27	AREA_Velachery	inf
28	AREA_Velchery	inf
29	BUILDTYPE_Comercial	inf
30	BUILDTYPE_Commercial	inf
31	BUILDTYPE_House	inf
32	BUILDTYPE_Other	inf
33	BUILDTYPE_Others	inf
34	STREET_Gravel	inf
35	STREET_No Access	inf
36	STREET_NoAccess	inf
37	STREET_Pavd	inf
38	STREET_Paved	inf
39	MZZONE_A	inf

```
40              MZZONE_C        inf
41              MZZONE_I        inf
42             MZZONE_RH        inf
43             MZZONE_RL        inf
44             MZZONE_RM        inf
```

```
plt.figure(figsize=(25,25))
sns.heatmap(X_1.corr(),annot=True);
```




```
features_to_drop_1=['N_BEDROOM','N_BATHROOM','N_ROOM']

features_2=list(set(X_1.columns)-set(features_to_drop_1))

features_2
```

```
['STREET_NoAccess',
 'AREA_Chrmpet',
 'AREA_Adyr',
 'AREA_Velachery',
 'BUILDTYPE_Others',
 'QS_ROOMS',
 'QS_BATHROOM',
 'BUILDTYPE_Comercial',
 'AREA_TNagar',
 'AREA_Karapakkam',
 'AREA_Adyar',
 'AREA_Chrompt',
 'STREET_Gravel',
 'AREA_Chrompet',
 'MZZONE_C',
 'MZZONE_RL',
 'AREA_Ana Nagar',
 'MZZONE_RM',
 'STREET_No Access',
 'DIST_MAINROAD',
 'AREA_KK Nagar',
 'MZZONE_I',
 'BUILDTYPE_Commercial',
 'QS_OVERALL',
 'AREA_Anna Nagar',
 'INT_SQFT',
 'COMMIS',
 'AREA_Chormpet',
 'BUILDTYPE_Other',
 'BUILDTYPE_House',
 'AREA_Velchery',
 'MZZONE_RH',
 'AREA_KKNagar',
 'PARK_FACIL',
 'STREET_Pavd',
 'AREA_Karapakam',
 'STREET_Paved',
 'const',
 'AREA_Ann Nagar',
 'QS_BEDROOM',
 'AREA_T Nagar',
 'MZZONE_A']
```

```
len(features_2)
```

```
42
```

The new feature set

```
X_2=X_1[features_2]
```

```
X_2
```

```
      STREET_NoAccess  AREA_Chrmpet  AREA_Adyr  AREA_Velachery  \
0                   0             0          0               0
1                   0             0          0               0
2                   0             0          0               0
3                   0             0          0               1
4                   0             0          0               0
...               ...           ...        ...             ...
7104                0             0          0               0
7105                0             0          0               1
7106                0             0          0               1
7107                0             0          0               0
7108                0             0          0               1

      BUILDTYPE_Others  QS_ROOMS  QS_BATHROOM  BUILDTYPE_Comercial  \
0                    0       4.0          3.9                    0
1                    0       4.9          4.2                    0
2                    0       4.1          3.8                    0
3                    1       4.7          3.9                    0
4                    1       3.0          2.5                    0
...                ...       ...          ...                  ...
7104                 1       3.0          2.2                    0
7105                 1       3.6          4.5                    0
7106                 0       4.3          4.2                    0
7107                 0       4.6          3.8                    0
7108                 1       3.1          3.5                    0

      AREA_TNagar  AREA_Karapakkam  ...  AREA_KKNagar  PARK_FACIL  \
0               0                1  ...             0           1
1               0                0  ...             0           0
2               0                0  ...             0           1
3               0                0  ...             0           0
4               0                1  ...             0           1
...           ...              ...  ...           ...         ...
7104            0                1  ...             0           0
7105            0                0  ...             0           1
7106            0                0  ...             0           0
7107            0                1  ...             0           1
7108            0                0  ...             0           1

      STREET_Pavd  AREA_Karapakam  STREET_Paved  const  AREA_Ann Nagar
\
0               0               0             1    1.0               0

1               0               0             0    1.0               0

2               0               0             0    1.0               0

3               0               0             1    1.0               0
```

```
4                0                0              0    1.0                0
...             ...              ...            ...   ...              ...
7104             0                0              0    1.0                0
7105             0                0              0    1.0                0
7106             0                0              0    1.0                0
7107             0                0              1    1.0                0
7108             0                0              1    1.0                0

      QS_BEDROOM  AREA_T Nagar  MZZONE_A
0            4.9             0         1
1            2.5             0         0
2            2.2             0         0
3            3.6             0         0
4            4.1             0         0
...          ...           ...       ...
7104         2.4             0         0
7105         3.3             0         0
7106         2.9             0         0
7107         4.1             0         0
7108         4.3             0         0

[7109 rows x 42 columns]
```

```
# Splitting X_2 to train and test

X_train_2,X_test_2,y_train_2,y_test_2=train_test_split(
                 X_2,y,test_size=0.2, random_state=10)

X_train_2.shape,X_test_2.shape,y_train_2.shape,y_test_2.shape
```

```
((5687, 42), (1422, 42), (5687,), (1422,))
```

```
X_train_2
```

```
      STREET_NoAccess  AREA_Chrmpet  AREA_Adyr  AREA_Velachery  \
591                 0             0          0               0
2651                0             0          0               0
5737                0             0          0               0
5601                0             0          0               0
1114                0             0          0               0
...               ...           ...        ...             ...
1180                0             0          0               0
3441                0             0          0               0
```

```
1344                0             0          0               1
4623                0             0          0               0
1289                0             0          0               0

      BUILDTYPE_Others  QS_ROOMS  QS_BATHROOM  BUILDTYPE_Comercial  \
591                  1       3.9          4.8                    0
2651                 0       4.0          2.7                    0
5737                 0       4.4          4.3                    0
5601                 1       3.4          4.6                    0
1114                 0       2.7          4.3                    0
...                ...       ...          ...                  ...
1180                 0       3.0          3.4                    0
3441                 0       2.7          4.8                    0
1344                 0       2.7          2.0                    0
4623                 0       2.1          3.7                    0
1289                 0       2.2          2.2                    0

      AREA_TNagar  AREA_Karapakkam  ...  AREA_KKNagar  PARK_FACIL  \
591             0                0  ...             0           0
2651            0                0  ...             0           0
5737            0                1  ...             0           1
5601            0                0  ...             0           1
1114            0                0  ...             0           0
...           ...              ...  ...           ...         ...
1180            0                1  ...             0           1
3441            0                1  ...             0           1
1344            0                0  ...             0           1
4623            0                0  ...             0           1
1289            0                0  ...             0           1

      STREET_Pavd  AREA_Karapakam  STREET_Paved  const  AREA_Ann Nagar
\
591             0               0             0    1.0               0

2651            0               0             1    1.0               0

5737            0               0             0    1.0               0

5601            0               0             0    1.0               0

1114            0               0             0    1.0               0

...           ...             ...           ...    ...             ...

1180            0               0             1    1.0               0

3441            0               0             1    1.0               0

1344            0               0             0    1.0               0
```

```
4623            0                0             0    1.0              0

1289            0                0             0    1.0              0


      QS_BEDROOM  AREA_T Nagar  MZZONE_A
591          4.2             0         0
2651         4.7             0         0
5737         5.0             0         1
5601         4.2             0         0
1114         2.4             0         0
...          ...           ...       ...
1180         2.7             0         0
3441         2.4             0         0
1344         3.1             0         0
4623         3.4             0         0
1289         5.0             0         0

[5687 rows x 42 columns]
```

```
# Building the model

mlr_2=sm.OLS(y_train_2,X_train_2)

# Fit

mlr_2=mlr_2.fit()

# Diagnosis

mlr_2.summary2()
```

```
<class 'statsmodels.iolib.summary2.Summary'>
"""
                             Results: Ordinary least squares
======================================================================
====================
Model:                      OLS                     Adj. R-squared:
0.951
Dependent Variable:         SALES_PRICE             AIC:
171190.4518
Date:                       2022-11-10 01:46        BIC:
171442.9974
No. Observations:           5687                    Log-Likelihood:
-85557.
Df Model:                   37                      F-statistic:
2999.
Df Residuals:               5649                    Prob (F-
statistic):        0.00
```

R-squared: 0.952 Scale: 6.8833e+11

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
STREET_NoAccess	457365.8240	284693.8407	1.6065	0.1082	-100743.4313	1015475.0794
AREA_Chrmpet	533891.4931	339315.5303	1.5734	0.1157	-131297.2496	1199080.2358
AREA_Adyr	1064968.5697	790815.7195	1.3467	0.1781	-485333.9281	2615271.0676
AREA_Velachery	-1410884.3476	117940.7831	-11.9627	0.0000	-1642093.5739	-1179675.1213
BUILDTYPE_Others	-694861.8194	111581.8528	-6.2274	0.0000	-913605.1004	-476118.5383
QS_ROOMS	-17817.0027	24531.5459	-0.7263	0.4677	-65908.2533	30274.2478
QS_BATHROOM	-15864.7060	26287.0808	-0.6035	0.5462	-67397.4791	35668.0671
BUILDTYPE_Comercial	2024077.3383	485922.5623	4.1654	0.0000	1071482.5129	2976672.1636
AREA_TNagar	2193080.4458	407799.8187	5.3778	0.0000	1393636.1986	2992524.6931
AREA_Karapakkam	-2182255.6491	116351.3429	-18.7557	0.0000	-2410348.9623	-1954162.3359
AREA_Adyar	369299.4733	119363.2056	3.0939	0.0020	135301.7526	603297.1939
AREA_Chrompt	167866.6891	298896.2801	0.5616	0.5744	-418084.8015	753818.1797
STREET_Gravel	1115206.0058	79219.8938	14.0773	0.0000	959904.5921	1270507.4194
AREA_Chrompet	-23031.3189	116138.4901	-0.1983	0.8428	-250707.3588	204644.7210
MZZONE_C	-403677.5243	42435.3928	-9.5128	0.0000	-486867.1901	-320487.8585
MZZONE_RL	1283375.1264	32425.7394	39.5789	0.0000	1219808.2251	1346942.0276
AREA_Ana Nagar	4717499.3646	790964.8076	5.9642	0.0000	3166904.5966	6268094.1326
MZZONE_RM	1894055.6775	32682.0972	57.9539	0.0000	1829986.2164	1958125.1386
STREET_No Access	6562.3025	79666.3930	0.0824	0.9344	-149614.4211	162739.0262
DIST_MAINROAD	-73.1462	192.0718	-0.3808	0.7033	-449.6807	303.3883
AREA_KK Nagar	-1592795.0856	118915.0614	-13.3944	0.0000	-1825914.2715	-1359675.8997

```
MZZONE_I               133679.4603  43999.0069   3.0382 0.0024
47424.5104   219934.4102
BUILDTYPE_Commercial  3149982.5340 112110.9485  28.0970 0.0000
2930202.0223  3369763.0458
QS_OVERALL              84157.9464  71337.2284   1.1797 0.2382   -
55690.4161   224006.3089
AREA_Anna Nagar       1854703.1179 118474.8331  15.6548 0.0000
1622446.9485  2086959.2873
INT_SQFT                 4145.0527     47.0418  88.1142 0.0000
4052.8327     4237.2728
COMMIS                      2.8435      0.2029  14.0165 0.0000
2.4458        3.2412
AREA_Chormpet          135147.3100 467725.5387   0.2889 0.7726  -
781774.3611  1052068.9810
BUILDTYPE_Other       -440033.9664 183053.9274  -2.4038 0.0163  -
798889.9602   -81177.9727
BUILDTYPE_House      -1367109.8480 111428.5860 -12.2689 0.0000 -
1585552.6672 -1148667.0287
AREA_Velchery        -1853425.3538 791470.6019  -2.3417 0.0192 -
3405011.6727  -301839.0349
MZZONE_RH              687374.7461  32149.9106  21.3803 0.0000
624348.5752   750400.9169
AREA_KKNagar         -1913331.1232 790634.2068  -2.4200 0.0156 -
3463277.7866  -363384.4598
PARK_FACIL             988397.8002  22262.7629  44.3969 0.0000
944754.2357  1032041.3648
STREET_Pavd            480414.7334 216396.3959   2.2201 0.0265
56194.6973   904634.7696
AREA_Karapakam       -2597074.5012 466350.7250  -5.5689 0.0000 -
3511301.0094 -1682847.9930
STREET_Paved           612505.3728  79259.7216   7.7278 0.0000
457125.8813   767884.8643
const                 2672054.2386 131633.0004  20.2993 0.0000
2414003.0083  2930105.4689
AREA_Ann Nagar        1196810.8868 564824.3208   2.1189 0.0341
89538.3152  2304083.4584
QS_BEDROOM               4701.5959  29394.3617   0.1599 0.8729   -
52922.6410    62325.8329
AREA_T Nagar          2011584.2670 121029.2920  16.6206 0.0000
1774320.3773  2248848.1567
MZZONE_A              -922753.2473  43119.4426 -21.3999 0.0000 -
1007283.9134  -838222.5812
----------------------------------------------------------------------
--------------------
Omnibus:                     215.321          Durbin-Watson:
2.042
Prob(Omnibus):               0.000            Jarque-Bera (JB):
317.681
Skew:                        0.366            Prob(JB):
0.000
```

```
Kurtosis:                   3.897           Condition No.:
22487865609621064
======================================================================
====================
* The condition number is large (2e+16). This might indicate
strong
multicollinearity or other numerical problems.
"""
```

```python
# VIF

var_inf_factor(X_2)
```

```
C:\Users\hp\anaconda3\lib\site-packages\statsmodels\stats\
outliers_influence.py:195: RuntimeWarning: divide by zero encountered
in double_scalars
  vif = 1. / (1. - r_squared_i)

                 Feature  VIF_Value
0        STREET_NoAccess        inf
1           AREA_Chrmpet        inf
2              AREA_Adyr        inf
3         AREA_Velachery        inf
4       BUILDTYPE_Others        inf
5               QS_ROOMS   4.011797
6            QS_BATHROOM   4.619991
7    BUILDTYPE_Comercial        inf
8            AREA_TNagar        inf
9         AREA_Karapakkam        inf
10            AREA_Adyar        inf
11          AREA_Chrompt        inf
12          STREET_Gravel       inf
13         AREA_Chrompet        inf
14              MZZONE_C        inf
15             MZZONE_RL        inf
16        AREA_Ana Nagar        inf
17             MZZONE_RM        inf
18      STREET_No Access        inf
19         DIST_MAINROAD   1.006590
20          AREA_KK Nagar       inf
21              MZZONE_I        inf
22  BUILDTYPE_Commercial        inf
23            QS_OVERALL  12.293344
24       AREA_Anna Nagar        inf
25              INT_SQFT   3.835332
26                COMMIS   2.108649
27         AREA_Chormpet        inf
28       BUILDTYPE_Other        inf
29       BUILDTYPE_House        inf
30         AREA_Velchery        inf
31             MZZONE_RH        inf
```

```
32           AREA_KKNagar        inf
33             PARK_FACIL   1.023054
34            STREET_Pavd        inf
35         AREA_Karapakam        inf
36           STREET_Paved        inf
37                  const   0.000000
38         AREA_Ann Nagar        inf
39             QS_BEDROOM   5.648738
40           AREA_T Nagar        inf
41               MZZONE_A        inf

C:\Users\hp\anaconda3\lib\site-packages\statsmodels\regression\
linear_model.py:1736: RuntimeWarning: divide by zero encountered in
double_scalars
  return 1 - self.ssr/self.centered_tss
```

```
## now there is no multicolinearity

# Features with p <0.05
features_3=['INT_SQFT','PARK_FACIL','DIST_MAINROAD']

X_3=X_2[features_3]

X_3
```

```
      INT_SQFT  PARK_FACIL  DIST_MAINROAD
0         1004           1            131
1         1986           0             26
2          909           1             70
3         1855           0             14
4         1226           1             84
...        ...         ...            ...
7104       598           0             51
7105      1897           1             52
7106      1614           0            152
7107       787           1             40
7108      1896           1            156

[7109 rows x 3 columns]
```

```
# Splitting

X_train_3,X_test_3,y_train_3,y_test_3=train_test_split(X_3,

y,train_size=0.8,random_state=10)

X_train_3.shape,X_test_3.shape,y_train_3.shape,y_test_3.shape
```

```
((5687, 3), (1422, 3), (5687,), (1422,))
```

```
# Buidling the model
mlr_3=sm.OLS(y_train_3,X_train_3)
```

```
mlr_3=mlr_3.fit()

mlr_3.summary2()

<class 'statsmodels.iolib.summary2.Summary'>
"""
                         Results: Ordinary least squares
======================================================================
===========
Model:                  OLS              Adj. R-squared (uncentered):
0.930
Dependent Variable:     SALES_PRICE      AIC:
185971.2208
Date:                   2022-11-10 01:46 BIC:
185991.1586
No. Observations:       5687             Log-Likelihood:
-92983.
Df Model:               3                F-statistic:
2.518e+04
Df Residuals:           5684             Prob (F-statistic):
0.00
R-squared (uncentered): 0.930            Scale:
9.3156e+12
----------------------------------------------------------------------
-----------
                   Coef.      Std.Err.     t     P>|t|     [0.025
0.975]
----------------------------------------------------------------------
-----------
INT_SQFT           6565.4930    52.8530 124.2218 0.0000    6461.8810
6669.1051
PARK_FACIL      1518613.2545 78038.3975  19.4598 0.0000 1365628.2292
1671598.2799
DIST_MAINROAD      8108.3462   622.4316  13.0269 0.0000    6888.1428
9328.5495
----------------------------------------------------------------------
-----------
Omnibus:                    93.877           Durbin-Watson:
1.988
Prob(Omnibus):              0.000            Jarque-Bera (JB):
98.255
Skew:                       0.320            Prob(JB):
0.000
Kurtosis:                   2.934            Condition No.:
2817
======================================================================
===========
* The condition number is large (3e+03). This might indicate
strong
multicollinearity or other numerical problems.
"""
```

```
var_inf_factor(X_3)
```

```
         Feature  VIF_Value
0       INT_SQFT   3.626530
1     PARK_FACIL   1.872617
2  DIST_MAINROAD   3.154200
```

```
features_4=['INT_SQFT','PARK_FACIL']
X_4=X_3[features_4]
```

```
X_4
```

```
      INT_SQFT  PARK_FACIL
0         1004           1
1         1986           0
2          909           1
3         1855           0
4         1226           1
...        ...         ...
7104       598           0
7105      1897           1
7106      1614           0
7107       787           1
7108      1896           1

[7109 rows x 2 columns]
```

```
# Splitting

X_train_4,X_test_4,y_train_4,y_test_4=train_test_split(X_4,
                                        y,
train_size=0.8,random_state=10)
```

```
X_train_4.shape,X_test_4.shape,y_train_4.shape,y_test_4.shape
```

```
((5687, 2), (1422, 2), (5687,), (1422,))
```

```
# Building
mlr_4=sm.OLS(y_train_4,X_train_4)
mlr_4=mlr_4.fit()
```

```
mlr_4.summary2()
```

```
<class 'statsmodels.iolib.summary2.Summary'>
"""
                         Results: Ordinary least squares
======================================================================
===========
Model:                  OLS              Adj. R-squared (uncentered):
0.928
Dependent Variable:     SALES_PRICE      AIC:
186136.5249
```

```
Date:                    2022-11-10 01:46 BIC:
186149.8168
No. Observations:        5687             Log-Likelihood:
-93066.
Df Model:                2                F-statistic:
3.660e+04
Df Residuals:            5685             Prob (F-statistic):
0.00
R-squared (uncentered): 0.928            Scale:
9.5920e+12
----------------------------------------------------------------------
------------
               Coef.       Std.Err.      t      P>|t|      [0.025
0.975]
----------------------------------------------------------------------
------------
INT_SQFT       7045.6051     38.4409  183.2843  0.0000     6970.2463
7120.9638
PARK_FACIL  1658157.6922  78438.2336   21.1397  0.0000  1504388.8413
1811926.5430
----------------------------------------------------------------------
-----------
Omnibus:                     59.799           Durbin-Watson:
1.983
Prob(Omnibus):               0.000            Jarque-Bera (JB):
58.749
Skew:                        0.227            Prob(JB):
0.000
Kurtosis:                    2.796            Condition No.:
2784
======================================================================
===========
* The condition number is large (3e+03). This might indicate
strong
multicollinearity or other numerical problems.
"""
```

```
var_inf_factor(X_4)
```

```
      Feature  VIF_Value
0    INT_SQFT   1.838724
1  PARK_FACIL   1.838724
```

```
mlr_4.params
```

```
INT_SQFT      7.045605e+03
PARK_FACIL    1.658158e+06
dtype: float64
```

The model is:

SOLD PRICE = INT_SQFT * 7.045605 + PARK_FACIL * 1.658158

Residual Analysis

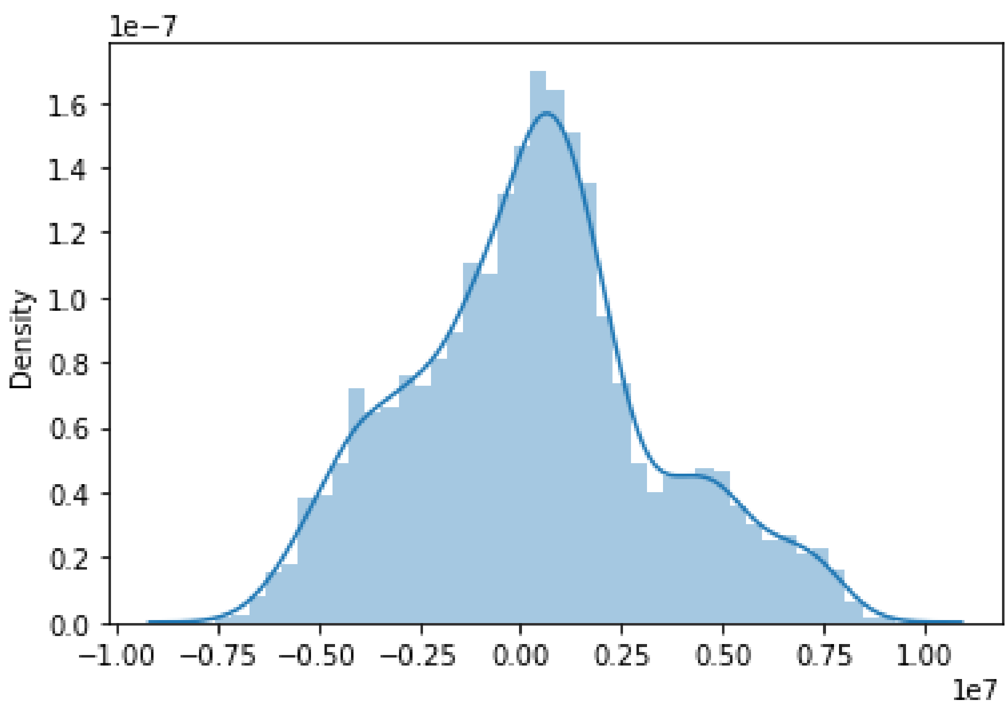
Normality

```
mlr_4.resid
```

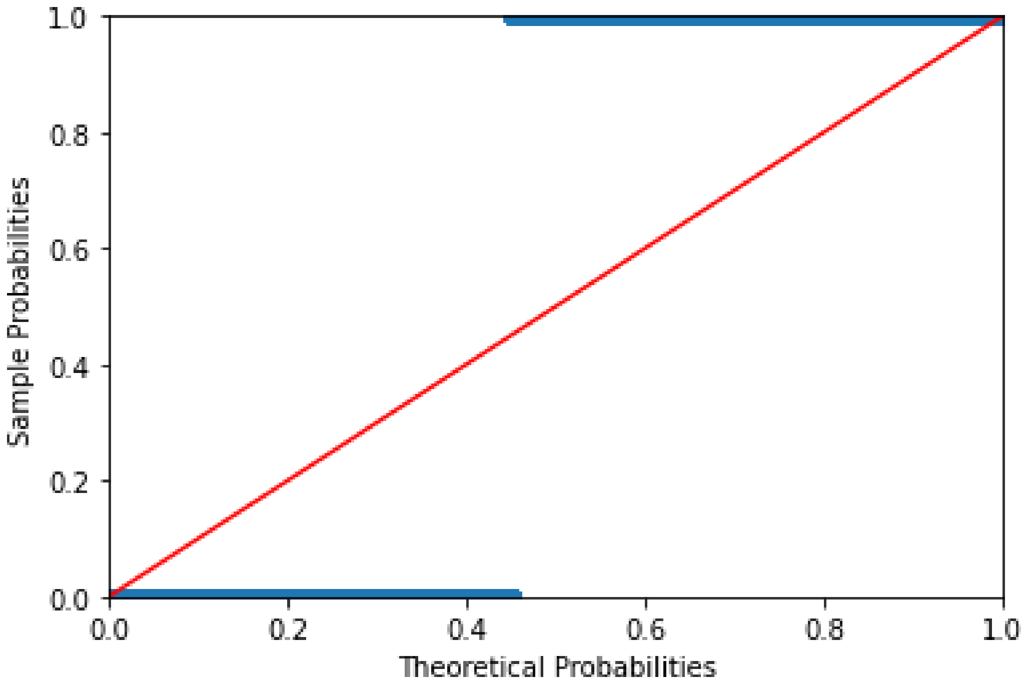
```
591     3.093338e+06
2651    2.186518e+05
5737   -6.441511e+06
5601   -1.482561e+06
1114    4.192007e+06
            ...
1180   -2.630428e+06
3441   -2.635413e+06
1344    1.122404e+06
4623    3.577222e+06
1289    9.189846e+05
Length: 5687, dtype: float64
```

```
sns.distplot(mlr_4.resid);
```

```
C:\Users\hp\anaconda3\lib\site-packages\seaborn\distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



```python
## Prob prob plot

def prob_prob_plot(model):
    probplot=sm.ProbPlot(model.resid)
    probplot.ppplot(line='45')
    plt.show();

prob_prob_plot(mlr_4)
```



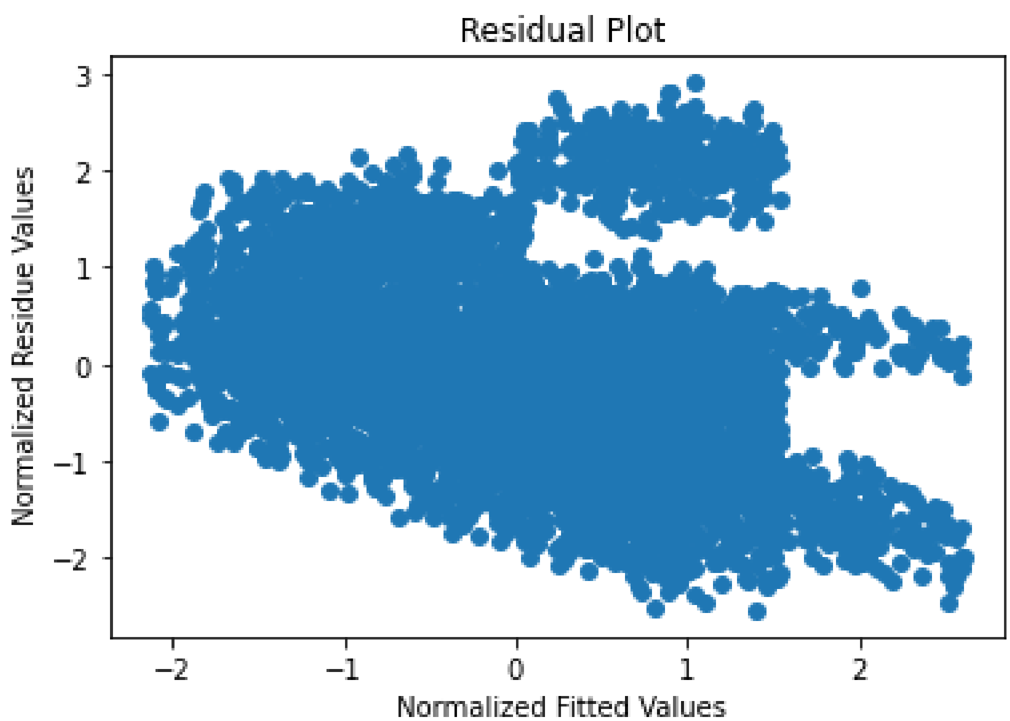

```python
X_train_4
```

```
      INT_SQFT  PARK_FACIL
591        764           0
2651      1848           0
5737      1592           1
5601      1303           1
1114       790           0
...        ...         ...
1180      1042           1
3441      1091           1
1344      1864           1
4623       917           1
1289       897           1

[5687 rows x 2 columns]
```

The residuals are not fully normally distributed

Homoscedasticity

```
def standard(data):
    return (data-data.mean())/data.std()

# Plotting residual plot

def residual_plot(model):
    plt.scatter(standard(model.fittedvalues),
                standard(model.resid))
    plt.xlabel(' Normalized Fitted Values')
    plt.ylabel(' Normalized Residue Values')
    plt.title(' Residual Plot')
    plt.show();

# Calling the function

residual_plot(mlr_4)
```



Checking the outliers

Z SCore

```
from scipy.stats import zscore

z_score=zscore(X_4)
```

```
z_score[z_score>3].count()
```

```
INT_SQFT      0
PARK_FACIL    0
dtype: int64
```

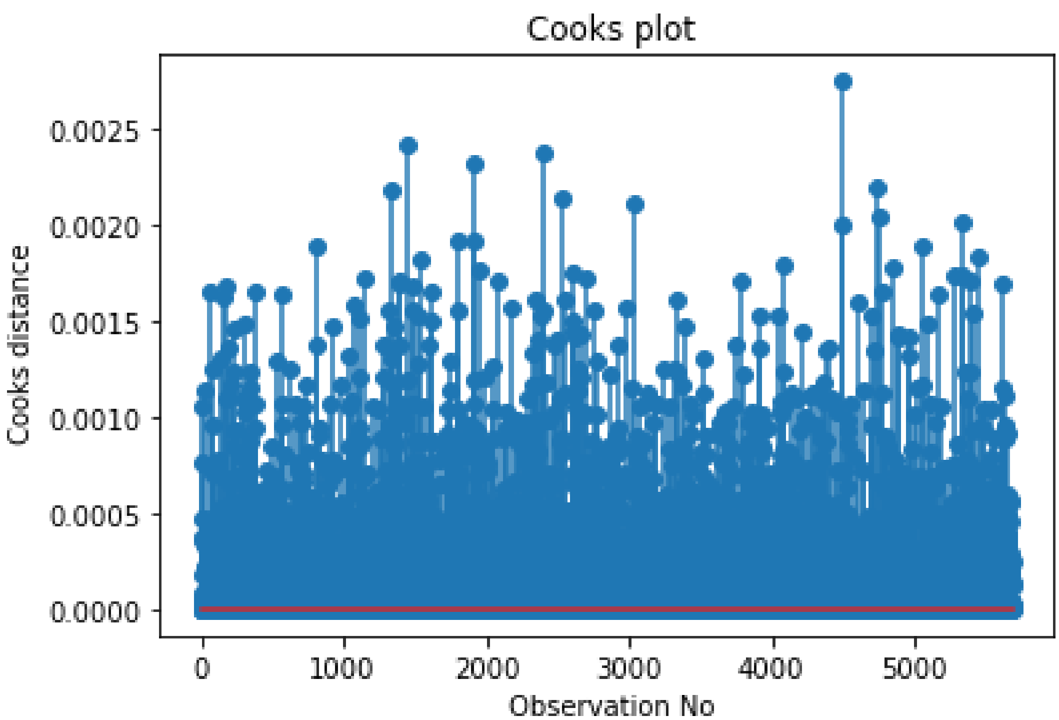
```
z_score[z_score <-3].count()
```

```
INT_SQFT      0
PARK_FACIL    0
dtype: int64
```

Cooks Distance

```
def cooks_dist(model):
    model_influence=model.get_influence()
    (c,_)=model_influence.cooks_distance
    plt.stem(np.arange(len(X_train_4)),c)
    plt.xlabel(' Observation No')
    plt.ylabel(' Cooks distance')
    plt.title(' Cooks plot')
    plt.show();
```

```
cooks_dist(mlr_4)
```



There is no observation with cooks distance > 1.

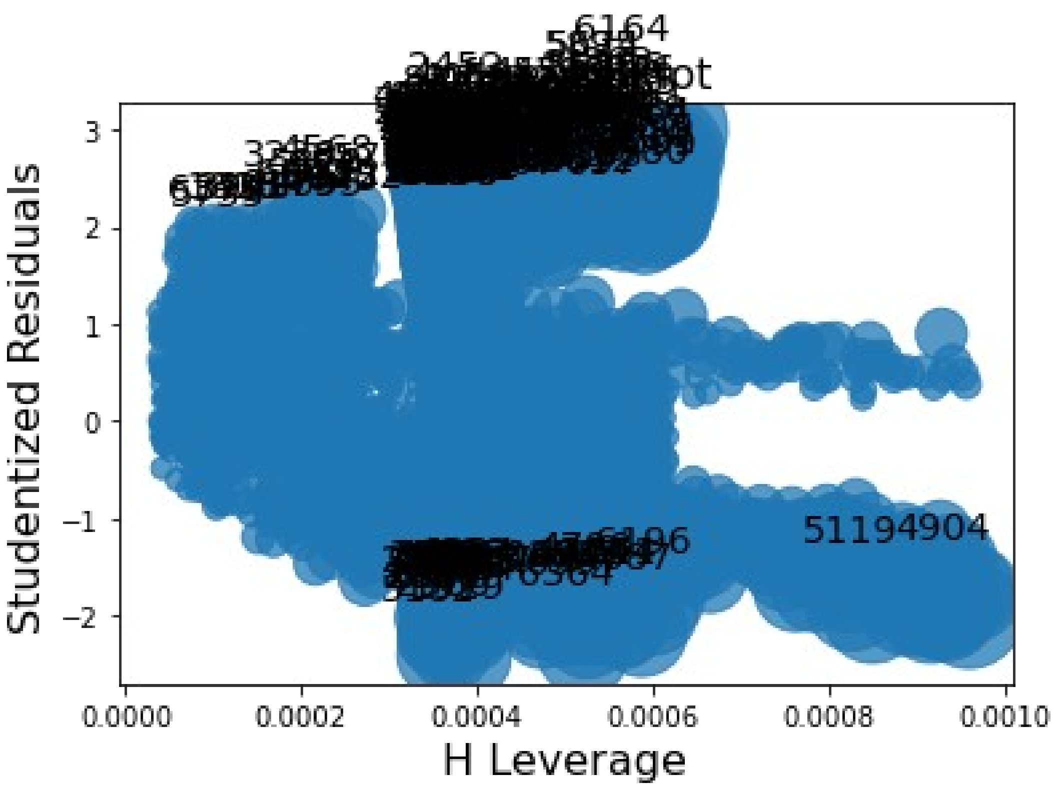
So, the measure says there is no outlier.

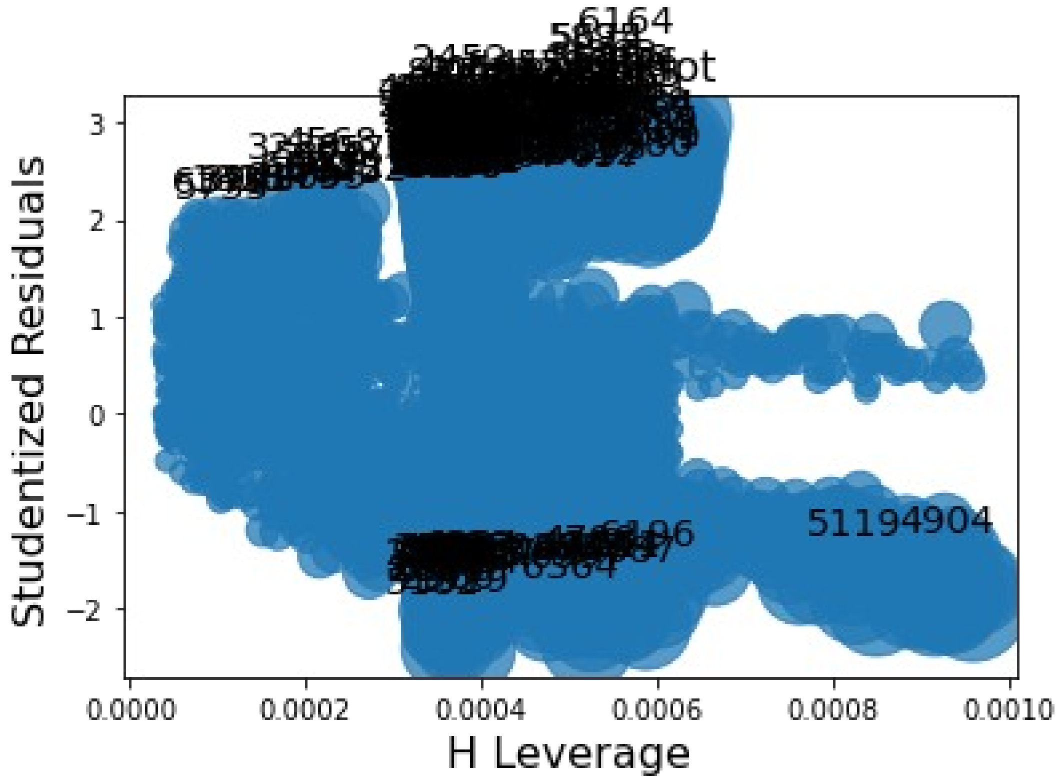
Leverage distance

```
n=5687 # No of training data

k= 2 # No of features in the model

lev_cutoff= (3*(k+1))/n

print(' The Leverage cut off:',lev_cutoff)
```

```
 The Leverage cut off: 0.0015825567082820467
```

```
from statsmodels.graphics.regressionplots import influence_plot
influence_plot(mlr_4)
```






```
X_train_4.iloc[4904]
```

```
INT_SQFT      1776
PARK_FACIL       0
Name: 1345, dtype: int64
```

```
y_pred= mlr_4.predict(X_test_4)
y_pred
```

```
461     1.267044e+07
3358    1.113206e+07
3751    1.438957e+07
2386    1.757878e+07
1125    1.000476e+07
            ...
6010    1.224526e+07
4903    1.356523e+07
6806    1.079631e+07
3832    1.362620e+07
364     5.650575e+06
Length: 1422, dtype: float64
```

```
X_train_4['INT_SQFT'].min(),X_train_4['INT_SQFT'].max()
```

```
(500, 2500)
```

```
X_5= X_4.drop([5119,4904])
```

```
X_5.shape
```

```
(7107, 2)
```

```
y_5=y.drop([5119,4904])
```

```
y_5.shape
```

```
(7107,)
```

```
X_train_5,X_test_5,y_train_5,y_test_5=train_test_split(X_5,
                                                        y_5,
train_size=0.8,random_state=10)
```

```
X_train_5.shape,X_test_5.shape,y_train_5.shape,y_test_5.shape
```

```
((5685, 2), (1422, 2), (5685,), (1422,))
```

```
mlr_5=sm.OLS(y_train_5,X_train_5)
```

```
mlr_5=mlr_5.fit()
```

```
mlr_5.summary2()
```

```
<class 'statsmodels.iolib.summary2.Summary'>
"""
                         Results: Ordinary least squares
======================================================================
===========
Model:                  OLS              Adj. R-squared (uncentered):
0.928
Dependent Variable:     SALES_PRICE      AIC:
186089.2541
Date:                   2022-11-10 01:49 BIC:
186102.5453
No. Observations:       5685             Log-Likelihood:
-93043.
Df Model:               2                F-statistic:
3.651e+04
Df Residuals:           5683             Prob (F-statistic):
0.00
R-squared (uncentered): 0.928            Scale:
9.6227e+12
----------------------------------------------------------------------
------------
               Coef.       Std.Err.      t      P>|t|      [0.025
0.975]
----------------------------------------------------------------------
------------
```

```
INT_SQFT       7043.2743     38.3709  183.5579  0.0000     6968.0527
7118.4958
PARK_FACIL  1673080.9715  78513.0265   21.3096  0.0000  1519165.4865
1826996.4565
----------------------------------------------------------------------
-----------
Omnibus:                      61.517           Durbin-Watson:
1.992
Prob(Omnibus):                0.000            Jarque-Bera (JB):
60.275
Skew:                         0.229            Prob(JB):
0.000
Kurtosis:                     2.789            Condition No.:
2784
======================================================================
===========
* The condition number is large (3e+03). This might indicate
strong
multicollinearity or other numerical problems.
"""
```

Predicting the value

```
y_pred= mlr_5.predict(X_test_5)
y_pred
```

```
461     1.268172e+07
1798    1.156184e+07
3655    6.814671e+06
5207    1.480879e+07
1125    1.000145e+07
            ...
5334    1.391429e+07
4302    9.300947e+06
3831    1.023066e+07
364     5.648706e+06
591     5.381062e+06
Length: 1422, dtype: float64
```

Performance of the model

```
from sklearn.metrics import r2_score,mean_squared_error

r2= r2_score(y_test_5, y_pred)
print('R2:', r2)

mse = mean_squared_error(y_test_5, y_pred)

rmse= np.sqrt(mse)

print(' RMSE:',rmse)
```

```
R2: 0.34575389208379126
 RMSE: 3093969.0027331156
```

Tramsform the target

```
y.min(),y.max()
```

```
(2156875, 23667340)
```

```
y_sq= np.sqrt(y)
```

```
y_sq.min(),y_sq.max()
```

```
(1468.6303142724516, 4864.909043342948)
```

```
# log
```

```
y_log= np.log(y)
```

```
y_log.min(),y_log.max()
```

```
(14.584170972834405, 16.97960659663699)
```