In [1]:
```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [4]:
```python
data = pd.read_csv("D:\\IPSITA\\Chennai houseing sale.csv")
```

In [5]:
```python
df=pd.DataFrame(data)
```

In [8]:
```python
df.isnull().sum
```

Out[8]:
```
<bound method NDFrame._add_numeric_operations.<locals>.sum of       PRT_ID   AREA  INT_SQFT  DATE_SALE  DIST_MAINROAD  N_BEDROOM  \
0      False  False     False      False          False      False
1      False  False     False      False          False      False
2      False  False     False      False          False      False
3      False  False     False      False          False      False
4      False  False     False      False          False      False
...      ...    ...       ...        ...            ...        ...
7104   False  False     False      False          False      False
7105   False  False     False      False          False      False
7106   False  False     False      False          False      False
7107   False  False     False      False          False      False
7108   False  False     False      False          False      False

      N_BATHROOM  N_ROOM  SALE_COND  PARK_FACIL  ...  UTILITY_AVAIL  STREET  \
0          False   False      False       False  ...          False   False
1          False   False      False       False  ...          False   False
2          False   False      False       False  ...          False   False
3          False   False      False       False  ...          False   False
4          False   False      False       False  ...          False   False
...          ...     ...        ...         ...  ...            ...     ...
7104       False   False      False       False  ...          False   False
7105       False   False      False       False  ...          False   False
7106       False   False      False       False  ...          False   False
7107       False   False      False       False  ...          False   False
7108       False   False      False       False  ...          False   False

      MZZONE  QS_ROOMS  QS_BATHROOM  QS_BEDROOM  QS_OVERALL  REG_FEE  COMMIS  \
0      False     False        False       False       False    False   False
1      False     False        False       False       False    False   False
2      False     False        False       False       False    False   False
3      False     False        False       False       False    False   False
4      False     False        False       False       False    False   False
...      ...       ...          ...         ...         ...      ...     ...
7104   False     False        False       False       False    False   False
7105   False     False        False       False       False    False   False
7106   False     False        False       False       False    False   False
7107   False     False        False       False       False    False   False
7108   False     False        False       False       False    False   False

      SALES_PRICE
0           False
1           False
2           False
3           False
4           False
...           ...
7104        False
7105        False
7106        False
7107        False
7108        False

[7109 rows x 22 columns]>
```

In [10]:
```python
df.fillna(-1, inplace=True)
```

In [11]:
```python
df.head()
```

Out[11]:

	PRT_ID	AREA	INT_SQFT	DATE_SALE	DIST_MAINROAD	N_BEDROOM	N_BATHROOM	N_ROOM	SALE_COND	PARK_FACIL	...	UTILITY_AVAIL	STREET	MZZONE	QS_ROOMS	QS_BATHROOM	QS_BEDROOM	QS_OV
0	P03210	Karapakkam	1004	04-05-2011	131	1.0	1.0	3	AbNormal	Yes	...	AllPub	Paved	A	4.0	3.9	4.9	
1	P09411	Anna Nagar	1986	19-12-2006	26	2.0	1.0	5	AbNormal	No	...	AllPub	Gravel	RH	4.9	4.2	2.5	
2	P01812	Adyar	909	04-02-2012	70	1.0	1.0	3	AbNormal	Yes	...	ELO	Gravel	RL	4.1	3.8	2.2	
3	P05346	Velachery	1855	13-03-2010	14	3.0	2.0	5	Family	No	...	NoSewr	Paved	I	4.7	3.9	3.6	
4	P06210	Karapakkam	1226	05-10-2009	84	1.0	1.0	3	AbNormal	Yes	...	AllPub	Gravel	C	3.0	2.5	4.1	

5 rows × 22 columns

In [13]:
```python
df['Total_price'] = df['REG_FEE'] + df['COMMIS'] + df['SALES_PRICE']
```

In [20]:
```python
pd.to_datetime(df['DATE_BUILD'])
```

```
C:\Users\tuhin\AppData\Local\Temp\ipykernel_11436\2327747234.py:1: UserWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was specified. This may lead to inconsistently parsed dates! Specify a format to ensure consistent parsing.
  pd.to_datetime(df['DATE_BUILD'])
```

Out[20]:
```
0      1967-05-15
1      1995-12-22
2      1992-09-02
3      1988-03-18
4      1979-10-13
          ...
7104   1962-01-15
7105   1995-11-04
7106   1978-01-09
7107   1977-11-08
7108   1961-07-24
Name: DATE_BUILD, Length: 7109, dtype: datetime64[ns]
```

In [37]:
```python
df['Year'] = df['DATE_BUILD'].apply(lambda x: x[6:10])
```

In [38]:
```python
print(df)
```

```
      PRT_ID        AREA  INT_SQFT   DATE_SALE  DIST_MAINROAD  N_BEDROOM  \
0     P03210  Karapakkam      1004  04-05-2011            131        1.0
1     P09411  Anna Nagar      1986  19-12-2006             26        2.0
2     P01812       Adyar       909  04-02-2012             70        1.0
3     P05346   Velachery      1855  13-03-2010             14        3.0
4     P06210  Karapakkam      1226  05-10-2009             84        1.0
...      ...         ...       ...         ...            ...        ...
7104  P03834  Karapakkam       598  03-01-2011             51        1.0
7105  P10000   Velachery      1897  08-04-2004             52        3.0
7106  P09594   Velachery      1614  25-08-2006            152        2.0
7107  P06508  Karapakkam       787  03-08-2009             40        1.0
7108  P09794   Velachery      1896  13-07-2005            156        3.0

      N_BATHROOM  N_ROOM    SALE_COND PARK_FACIL  ... MZZONE QS_ROOMS  \
0            1.0       3     AbNormal        Yes  ...      A      4.0
1            1.0       5     AbNormal         No  ...     RH      4.9
2            1.0       3     AbNormal        Yes  ...     RL      4.1
3            2.0       5       Family         No  ...      I      4.7
4            1.0       3     AbNormal        Yes  ...      C      3.0
...          ...     ...          ...        ...  ...    ...      ...
7104         1.0       2      AdjLand         No  ...     RM      3.0
7105         2.0       5       Family        Yes  ...     RH      3.6
7106         1.0       4  Normal Sale         No  ...      I      4.3
7107         1.0       2      Partial        Yes  ...     RL      4.6
7108         2.0       5      Partial        Yes  ...      I      3.1

     QS_BATHROOM QS_BEDROOM QS_OVERALL  REG_FEE  COMMIS  SALES_PRICE  \
0            3.9        4.9      4.330   380000  144400      7600000
1            4.2        2.5      3.765   760122  304049     21717770
2            3.8        2.2      3.090   421094   92114     13159200
3            3.9        3.6      4.010   356321   77042      9630290
4            2.5        4.1      3.290   237000   74063      7406250
...          ...        ...        ...      ...     ...          ...
7104         2.2        2.4      2.520   208767  107060      5353000
7105         4.5        3.3      3.920   346191  205551     10818480
7106         4.2        2.9      3.840   317354  167028      8351410
7107         3.8        4.1      4.160   425350  119098      8507000
7108         3.5        4.3      3.640   349177   79812      9976480

      Total_price  Year
0         8124400  1967
1        22781941  1995
2        13672408  1992
3        10063653  1988
4         7717313  1979
...           ...   ...
7104      5668827  1962
7105     11370222  1995
7106      8835792  1978
7107      9051448  1977
7108     10405469  1961

[7109 rows x 24 columns]
```

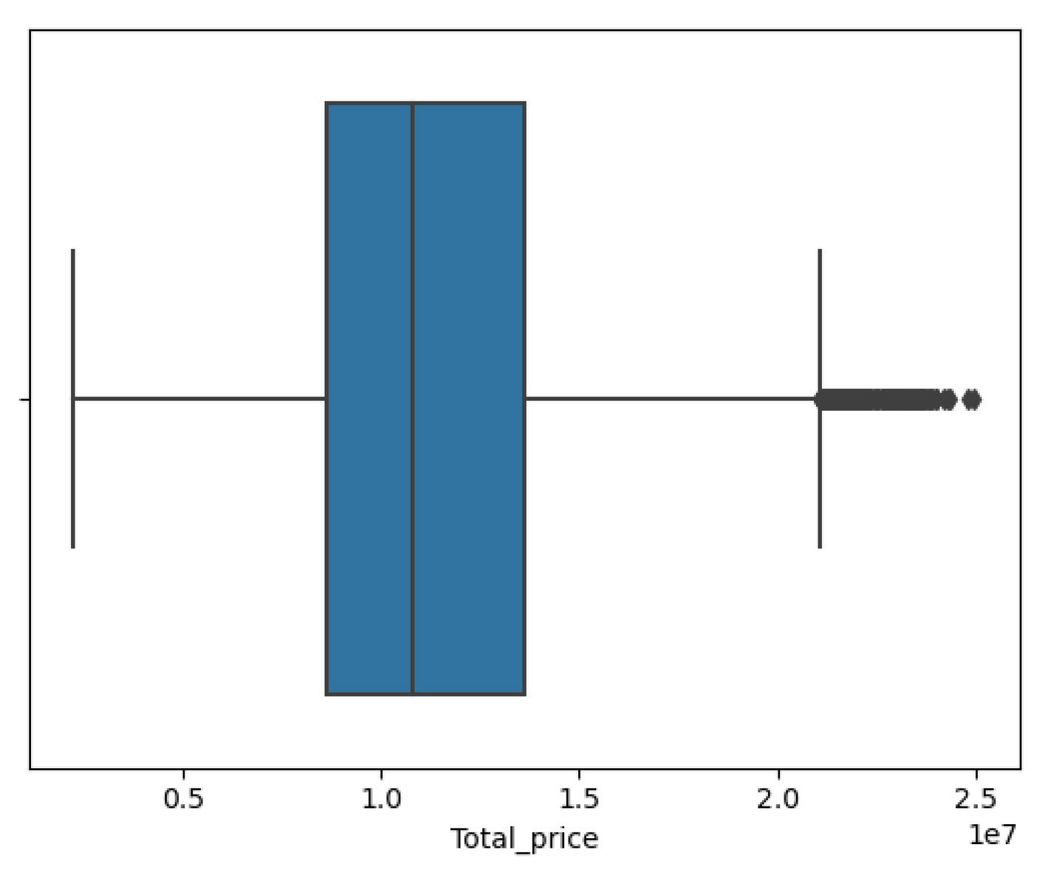
In [42]:
```python
df['Year'] = df['Year'].astype(int)
```

In [43]:
```python
df['AgeOfProperty'] = 2023 - df['Year']
```

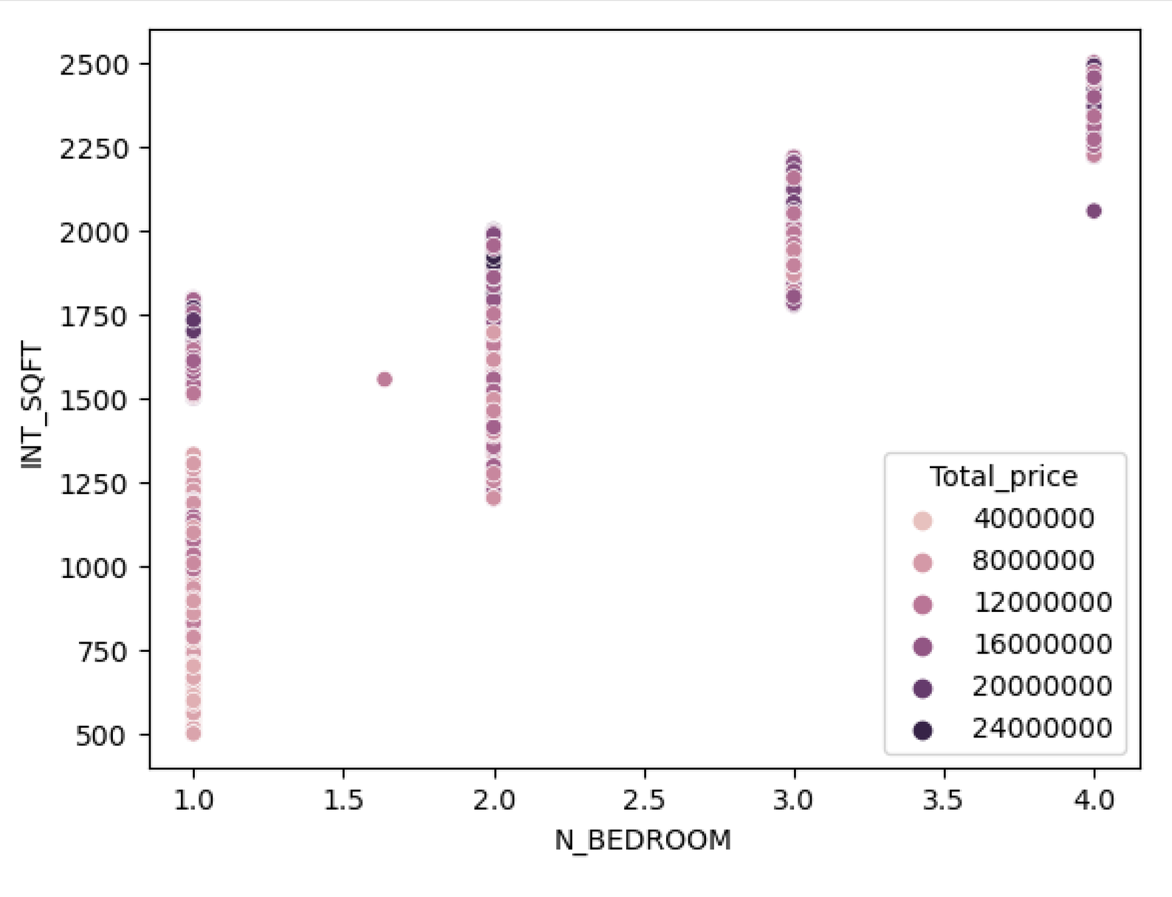
In [44]:
```python
print(df)
```

```
      PRT_ID        AREA  INT_SQFT   DATE_SALE  DIST_MAINROAD  N_BEDROOM  \
0     P03210  Karapakkam      1004  04-05-2011            131        1.0
1     P09411  Anna Nagar      1986  19-12-2006             26        2.0
2     P01812       Adyar       909  04-02-2012             70        1.0
3     P05346   Velachery      1855  13-03-2010             14        3.0
4     P06210  Karapakkam      1226  05-10-2009             84        1.0
...      ...         ...       ...         ...            ...        ...
7104  P03834  Karapakkam       598  03-01-2011             51        1.0
7105  P10000   Velachery      1897  08-04-2004             52        3.0
7106  P09594   Velachery      1614  25-08-2006            152        2.0
7107  P06508  Karapakkam       787  03-08-2009             40        1.0
7108  P09794   Velachery      1896  13-07-2005            156        3.0

      N_BATHROOM  N_ROOM    SALE_COND PARK_FACIL  ... QS_ROOMS QS_BATHROOM  \
0            1.0       3     AbNormal        Yes  ...      4.0         3.9
1            1.0       5     AbNormal         No  ...      4.9         4.2
2            1.0       3     AbNormal        Yes  ...      4.1         3.8
3            2.0       5       Family         No  ...      4.7         3.9
4            1.0       3     AbNormal        Yes  ...      3.0         2.5
...          ...     ...          ...        ...  ...      ...         ...
7104         1.0       2      AdjLand         No  ...      3.0         2.2
7105         2.0       5       Family        Yes  ...      3.6         4.5
7106         1.0       4  Normal Sale         No  ...      4.3         4.2
7107         1.0       2      Partial        Yes  ...      4.6         3.8
7108         2.0       5      Partial        Yes  ...      3.1         3.5

     QS_BEDROOM QS_OVERALL REG_FEE  COMMIS  SALES_PRICE  Total_price  Year  \
0           4.9      4.330  380000  144400      7600000      8124400  1967
1           2.5      3.765  760122  304049     21717770     22781941  1995
2           2.2      3.090  421094   92114     13159200     13672408  1992
3           3.6      4.010  356321   77042      9630290     10063653  1988
4           4.1      3.290  237000   74063      7406250      7717313  1979
...         ...        ...     ...     ...          ...          ...   ...
7104        2.4      2.520  208767  107060      5353000      5668827  1962
7105        3.3      3.920  346191  205551     10818480     11370222  1995
7106        2.9      3.840  317354  167028      8351410      8835792  1978
7107        4.1      4.160  425350  119098      8507000      9051448  1977
7108        4.3      3.640  349177   79812      9976480     10405469  1961

      AgeOfProperty
0                56
1                28
2                31
3                35
4                44
...             ...
7104             61
7105             28
7106             45
7107             46
7108             62

[7109 rows x 25 columns]
```

In [45]:
```python
sns.boxplot(x=df['Total_price'])
plt.show()
```



In [47]:
```python
sns.scatterplot(x='N_BEDROOM', y='INT_SQFT', hue='Total_price', data=df)
plt.show()
```

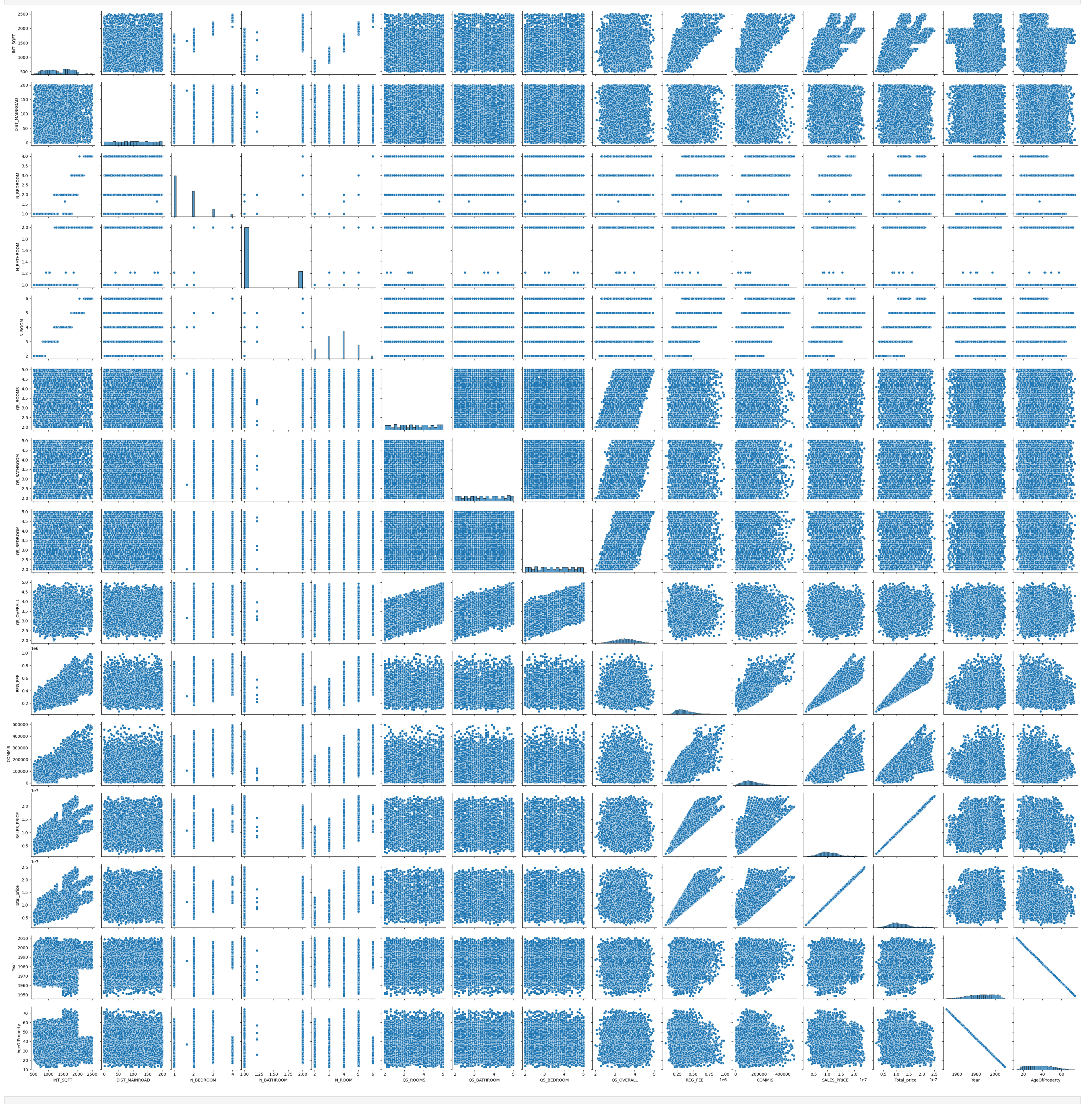


In [49]:
```python
correlation_matrix = df.corr()
print("Correlation Matrix:")
print(correlation_matrix)
```

```
Correlation Matrix:
               INT_SQFT  DIST_MAINROAD  N_BEDROOM  N_BATHROOM    N_ROOM  \
INT_SQFT       1.000000       0.002022   0.786340    0.515261  0.951279
DIST_MAINROAD  0.002022       1.000000  -0.002565    0.001969  0.002301
N_BEDROOM      0.786340      -0.002565   1.000000    0.755055  0.840280
N_BATHROOM     0.515261       0.001969   0.755055    1.000000  0.568564
N_ROOM         0.951279       0.002301   0.840280    0.568564  1.000000
QS_ROOMS       0.019850       0.002237   0.015109    0.013108  0.016524
QS_BATHROOM   -0.008337      -0.029468  -0.007635   -0.012055 -0.007545
QS_BEDROOM     0.008865       0.001554   0.015566    0.013190  0.015072
QS_OVERALL     0.013989      -0.017255   0.014172    0.007877  0.015418
REG_FEE        0.657544       0.011600   0.455351    0.260249  0.630932
COMMIS         0.571076       0.010994   0.430418    0.256453  0.533343
SALES_PRICE    0.612125       0.018783   0.330999    0.108865  0.602760
Total_price    0.620010       0.018584   0.341310    0.118558  0.609345
Year          -0.009301       0.005706   0.012426    0.001344  0.013409
AgeOfProperty  0.009301      -0.005706  -0.012426   -0.001344 -0.013409

               QS_ROOMS  QS_BATHROOM  QS_BEDROOM  QS_OVERALL   REG_FEE  \
INT_SQFT       0.019850    -0.008337    0.008865    0.013989  0.657544
DIST_MAINROAD  0.002237    -0.029468    0.001554   -0.017255  0.011600
N_BEDROOM      0.015109    -0.007635    0.015566    0.014172  0.455351
N_BATHROOM     0.013108    -0.012055    0.013190    0.007877  0.260249
N_ROOM         0.016524    -0.007545    0.015072    0.015418  0.630932
QS_ROOMS       1.000000     0.008828    0.007789    0.516147  0.019739
QS_BATHROOM    0.008828     1.000000   -0.011745    0.549056 -0.006719
QS_BEDROOM     0.007789    -0.011745    1.000000    0.628168  0.021526
QS_OVERALL     0.516147     0.549056    0.628168    1.000000  0.022410
REG_FEE        0.019739    -0.006719    0.021526    0.022410  1.000000
COMMIS         0.009377    -0.000471    0.020118    0.016946  0.659903
SALES_PRICE    0.021967    -0.011377    0.018804    0.020485  0.878148
Total_price    0.021888    -0.011122    0.019146    0.020720  0.888335
Year          -0.018704    -0.021168    0.018011   -0.012200  0.108196
AgeOfProperty  0.018704     0.021168   -0.018011    0.012200 -0.108196

                 COMMIS  SALES_PRICE  Total_price      Year  AgeOfProperty
INT_SQFT       0.571076     0.612125     0.620010 -0.009301       0.009301
DIST_MAINROAD  0.010994     0.018783     0.018584  0.005706      -0.005706
N_BEDROOM      0.430418     0.330999     0.341310  0.012426      -0.012426
N_BATHROOM     0.256453     0.108865     0.118558  0.001344      -0.001344
N_ROOM         0.533343     0.602760     0.609345  0.013409      -0.013409
QS_ROOMS       0.009377     0.021967     0.021888 -0.018704       0.018704
QS_BATHROOM   -0.000471    -0.011377    -0.011122 -0.021168       0.021168
QS_BEDROOM     0.020118     0.018804     0.019146  0.018011      -0.018011
QS_OVERALL     0.016946     0.020485     0.020720 -0.012200       0.012200
REG_FEE        0.659903     0.878148     0.888335  0.108196      -0.108196
COMMIS         1.000000     0.626275     0.642178  0.087359      -0.087359
SALES_PRICE    0.626275     1.000000     0.999649  0.116043      -0.116043
Total_price    0.642178     0.999649     1.000000  0.116524      -0.116524
Year           0.087359     0.116043     0.116524  1.000000      -1.000000
AgeOfProperty -0.087359    -0.116043    -0.116524 -1.000000       1.000000
```

```
C:\Users\tuhin\AppData\Local\Temp\ipykernel_11436\4063387690.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  correlation_matrix = df.corr()
```

In [51]:
```python
sns.pairplot(df)
plt.show()
```



In [ ]: