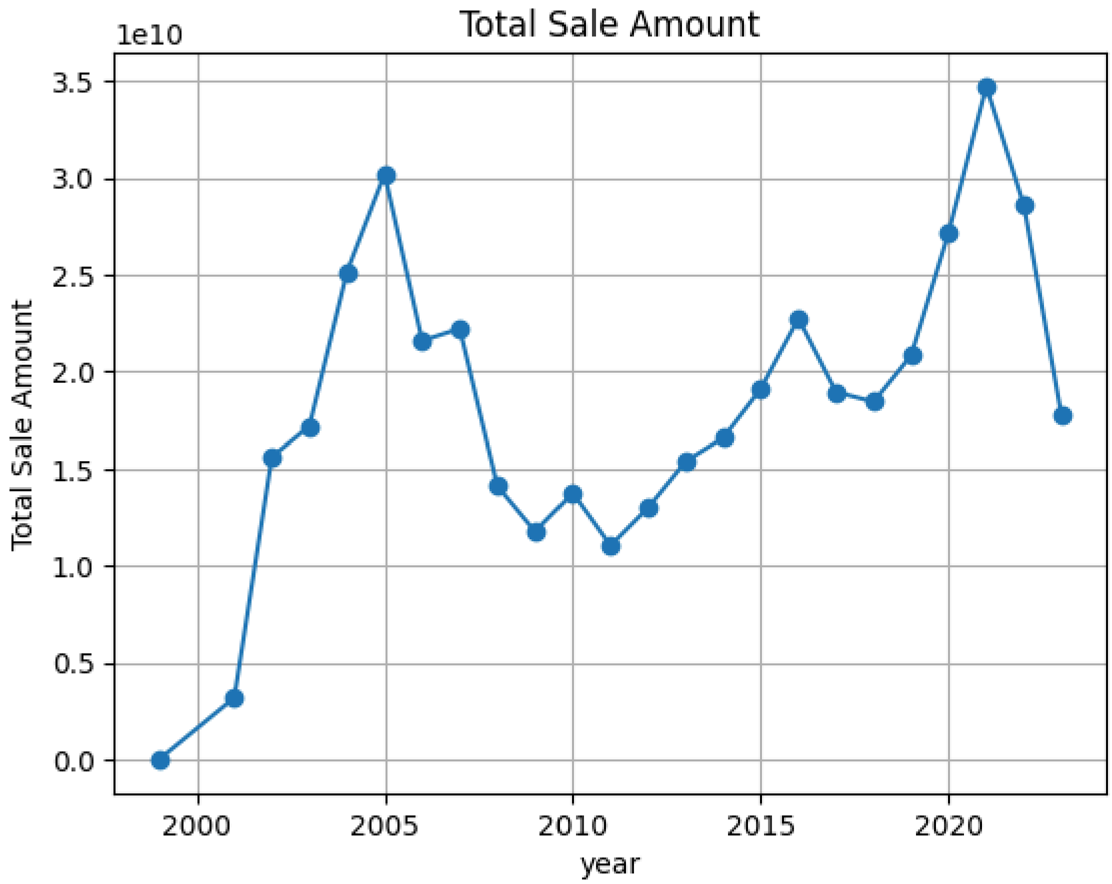
In [10]:

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Real_Estate_Sales.csv')
incorrect_rows = df[df.apply(lambda x: x.astype(str).str.contains('incorrect', c
print(incorrect_rows)
df = df[~df.apply(lambda x: x.astype(str).str.contains('incorrect', case=False,

print(df.info())

df['Date Recorded'] = pd.to_datetime(df['Date Recorded'], errors='coerce')

df['year']  = df['Date Recorded'].dt.year

sales_by_years = df.groupby('year')['Sale Amount'].sum()

sales_by_years.plot(kind='line', marker='o')
plt.title('Total Sale Amount')
plt.xlabel('year')
plt.ylabel('Total Sale Amount')
plt.grid()
plt.show()

bins = [0,200000, 500000, 1000000, 5000000]
labels = ['0-200k', '200k-500k', '500k-1M', '1M-5M']

df['Amount Range'] =  pd.cut(df['Sale Amount'], bins=bins, labels=labels)

sales_by_range = df.groupby(['year', 'Amount Range'])['Sale Amount'].sum().unsta
print(sales_by_range)

sales_by_range.plot(kind='line', marker='o', figsize = (10,6))
plt.title('Sale Amount Trends by Range')
plt.xlabel('Year')
plt.ylabel('Total Sale Amount')
plt.legend(title = 'Amount Range')
plt.grid()
plt.show()


sales_by_town = df.groupby(['year', 'Town'])['Sale Amount'].sum().unstack()

top_towns = sales_by_town.sum().sort_values(ascending=False).head(5).index

sales_by_town[top_towns].plot(kind='line', figsize=(10,6))

plt.title('Sale Amount Trends by Town')
plt.xlabel('Year')
plt.ylabel('Total Sale Amount')
plt.legend(title='Towns')
plt.grid()
plt.show()


key_towns = ['Greenwich', 'Stamford']
sales_by_town[key_towns].plot(kind='line', figsize=(10,6), marker='o')
plt.title('Sale Amount Trends: Greenwich vs Stamford')
plt.xlabel('Year')
plt.ylabel('Total Sale Amount')
plt.legend(title='Towns')
plt.grid()
plt.show()
```

```
C:\Users\hector\AppData\Local\Temp\ipykernel_10668\516065680.py:4: DtypeWarning:
Columns (8,9,10,11,12) have mixed types. Specify dtype option on import or set lo
w_memory=False.
  df = pd.read_csv('Real_Estate_Sales.csv')
```

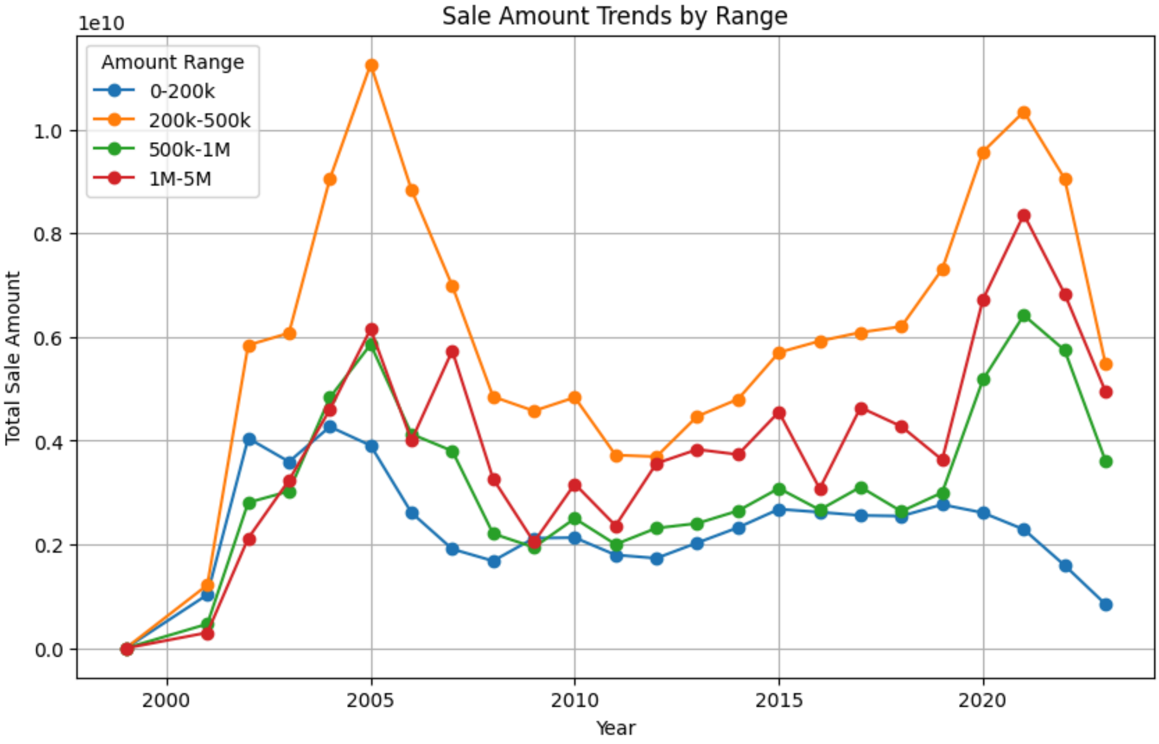
```
         Serial Number  List Year Date Recorded           Town  \
933             200562       2020    02/03/2021        Danbury   
3905             20880       2020    06/15/2021       Hartford   
4942            200151       2020    12/21/2020       Guilford   
5899            200133       2020    05/03/2021   East Windsor   
6737            200041       2020    12/01/2020   East Windsor   
...                ...        ...           ...            ...   
1003224         220038       2022    10/18/2022       Hartford   
1016523         210533       2021    06/07/2022   East Windsor   
1031076         210059       2021    06/22/2022    Middlefield   
1058529         210412       2021    05/26/2022       Westport   
1061891         210001       2021    10/01/2021  Windsor Locks   

                    Address  Assessed Value  Sale Amount  Sales Ratio  \
933             19  MILL RD        263600.0     415000.0     0.635181   
3905          12 BALDWIN ST         62402.0     200000.0     0.312000   
4942      33 INDIAN COVE RD        241200.0     920000.0     0.262174   
5899      47 BARBER HILL RD        261100.0     566895.0     0.460579   
6737        90 ROCKVILLE RD        130880.0     295000.0     0.443600   
...                     ...             ...          ...          ...   
1003224     211 STANDISH ST         58653.0     170000.0     0.345000   
1016523         18 WELLS RD         58400.0      60000.0     0.973300   
1031076  168 BAILEYVILLE RD        338200.0     825000.0     0.409939   
1058529    36 BRIDGE STREET        670400.0    1899000.0     0.353028   
1061891   28 LIN-SAL STREET        120890.0     235000.0     0.514400   

        Property Type Residential Type             Non Use Code  \
933       Residential    Single Family               25 - Other   
3905      Residential    Single Family                      NaN   
4942      Residential    Single Family  07 - Change in Property   
5899      Residential    Single Family  07 - Change in Property   
6737      Residential    Single Family  07 - Change in Property   
...               ...              ...                      ...   
1003224   Residential    Single Family         14 - Foreclosure   
1016523   Vacant Land              NaN                      NaN   
1031076   Residential    Single Family               25 - Other   
1058529   Residential    Single Family  07 - Change in Property   
1061891   Residential    Single Family                      NaN   

                                          Assessor Remarks  \
933                      AFFORDABLE HOUSING / B15001-20-19   
3905                            TOWN CLERK INCORRECT ORDER   
4942                                                   NaN   
5899                               SKETCH ON PRC INCORRECT   
6737                                  SKETCHED INCORRECTLY   
...                                                    ...   
1003224  Sale of a previous foreclosure/deed error inco...   
1016523                 DEED READS 20 WELLS RD INCORRECTLY   
1031076  ACREAGE INCORRECT IN CAMA 3.36 AC NOT 2.72 AC;...   
1058529     Incorrect home features listed at time of sale   
1061891  GRANTOR NAME (POLASKI) SPELLED INCORRECTLY WAI...   

                                               OPM remarks  \
933                        INCORRECT DATA PER TOWN RECORDS   
3905                                                   NaN   
4942     19x8 AREA OF TOWN SKETCH IS INCORRECT PER MLS ...   
5899                                                   NaN   
6737                                                   NaN   
...                                                    ...   
1003224                                                NaN   
```

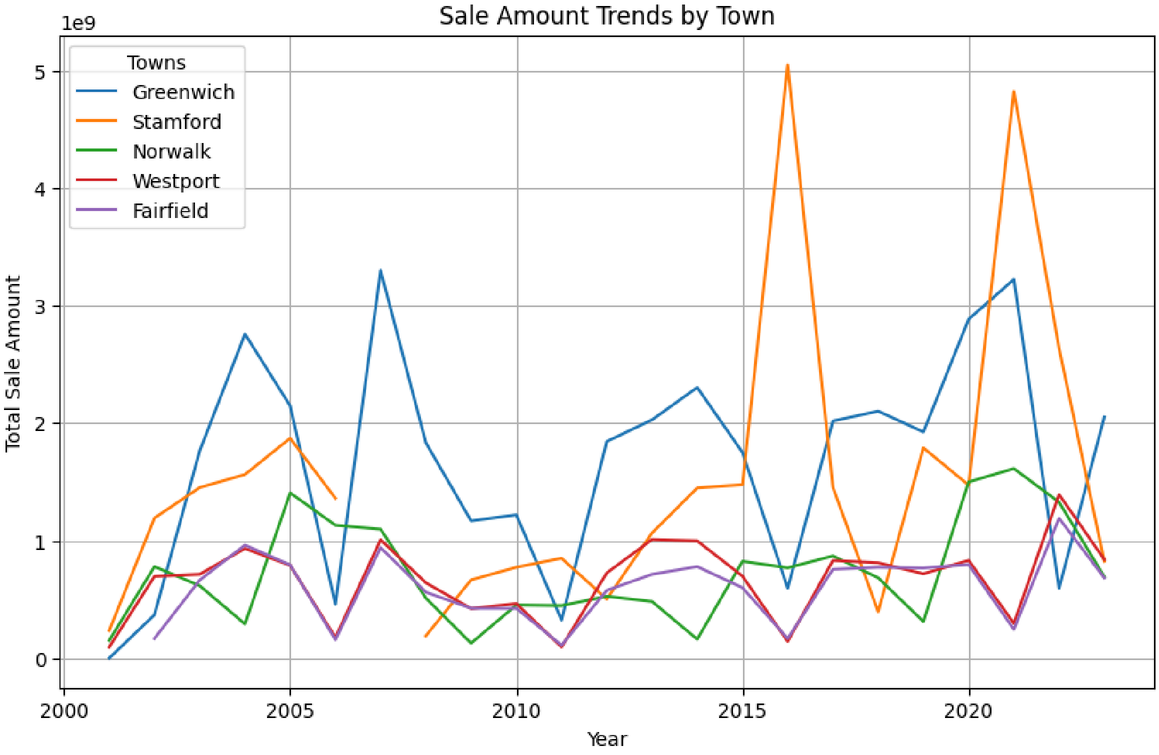
```
1016523                                                NaN
1031076                                                NaN
1058529                                                NaN
1061891                                                NaN

                                   Location
933              POINT (-73.53692 41.38822)
3905             POINT (-72.67801 41.73461)
4942             POINT (-72.68572 41.25516)
5899                                    NaN
6737                                    NaN
...                                     ...
1003224   POINT (-72.680506032 41.73528201)
1016523  POINT (-72.567789685 41.923347946)
1031076  POINT (-72.727145959 41.513904003)
1058529  POINT (-73.363145997 41.122312006)
1061891  POINT (-72.634101972 41.908486999)

[92 rows x 14 columns]
<class 'pandas.core.frame.DataFrame'>
Index: 1097537 entries, 0 to 1097628
Data columns (total 14 columns):
 #   Column            Non-Null Count    Dtype  
---  ------            --------------    -----  
 0   Serial Number     1097537 non-null  int64  
 1   List Year         1097537 non-null  int64  
 2   Date Recorded     1097535 non-null  object 
 3   Town              1097537 non-null  object 
 4   Address           1097486 non-null  object 
 5   Assessed Value    1097537 non-null  float64
 6   Sale Amount       1097537 non-null  float64
 7   Sales Ratio       1097537 non-null  float64
 8   Property Type     715106 non-null   object 
 9   Residential Type  699165 non-null   object 
 10  Non Use Code      313370 non-null   object 
 11  Assessor Remarks  171148 non-null   object 
 12  OPM remarks       13013 non-null    object 
 13  Location          298088 non-null   object 
dtypes: float64(3), int64(2), object(9)
memory usage: 125.6+ MB
None
```

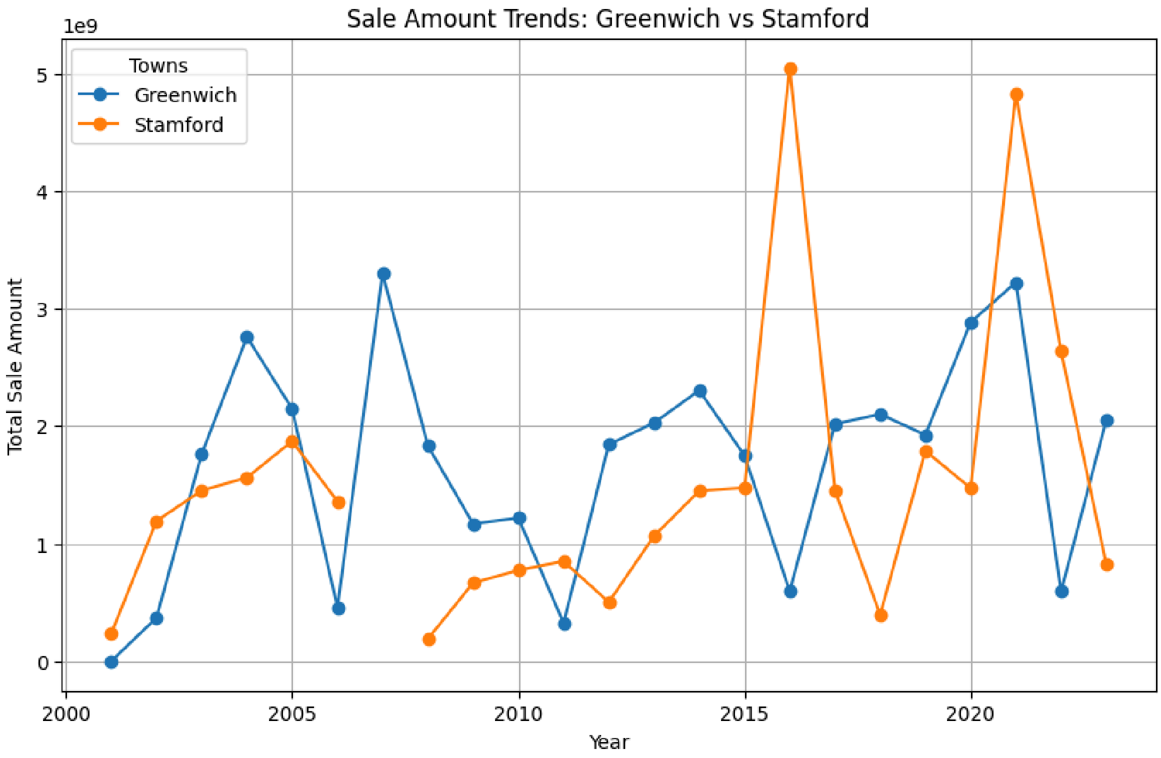


```
C:\Users\hector\AppData\Local\Temp\ipykernel_10668\516065680.py:29: FutureWarnin
g: The default of observed=False is deprecated and will be changed to True in a f
uture version of pandas. Pass observed=False to retain current behavior or observ
ed=True to adopt the future default and silence this warning.
  sales_by_range = df.groupby(['year', 'Amount Range'])['Sale Amount'].sum().unst
ack()
```

```
Amount Range        0-200k     200k-500k       500k-1M         1M-5M
year
1999.0        9.500000e+04  0.000000e+00  0.000000e+00  0.000000e+00
2001.0        1.032246e+09  1.218728e+09  4.676976e+08  3.007207e+08
2002.0        4.047768e+09  5.842831e+09  2.807996e+09  2.118725e+09
2003.0        3.595137e+09  6.075020e+09  3.024984e+09  3.238053e+09
2004.0        4.270330e+09  9.045554e+09  4.831042e+09  4.600120e+09
2005.0        3.907878e+09  1.125598e+10  5.855934e+09  6.153066e+09
2006.0        2.612561e+09  8.837296e+09  4.128769e+09  4.006803e+09
2007.0        1.912068e+09  6.985225e+09  3.803090e+09  5.735626e+09
2008.0        1.682887e+09  4.848626e+09  2.208657e+09  3.266624e+09
2009.0        2.122136e+09  4.574242e+09  1.945040e+09  2.046261e+09
2010.0        2.132318e+09  4.835148e+09  2.500625e+09  3.166788e+09
2011.0        1.798131e+09  3.722669e+09  2.002610e+09  2.363148e+09
2012.0        1.735201e+09  3.693436e+09  2.315252e+09  3.562643e+09
2013.0        2.032208e+09  4.470139e+09  2.406266e+09  3.832090e+09
2014.0        2.326717e+09  4.801572e+09  2.649190e+09  3.734251e+09
2015.0        2.680581e+09  5.701196e+09  3.082860e+09  4.554719e+09
2016.0        2.622656e+09  5.921315e+09  2.667630e+09  3.075458e+09
2017.0        2.560232e+09  6.088658e+09  3.107155e+09  4.640015e+09
2018.0        2.547569e+09  6.201392e+09  2.638059e+09  4.275537e+09
2019.0        2.769587e+09  7.304794e+09  3.000134e+09  3.631036e+09
2020.0        2.611162e+09  9.572224e+09  5.179293e+09  6.725883e+09
2021.0        2.289640e+09  1.034153e+10  6.419401e+09  8.361225e+09
2022.0        1.597184e+09  9.050773e+09  5.743202e+09  6.830642e+09
2023.0        8.486588e+08  5.483864e+09  3.613517e+09  4.950068e+09
```


Sale Amount Trends by Range
1e10
Amount Range
0-200k
200k-500k
500k-1M
1M-5M
Total Sale Amount
0.0
0.2
0.4
0.6
0.8
1.0
2000
2005
2010
2015
2020
Year

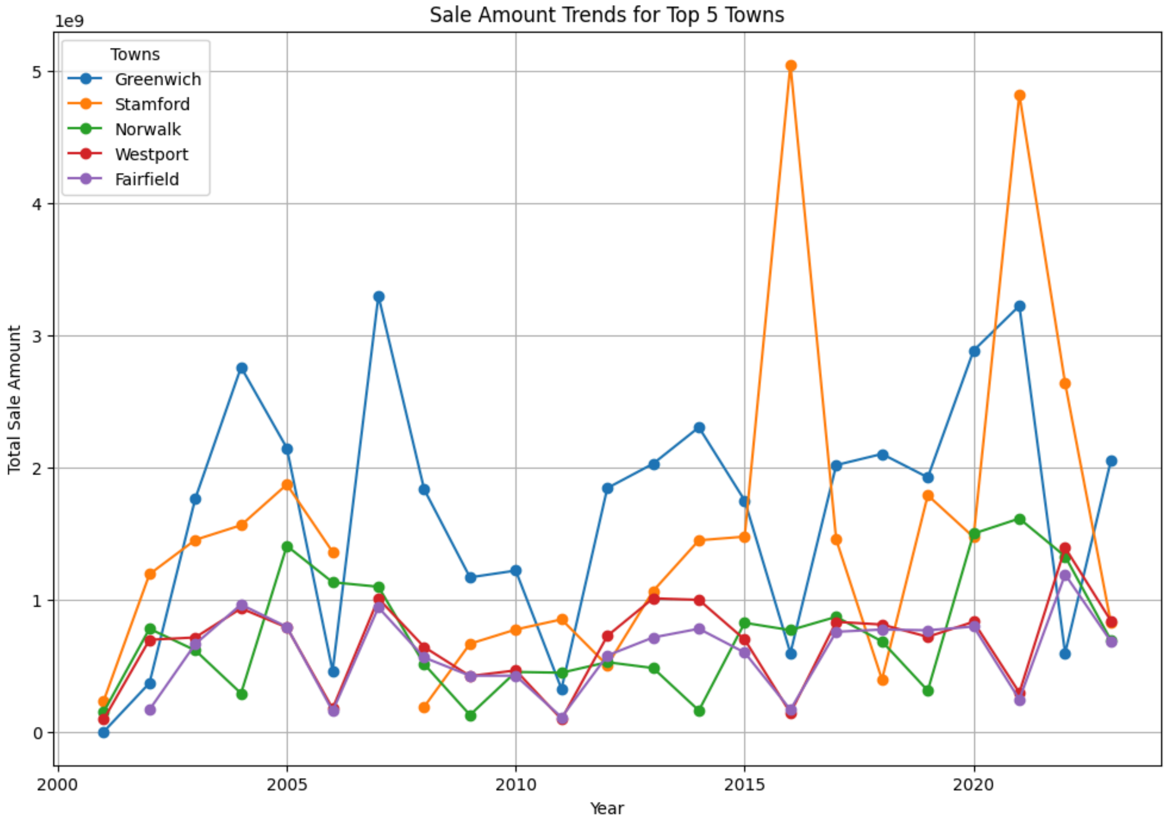

Sale Amount Trends by Town
1e9
Towns
Greenwich
Stamford
Norwalk
Westport
Fairfield
Total Sale Amount
0
1
2
3
4
5
2000
2005
2010
2015
2020
Year



Observetion:

the data for stamford shows gaps betweem 2005 and 2010, which may affect the trend accuracy for this period.

```
In [11]: sales_by_town[top_towns].plot(kind='line',figsize=(12,8), marker='o')
         plt.title('Sale Amount Trends for Top 5 Towns')
         plt.xlabel('Year')
         plt.ylabel('Total Sale Amount')
         plt.legend(title='Towns')
         plt.grid()
         plt.show()
```

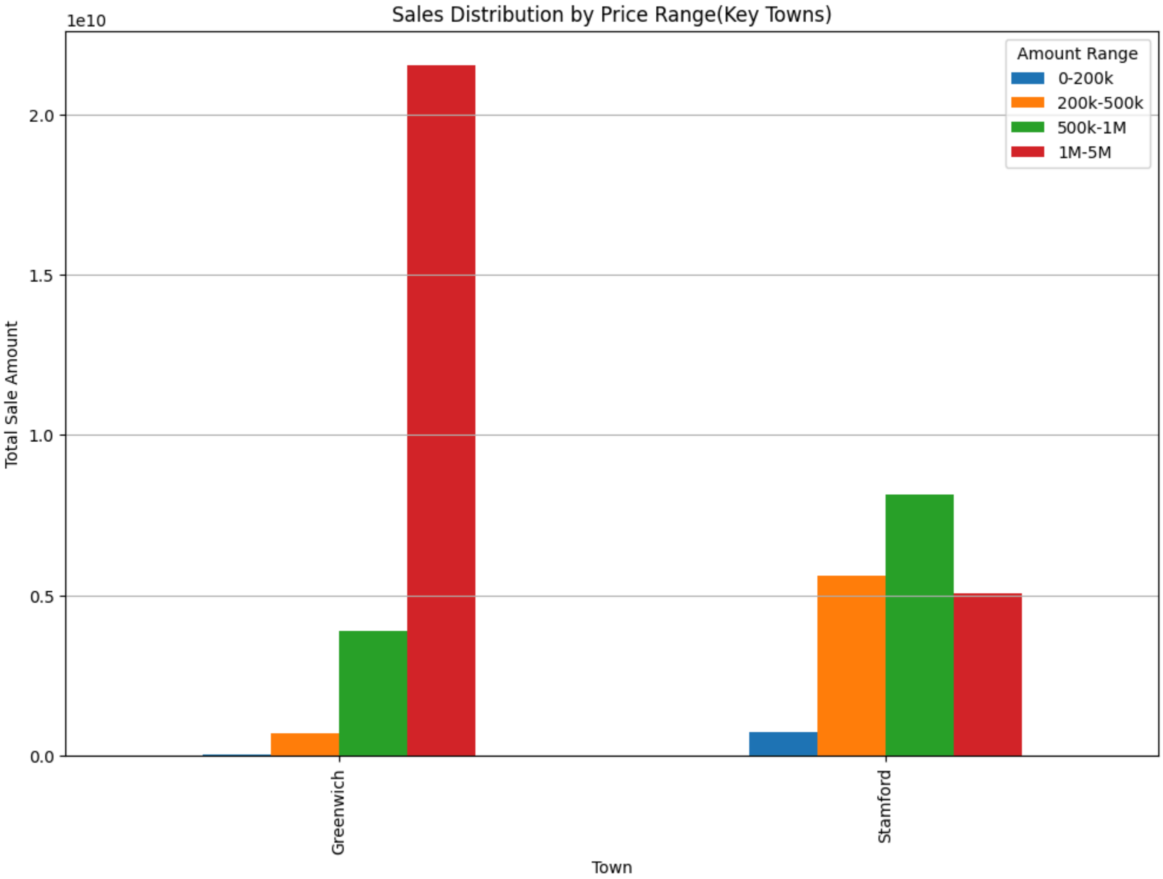


Greenwich: The real estate market in Greenwich shows periodic highs and lows, possibly due to fluctuating demand for high-value properties. It may indicate an inconsistent Luxury Market. **Stamford**: Stamford exhibits significant oscillations in sale amounts over time, suggesting volatility in property sales or market conditions. **Norwalk, Westport, Fairfield**: The analysis of Norwalk, Westport, and Fairfield demonstrates stable trends, with fewer fluctuations, indicating balanced or predictable markets."

In [12]:

```python
sales_by_town_range = df.groupby(['Town', 'Amount Range'])['Sale Amount'].sum().

sales_by_town_range.loc[key_towns].plot(kind='bar', figsize= (12,8))
plt.title('Sales Distribution by Price Range(Key Towns)')
plt.xlabel('Town')
plt.ylabel('Total Sale Amount')
plt.legend(title='Amount Range')
plt.grid(axis='y')
plt.show()
```

```
C:\Users\hector\AppData\Local\Temp\ipykernel_10668\2687581892.py:1: FutureWarnin
g: The default of observed=False is deprecated and will be changed to True in a f
uture version of pandas. Pass observed=False to retain current behavior or observ
ed=True to adopt the future default and silence this warning.
  sales_by_town_range = df.groupby(['Town', 'Amount Range'])['Sale Amount'].sum
().unstack()
```



Sales in the 200k-500k range dominate across most towns, but the 1M-5M range contributes significantly in Greenwich.

Final Summary

this analysis of real estate sales trends reveal several key insights:

1. **Overall Market Trends**: Sale amounts have shown a steady increase over time, with occasional dips in certain years.
2. **Price Range Analysis**: The majority of sales occur in the 0-200k range, but high-value properties (1M-5M) contribute significantly in towns like Greenwich.
3. **Town-Specific Trends**: -**Greenwich**: Periodic highs and lows, indicating a fluctuating Luxury Market. -**Stamford**: High volatility, possibly due to varying property types and inconsistent demand. -**Norwalk, Westport, Fairfield**: Stable trends with smaller oscillations, representing balanced markets.

These findings can guide real estate firms in tailoring their strategies to specific towns and price ranges, enabling more effective marketing and investment decisions.

```
In [ ]:
```