Sales DATA Analysis ¶

import libraries

In [64]:

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

Explonatory Data Analysis

In [65]:

```
df=pd.read_excel("D:\datset\ECOMM DATA.xlsx")
```

In [66]:

```
df.head()
```

Out[66]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	
0	32298	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen	Consumer	New York City	N
1	26341	IN-2013-77878	2013-02-05	2013-02-07	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	Ne
2	25330	IN-2013-71249	2013-10-17	2013-10-18	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Que
3	13524	ES-2013-1579342	2013-01-28	2013-01-30	First Class	KM-16375	Katherine Murray	Home Office	Berlin	
4	47221	SG-2013-4320	2013-11-05	2013-11-06	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	

5 rows × 24 columns

```
In [67]: df.tail()
```

Out[67]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	
51285	29002	IN-2014-62366	2014-06-19	2014-06-19	Same Day	KE-16420	Katrina Edelman	Corporate	Kure	Hi
51286	35398	US-2014-102288	2014-06-20	2014-06-24	Standard Class	ZC-21910	Zuschuss Carroll	Consumer	Houston	
51287	40470	US-2013-155768	2013-12-02	2013-12-02	Same Day	LB-16795	Laurel Beltran	Home Office	Oxnard	C
51288	9596	MX-2012-140767	2012-02-18	2012-02-22	Standard Class	RB-19795	Ross Baird	Home Office	Valinhos	
51289	6147	MX-2012-134460	2012-05-22	2012-05-26	Second Class	MC-18100	Mick Crebagga	Consumer	Tipitapa	M

5 rows × 24 columns

In [68]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51290 entries, 0 to 51289
Data columns (total 24 columns):
 #   Column          Non-Null Count  Dtype         
---  ------          --------------  -----         
 0   Row ID          51290 non-null  int64         
 1   Order ID        51290 non-null  object        
 2   Order Date      51290 non-null  datetime64[ns]
 3   Ship Date       51290 non-null  datetime64[ns]
 4   Ship Mode       51290 non-null  object        
 5   Customer ID     51290 non-null  object        
 6   Customer Name   51290 non-null  object        
 7   Segment         51290 non-null  object        
 8   City            51290 non-null  object        
 9   State           51290 non-null  object        
 10  Country         51290 non-null  object        
 11  Postal Code     9994 non-null   float64       
 12  Market          51290 non-null  object        
 13  Region          51290 non-null  object        
 14  Product ID      51290 non-null  object        
 15  Category        51290 non-null  object        
 16  Sub-Category    51290 non-null  object        
 17  Product Name    51290 non-null  object        
 18  Sales           51290 non-null  float64       
 19  Quantity        51290 non-null  int64         
 20  Discount        51290 non-null  float64       
 21  Profit          51290 non-null  float64       
 22  Shipping Cost   51290 non-null  float64       
 23  Order Priority  51290 non-null  object        
dtypes: datetime64[ns](2), float64(5), int64(2), object(15)
memory usage: 9.4+ MB
```

In [69]:
```
df.shape
```

Out[69]:
```
(51290, 24)
```

In [70]:
```
df.isnull().sum()
```

Out[70]:
```
Row ID                0
Order ID              0
Order Date            0
Ship Date             0
Ship Mode             0
Customer ID           0
Customer Name         0
Segment               0
City                  0
State                 0
Country               0
Postal Code       41296
Market                0
Region                0
Product ID            0
Category              0
Sub-Category          0
Product Name          0
Sales                 0
Quantity              0
Discount              0
Profit                0
Shipping Cost         0
Order Priority        0
dtype: int64
```

In [71]:
```
df.drop(columns=['Postal Code'],inplace=True) # postel code has large amou
```

In [72]:
```
df.isna().sum()
```

Out[72]:
```
Row ID            0
Order ID          0
Order Date        0
Ship Date         0
Ship Mode         0
Customer ID       0
Customer Name     0
Segment           0
City              0
State             0
Country           0
Market            0
Region            0
Product ID        0
Category          0
Sub-Category      0
Product Name      0
Sales             0
Quantity          0
Discount          0
Profit            0
Shipping Cost     0
Order Priority    0
dtype: int64
```

print total sales

In [73]:

```
total_sales = df['Sales'].sum()
print("toatl sales = ",total_sales)
```

```
toatl sales =  12642501.909880001
```

converting the order date to date time format

In [74]:

```
df['Order Date'] = pd.to_datetime(df['Order Date'])
```

set order date as index

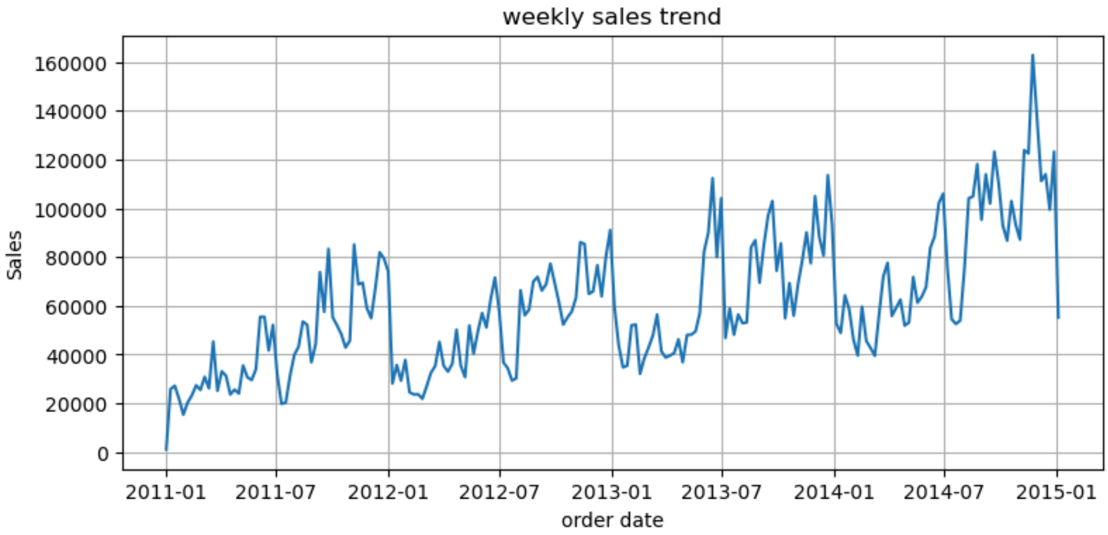
In [77]:

```
df.set_index('Order Date', inplace=True)
```

weekly sales trend

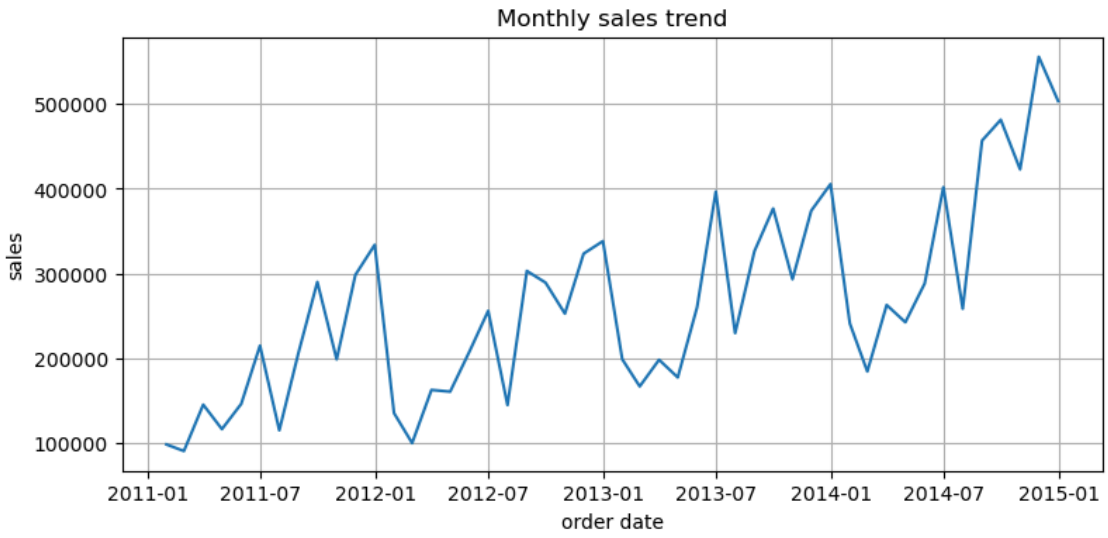
In [78]:

```
weekly_sales=df['Sales'].resample('W').sum() #Resample the data to weakly
plt.figure(figsize=(9,4))
plt.plot(weekly_sales.index,weekly_sales.values)
plt.title('weekly sales trend')
plt.xlabel('order date')
plt.ylabel('Sales')
plt.grid(True)
plt.show()
```



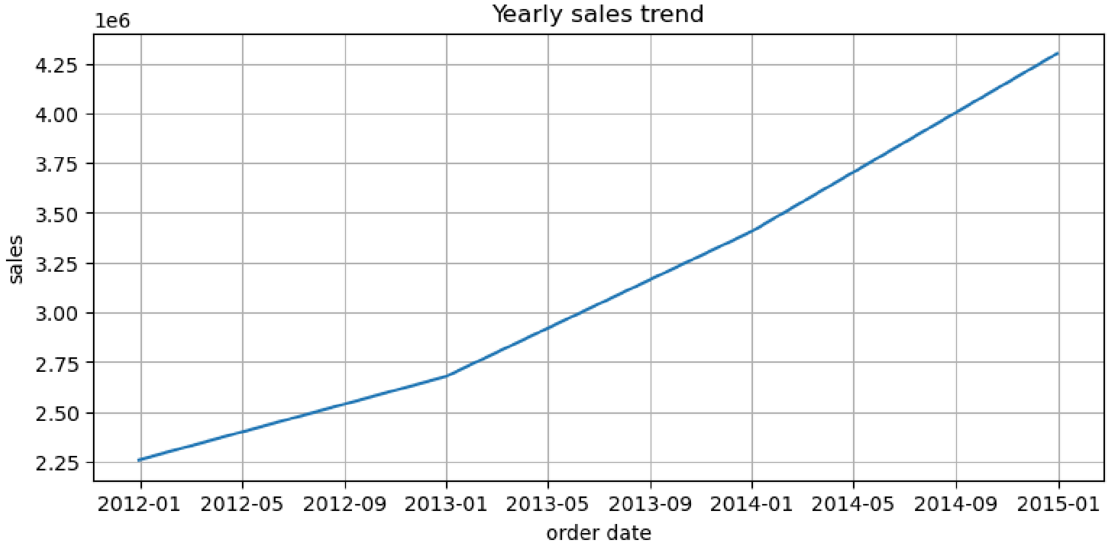
Monthly sales trend

In [79]:
```
monthly_sales=df['Sales'].resample('M').sum() # Resample the data to Month
plt.figure(figsize=(9,4))
plt.plot(monthly_sales.index,monthly_sales.values)
plt.title('Monthly sales trend')
plt.xlabel('order date')
plt.ylabel('sales')
plt.grid(True)
plt.show()
```



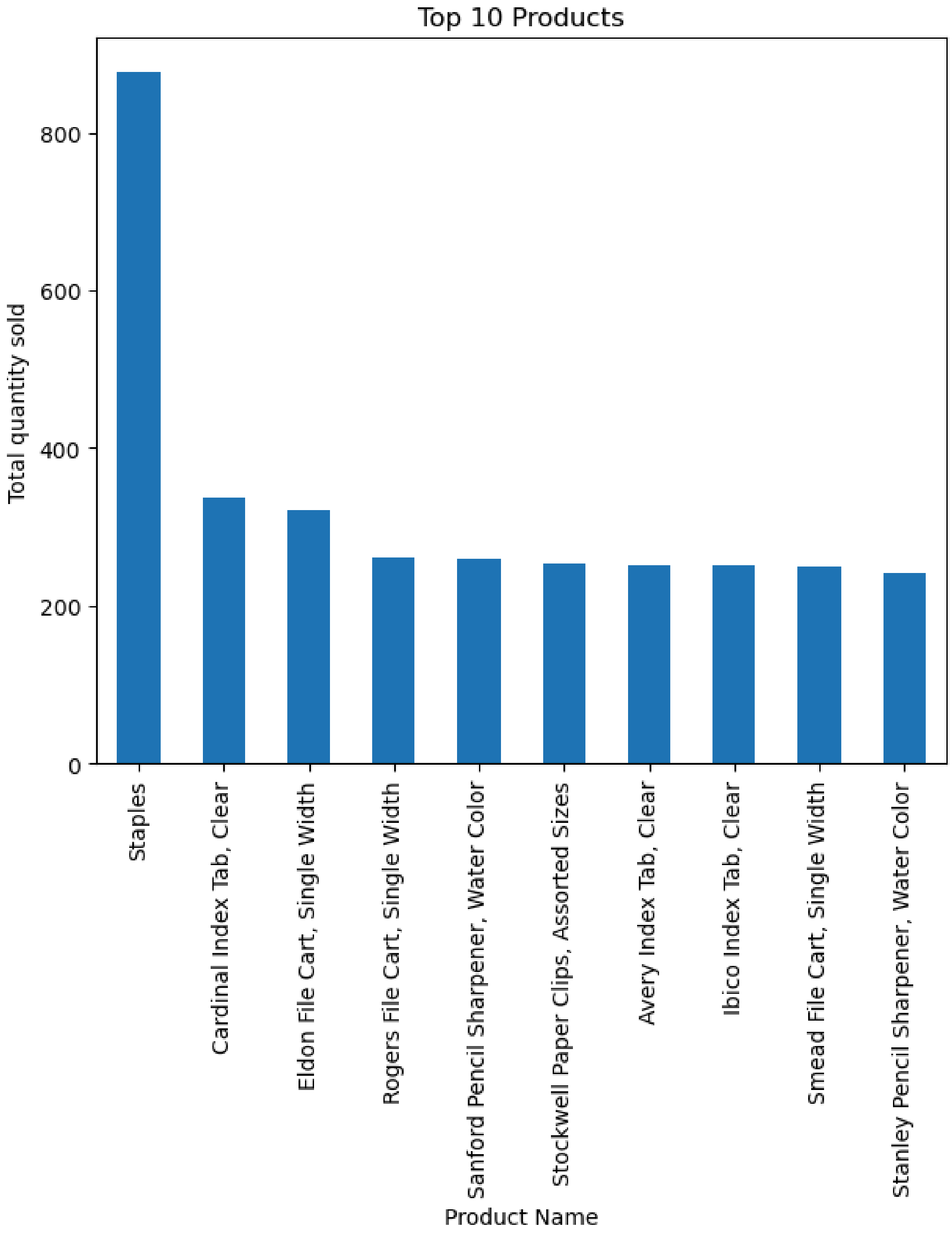
Yearly sales trend

In [80]:
```
yearly_sales=df['Sales'].resample('Y').sum() # Resample the data to Monthl
plt.figure(figsize=(9,4))
plt.plot(yearly_sales.index,yearly_sales.values)
plt.title('Yearly sales trend')
plt.xlabel('order date')
plt.ylabel('sales')
plt.grid(True)
plt.show()
```



TOP 10 SELLING PRODUCTS

```
In [81]: product_sales=df.groupby('Product Name')['Quantity'].sum().sort_values(asc
         plt.figure(figsize=(7,6))
         top_10_products=product_sales.head(10)
         top_10_products.plot(kind='bar')
         plt.title('Top 10 Products')
         plt.xlabel('Product Name')
         plt.ylabel('Total quantity sold')
         plt.show()
```


Top 10 Products
Total quantity sold
Staples
Cardinal Index Tab, Clear
Eldon File Cart, Single Width
Rogers File Cart, Single Width
Sanford Pencil Sharpener, Water Color
Stockwell Paper Clips, Assorted Sizes
Avery Index Tab, Clear
Ibico Index Tab, Clear
Smead File Cart, Single Width
Stanley Pencil Sharpener, Water Color
Product Name