

Name: - L Prathyusha

TASK 1 - Data science and Business Analytics Internship

In [64]:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

task: -

1. Perform 'Exploratory Data Analysis' on dataset 'SampleSuperstore'
2. As a business manager, try to find out the weak areas where you can work to make more profit.
3. What all business problems you can derive by exploring the data?

In [69]:

```
sales=pd.read_csv('C:/Users/Prathyu Lachireddy/Desktop/SampleSuperstore - SampleSuperstore.
sales
```

Out[69]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category
0	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases
1	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs
2	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Laboratory Equipment
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables
4	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage
...
9989	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings
9990	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings
9991	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones
9992	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper
9993	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances

9994 rows × 13 columns



In [4]:

```
sales.head()
```

Out[4]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	
0	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	26
1	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	73
2	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	1
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	95
4	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	2

In [5]:

```
sales.tail()
```

Out[5]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	
9989	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishin	
9990	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishin	
9991	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phon	
9992	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Pa	
9993	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Applianc	

In [6]:

sales.nunique

Out[6]:

```
<bound method DataFrame.nunique of
untry      City      State \
0      Second Class  Consumer United States      Henderson      Kentucky
1      Second Class  Consumer United States      Henderson      Kentucky
2      Second Class  Corporate United States      Los Angeles      California
3      Standard Class  Consumer United States      Fort Lauderdale      Florida
4      Standard Class  Consumer United States      Fort Lauderdale      Florida
...      ...      ...      ...      ...      ...
9989     Second Class  Consumer United States      Miami      Florida
9990     Standard Class  Consumer United States      Costa Mesa      California
9991     Standard Class  Consumer United States      Costa Mesa      California
9992     Standard Class  Consumer United States      Costa Mesa      California
9993     Second Class  Consumer United States      Westminster      California

Postal Code Region      Category Sub-Category      Sales      Quantity
\
0      42420      South      Furniture      Bookcases      261.9600      2
1      42420      South      Furniture      Chairs      731.9400      3
2      90036      West      Office Supplies      Labels      14.6200      2
3      33311      South      Furniture      Tables      957.5775      5
4      33311      South      Office Supplies      Storage      22.3680      2
...      ...      ...      ...      ...      ...
9989     33180      South      Furniture      Furnishings      25.2480      3
9990     92627      West      Furniture      Furnishings      91.9600      2
9991     92627      West      Technology      Phones      258.5760      2
9992     92627      West      Office Supplies      Paper      29.6000      4
9993     92683      West      Office Supplies      Appliances      243.1600      2

Discount      Profit
0      0.00      41.9136
1      0.00      219.5820
2      0.00      6.8714
3      0.45      -383.0310
4      0.20      2.5164
...      ...      ...
9989     0.20      4.1028
9990     0.00      15.6332
9991     0.20      19.3932
9992     0.00      13.3200
9993     0.00      72.9480
```

[9994 rows x 13 columns]>

In [10]:

```
sales.dtypes
```

Out[10]:

```
Ship Mode      object
Segment        object
Country        object
City           object
State          object
Postal Code    int64
Region         object
Category       object
Sub-Category   object
Sales          float64
Quantity       int64
Discount       float64
Profit         float64
dtype: object
```

In [11]:

```
sales.corr()
```

Out[11]:

	Postal Code	Sales	Quantity	Discount	Profit
Postal Code	1.000000	-0.023854	0.012761	0.058443	-0.029961
Sales	-0.023854	1.000000	0.200795	-0.028190	0.479064
Quantity	0.012761	0.200795	1.000000	0.008623	0.066253
Discount	0.058443	-0.028190	0.008623	1.000000	-0.219487
Profit	-0.029961	0.479064	0.066253	-0.219487	1.000000

In [14]:

```
sales.shape
```

Out[14]:

```
(9994, 13)
```

In [16]:

```
sales.describe()
```

Out[16]:

	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	55190.379428	229.858001	3.789574	0.156203	28.656896
std	32063.693350	623.245101	2.225110	0.206452	234.260108
min	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	90008.000000	209.940000	5.000000	0.200000	29.364000
max	99301.000000	22638.480000	14.000000	0.800000	8399.976000

In [17]:

```
sales.isnull().sum()
```

Out[17]:

Ship Mode 0
Segment 0
Country 0
City 0
State 0
Postal Code 0
Region 0
Category 0
Sub-Category 0
Sales 0
Quantity 0
Discount 0
Profit 0
dtype: int64

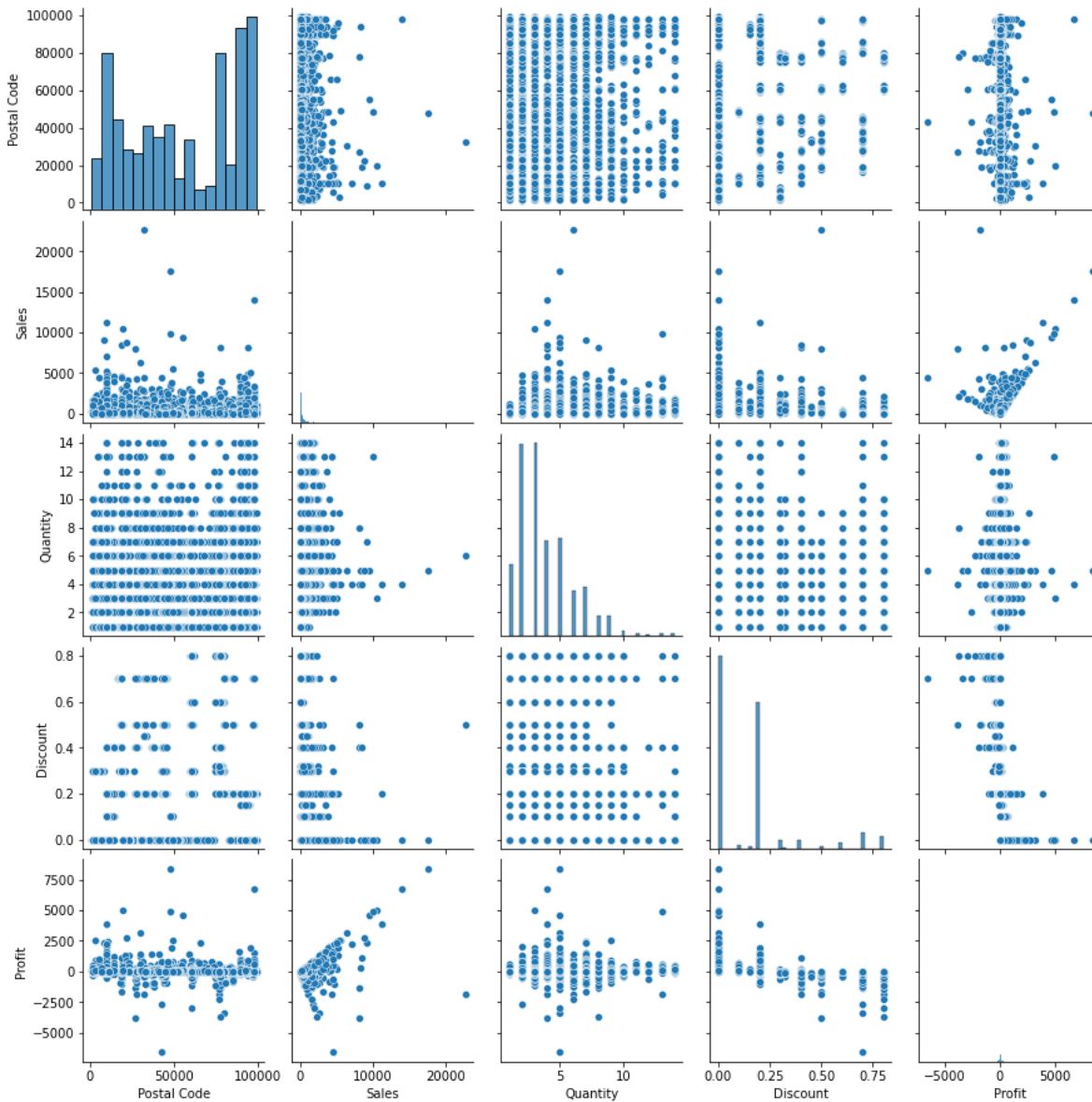
Data Visualisation

In [19]:

```
sns.pairplot(sales)
```

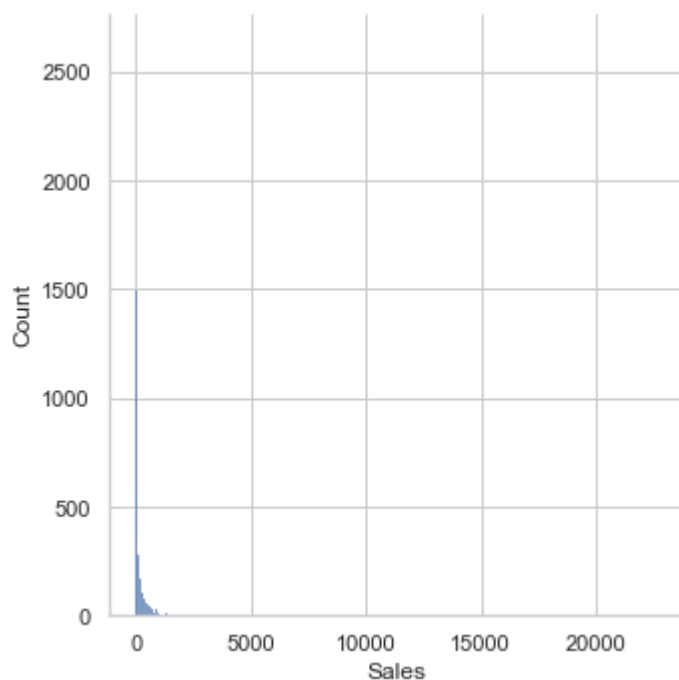
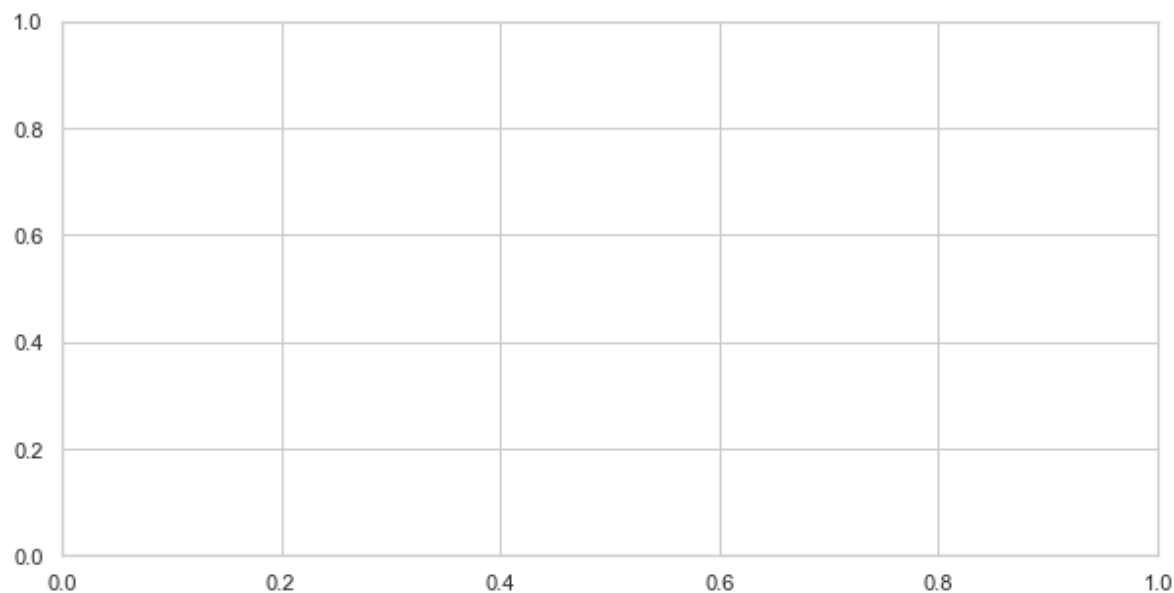
Out[19]:

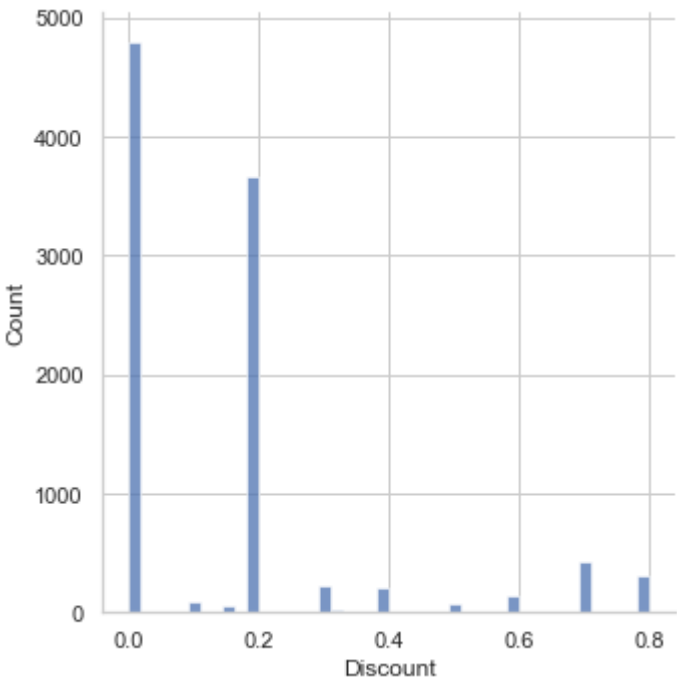
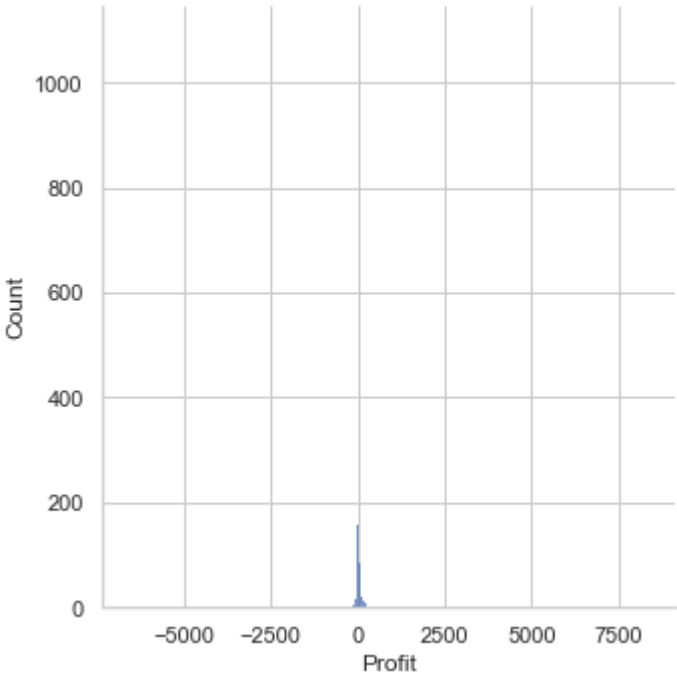
<seaborn.axisgrid.PairGrid at 0x1f3dd85fa60>

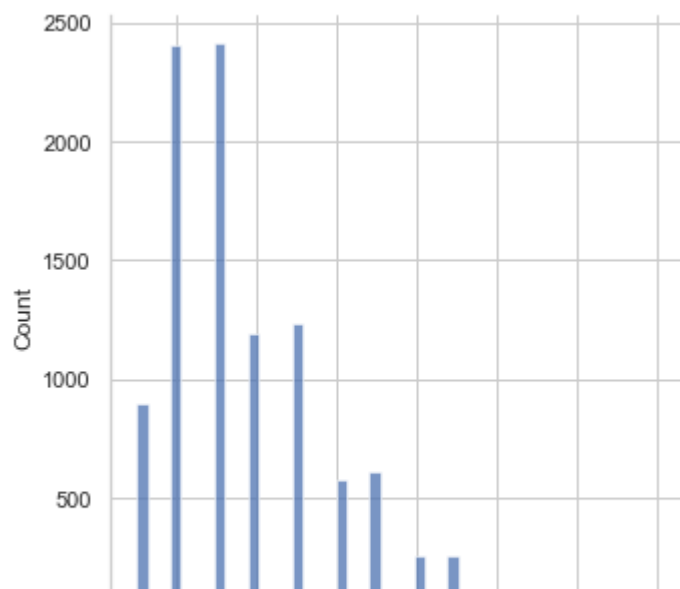


In [67]:

```
fig_dims = (10, 5)
fig, ax = plt.subplots(figsize=fig_dims)
sns.displot(sales['Sales'],ax=axes[0,0])
sns.displot(sales['Profit'],ax=axes[0,1])
sns.displot(sales['Discount'],ax=axes[1,0])
sns.displot(sales['Quantity'],ax=axes[1,1])
plt.show()
```

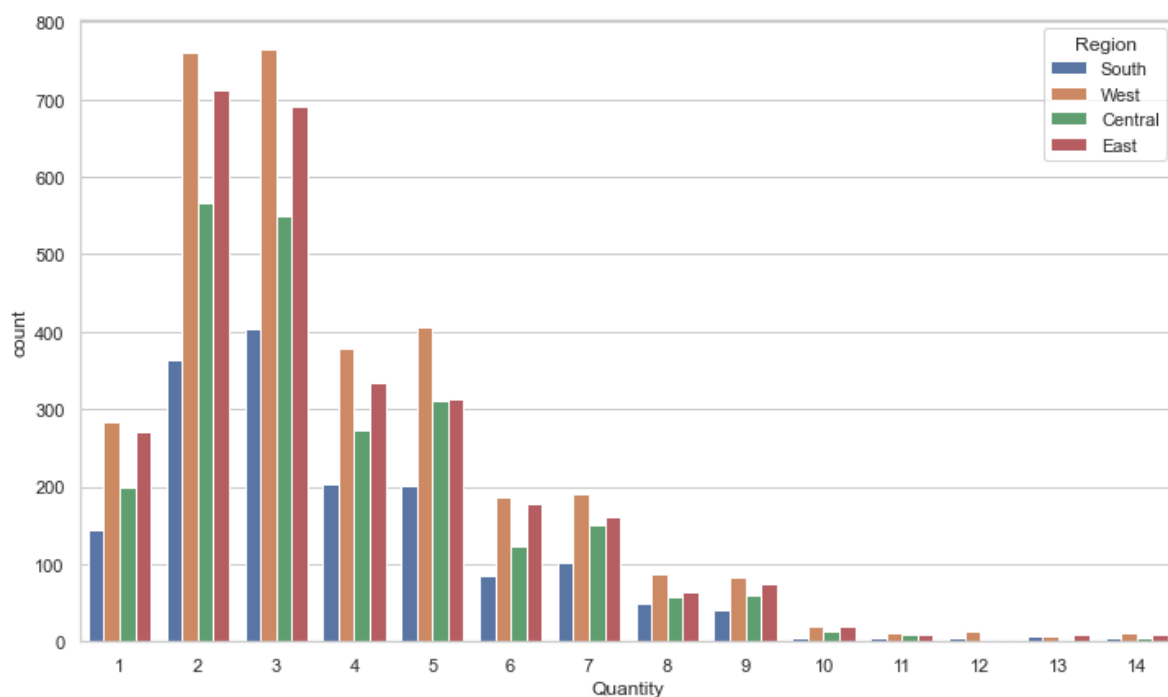






In [66]:

```
fig,ax=plt.subplots(1,1,figsize=(12,7))
sns.countplot(sales['Quantity'],hue=sales['Region'])
plt.show()
```

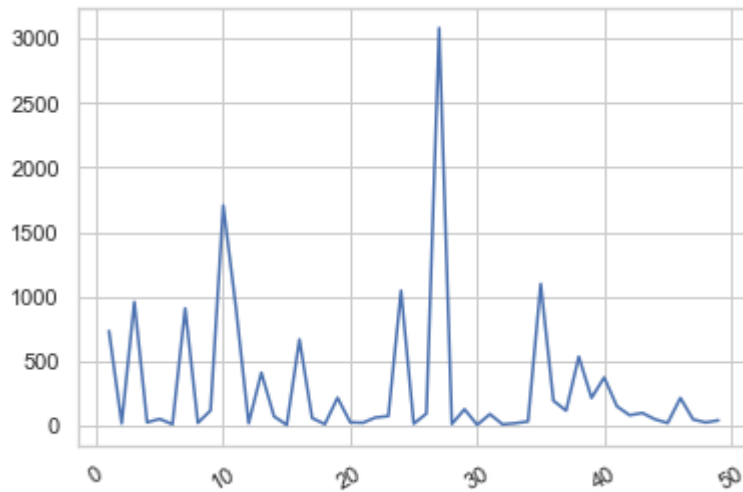


In [70]:

```
plt.plot(sales[1:50]['Sales'])  
plt.xticks(rotation=30)  
plt.show
```

Out[70]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

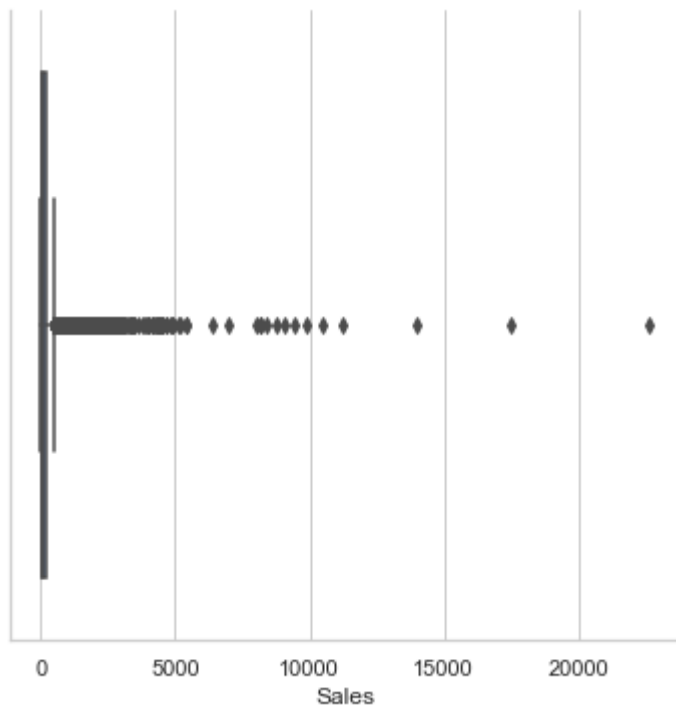


In [72]:

```
sns.catplot(x='Sales', kind='box', data=sales)
```

Out[72]:

<seaborn.axisgrid.FacetGrid at 0x1f3ee3e5370>



In []: