

```
In [101... import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

## DATASET LOADING

```
In [102... df=pd.read_csv("C:\\Users\\hp\\Downloads\\financial sample dataset.csv")
df
```

Out[102]:

|     | Segment          | Country                  | Product   | Discount Band | Units Sold | Manufacturing Price | Sale Price | Gross Sales |     |
|-----|------------------|--------------------------|-----------|---------------|------------|---------------------|------------|-------------|-----|
| 0   | Government       | Canada                   | Carretera | None          | 1618.5     | 3                   | 20         | 32370.0     | 1   |
| 1   | Government       | Germany                  | Carretera | None          | 1321.0     | 3                   | 20         | 26420.0     | 1   |
| 2   | Midmarket        | France                   | Carretera | None          | 2178.0     | 3                   | 15         | 32670.0     | 1   |
| 3   | Midmarket        | Germany                  | Carretera | None          | 888.0      | 3                   | 15         | 13320.0     | 1   |
| 4   | Midmarket        | Mexico                   | Carretera | None          | 2470.0     | 3                   | 15         | 37050.0     | 1   |
| ... | ...              | ...                      | ...       | ...           | ...        | ...                 | ...        | ...         | ... |
| 695 | Small Business   | France                   | Amarilla  | High          | 2475.0     | 260                 | 300        | 742500.0    | 11  |
| 696 | Small Business   | Mexico                   | Amarilla  | High          | 546.0      | 260                 | 300        | 163800.0    | 2   |
| 697 | Government       | Mexico                   | Montana   | High          | 1368.0     | 5                   | 7          | 9576.0      |     |
| 698 | Government       | Canada                   | Paseo     | High          | 723.0      | 10                  | 7          | 5061.0      |     |
| 699 | Channel Partners | United States of America | VTT       | High          | 1806.0     | 250                 | 12         | 21672.0     |     |

700 rows × 16 columns

## EXPLORING DATA

```
In [103... df.shape
```

Out[103]: (700, 16)

```
In [104... df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Segment                700 non-null    object
1   Country                700 non-null    object
2   Product                700 non-null    object
3   Discount Band          700 non-null    object
4   Units Sold             700 non-null    float64
5   Manufacturing Price     700 non-null    int64
6   Sale Price             700 non-null    int64
7   Gross Sales            700 non-null    float64
8   Discounts              700 non-null    float64
9   Sales                  700 non-null    float64
10  COGS                   700 non-null    float64
11  Profit                 700 non-null    float64
12  Date                   700 non-null    object
13  Month Number           700 non-null    int64
14  Month Name             700 non-null    object
15  Year                   700 non-null    int64
dtypes: float64(6), int64(4), object(6)
memory usage: 87.6+ KB
```

In [105...

df.describe()

Out[105]:

|              | Units Sold  | Manufacturing Price | Sale Price | Gross Sales  | Discounts     | Sales        |
|--------------|-------------|---------------------|------------|--------------|---------------|--------------|
| <b>count</b> | 700.000000  | 700.000000          | 700.000000 | 7.000000e+02 | 700.000000    | 7.000000e+02 |
| <b>mean</b>  | 1608.294286 | 96.477143           | 118.428571 | 1.827594e+05 | 14227.586475  | 1.696091e+05 |
| <b>std</b>   | 867.427859  | 108.602612          | 136.775515 | 2.542623e+05 | 22651.928275  | 2.367263e+05 |
| <b>min</b>   | 200.000000  | 3.000000            | 7.000000   | 1.799000e+03 | 18.410000     | 1.655080e+03 |
| <b>25%</b>   | 905.000000  | 5.000000            | 12.000000  | 1.739175e+04 | 1188.960000   | 1.592800e+04 |
| <b>50%</b>   | 1542.500000 | 10.000000           | 20.000000  | 3.798000e+04 | 3872.000000   | 3.554020e+04 |
| <b>75%</b>   | 2229.125000 | 250.000000          | 300.000000 | 2.790250e+05 | 15956.347500  | 2.610775e+05 |
| <b>max</b>   | 4492.500000 | 260.000000          | 350.000000 | 1.207500e+06 | 149677.500000 | 1.159200e+06 |

In [106...

df.columns

```
Out[106]: Index(['Segment', 'Country', 'Product', 'Discount Band', 'Units Sold',
                'Manufacturing Price', 'Sale Price', 'Gross Sales', 'Discounts',
                'Sales', 'COGS', 'Profit', 'Date', 'Month Number',
                'Month Name', 'Year'],
                dtype='object')
```

In [107...

df.dtypes

```
Out[107]: Segment          object
          Country          object
          Product          object
          Discount Band    object
          Units Sold       float64
          Manufacturing Price int64
          Sale Price       int64
          Gross Sales      float64
          Discounts        float64
          Sales            float64
          COGS             float64
          Profit           float64
          Date             object
          Month Number     int64
          Month Name       object
          Year             int64
          dtype: object
```

```
In [108... df.isnull().sum()
```

```
Out[108]: Segment          0
          Country          0
          Product          0
          Discount Band    0
          Units Sold       0
          Manufacturing Price int64
          Sale Price       0
          Gross Sales      0
          Discounts        0
          Sales            0
          COGS             0
          Profit           0
          Date             0
          Month Number     0
          Month Name       0
          Year             0
          dtype: int64
```

```
In [109... df.duplicated().sum()
```

```
Out[109]: 0
```

```
In [110... for column in df.columns:
            print(f"Value Counts for {column}:")
            print(df[column].value_counts())
            print()
```

## Value Counts for Segment:

## Segment

|                  |     |
|------------------|-----|
| Government       | 300 |
| Midmarket        | 100 |
| Channel Partners | 100 |
| Enterprise       | 100 |
| Small Business   | 100 |

Name: count, dtype: int64

## Value Counts for Country:

## Country

|                          |     |
|--------------------------|-----|
| Canada                   | 140 |
| Germany                  | 140 |
| France                   | 140 |
| Mexico                   | 140 |
| United States of America | 140 |

Name: count, dtype: int64

## Value Counts for Product :

## Product

|           |     |
|-----------|-----|
| Paseo     | 202 |
| Velo      | 109 |
| VTT       | 109 |
| Amarilla  | 94  |
| Carretera | 93  |
| Montana   | 93  |

Name: count, dtype: int64

## Value Counts for Discount Band :

## Discount Band

|        |     |
|--------|-----|
| High   | 245 |
| Medium | 242 |
| Low    | 160 |
| None   | 53  |

Name: count, dtype: int64

## Value Counts for Units Sold:

## Units Sold

|        |   |
|--------|---|
| 727.0  | 5 |
| 2844.0 | 4 |
| 1916.0 | 4 |
| 663.0  | 4 |
| 1743.0 | 4 |
| ..     |   |
| 2723.0 | 1 |
| 1460.0 | 1 |
| 980.0  | 1 |
| 887.0  | 1 |
| 1806.0 | 1 |

Name: count, Length: 510, dtype: int64

## Value Counts for Manufacturing Price :

## Manufacturing Price

|     |     |
|-----|-----|
| 10  | 202 |
| 120 | 109 |
| 250 | 109 |

|     |    |
|-----|----|
| 260 | 94 |
|-----|----|

|   |    |
|---|----|
| 3 | 93 |
|---|----|

|   |    |
|---|----|
| 5 | 93 |
|---|----|

Name: count, dtype: int64

Value Counts for Sale Price :

Sale Price

|    |     |
|----|-----|
| 20 | 100 |
|----|-----|

|    |     |
|----|-----|
| 15 | 100 |
|----|-----|

|     |     |
|-----|-----|
| 350 | 100 |
|-----|-----|

|    |     |
|----|-----|
| 12 | 100 |
|----|-----|

|     |     |
|-----|-----|
| 125 | 100 |
|-----|-----|

|     |     |
|-----|-----|
| 300 | 100 |
|-----|-----|

|   |     |
|---|-----|
| 7 | 100 |
|---|-----|

Name: count, dtype: int64

Value Counts for Gross Sales :

Gross Sales

|         |   |
|---------|---|
| 37050.0 | 3 |
|---------|---|

|          |   |
|----------|---|
| 738000.0 | 3 |
|----------|---|

|         |   |
|---------|---|
| 22710.0 | 3 |
|---------|---|

|        |   |
|--------|---|
| 4404.0 | 3 |
|--------|---|

|         |   |
|---------|---|
| 82875.0 | 3 |
|---------|---|

..

|          |   |
|----------|---|
| 110875.0 | 1 |
|----------|---|

|         |   |
|---------|---|
| 30450.0 | 1 |
|---------|---|

|        |   |
|--------|---|
| 6744.0 | 1 |
|--------|---|

|         |   |
|---------|---|
| 22350.0 | 1 |
|---------|---|

|         |   |
|---------|---|
| 21672.0 | 1 |
|---------|---|

Name: count, Length: 550, dtype: int64

Value Counts for Discounts :

Discounts

|            |    |
|------------|----|
| 14227.5900 | 51 |
|------------|----|

|           |   |
|-----------|---|
| 5690.0000 | 3 |
|-----------|---|

|            |   |
|------------|---|
| 20139.0000 | 3 |
|------------|---|

|           |   |
|-----------|---|
| 1218.6000 | 3 |
|-----------|---|

|            |   |
|------------|---|
| 14227.5862 | 2 |
|------------|---|

..

|           |   |
|-----------|---|
| 1965.6000 | 1 |
|-----------|---|

|           |   |
|-----------|---|
| 2567.6000 | 1 |
|-----------|---|

|           |   |
|-----------|---|
| 1582.5600 | 1 |
|-----------|---|

|          |   |
|----------|---|
| 610.6800 | 1 |
|----------|---|

|           |   |
|-----------|---|
| 3250.8000 | 1 |
|-----------|---|

Name: count, Length: 516, dtype: int64

Value Counts for Sales :

Sales

|          |   |
|----------|---|
| 20794.80 | 2 |
|----------|---|

|           |   |
|-----------|---|
| 334302.50 | 2 |
|-----------|---|

|           |   |
|-----------|---|
| 136560.00 | 2 |
|-----------|---|

|          |   |
|----------|---|
| 26945.60 | 2 |
|----------|---|

|         |   |
|---------|---|
| 9662.40 | 2 |
|---------|---|

..

|          |   |
|----------|---|
| 28623.00 | 1 |
|----------|---|

|         |   |
|---------|---|
| 6339.36 | 1 |
|---------|---|

```
21009.00    1
53594.10    1
18421.20    1
Name: count, Length: 559, dtype: int64
```

Value Counts for COGS :

```
COGS
17430.0    4
8655.0     3
24700.0    3
1101.0     3
15140.0    3
..
20300.0    1
1686.0     1
11175.0    1
38010.0    1
5418.0     1
Name: count, Length: 545, dtype: int64
```

Value Counts for Profit :

```
Profit
24133.86    5
10768.80    2
6822.50     2
7829.35     2
11635.60    2
..
-2217.50    1
415.54      1
8323.00     1
4653.36     1
13003.20    1
Name: count, Length: 557, dtype: int64
```

Value Counts for Date:

```
Date
01-06-2014    70
01-12-2014    70
01-10-2013    70
01-10-2014    70
01-01-2014    35
01-03-2014    35
01-07-2014    35
01-08-2014    35
01-09-2014    35
01-02-2014    35
01-09-2013    35
01-11-2013    35
01-12-2013    35
01-04-2014    35
01-05-2014    35
01-11-2014    35
Name: count, dtype: int64
```

Value Counts for Month Number:

Month Number

10 140

12 105

6 70

9 70

11 70

1 35

3 35

7 35

8 35

2 35

4 35

5 35

Name: count, dtype: int64

Value Counts for Month Name :

Month Name

October 140

December 105

June 70

September 70

November 70

January 35

March 35

July 35

August 35

February 35

April 35

May 35

Name: count, dtype: int64

Value Counts for Year:

Year

2014 525

2013 175

Name: count, dtype: int64

In [114...

```
number_columns=df.drop(['Segment',' Product ', ' Discount Band ','Date', ' Month Na  
number_columns
```

Out[114]:

|     | Units Sold | Manufacturing Price | Sale Price | Gross Sales | Sales     | COGS     | Profit   | Month Number | Year |
|-----|------------|---------------------|------------|-------------|-----------|----------|----------|--------------|------|
| 0   | 1618.5     | 3                   | 20         | 32370.0     | 32370.00  | 16185.0  | 16185.00 | 1            | 2014 |
| 1   | 1321.0     | 3                   | 20         | 26420.0     | 26420.00  | 13210.0  | 13210.00 | 1            | 2014 |
| 2   | 2178.0     | 3                   | 15         | 32670.0     | 32670.00  | 21780.0  | 10890.00 | 6            | 2014 |
| 3   | 888.0      | 3                   | 15         | 13320.0     | 13320.00  | 8880.0   | 4440.00  | 6            | 2014 |
| 4   | 2470.0     | 3                   | 15         | 37050.0     | 37050.00  | 24700.0  | 12350.00 | 6            | 2014 |
| ... | ...        | ...                 | ...        | ...         | ...       | ...      | ...      | ...          | ...  |
| 695 | 2475.0     | 260                 | 300        | 742500.0    | 631125.00 | 618750.0 | 12375.00 | 3            | 2014 |
| 696 | 546.0      | 260                 | 300        | 163800.0    | 139230.00 | 136500.0 | 2730.00  | 10           | 2014 |
| 697 | 1368.0     | 5                   | 7          | 9576.0      | 8139.60   | 6840.0   | 1299.60  | 2            | 2014 |
| 698 | 723.0      | 10                  | 7          | 5061.0      | 4301.85   | 3615.0   | 686.85   | 4            | 2014 |
| 699 | 1806.0     | 250                 | 12         | 21672.0     | 18421.20  | 5418.0   | 13003.20 | 5            | 2014 |

700 rows × 9 columns

In [141...

number\_columns.dtypes

Out[141]:

```
Units Sold          float64
Manufacturing Price int64
Sale Price          int64
Gross Sales         float64
Sales               float64
COGS                float64
Profit              float64
Month Number        int64
Year                int64
dtype: object
```

In [117...

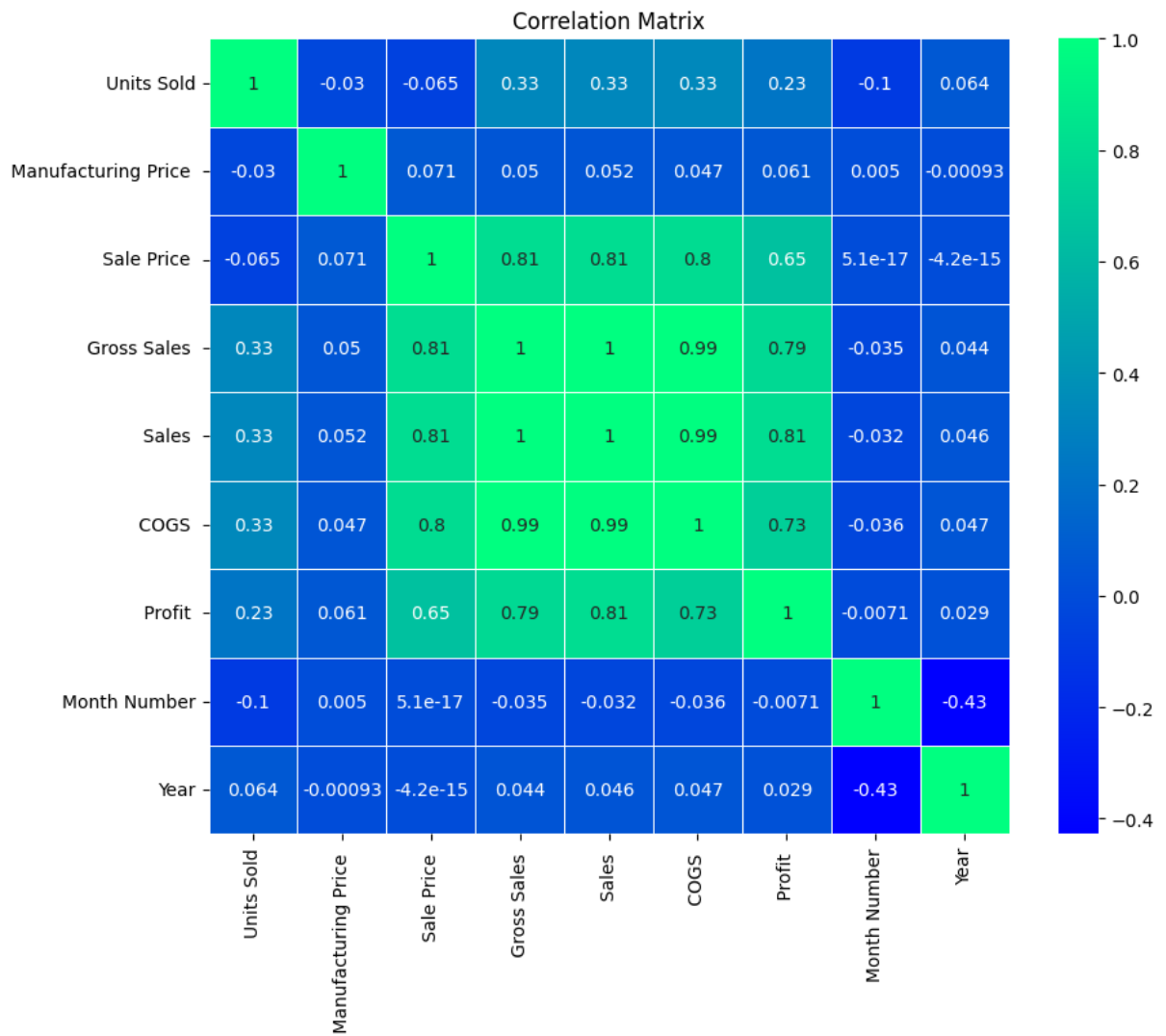
correlation\_matrix=number\_columns.corr()

## RELATIONSHIP PLOTTING

In [190...

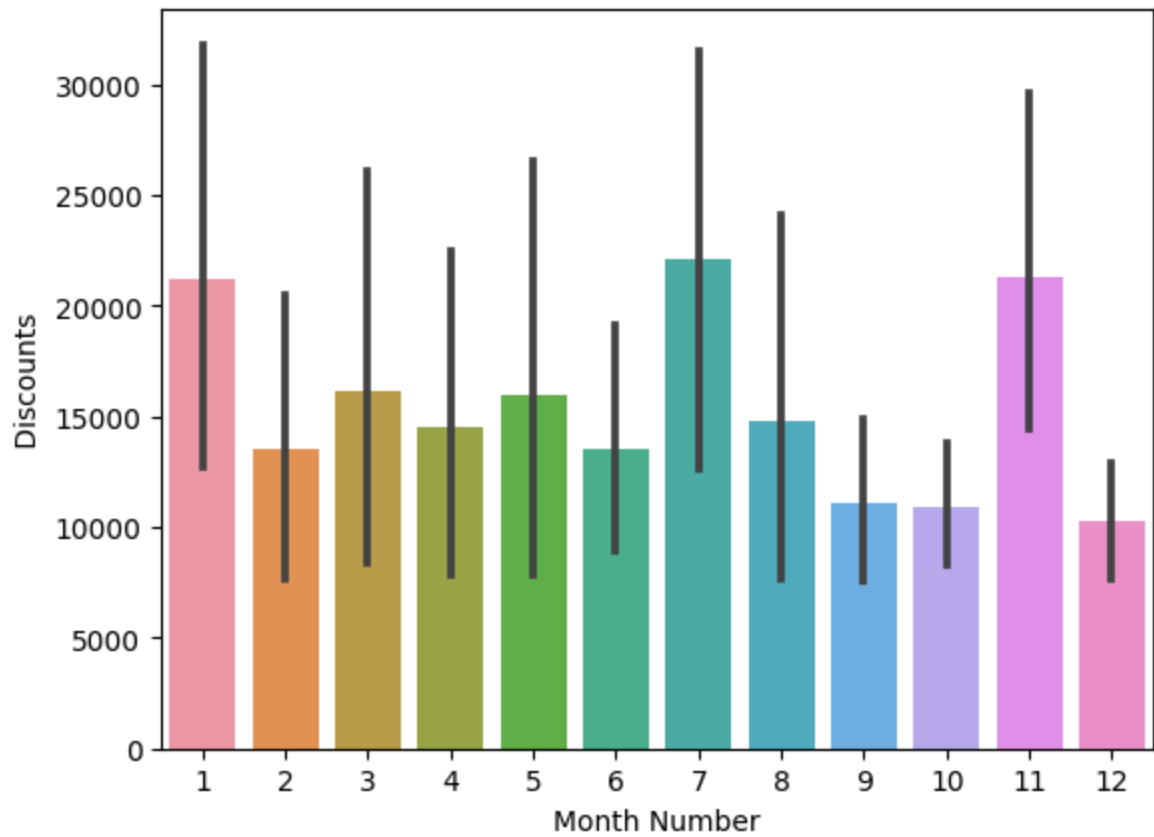
```
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='winter', linewidths=0.5)
plt.title('Correlation Matrix')
plt.show()
```



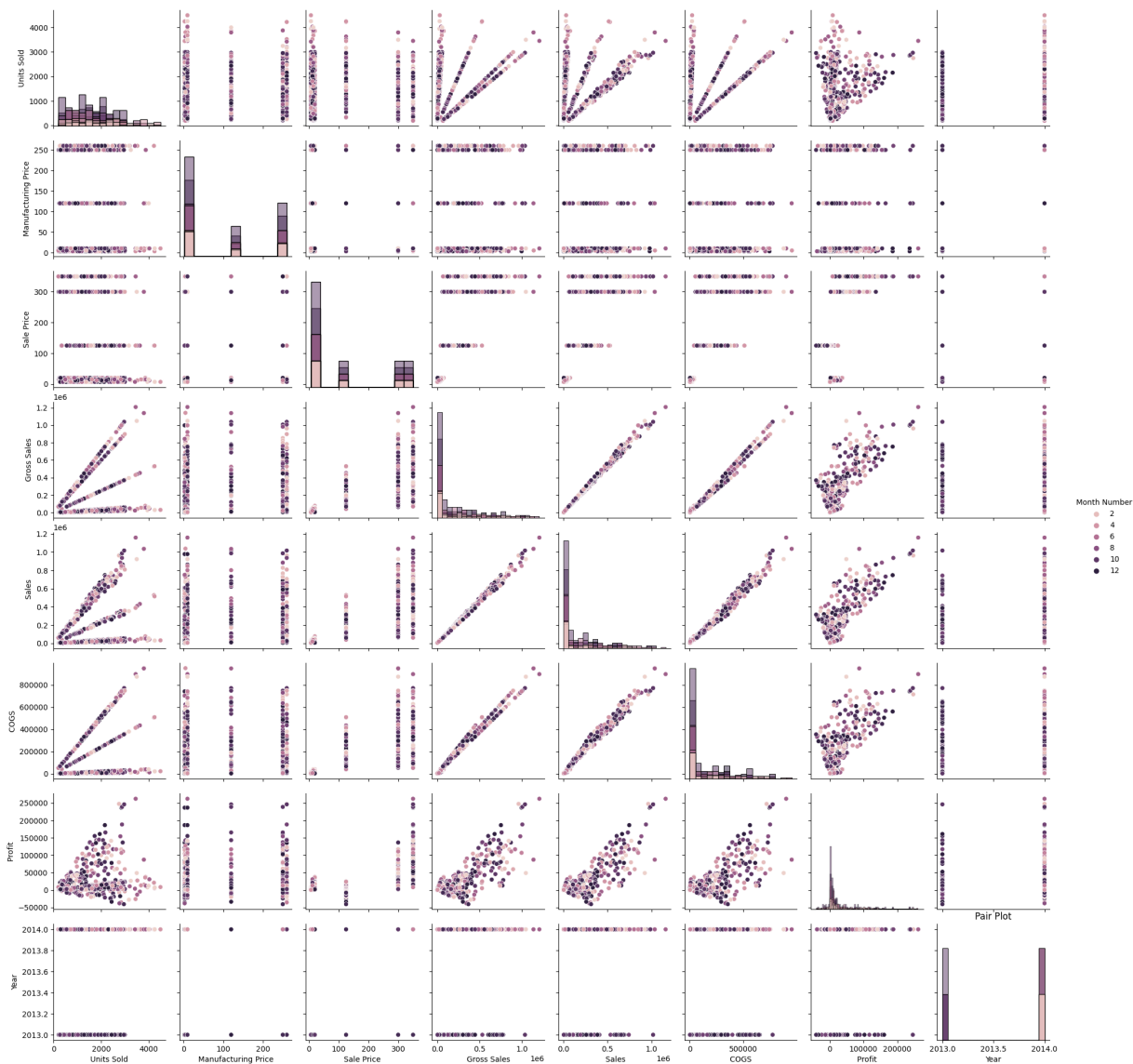


```
In [175]: sns.barplot(x='Month Number',y=' Discounts ',data=df)
```

```
Out[175]: <Axes: xlabel='Month Number', ylabel=' Discounts ' >
```

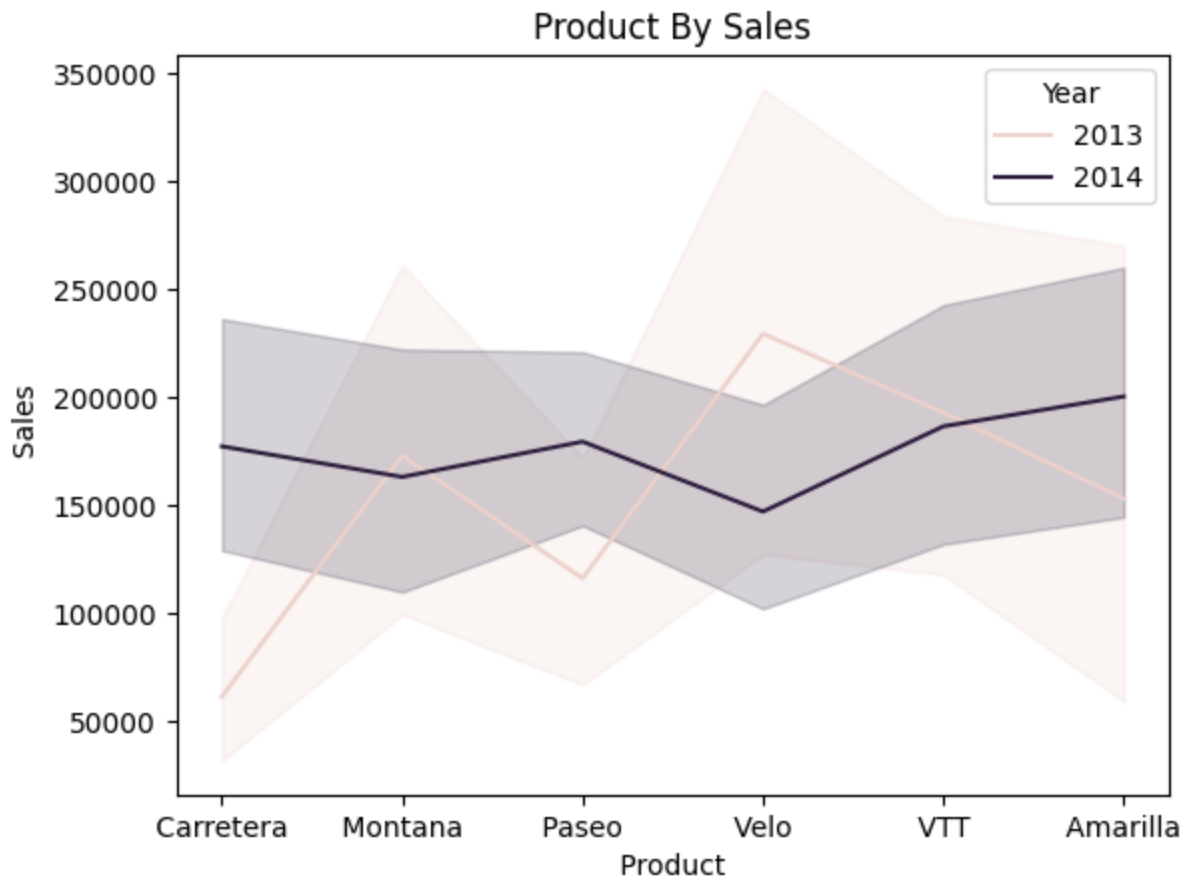


```
In [164... sns.pairplot(number_columns,hue='Month Number',diag_kind='hist')  
plt.title('Pair Plot')  
plt.show()
```



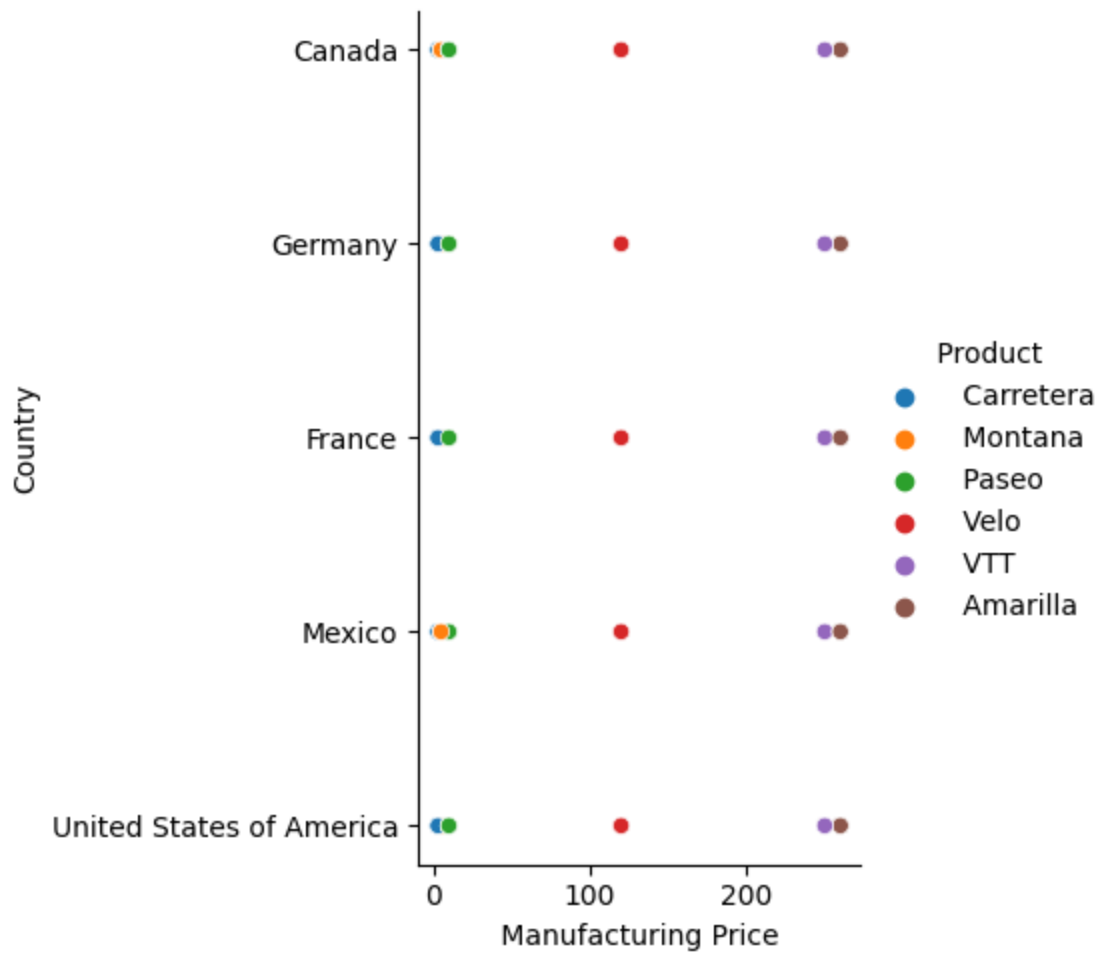
In [131...

```
sns.lineplot(x=' Product ',y=' Sales ',data=df,hue='Year')
plt.title("Product By Sales")
plt.show()
```



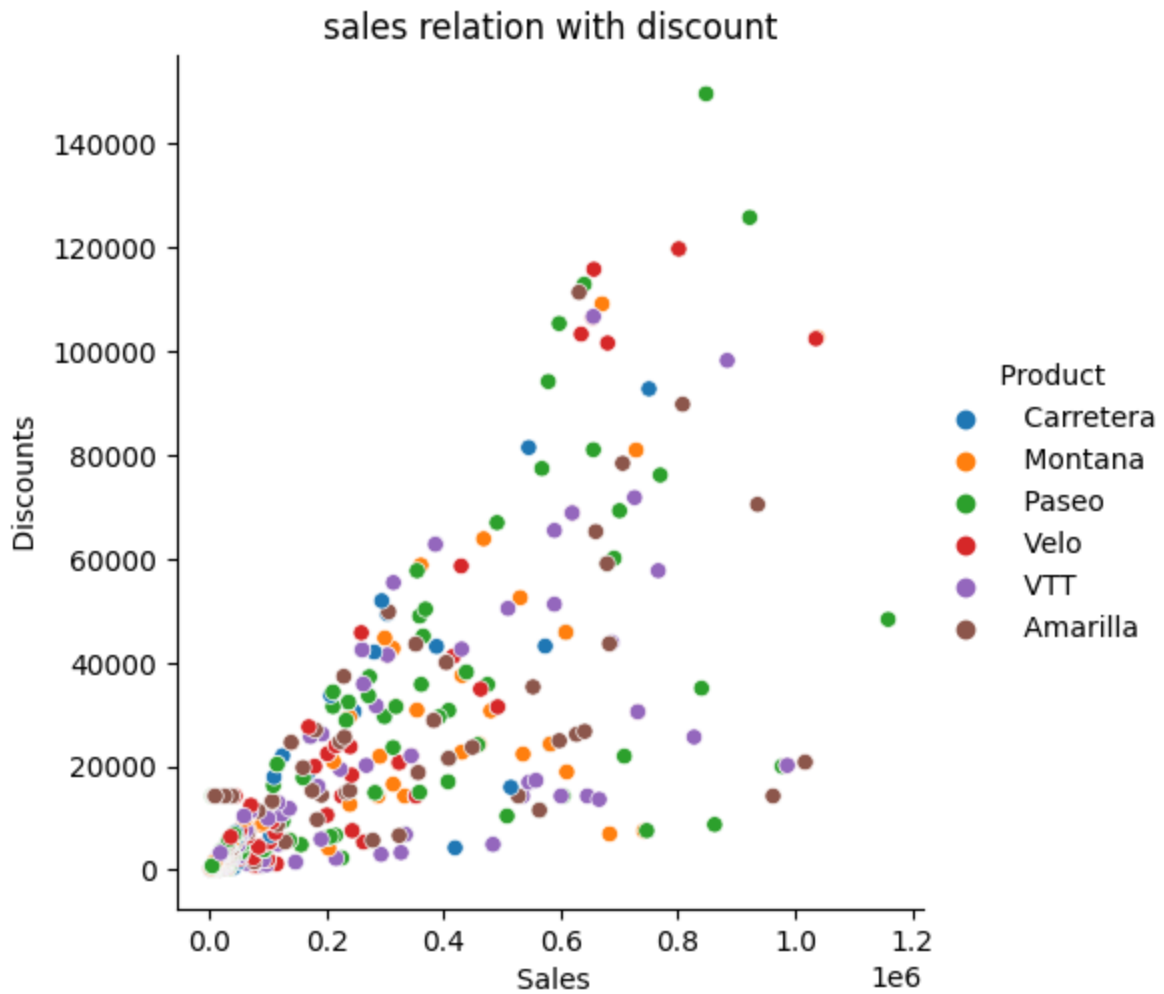
```
In [171...] sns.relplot(x=' Manufacturing Price ',y='Country',hue=' Product ',data=df)
```

```
Out[171]: <seaborn.axisgrid.FacetGrid at 0x1d709320390>
```

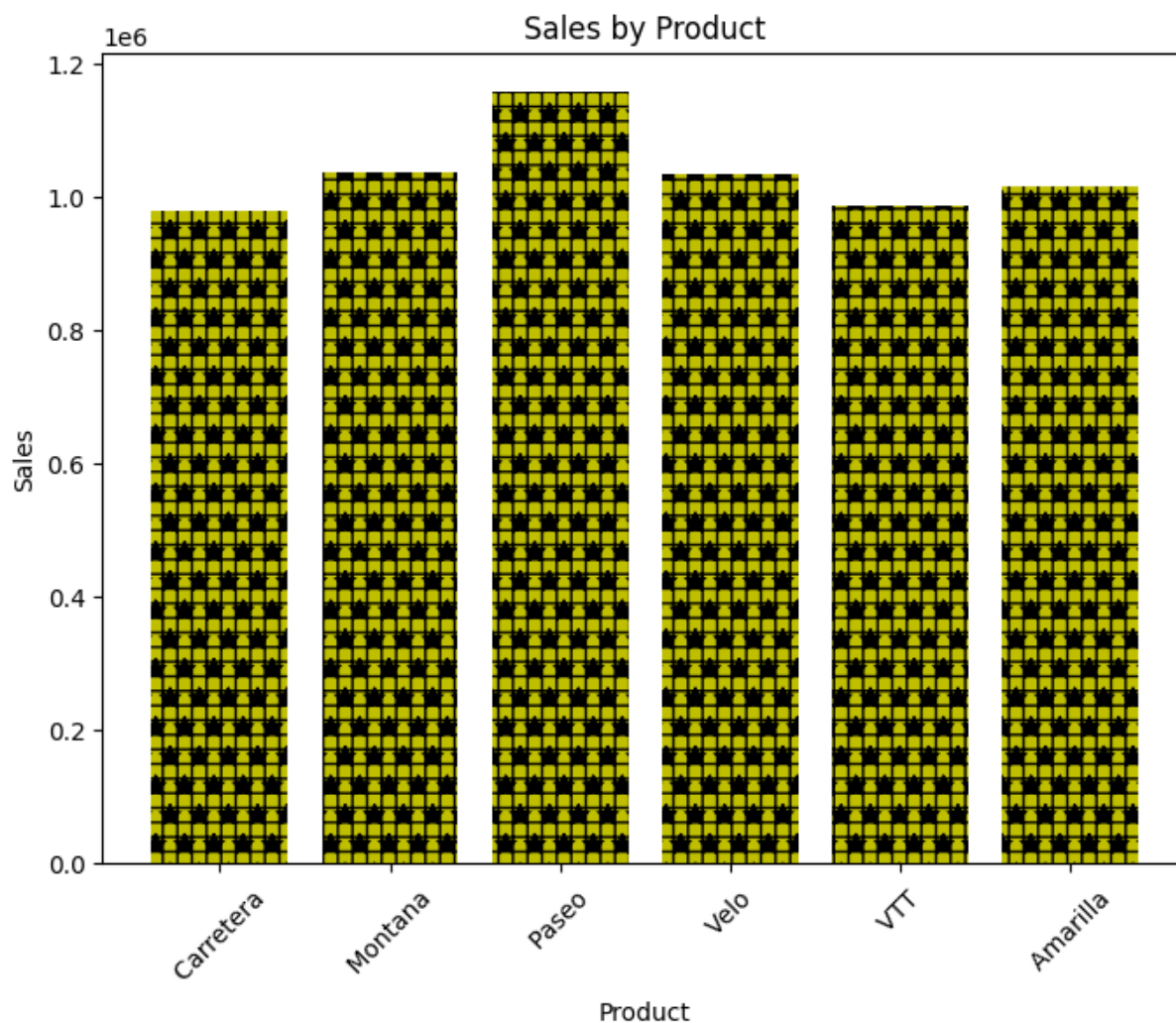


```
In [165... sns.relplot(x=' Sales ',y=' Discounts ',hue=' Product ',data=df)
plt.title('sales relation with discount')
```

```
Out[165]: Text(0.5, 1.0, 'sales relation with discount')
```

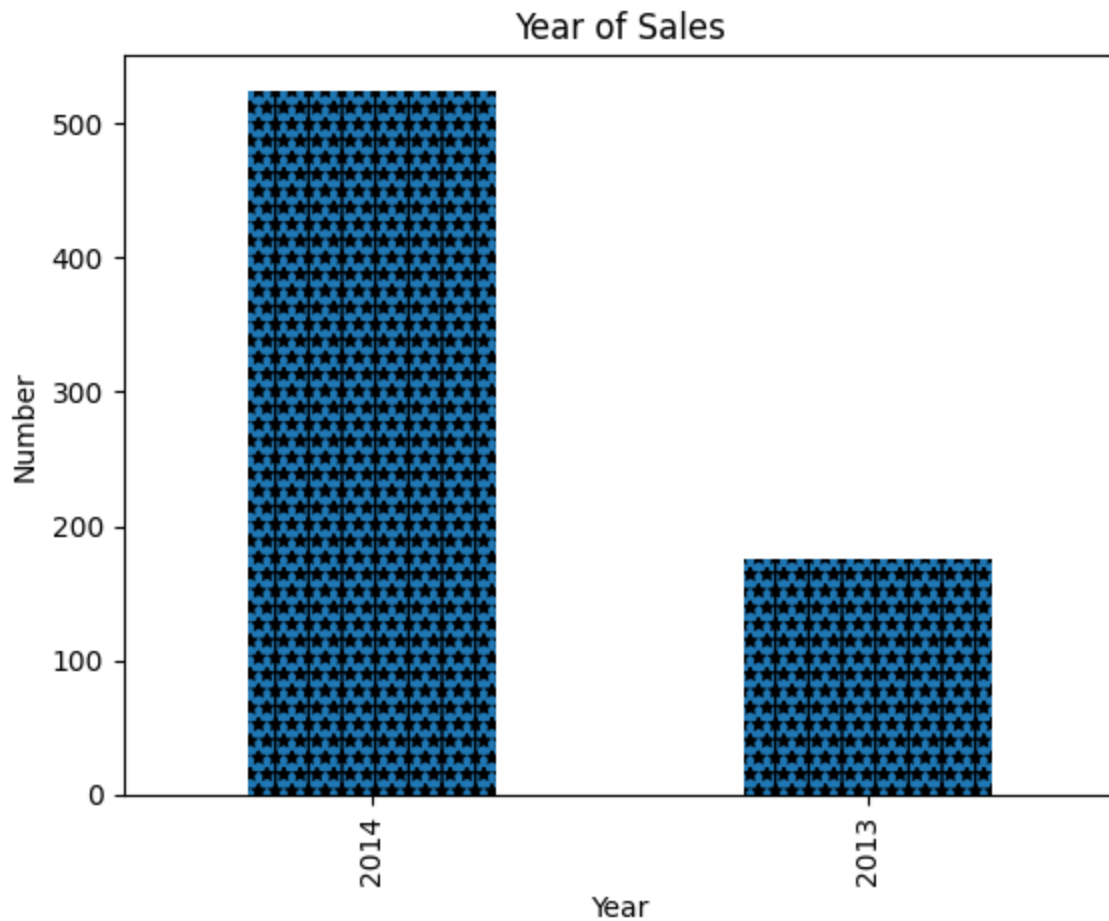


```
In [113... plt.figure(figsize=(8, 6))
plt.bar(' Product ', ' Sales ',data=df,fill='True',hatch='|-*|',color='y')
plt.xlabel('Product')
plt.ylabel('Sales')
plt.title('Sales by Product')
plt.xticks(rotation=45)
plt.show()
```



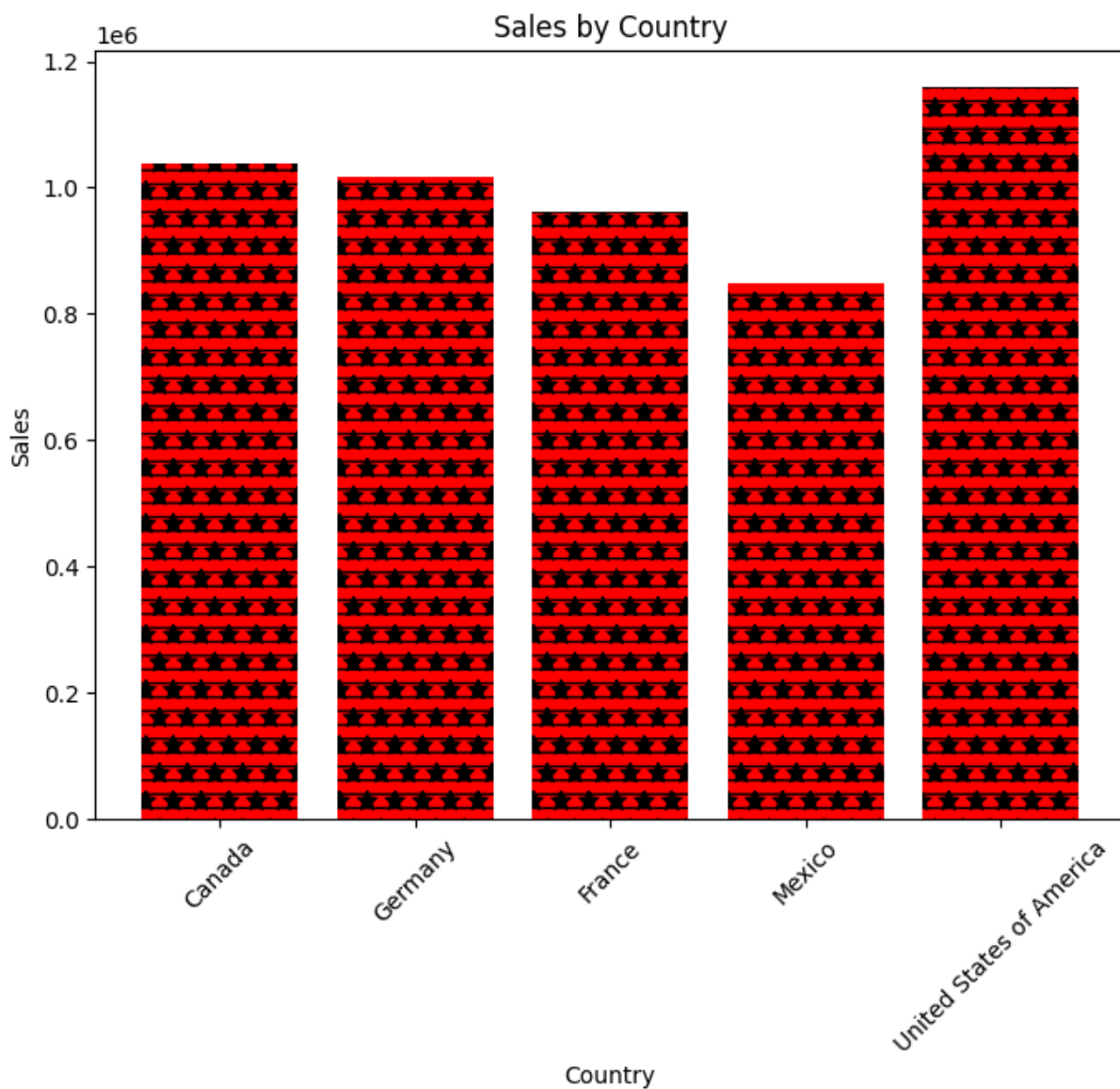
```
In [111...] ax = df['Year'].value_counts() \
              .plot(kind='bar', title = 'Year of Sales', fill='True', hatch="*|*")
ax.set_xlabel('Year')
ax.set_ylabel('Number')
```

```
Out[111]: Text(0, 0.5, 'Number')
```



```
In [112... # Plotting the bar chart of country over sales
plt.figure(figsize=(8, 6))
plt.bar('Country', ' Sales ',data=df,fill='True',hatch='-*-',color='r')
plt.xlabel('Country')
plt.ylabel('Sales')
plt.title('Sales by Country')
plt.xticks(rotation=45)
plt.show()
```





In [136... `df.columns`

Out[136]: Index(['Segment', 'Country', ' Product ', ' Discount Band ', 'Units Sold',  
 ' Manufacturing Price ', ' Sale Price ', ' Gross Sales ', ' Discounts ',  
 ' Sales ', ' COGS ', ' Profit ', 'Date', 'Month Number',  
 ' Month Name ', 'Year'],  
 dtype='object')

In [ ]:

In [ ]: