

# Data Visualization with Matplotlib - Exercises 2

จงทำตามค่าสั่งต่อไปนี้ด้วย data ที่กำหนดให้ต่อไปนี้

6521603795 Patiphat Aiamram

```
In [1]: import matplotlib.pyplot as plt  
import numpy as np  
import pandas as pd
```

อ่านไฟล์ Superstore.csv

```
In [241]: df = pd.read_csv('Superstore.csv', encoding = 'iso-8859-1')
```

```
In [3]: df.head()
```

	Order ID	Customer Name	Segment	Day	Month	Year	Ship Mode	City	State	Category
0	CA-2016-152156	Claire Gute	Consumer	8	11	2016	Second Class	Henderson	Kentucky	Furniture
1	CA-2016-152156	Claire Gute	Consumer	8	11	2016	Second Class	Henderson	Kentucky	Furniture
2	CA-2016-138688	Darrin Van Huff	Corporate	12	6	2016	Second Class	Los Angeles	California	Office Supplies
3	US-2015-108966	Sean O'Donnell	Consumer	11	10	2015	Standard Class	Fort Lauderdale	Florida	Furniture
4	US-2015-108966	Sean O'Donnell	Consumer	11	10	2015	Standard Class	Fort Lauderdale	Florida	Office Supplies

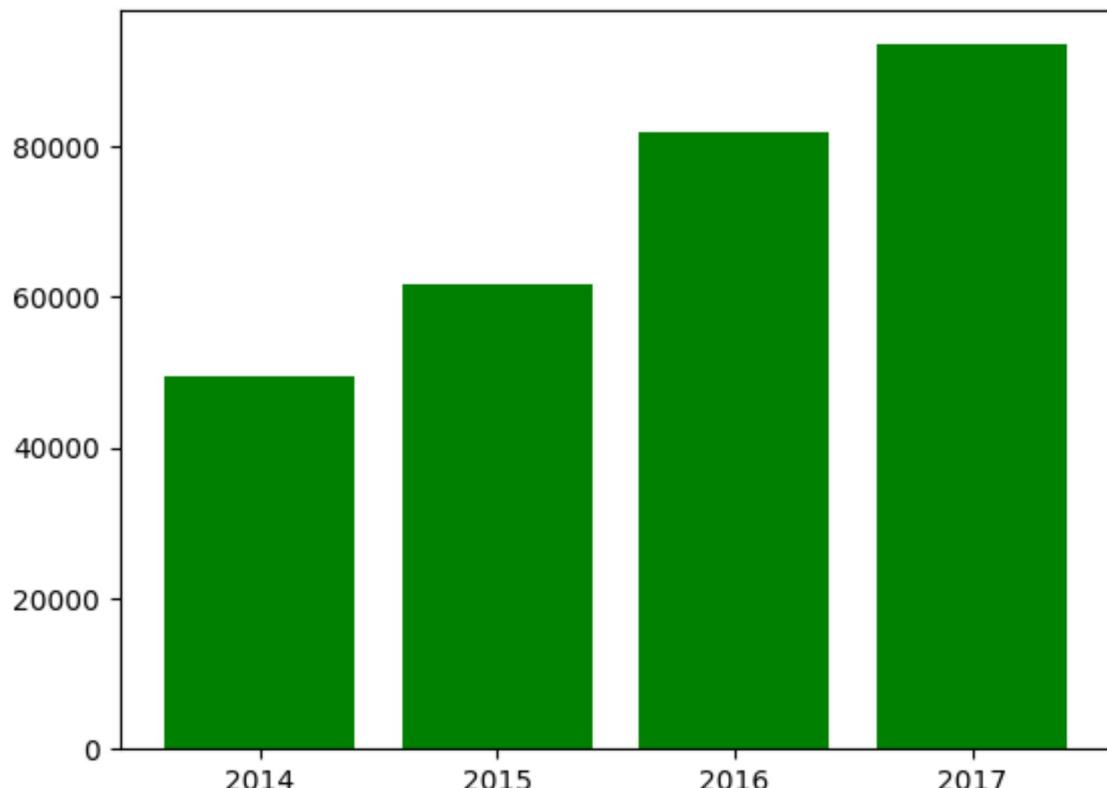
```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 16 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Order ID          9994 non-null   object  
 1   Customer Name     9994 non-null   object  
 2   Segment           9994 non-null   object  
 3   Day               9994 non-null   int64   
 4   Month              9994 non-null   int64   
 5   Year               9994 non-null   int64   
 6   Ship Mode         9994 non-null   object  
 7   City               9994 non-null   object  
 8   State              9994 non-null   object  
 9   Category           9994 non-null   object  
 10  Sub-Category      9994 non-null   object  
 11  Product Name      9994 non-null   object  
 12  Sales              9994 non-null   float64 
 13  Quantity           9994 non-null   int64   
 14  Discount            9994 non-null   float64 
 15  Profit              9994 non-null   float64 
dtypes: float64(3), int64(4), object(9)
memory usage: 1.2+ MB
```

## Exercise 1

จงวาดกราฟแท่งแสดงรายได้ของปี 2014 - 2017 และตกแต่งให้สวยงาม

```
In [60]: profit_year = df.groupby(['Year'])['Profit'].sum()
x = profit_year.index
y = profit_year
plt.bar(x, y, color = 'g' )
plt.xticks([2014, 2015, 2016, 2017])
plt.show()
```



## Exercise 2

จงวัดกราฟวงกลม และงบเปอร์เซ็นต์การขนส่งแต่ละแบบ ( Ship Mode ) พร้อมตกลงให้สวยงาม

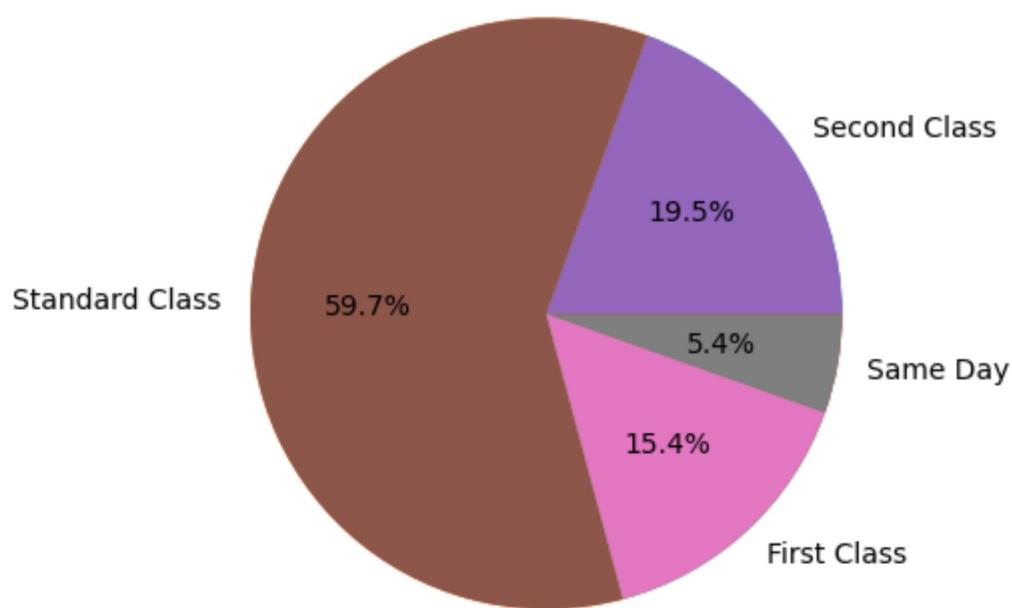
In [47]:

Out[47]: np.int64(1945)

In [70]:

```
a = df[df['Ship Mode'] == 'Second Class']['Ship Mode'].count()
b = df[df['Ship Mode'] == 'Standard Class']['Ship Mode'].count()
c = df[df['Ship Mode'] == 'First Class']['Ship Mode'].count()
d = df[df['Ship Mode'] == 'Same Day']['Ship Mode'].count()
data = [a, b, c, d]
lb = ['Second Class', 'Standard Class', 'First Class', 'Same Day']
plt.title('Ship Mode Pie chart')
plt.pie(data, labels=lb, autopct='%.1f%%')
plt.show()
```

Ship Mode Pie chart



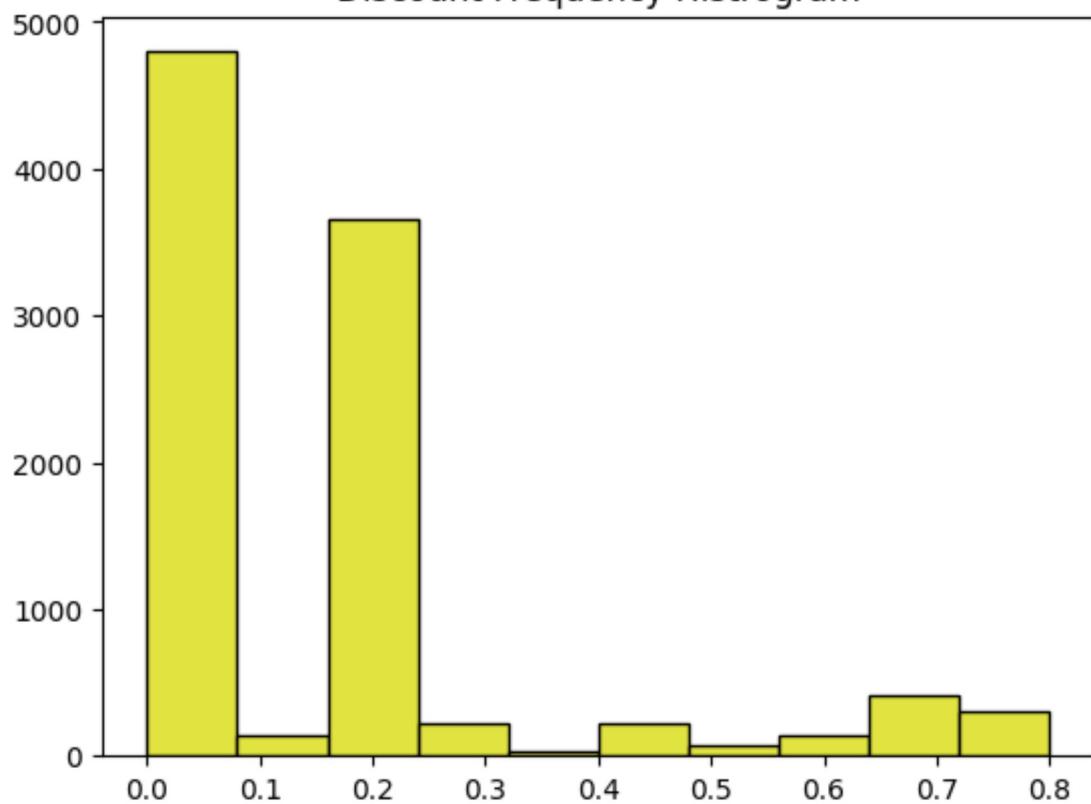
## Exercise 3

จงวัดกราฟความถี่นับจำนวนส่วนลด (Discount) พร้อมตกลงให้สวยงาม

In [86]:

```
plt.title('Discount Frequency Histogram')
plt.hist(df['Discount'], ec='k', color='#e0e340')
plt.show()
```

### Discount Frequency Histogram

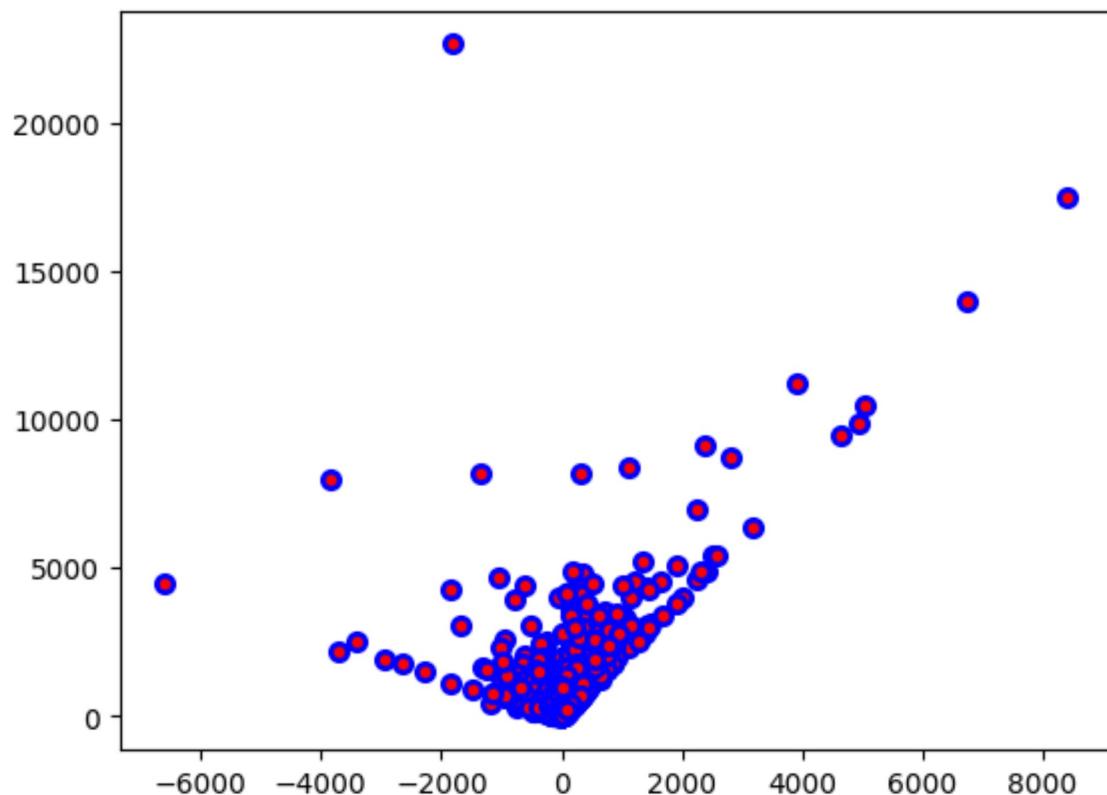


## Exercise 4

จงวาดกราฟจุด(Scatter) แสดงราคาขายกับกำไรที่ได้ (Sales , Profit) พร้อมตกแต่งให้สวยงาม

```
In [96]: x = df['Profit']
y = df['Sales']
plt.scatter(x , y, marker = 'o', c ='r', ec = 'b', lw = 2)
plt.title('Scatter between Sales and Profit')
plt.show()
```

### Scatter between Sales and Profit



In [ ]:

## Exercise 5

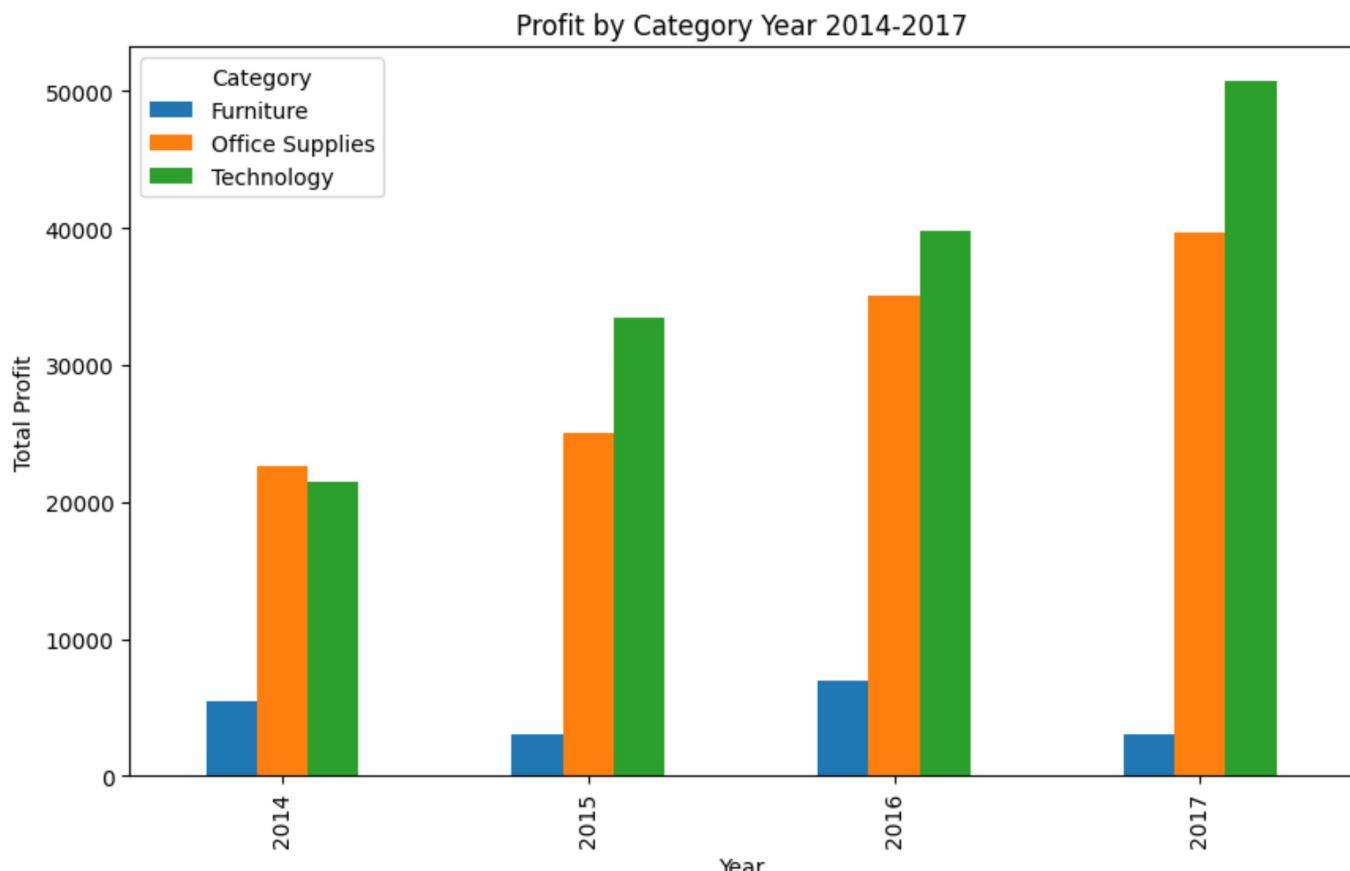
จงวาดกราฟแท่งแสดงรายได้ของปี 2014 - 2017 ของแต่ละหมวดหมู่ในกราฟเดียว(แกน x เป็นปี) พร้อมตกแต่งให้สวยงาม

In [247]:

```
profit_by_category_year = df.groupby(['Year', 'Category'])['Profit'].sum().unstack()
profit_by_category_year.plot(kind='bar', figsize=(10, 6))

plt.title('Profit by Category Year 2014-2017')
plt.xlabel('Year')
plt.ylabel('Total Profit')

plt.legend(title='Category')
plt.show()
```



In [ ]:

## Exercise 6

จงวาดกราฟแท่งแสดงรายได้ของแต่ละหมวด ตั้งแต่ปี 2014-2017 แบบแยกกราฟ (1หมวด 1กราฟ) พร้อมตกแต่งให้สวยงาม

In [208]:

```
df[df['Category'] == 'Furniture']['Profit'].sum()
df[df['Category'] == 'Office Supplies']['Profit'].sum()
df[df['Category'] == 'Technology']['Profit'].sum()
```

Out[208]:

```
np.float64(145454.9481)
```

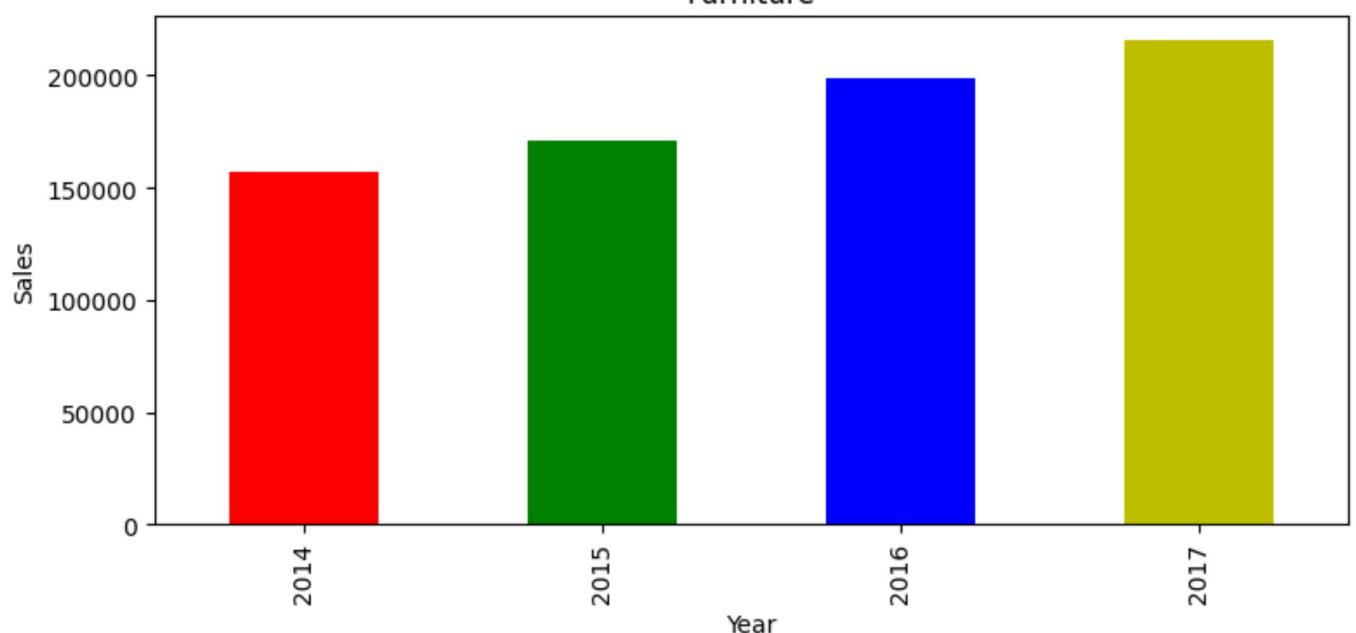
In [275]:

```
category_sales = df.groupby(["Category", "Year"])["Sales"].sum().unstack()

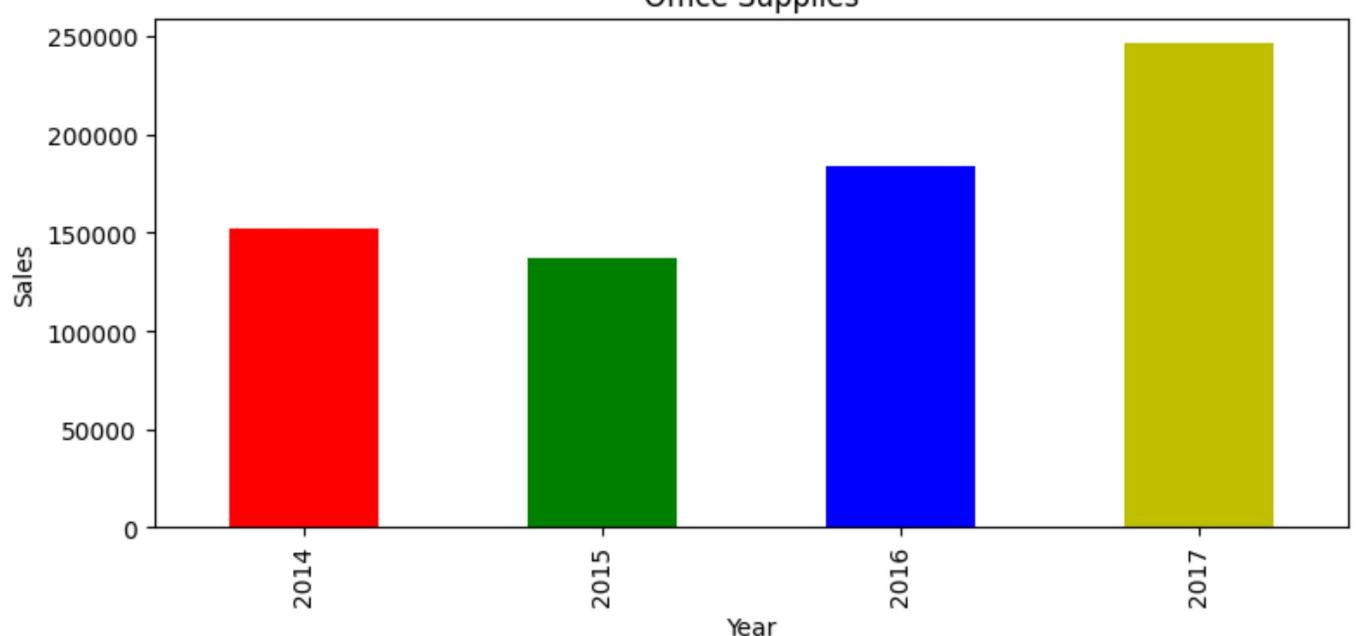
fig, axes = plt.subplots(nrows=3, figsize=(8, 12))
```

```
for ax, (category, sales) in zip(axes, category_sales.iterrows()):  
    sales.plot(kind="bar", ax=ax, color=['r','g','b','y'] )  
    ax.set_title(f"{category}")  
    ax.set_ylabel("Sales")  
    ax.set_xlabel("Year")  
  
plt.tight_layout()  
plt.show()
```

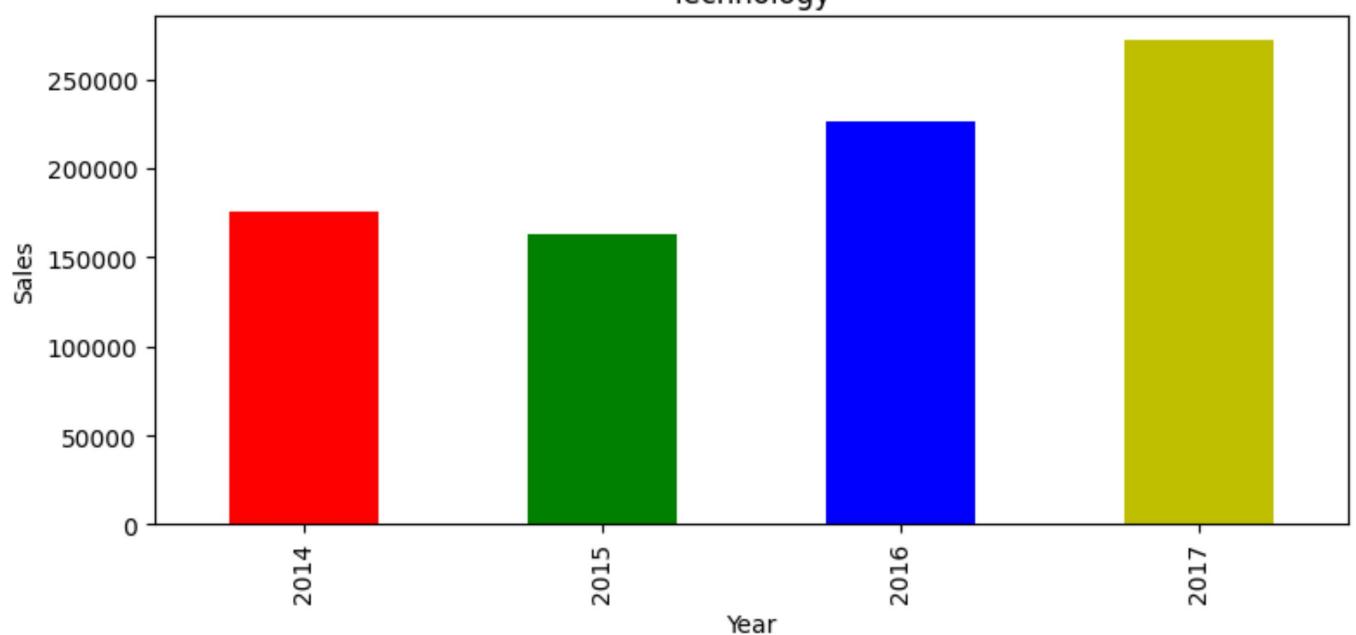
### Furniture



### Office Supplies



### Technology



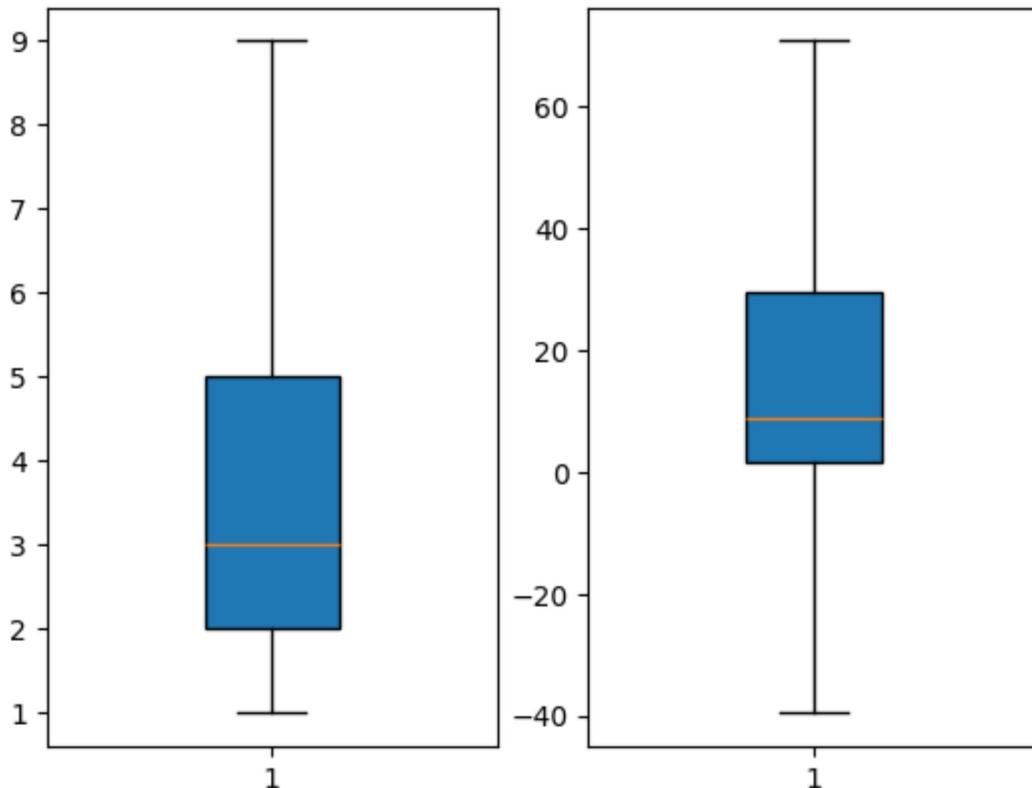
In [ ]:

จงวาด Box plot ของยอดขาย และกำไร พร้อมตกลงให้สวยงาม (แนะนำให้แยกกราฟ)

In [276...]: df['Quantity']

Out[276...]: np.int64(37873)

In [272...]:  
x = df['Quantity']  
y = df['Profit']  
data = [x,y]  
fig, axes = plt.subplots(ncols=2)  
axes[0].boxplot(x, sym='', vert=True, widths=0.3, label='Quantity', patch\_artist=True)  
axes[1].boxplot(y, sym='', vert=True, widths=0.3, label='Profit', patch\_artist=True)  
plt.show()



In [ ]:

## Exercise 8

จงดูกราฟแท่งแสดงจำนวน Furniture ที่ขายได้ในแต่ละไตรมาส (จำนวนให้นับจาก Quantity)

In [264...]:dffur = df[df['Category'] == 'Furniture']  
dffur

Out[264...]

	Order ID	Customer Name	Segment	Day	Month	Year	Ship Mode	City	State	C
0	CA-2016-152156	Claire Gute	Consumer	8	11	2016	Second Class	Henderson	Kentucky	F
1	CA-2016-152156	Claire Gute	Consumer	8	11	2016	Second Class	Henderson	Kentucky	F
3	US-2015-108966	Sean O'Donnell	Consumer	11	10	2015	Standard Class	Fort Lauderdale	Florida	F
5	CA-2014-115812	Brosina Hoffman	Consumer	9	6	2014	Standard Class	Los Angeles	California	F
10	CA-2014-115812	Brosina Hoffman	Consumer	9	6	2014	Standard Class	Los Angeles	California	F
...	...	...	...	...	...	...	...	...	...	...
9962	CA-2015-168088	Corinna Mitchell	Home Office	19	3	2015	First Class	Houston	Texas	F
9964	CA-2016-146374	Harold Engle	Corporate	5	12	2016	Second Class	Newark	Delaware	F
9980	US-2015-151435	Shaun Weien	Consumer	6	9	2015	Second Class	Lafayette	Louisiana	F
9989	CA-2014-110422	Tom Boeckenhauer	Consumer	21	1	2014	Second Class	Miami	Florida	F
9990	CA-2017-121258	Dave Brooks	Consumer	26	2	2017	Standard Class	Costa Mesa	California	F

2121 rows × 17 columns

In [265...]

```
def set_quarter(a):
    if a in range(1 ,4):
        return 'Quarter 1'
    elif a in range(4, 7):
        return 'Quarter 2'
```

```
    elif a in range(7, 10):
        return 'Quarter 3'
    elif a in range(10, 13):
        return 'Quarter 4'

dffur['Quarter'] = dffur['Month'].apply(set_quarter)
dffur
```

C:\Users\francis\AppData\Local\Temp\ipykernel\_13928\748045746.py:11: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
dffur['Quarter'] = dffur['Month'].apply(set\_quarter)

Out[265...]

	Order ID	Customer Name	Segment	Day	Month	Year	Ship Mode	City	State	C
0	CA-2016-152156	Claire Gute	Consumer	8	11	2016	Second Class	Henderson	Kentucky	F
1	CA-2016-152156	Claire Gute	Consumer	8	11	2016	Second Class	Henderson	Kentucky	F
3	US-2015-108966	Sean O'Donnell	Consumer	11	10	2015	Standard Class	Fort Lauderdale	Florida	F
5	CA-2014-115812	Brosina Hoffman	Consumer	9	6	2014	Standard Class	Los Angeles	California	F
10	CA-2014-115812	Brosina Hoffman	Consumer	9	6	2014	Standard Class	Los Angeles	California	F
...	...	...	...	...	...	...	...	...	...	...
9962	CA-2015-168088	Corinna Mitchell	Home Office	19	3	2015	First Class	Houston	Texas	F
9964	CA-2016-146374	Harold Engle	Corporate	5	12	2016	Second Class	Newark	Delaware	F
9980	US-2015-151435	Shaun Weien	Consumer	6	9	2015	Second Class	Lafayette	Louisiana	F
9989	CA-2014-110422	Tom Boeckenhauer	Consumer	21	1	2014	Second Class	Miami	Florida	F
9990	CA-2017-121258	Dave Brooks	Consumer	26	2	2017	Standard Class	Costa Mesa	California	F

2121 rows × 17 columns

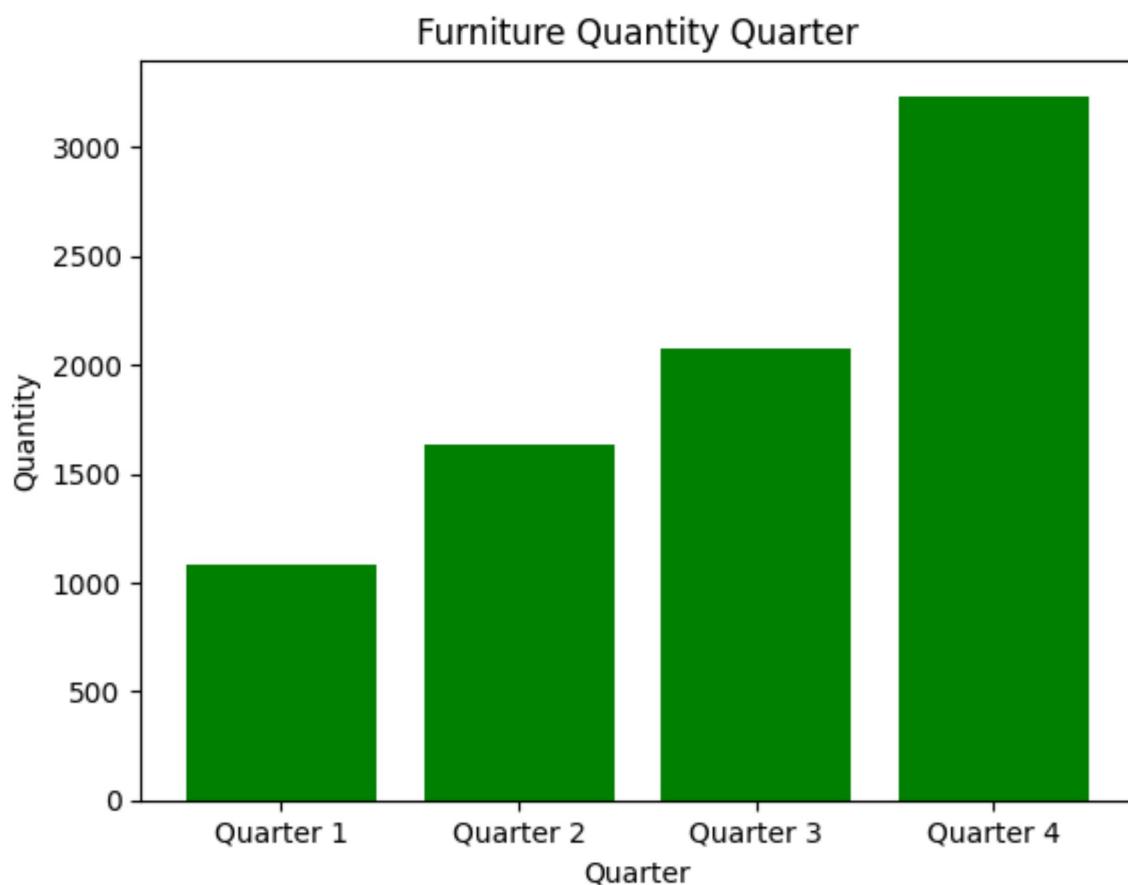
In [270...]

```

data = dffur.groupby(['Quarter'])['Quantity'].sum()
x = data.index
y = data
plt.xticks([0,1,2,3])
plt.xlabel('Quarter')

```

```
plt.ylabel('Quantity')
plt.title('Furniture Quantity Quarter')
plt.bar(x,y, color='g')
plt.show()
```



In [ ]: