SUPERMARKET SALES ANALYSIS - 2019
(JANUARY - MARCH)

INDEX

INTRODUCTION

As we all know a large retail market that sells food and other household goods and that is usually operated on a self-service basis is called supermarket. The growth of supermarkets in most populated cities are increasing and market competitions are also high.

This dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data. This supermarket wants to know how to improve their business and what are steps they need to carry out to run a successful business.

To find answer for all questions, data analysis and visualization is going to be done by using the effective libraries given by python.

CODE AND RESOURCES

Importing Libraries and Loading Data into Pandas

- We are going to import all the libraries required to analyze and visualize
- Load data using pandas

```
In [2]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns

        import warnings
        warnings.filterwarnings('ignore')
        %matplotlib inline
```

```
In [3]: data = pd.read_excel('supermarket.xls')  # Load dataset
```

```
In [4]: data
```

Out[4]:

	invoiceID	branch	city	cust_type	gender	type	unit_price	quantity	date	time	payment	cost	gross income	rating
0	101-17-6199	A	Cairo	Normal	Male	Food and beverages	45.79	7	2019-03-13	19:44:00	Credit card	320.53	16.0265	7.0
1	101-81-4070	C	Alexandria	Member	Female	Health and beauty	62.82	2	2019-01-17	12:36:00	eWallet	125.64	6.2820	4.9
2	102-06-2002	C	Alexandria	Member	Male	Sports and travel	25.25	5	2019-03-20	17:52:00	Cash	126.25	6.3125	6.1
3	102-77-2261	C	Alexandria	Member	Male	Health and beauty	65.31	7	2019-03-05	18:02:00	Credit card	457.17	22.8585	4.2
4	105-10-6182	A	Cairo	Member	Male	Fashion accessories	21.48	2	2019-02-27	12:22:00	eWallet	42.96	2.1480	6.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	894-41-5205	C	Alexandria	Normal	Female	Food and beverages	43.18	8	2019-01-19	19:39:00	Credit card	345.44	17.2720	8.3
996	895-03-6665	B	Ismailia	Normal	Female	Fashion accessories	36.51	9	2019-02-16	10:52:00	Cash	328.59	16.4295	4.2
997	895-66-0685	B	Ismailia	Member	Male	Food and beverages	18.08	3	2019-03-05	19:46:00	eWallet	54.24	2.7120	8.0
998	896-34-0956	A	Cairo	Normal	Male	Fashion accessories	21.32	1	2019-01-26	12:43:00	Cash	21.32	1.0660	5.9
999	898-04-2717	A	Cairo	Normal	Male	Fashion accessories	76.40	9	2019-03-19	15:49:00	eWallet	687.60	34.3800	7.5

1000 rows × 14 columns

Supermarket branch names

```
In [6]: data['branch'].value_counts().index.tolist()
```
```
Out[6]: ['A', 'B', 'C']
```

Location of branches

```
In [7]: data['city'].value_counts().index.tolist()

Out[7]: ['Cairo', 'Ismailia', 'Alexandria']
```

Available payment methods

```
In [8]: data['payment'].value_counts().index.tolist()

Out[8]: ['eWallet', 'Cash', 'Credit card']
```

List of products:

```
In [10]: data['type'].unique()

Out[10]: array(['Food and beverages', 'Health and beauty', 'Sports and travel',
               'Fashion accessories', 'Home and lifestyle',
               'Electronic accessories'], dtype=object)
```

- This supermarket has three branches in three different cities and all branches supports 'eWallet', 'Cash', 'Credit card' payment methods.
- They are selling products such as 'Fashion accessories', 'Food and beverages', 'Electronic accessories', 'Sports and travel', 'Home and lifestyle' and 'Health and beauty'.

Profit based on products

```
In [19]: d2=data.copy()
         d2=d2.groupby(['type'])['gross_income'].sum()
         d2=pd.DataFrame(d2)
         d2
```

Out[19]:

type	gross_income
Electronic accessories	2587.5015
Fashion accessories	2585.9950
Food and beverages	2673.5640
Health and beauty	2342.5590
Home and lifestyle	2564.8530
Sports and travel	2624.8965

Profit based on cities

```
In [14]: d1=data.copy()
         d1=d1.pivot_table('gross income',columns='city',aggfunc='sum')
         d1
```

Out[14]:

city	Alexandria	Cairo	Ismailia
gross income	5265.1765	5057.1605	5057.032

Sales based on products

```
In [20]: d2=data.copy()
         d2=d2.groupby(['type'])['cost'].sum()
         d2=pd.DataFrame(d2)
         d2
```

Out[20]:

type	cost
Electronic accessories	51750.03
Fashion accessories	51719.90
Food and beverages	53471.28
Health and beauty	46851.18
Home and lifestyle	51297.06
Sports and travel	52497.93

Products sold based on city

```
In [22]: pd.crosstab(data['city'],data['type'])
```

Out[22]:

| type / city	Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
Alexandria	55	65	66	52	45	45
Cairo	60	51	58	47	65	59
Ismailia	55	62	50	53	50	62

VIZUALIZATION

Products sold in branch wise

```python
# Calculating for Branch A
plt.title('Products Sold in branch A',fontsize=20)
a=data['branch']=='A'
b=data[a]
b['type'].value_counts().plot.pie(autopct='%1.1f%%' ,shadow=True)
plt.show()

# Calculating for Branch B
plt.title('Products Sold in branch B',fontsize=20)
c=data['branch']=='B'
d=data[c]
d['type'].value_counts().plot.pie(autopct='%1.1f%%' ,shadow=True)
plt.show()

# Calculating for Branch C
plt.title('Products Sold in branch C',fontsize=20)
e=data['branch']=='C'
f=data[e]
f['type'].value_counts().plot.pie(autopct='%1.1f%%' ,shadow=True)
plt.show()
```
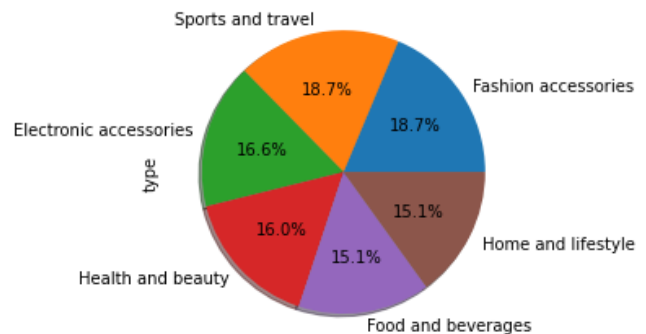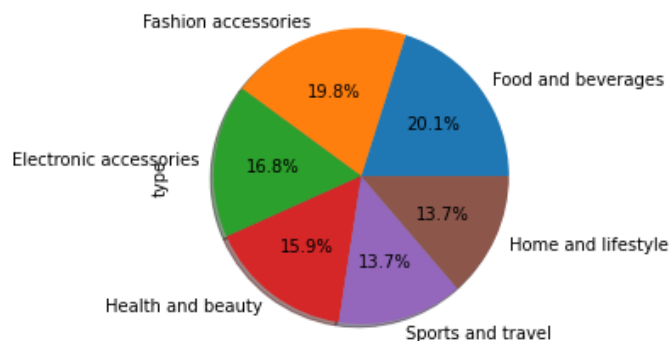

Products Sold in branch A
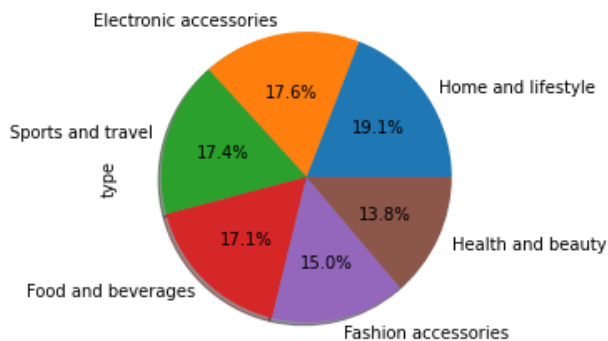
Products Sold in branch B

Products Sold in branch C

- According to above charts highest products sold in each branch are:
  - Branch "A" – Home and Lifestyle
  - Branch "B" – Fashion Accessories, Sports and Travel
  - Branch "C" – Fashion Accessories, Food and Beverages

Preferred Payment method

```python
#counting ewallet payments
ewallet= data['payment'] == "eWallet"
w=data[ewallet].shape[0]

#counting cash payments
Cash= data['payment'] == "Cash"
c=data[Cash].shape[0]

#counting credit card payments
card= data['payment'] == "Credit card"
cc=data[card].shape[0]

def addlabels(x,y):
    for i in range(len(x)):
        plt.text(i, y[i]//2,y[i], ha = 'center',
                 Bbox = dict(facecolor = 'white', alpha = .5))
# Horizontal Bar Plot
colors = ['green','blue','gold']
plt.bar(['eWallet','Cash','Credit card'],[w,c,cc],
        label ='payment',color=colors,width=0.2)
plt.title(label="Payment Methods",loc="center",
          fontstyle='italic',size=20)
plt.xlabel("Payment Type")
plt.ylabel("No of customers used",size=14)
addlabels(['eWallet','Cash','Credit card'],[w,c,cc])

# Show Plot
plt.show()
```
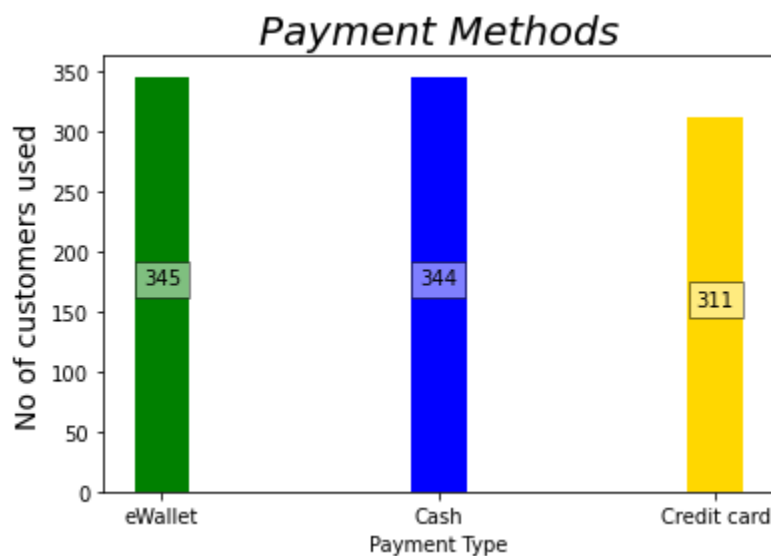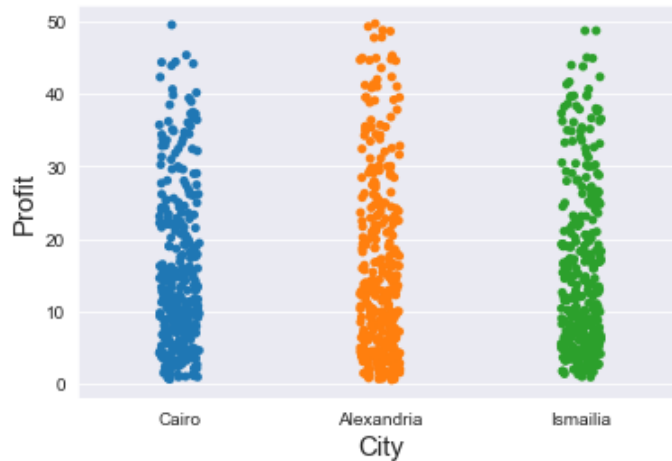

- Customers are preferring eWallet and Cash methods to pay the bills.

Profit based on City

```python
sns.set_style("darkgrid")
sns.stripplot(x="city", y="gross income", data=data)
plt.xlabel ('City',fontsize=15)
plt.ylabel('Profit', fontsize=15)
plt.show()
```



```python
plt.figure(figsize=(10,6))
sns.set_style("darkgrid")
sns.stripplot(x="city", y="gross income", hue="cust_type",
              jitter=0.30,dodge=True,data=data)
plt.xlabel ('City',fontsize=15)
plt.ylabel('Profit', fontsize=15)
plt.show()
```
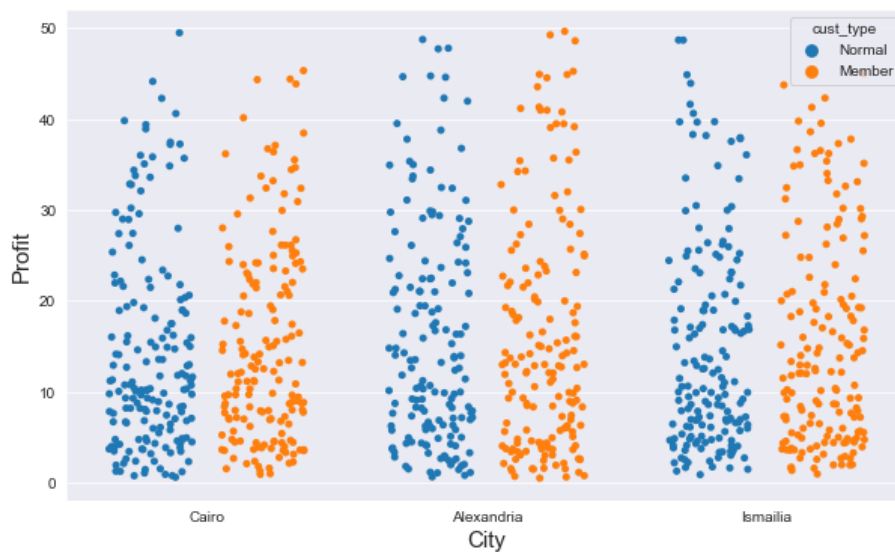
- The supermarket in "Alexandria" is giving highest profit and also the amount of members is larger than non-members.

Customers visited supermarket

```
: #counting Female Customers
  female= data['gender'] == "Female"
  f=data[female].shape[0]

  #counting Female Customers
  male= data['gender'] == "Male"
  m=data[male].shape[0]

  labels=['Female','Male']
  colors = ['lightskyblue', 'green']

  # displaying the title
  plt.title(label="Customers",loc="left",fontstyle='italic')

  plt.pie([f,m], colors=colors, labels=labels,startangle=50, autopct='%.1f%%')
  plt.show()
```
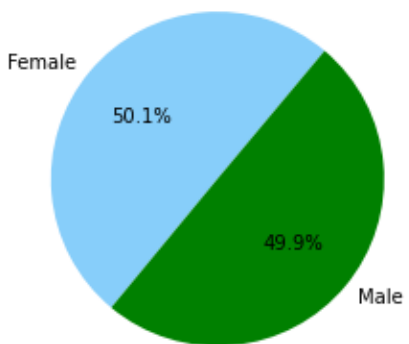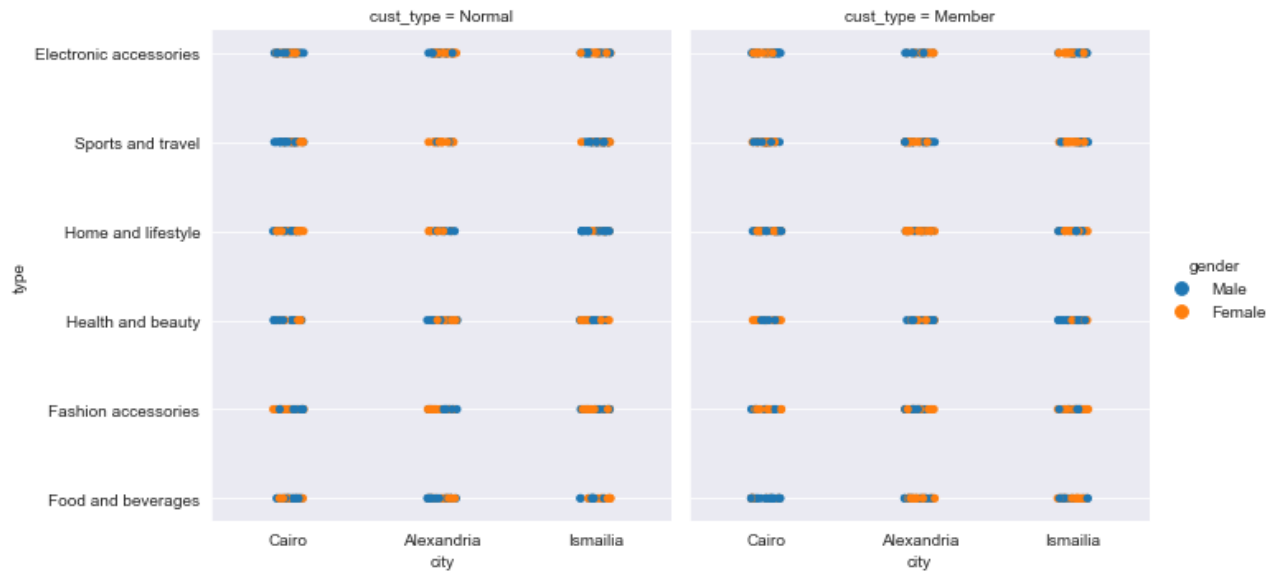
Customers

- Female consumers are greater than male.

Products sold in each city based on customers

```
sns.catplot(x='city', y='type', col='cust_type', hue='gender', data=data)
plt.show()
```
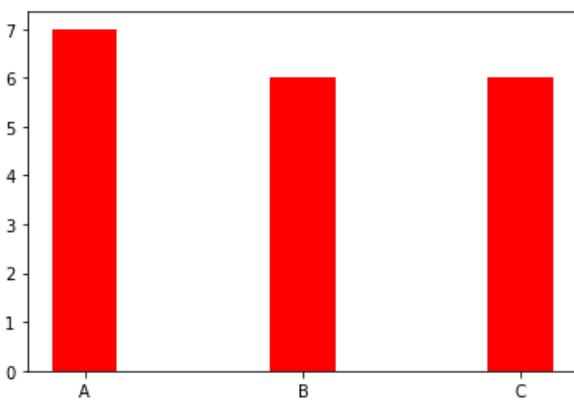
Rating of branches

```python
# Calculating for Branch A
a=data['branch']=='A'
b=data[a]
A=int(b['rating'].mean())

# Calculating for Branch B
c=data['branch']=='B'
d=data[c]
B=int(d['rating'].mean())

# Calculating for Branch C
e=data['branch']=='C'
f=data[c]
C=int(f['rating'].mean())

branch=["A","B","C"]
rating=[A,B,C]

plt.bar(branch, rating, color="red",width=0.3)
plt.show()
```

SUMMARY

The Analysis of the supermarket sales data depicts that among the 3 branches the branch "A" which is located at city "Cairo" is providing good service compared to remaining branches. The supermarket is getting highest profit in a branch which is located in Alexandria and the products that helping to achieve this profit is "Food and beverages" and "Fashion Accessories".

 Home and Lifestyle products are selling more in branch "A" and Fashion Accessories, Sports and Travel products are selling more in Branch "B".

Most of the customers using eWallet to pay their bills. The number of member and non-member counts are equal. The amount of female buying products is 0.1% larger than male.

Credit card users are less compare to cash payment customers.

Conclusion:

- The services providing by branch B and C can be increased to increase the sales.

- We can take survey in branch B and C to provide better service

- Most of the consumers are female, so we can increase existing products and introduce new products which are related to female customers.

- Giving attractive discounts to members will turn the normal consumers into members.

REFERENCE

- **www.kaggle.com**