```
In [1]: import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib
```

載入資料

```
In [2]: train_df=pd.read_csv("dataset/商品對照表.__商品對照表.csv")
#train_df=pd.read_csv("dataset/家樂福數據集.csv")
# Create a dictionary from 'Name' and 'Age' columns
pid_to_name_dict = train_df.set_index('product')['product_name'].to_dict()
name_to_pid_dict = {v:k for k,v in pid_to_name_dict.items()}

train_df.head(5)
```

```
Out[2]:
                                                                                            department
                        id
                           order_date product_name
                                                            product sales_price quantity
                4dd5ac40-
                                                          003dc937-
                                                                                               22f2e7c7-
                                         絲花極柔化妝
                                                                                              6679-41fd-
                a1b3-43a2-
                            2021/6/10
                                                         e898-4259-
                                                                                                            1
          0
                                                                             96
                                                                                       2
                    be16-
                                 22:57
                                              棉66片
                                                              870a-
                                                                                                  9c1d-
             aa1392a4fde0
                                                       4a9afe2eacd6
                                                                                          6e9562b51bce
                                                                                                           19
                181be294-
                                                          004e97ea-
                                                                                               f57143f5-
                                         光泉茉莉茶園
               2190-47db-
                            2021/5/27
                                                         3847-4b71-
                                                                                              672f-4c78-
                                                                                                            1
                                               蘋果紅
                                                                             48
                    bab6-
                                  2:42
                                                              9b6d-
                                                                                                   89f1-
                                            茶-250ml
                                                       380fbc2067bd
              519f76ccf681
                                                                                           242f7a2d8a65
                                                                                                           19
                4ec0b688-
                                                          004e97ea-
                                                                                               f57143f5-
                                         光泉茉莉茶園
                36e0-4018-
                              2021/9/9
                                                         3847-4b71-
                                                                                              672f-4c78-
                                                                                                            a!
          2
                                               蘋果紅
                                                                             48
                                                                                       1
                     99fa-
                                 23:16
                                                              9b6d-
                                                                                                   89f1-
                                             茶-250ml
                                                       380fbc2067bd
              045f6ca1e3cf
                                                                                           242f7a2d8a65
                                                                                                          a5b
                11d772a7-
                                                          00818a2f-
                                                                                               f57143f5-
                                         新東陽雪花金
                5a16-4e87-
                             2021/7/2
                                                         857a-402e-
                                                                                              672f-4c78-
                                                                                                            30
          3
                                                                            139
                                             錢豬肉乾
                     8ac2-
                                  9:56
                                                              97e5-
                                                                                                   89f1-
             0c438b1ada18
                                                       fed82eb24f50
                                                                                           242f7a2d8a65
                                                                                                          afe2
                                                          00c7a05e-
                a9174342-
                                                                                               f57143f5-
                                                                                                            7
                                         義美天然取向
                0889-4d43-
                           2020/11/21
                                                         aa41-47de-
                                                                                              672f-4c78-
                                                                                                            3
                                                                             71
                                 14:35 蘇打餅乾(紫菜)
                                                                                                   89f1-
                    b111-
                                                              9bda-
             d604c05b28ee
                                                       ee24f3cb224f
                                                                                           242f7a2d8a65 4729
```

```
In [3]: print(name_to_pid_dict['絲花極柔化妝棉66片']) print(pid_to_name_dict['003dc937-e898-4259-870a-4a9afe2eacd6'])
```

003dc937-e898-4259-870a-4a9afe2eacd6 絲花極柔化妝棉66片

In [4]: train_df.describe()

Out[4]:		sales_price	quantity	age_group
	count	799.000000	799.000000	798.000000
	mean	113.942428	1.426783	4.502506
	std	127.990089	0.917299	0.623157
	min	0.000000	1.000000	3.000000

```
sales_price
                     quantity age_group
25%
        49.000000
                     1.000000
                                 4.000000
50%
        80.000000
                     1.000000
                                 5.000000
75%
       128.000000
                     2.000000
                                 5.000000
max 1499.000000
                     8.000000
                                 5.000000
```

```
In [5]: train_df.dtypes
```

```
object
Out[5]: id
        order date
                         object
                         object
        product name
                         object
        product
                          int64
        sales price
                          int64
        quantity
                         object
        department
                         object
        store
        city
                         object
        district
                         object
        customer
                         object
                         object
                         float64
        age_group
        dtype: object
```

```
In [6]: train_df.isna().sum() #Note: 原始 age_group 沒有缺資料, 但這裡存在1筆資料沒有age_group
```

```
Out[6]: id
                           0
        order_date
                           0
        product_name
                           0
        product
                           0
        sales_price
                           0
        quantity
                           0
        department
                           0
        store
                           0
        city
                           0
        district
                           0
        customer
                           0
                         766
        sex
                           1
        age_group
        dtype: int64
```

```
In [7]: # train
    date = pd.to_datetime(train_df['order_date'])
    train_df['month']=date.dt.month
    train_df['year']=date.dt.year
    train_df['YearMonth'] = date.dt.strftime('%Y-%m')
    train_df['day_name']=date.dt.day_name()
    train_df['day']=date.dt.day
```

In [8]: | train_df.head(5)

Out[8]:		id	order_date	product_name	product	sales_price	quantity	department	
	0	4dd5ac40- a1b3-43a2- be16- aa1392a4fde0	2021/6/10 22:57	絲花極柔化妝 棉66片	003dc937- e898-4259- 870a- 4a9afe2eacd6	96	2	22f2e7c7- 6679-41fd- 9c1d- 6e9562b51bce	1
	1	181be294- 2190-47db- bab6- 519f76ccf681	2021/5/27 2:42	光泉茉莉茶園 蘋果紅 茶-250ml	004e97ea- 3847-4b71- 9b6d- 380fbc2067bd	48	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	1 19

	id	order_date	product_name	product	sales_price	quantity	department	
2	4ec0b688- 36e0-4018- 99fa- 045f6ca1e3cf	2021/9/9 23:16	光泉茉莉茶園 蘋果紅 茶-250ml	004e97ea- 3847-4b71- 9b6d- 380fbc2067bd	48	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	a! a5b
3	11d772a7- 5a16-4e87- 8ac2- 0c438b1ada18	2021/7/2 9:56	新東陽雪花金 錢豬肉乾	00818a2f- 857a-402e- 97e5- fed82eb24f50	139	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	3c afe2
4	a9174342- 0889-4d43- b111- d604c05b28ee	2020/11/21 14:35	義美天然取向 蘇打餅乾(紫菜)	00c7a05e- aa41-47de- 9bda- ee24f3cb224f	71	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	4729

In [9]: len(train_df)

Out[9]: **799**

商品的銷數量及金額

Note: 依據 product ID 有482筆, 但是中文 product_name 確有 485筆 ??

In [10]: df_group_by_prod=train_df.groupby('product').sum().reset_index()
 df_group_by_prod=df_group_by_prod[['product','sales_price','quantity']]
 df_group_by_prod

Out[10]:		product	sales_price	quantity
	0	003dc937-e898-4259-870a-4a9afe2eacd6	96	2
	1	004e97ea-3847-4b71-9b6d-380fbc2067bd	96	2
	2	00818a2f-857a-402e-97e5-fed82eb24f50	139	1
	3	00c7a05e-aa41-47de-9bda-ee24f3cb224f	213	5
	4	04971b9d-7d9d-47eb-8997-90bdb4404599	176	2
	•••			
	477	fcdebb1b-4e4a-451d-bc3e-ee957560f1b6	185	1
	478	fe0729cd-93b7-4228-bb0b-2b0360aa95cb	197	3
	479	fe3f7e8a-5c95-4d88-bdb6-69b79a34abb8	159	1
	480	ff0423d0-2e6d-41ec-98ab-60beb68106ef	219	2
	481	ff34d739-b1e8-4d8f-ad26-26d478659d3a	26	1

482 rows × 3 columns

```
In [11]: def query_name_by_pid(pid):
    return pid_to_name_dict[pid]
```

```
In [12]: train_df
```

	id	order_date	product_name	product	sales_price	quantity	department	
0	4dd5ac40- a1b3-43a2- be16- aa1392a4fde0	2021/6/10 22:57	絲花極柔化妝 棉66片	003dc937- e898-4259- 870a- 4a9afe2eacd6	96	2	22f2e7c7- 6679-41fd- 9c1d- 6e9562b51bce	
1	181be294- 2190-47db- bab6- 519f76ccf681	2021/5/27 2:42	光泉茉莉茶園 蘋果紅 茶-250ml	004e97ea- 3847-4b71- 9b6d- 380fbc2067bd	48	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	
2	4ec0b688- 36e0-4018- 99fa- 045f6ca1e3cf	2021/9/9 23:16	光泉茉莉茶園 蘋果紅 茶-250ml	004e97ea- 3847-4b71- 9b6d- 380fbc2067bd	48	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	i
3	11d772a7- 5a16-4e87- 8ac2- 0c438b1ada18	2021/7/2 9:56	新東陽雪花金 錢豬肉乾	00818a2f- 857a-402e- 97e5- fed82eb24f50	139	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	i
4	a9174342- 0889-4d43- b111- d604c05b28ee	2020/11/21 14:35	義美天然取向 蘇打餅乾(紫菜)	00c7a05e- aa41-47de- 9bda- ee24f3cb224f	71	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	4
•••								
794	c33d84e1- c1f9-47b4- a3b5- 76b5253de366	2021/7/29 1:23	KONOMI相撲 手脆紫菜-山葵 風味	fe0729cd- 93b7-4228- bb0b- 2b0360aa95cb	59	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	
795	181be294- 2190-47db- bab6- 519f76ccf681	2021/5/27 2:42	金車麥根沙士 can-330ml	fe0729cd- 93b7-4228- bb0b- 2b0360aa95cb	79	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	
796	7608d508- 9e41-451b- 9794- 30f0b5deeed7	2021/6/6 16:41	白蘭新動力配 方洗碗精-活力 鮮柚-2.8kg	fe3f7e8a- 5c95-4d88- bdb6- 69b79a34abb8	159	1	451ffc74- a780-4ae0- adba- ff8caa8f4b91	i
797	44b1ad0f- c9c7-4308- 872b- 70db6f741319	2020/10/29 19:12	扇屋 一口起司 鳕魚條 115.2g	ff0423d0- 2e6d-41ec- 98ab- 60beb68106ef	219	2	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	
798	181be294- 2190-47db- bab6- 519f76ccf681	2021/5/27 2:42	家樂福香蔥脆 餅100g	ff34d739- b1e8-4d8f- ad26- 26d478659d3a	26	1	f57143f5- 672f-4c78- 89f1- 242f7a2d8a65	

799 rows × 18 columns

銷售額前K名商品

df_prod_sales_price=df_group_by_prod.sort_values(by='sales_price',ascending=False) In [13]: df_prod_sales_price

Out[13]: product sales_price quantity

Out[12]:

	product	sales_price	quantity
42	17034cf4-9078-401e-b736-2c75c911bcd7	2260	5
296	a2499275-bd05-43b0-8b3e-4af30571ea54	1976	2
452	eedcdefa-3997-4796-bf72-2b23cc8abc07	1499	1
366	c7e01b71-ff4b-439c-83cd-d1ff956ba4c1	1444	2
•••			
245	8071720b-fdbe-409c-8684-b7c604bd082e	25	2
471	f82164d5-704f-4c03-a2ae-2100b8807d58	25	2
365	c639e0fb-ef07-4502-aa02-de8203701fe0	24	1
404	d7694980-815f-4a87-bd06-805d5e072a0f	19	1
57	1b1d0ba6-71b5-4d00-b0b7-960b03fd28d1	18	1

482 rows × 3 columns

```
In [14]: df_prod_sales=df_group_by_prod.sort_values(by='sales_price',ascending=False)
    df_prod_sales
```

Out[14]:		product	sales_price	quantity
	135	44efb156-79e0-43e9-bab8-9d377f2e2185	2883	9
	42	17034cf4-9078-401e-b736-2c75c911bcd7	2260	5
	296	a2499275-bd05-43b0-8b3e-4af30571ea54	1976	2
	452	eedcdefa-3997-4796-bf72-2b23cc8abc07	1499	1
	366	c7e01b71-ff4b-439c-83cd-d1ff956ba4c1	1444	2
	•••			
	245	8071720b-fdbe-409c-8684-b7c604bd082e	25	2
	471	f82164d5-704f-4c03-a2ae-2100b8807d58	25	2
	365	c639e0fb-ef07-4502-aa02-de8203701fe0	24	1
	404	d7694980-815f-4a87-bd06-805d5e072a0f	19	1
	57	1b1d0ba6-71b5-4d00-b0b7-960b03fd28d1	18	1

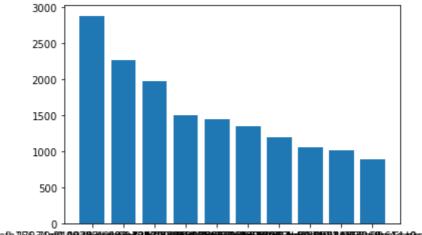
482 rows × 3 columns

```
In [15]:

def get_top_k_product(df,k,by_column='sales_price'):
    df_group_by_prod=df.groupby('product').sum().reset_index()
    df_group_by_prod=df_group_by_prod[['product','sales_price','quantity']]
    df_prod_sales=df_group_by_prod.sort_values(by=by_column,ascending=False)[:k]
    plt.bar(df_prod_sales['product'], df_prod_sales[by_column])
    plt.show()
    result=[]
    for i in range(len(df_prod_sales)):
        pid = df_prod_sales.iloc[i]['product']
        price = df_prod_sales.iloc[i]['sales_price']
        quantity = df_prod_sales.iloc[i]['quantity']
        #print(pid,query_name_by_pid(pid),price,quantity)
        result.append((pid,query_name_by_pid(pid),price,quantity))
```

In [16]:

```
top K prod=get top k product(train df,10,'sales price')
for i in top_K_prod: print(i)
```



44efb 15**6.34±919932431101472643 bazeara sanas s**

```
('44efb156-79e0-43e9-bab8-9d377f2e2185', 'OLAY防曬淨白乳液(UV)SPF19-150ml', 2883, 9)
('17034cf4-9078-401e-b736-2c75c911bcd7', 'Saborino晚安面膜28PC', 2260, 5)
('a2499275-bd05-43b0-8b3e-4af30571ea54', '聲寶TQ-BA30C日式多功能料理鍋3L', 1976, 2) ('eedcdefa-3997-4796-bf72-2b23cc8abc07', '歐樂B PRO 1 3D電動牙刷-白色', 1499, 1) ('c7e01b71-ff4b-439c-83cd-d1ff956ba4c1', '晶工牌JK-103多功能不鏽鋼電碗1.5L', 1444, 2)
('b529c84a-8085-444b-b200-feec78ff54be', '冷藏澳洲黑傑克板腱牛排(每盒約200g)', 1350, 3)
('46b068d9-0683-480c-acfa-d4f574868a99', '德國百靈Oral-B D100活力亮潔電動牙刷', 1190,
1)
('e413ee9a-61a6-4a59-9a93-d30c68ea6cc0', '【安心價】家樂福有機特級杏鮑菇(每包約1公斤)',
('78f999a6-daf4-4b28-95e2-db614e0c8123', 'CAB冷藏美國安格斯菲力牛排180g(貼體包裝)', 102
0, 2)
('f627e52a-9714-430a-ba4a-bcc2af0ecc86', '好神拖輕巧手壓旋轉拖把組', 888, 1)
```

銷售量前K名的商品

In [17]:

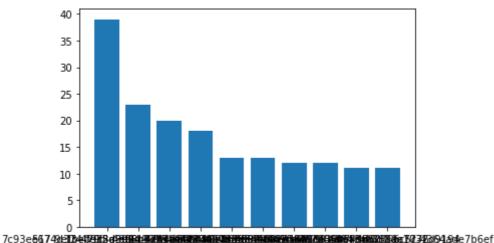
df_group_by_prod.sort_values(by='quantity',ascending=False)

_			-
\cap	14-	117	
UU	I L	1 1	

	product	sales_price	quantity
238	7c93ee17-9ef3-479c-a9e2-c4791ae422da	678	39
163	5674d374-fe32-4898-991e-e30bb92e8b2c	116	23
60	1be09c5d-ef6d-4ef5-8ba9-4b17c3f937a2	372	20
30	0f8d3904-374d-4443-a0b0-5e6d9002416a	600	18
331	b6b9d87b-7fb4-424d-860c-baae35008533	425	13
•••			
305	aaa5a5f8-99af-4744-b440-27e4fbad18e9	95	1
306	aaaf8829-150e-4b2d-9c81-58ef96f73689	79	1
307	aac1198c-dba3-4a72-bbd2-526df66f3d6f	155	1
308	aada8ad0-3638-4b06-abc6-5ee65877439c	35	1
481	ff34d739-b1e8-4d8f-ad26-26d478659d3a	26	1

482 rows × 3 columns

top_K_prod=get_top_k_product(train_df,10,'quantity') In [18]: for i in top K prod: print(i)



('7c93ee17-9ef3-479c-a9e2-c4791ae422da', '新東陽原味德式香陽-160g', 678, 39)

('5674d374-fe32-4898-991e-e30bb92e8b2c', 'LOTTE樂天粒粒葡萄汁', 116, 23)

('1be09c5d-ef6d-4ef5-8ba9-4b17c3f937a2', '華元海蝦蝦餅甜辣口味', 372, 20) ('0f8d3904-374d-4443-a0b0-5e6d9002416a', '義美寶吉葡萄果汁QQ糖-176g', 600, 18)

('b6b9d87b-7fb4-424d-860c-baae35008533', '米小餅-金薯口味(米果)', 425, 13)

('14fb6180-07f9-4699-98ec-e857b058e7dd', '安佳鮮奶油250ml', 178, 13)

('c170b78a-fe4e-461c-aa19-55bca36b2366', '【全素】隨緣素肉骨茶麵(袋) 90g', 414, 12)

('b3f8864b-45fe-49c2-aaf7-cba8889a7776', '海太脆皮黑森林蛋糕', 467, 12)

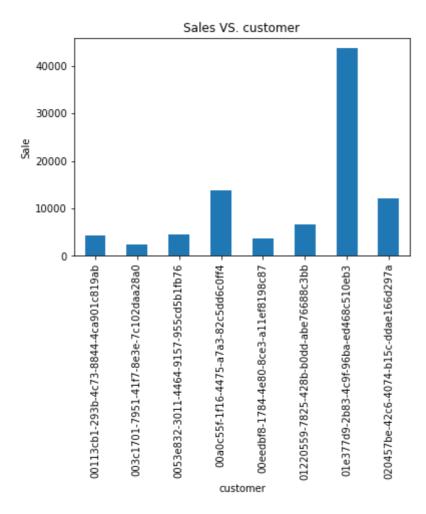
('76f0d2df-6810-446f-9b50-1ac1032d9194','【Nakery裸焙坊】法式香蒜麵包 (每條約162g)',1 95, 11)

('51871ce3-8849-40d1-bc5c-4364ade7b6ef', '安佳鮮奶油250ml', 645, 11)

整體銷售狀況

```
In [19]:
         total records=len(train df)
         total_transactions=len(train_df.groupby('id'))
         total_customer=len(train_df.groupby('customer'))
         #所有發票上的金額 ## same as train_df['sales_price'].sum()
```

```
total_sales=train_df.groupby('id')['sales_price'].sum().sum()
          total_products=len(train_df.groupby('product'))
         print('發票數量:',total transactions)
In [20]:
         print('客戶數量:',total customer)
         print('平均客戶消費次數:',total_transactions/total_customer)
          print('平均發票金額:',total_sales/total_transactions)
          print('平均每筆發票的商品種類:',total_records/total_transactions)
          print('商品種類:',total_products)
         發票數量:86
         客戶數量:8
         平均客戶消費次數: 10.75
         平均發票金額: 1058.6046511627908
         平均每筆發票的商品種類: 9.290697674418604
         商品種類: 482
        每個顧客的購買總金額
In [21]: #train_df[train_df['customer']=='003c1701-7951-41f7-8e3e-7c102daa28a0']
In [22]:
         df_group_by_customer=train_df.groupby('customer')['sales_price'].agg(['sum'])
          df_group_by_customer
Out[22]:
                                            sum
                                  customer
          00113cb1-293b-4c73-8844-4ca901c819ab
                                            4380
          003c1701-7951-41f7-8e3e-7c102daa28a0
                                            2447
         0053e832-3011-4464-9157-955cd5b1fb76
                                            4415
           00a0c55f-1f16-4475-a7a3-82c5dd6c0ff4 13805
          00eedbf8-1784-4e80-8ce3-a11ef8198c87
                                            3572
         01220559-7825-428b-b0dd-abe76688c3bb
                                            6537
          01e377d9-2b83-4c9f-96ba-ed468c510eb3 43703
         020457be-42c6-4074-b15c-ddae166d297a 12181
         #長條圖:每個顧客的購買總金額
In [23]:
          df_group_by_customer['sum'].plot(kind='bar');
          plt.title("Sales VS. customer")
          plt.xlabel("customer")
          plt.ylabel("Sale")
          plt.show()
```



每筆發票的金額

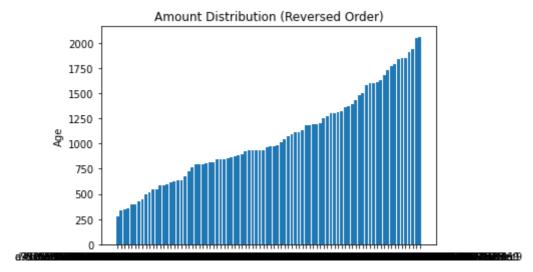
In [24]: df_group_by_id=train_df.groupby('id')['sales_price'].sum().reset_index()
 df_group_by_id=df_group_by_id.sort_values(by='sales_price') #ascending=False
 df_group_by_id

Out[24]:		id	sales_price
	70	d990fe98-8dff-4d7e-b686-d79ddda05e2f	280
	52	a70c6a3c-d8b8-4280-b85f-56eeb9621006	341
	21	410aff17-3fbc-4829-a121-496182f4dc8d	343
	65	ccefe038-e08f-409b-856e-b3bce9b975e6	354
	28	536bc374-20f1-46d4-a02c-63799188a9b6	392
	•••		
	8	1ea71713-8cbf-4726-8a3b-58372872e37c	1853
	77	e8885fa7-f816-45e1-8e4c-9445583a8e01	1912
	25	4dd5ac40-a1b3-43a2-be16-aa1392a4fde0	1943
	75	e3a90bce-064b-447b-a4b9-2e6bf1323ec9	2051
	73	df63c19c-faaf-45b6-81cd-7a57fc52f769	2061

86 rows × 2 columns

```
In [25]: plt.bar(df_group_by_id['id'], df_group_by_id['sales_price'])
plt.ylabel('Age')
```

```
plt.title('Amount Distribution (Reversed Order)')
#plt.xticks(rotation=45)
#plt.tight_layout()
plt.show()
```



查詢某客戶的購買歷史記錄(消費及其金額)

```
def show_purchase_by_customer(df,cid):
In [26]:
              dataset=df[train_df['customer']==cid].groupby('id')['sales_price'].agg(['sum']).
              total_amounts=dataset['sum'].sum()
              print(f'客戶({customer_id})消費筆數:{len(dataset)}, 累積消費金額為 {total_amounts}
              return dataset, total amounts
          customer id='00113cb1-293b-4c73-8844-4ca901c819ab'
In [27]:
          dataset,total_amounts=show_purchase_by_customer(train_df,customer_id)
          dataset
          客戶(00113cb1-293b-4c73-8844-4ca901c819ab)消費筆數:6, 累積消費金額為 4380元,平均每次消費
         730.0
Out[27]:
                                           id
                                              sum
             65f9657e-a473-46be-bf6c-ebbc508297a9
                                              1611
            a70c6a3c-d8b8-4280-b85f-56eeb9621006
                                               341
            a9a35c16-433c-47a3-a965-5a354b51142c
                                               624
             ccefe038-e08f-409b-856e-b3bce9b975e6
                                               354
          3
              cf8105d9-95e9-418e-85e7-bf2f7a8ae2cc
                                               934
            f2030d62-55df-4399-9dcb-199b1a04d440
                                               516
          dataset,total_amounts=show_purchase_by_customer(train_df,'003c1701-7951-41f7-8e3e-7c
In [28]:
          dataset
```

客戶(00113cb1-293b-4c73-8844-4ca901c819ab)消費筆數:3,累積消費金額為 2447元,平均每次消費 815.666666666666

Out[28]:		id	sum
	0	1b5ae746-37f0-442e-aa58-ffee397a2c6a	795
	1	3ecffe33-4293-423d-a533-c115ebf0fab4	845
	2	d96c447d-8c73-45dd-8cb9-1cde700485a3	807

查詢發票的交易明細

```
dataset=train_df[train_df['id']=='65f9657e-a473-46be-bf6c-ebbc508297a9'][['product_n
In [29]:
In [30]:
          dataset
Out[30]:
           index
                                      product_name sales_price quantity
         0
             476
                                    台糖精選魚油膠囊
                                                        397
                                                                  1
                 家樂福BIO有機紅肉火龍果(每盒約850克/2粒)
         1
             485
                                                         99
                                                                  1
         2
                              潔霜芳香浴廁清潔劑--檸檬
             601
                                                         39
                                                                  1
                             小瓜呆脆笛酥-草莓口味195G
         3
             731
                                                         89
         4
             768
                              好神拖輕巧手壓旋轉拖把組
                                                        888
                                                                  1
                                  中興履歷無洗米1.5Kg
         5
             775
                                                         99
                                                                  1
          def show_invoice(df,tid,show_detail=True):
In [31]:
             data=df[df['id']==tid][['product_name','sales_price','quantity']].reset_index()
             amount=data['sales_price'].sum()
             counts=len(data)
             if show detail:
                 print(f'發票號碼:{tid} 消費總金額: {amount} 元, 共 {counts} 樣商品·明細如下:')
                 print(f'
                            {"品項":30},{"數量":3},{"金額":5}')
                 for i in range(len(data)):
                     item = data.iloc[i]['product name']
                     q = data.iloc[i]['quantity']
                     price=data.iloc[i]['sales_price']
                     print(f'{i+1}:{item:<30}{q:3}{price:>5}')
             return amount,len(data),data
         show_invoice(train_df,'65f9657e-a473-46be-bf6c-ebbc508297a9',show_detail=False)
In [32]:
         (1611,
Out[32]:
          6,
             index
                                product name sales price quantity
          0
              476
                                    台糖精選魚油膠囊
                                                           397
                                                                       1
                   家樂福BIO有機紅肉火龍果(每盒約850克/2粒)
          1
              485
                                                                 99
                                                                            1
                               潔霜芳香浴廁清潔劑--檸檬
          2
              601
                                                              39
                                                                         1
                             小瓜呆脆笛酥-草莓口味195G
          3
              731
                                                              89
                                                                        1
                                好神拖輕巧手壓旋轉拖把組
          4
              768
                                                              888
                                                                          1
                                中興履歷無洗米1.5Kg
          5
              775
                                                           99
                                                                      1)
          show invoice(train df,'a70c6a3c-d8b8-4280-b85f-56eeb9621006',show detail=True)
In [33]:
```

```
發票號碼:a70c6a3c-d8b8-4280-b85f-56eeb9621006 消費總金額: 341 元, 共 4 樣商品‧明細如下:
           品項
                                     ,數量,金額
        1:威尼斯冷水壺1L
                                        2 148
        2:【火鍋好物】西北日式手工蝦餃
                                                 39
                                             2
        3:百研三合一瓶蓋清潔刷
                                          1 75
                                            79
        4:家樂福豬肉火鍋片-180g
                                         2
Out[33]: (341,
           index
                  product_name sales_price quantity
             179
                      威尼斯冷水壺1L
                                                   2
        0
                 【火鍋好物】西北日式手工蝦餃
        1
             554
                                               39
                                                        2
                                            75
        2
            754
                    百研三合一瓶蓋清潔刷
                                                     1
             782
                  家樂福豬肉火鍋片-180g
        3
                                           79
                                                    2)
        df_group_by_id=train_df.groupby('id').sum().reset_index()
In [34]:
        df_group_by_id
```

Out[34]:		id	sales_price	quantity	age_group	month	year	day
	0	0ad57976-5aca-4218-a907-aef772fc1039	1305	24	68.0	85	34357	510
	1	11d772a7-5a16-4e87-8ac2-0c438b1ada18	1309	12	44.0	77	22231	22
	2	14317593-ab59-43c5-af1e-a9b557e83e0f	892	9	20.0	35	10100	145
	3	17b18cda-3072-4355-a527-080f070fae7d	1095	17	28.0	35	14140	84
	4	181be294-2190-47db-bab6-519f76ccf681	1850	25	110.0	110	44462	594
	81	f2030d62-55df-4399-9dcb-199b1a04d440	516	6	18.0	42	12120	18
	82	fa56589e-1207-405e-b150-7869d50964f3	393	14	40.0	50	20210	230
	83	fb8911a9-8ed2-4219-b629- 24d01db08b18	1181	18	85.0	85	34357	289
	84	fc2a99ec-19cd-4561-b050-efdd18c5c0a6	1323	18	90.0	90	36378	558
	85	fe441d97-e167-4742-95b1-3fb63181a0b4	929	4	20.0	40	8080	44

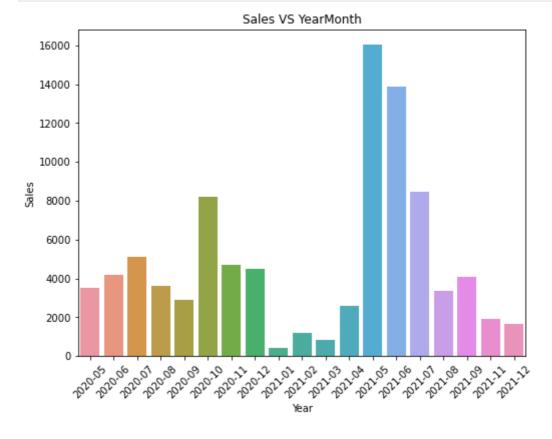
86 rows × 7 columns

每月總銷售額

Out[35]:		YearMonth	sum
	0	2020-05	3488
	1	2020-06	4207
	2	2020-07	5092
	3	2020-08	3615
	4	2020-09	2897
	5	2020-10	8185
	6	2020-11	4694

	YearMonth	sum
7	2020-12	4481
8	2021-01	426
9	2021-02	1204
10	2021-03	845
11	2021-04	2607
12	2021-05	16019
13	2021-06	13853
14	2021-07	8450
15	2021-08	3377
16	2021-09	4057
17	2021-11	1912
18	2021-12	1631

```
In [36]: plt.figure(figsize=(8, 6))
    sns.barplot(data=sales_per_month,x='YearMonth',y='sum')
    plt.title('Sales VS YearMonth')
    plt.ylabel('Sales')
    plt.xlabel("Year")
    plt.xticks(rotation=45)
    plt.show()
```



每月各分店的銷售額

```
In [37]: print('共有店家:',len (train_df.groupby('store')))
```

共有店家: 12

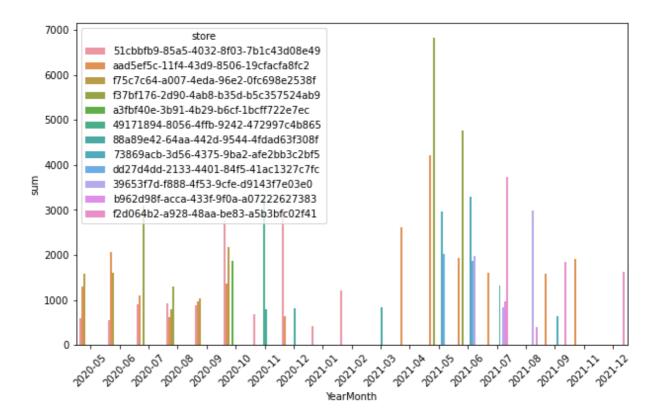
Out[38]: sum

YearMonth	store	
2020-05	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	591
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1305
	f75c7c64-a007-4eda-96e2-0fc698e2538f	1592
2020-06	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	548
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	2061
	f75c7c64-a007-4eda-96e2-0fc698e2538f	1598
2020-07	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	892
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1108
	f37bf176-2d90-4ab8-b35d-b5c357524ab9	3092
2020-08	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	919
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	614
	f37bf176-2d90-4ab8-b35d-b5c357524ab9	1288
	f75c7c64-a007-4eda-96e2-0fc698e2538f	794
2020-09	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	888
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	967
	f75c7c64-a007-4eda-96e2-0fc698e2538f	1042
2020-10	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	2774
	a3fbf40e-3b91-4b29-b6cf-1bcff722e7ec	1858
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1370
	f75c7c64-a007-4eda-96e2-0fc698e2538f	2183
2020-11	49171894-8056-4ffb-9242-472997c4b865	3225
	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	674
	88a89e42-64aa-442d-9544-4fdad63f308f	795
2020-12	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	3039
	88a89e42-64aa-442d-9544-4fdad63f308f	807
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	635
2021-01	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	426
2021-02	51cbbfb9-85a5-4032-8f03-7b1c43d08e49	1204
2021-03	88a89e42-64aa-442d-9544-4fdad63f308f	845
2021-04	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	2607
2021-05	73869acb-3d56-4375-9ba2-afe2bb3c2bf5	2975
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	4210
	dd27d4dd-2133-4401-84f5-41ac1327c7fc	2018

sum

YearMonth	store	
	f37bf176-2d90-4ab8-b35d-b5c357524ab9	6816
2021-06	39653f7d-f888-4f53-9cfe-d9143f7e03e0	1983
	73869acb-3d56-4375-9ba2-afe2bb3c2bf5	3289
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1943
	dd27d4dd-2133-4401-84f5-41ac1327c7fc	1864
	f37bf176-2d90-4ab8-b35d-b5c357524ab9	4774
2021-07	39653f7d-f888-4f53-9cfe-d9143f7e03e0	839
	73869acb-3d56-4375-9ba2-afe2bb3c2bf5	1309
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1602
	b962d98f-acca-433f-9f0a-a07222627383	978
	f2d064b2-a928-48aa-be83-a5b3bfc02f41	3722
2021-08	39653f7d-f888-4f53-9cfe-d9143f7e03e0	2985
	f2d064b2-a928-48aa-be83-a5b3bfc02f41	392
2021-09	73869acb-3d56-4375-9ba2-afe2bb3c2bf5	637
	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1581
	f2d064b2-a928-48aa-be83-a5b3bfc02f41	1839
2021-11	aad5ef5c-11f4-43d9-8506-19cfacfa8fc2	1912
2021-12	f2d064b2-a928-48aa-be83-a5b3bfc02f41	1631

每月各分店銷售狀況



分店每月銷售狀況

```
In [40]:

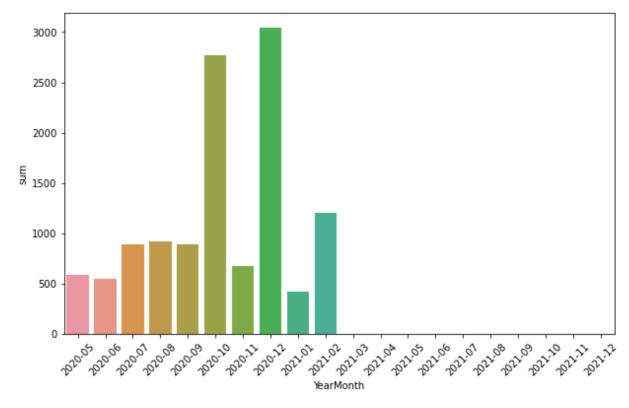
def analyze_store_sales(df,store_id):
    sales_per_month_store=df.groupby(['YearMonth','store'])['sales_price'].agg(['sum data=sales_per_month_store[sales_per_month_store['store']==store_id]
    data=data[['YearMonth','sum']]

# Create a DataFrame with all months (including missing ones) and fill missing v
    all_months = pd.date_range(start='2020-05', end='2022-01', freq='M')
    df_m = pd.DataFrame({'YearMonth': all_months})
    df_m['YearMonth'] = pd.to_datetime(df_m['YearMonth']).dt.to_period('M')
    df_m['YearMonth'] = df_m['YearMonth'].apply(str)
    df_data = pd.merge(df_m,data, on=['YearMonth'], how='left').fillna(0)
    plt.figure(figsize=(10, 6))
    plt.xticks(rotation=45)
    sns.barplot(x='YearMonth', y='sum', data=df_data)
    return df_data
```

In [41]: analyze_store_sales(train_df,'51cbbfb9-85a5-4032-8f03-7b1c43d08e49')

Out[41]:		YearMonth	sum
	0	2020-05	591.0
	1	2020-06	548.0
	2	2020-07	892.0
	3	2020-08	919.0
	4	2020-09	888.0
	5	2020-10	2774.0
	6	2020-11	674.0
	7	2020-12	3039.0
	8	2021-01	426.0

	YearMonth	sum
9	2021-02	1204.0
10	2021-03	0.0
11	2021-04	0.0
12	2021-05	0.0
13	2021-06	0.0
14	2021-07	0.0
15	2021-08	0.0
16	2021-09	0.0
17	2021-10	0.0
18	2021-11	0.0
19	2021-12	0.0

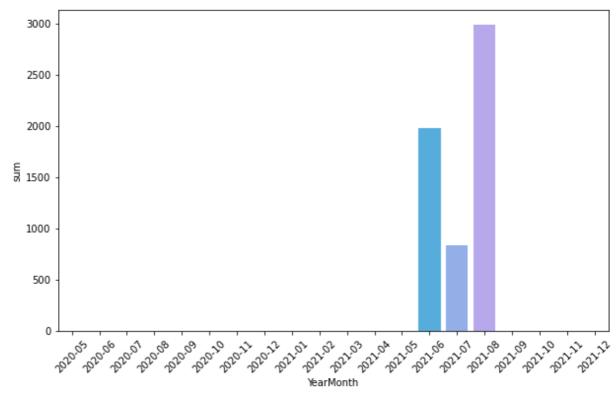


In [42]: analyze_store_sales(train_df,'39653f7d-f888-4f53-9cfe-d9143f7e03e0')

\cap		1/2	
VL	HL.	144	

	YearMonth	sum
0	2020-05	0.0
1	2020-06	0.0
2	2020-07	0.0
3	2020-08	0.0
4	2020-09	0.0
5	2020-10	0.0
6	2020-11	0.0
7	2020-12	0.0

	YearMonth	sum
8	2021-01	0.0
9	2021-02	0.0
10	2021-03	0.0
11	2021-04	0.0
12	2021-05	0.0
13	2021-06	1983.0
14	2021-07	839.0
15	2021-08	2985.0
16	2021-09	0.0
17	2021-10	0.0
18	2021-11	0.0
19	2021-12	0.0



Exercise

分析顧客購買商品的單價區間(最大,最小,平均)

In []:

分析客戶購買時段 (上午: 07:00-11:00,中午11:00-14:00,下午:14:00-21:00, 晚上:21:00-07:00)

In []: