In [89]:

```
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import calendar
import seaborn as sns
plt.rcParams['font.family']= ['Microsoft JhengHei']
```

In [2]:

```
df1 = pd.read_excel('CUST_PROPERTY_FIN_1.xlsx')
```

In [3]:

```
df2 = pd.read_excel('CUST_PROPERTY_FIN_2.xlsx')
```

In [219]:

```
df2.index = [65000+i for i in range(len(df2.index))]
```

In [220]:

```
frames = [df1,df2]
df = pd.concat(frames)
```

In [223]:

```
df.tail()
```

Out[223]:

	CUST_RK	ternure_m	recency_m	SIN	SIN_his	REG	REG_his	ILP	ILP_his	AHa	...
130482	251944	271	4	1	1	1	1	0	0	1	...
130483	251945	255	54	0	0	1	1	0	0	1	...
130484	251947	135	102	0	0	1	1	0	0	1	...
130485	251951	125	34	0	0	1	1	0	0	1	...
130486	251954	297	68	0	0	1	1	0	0	1	...

5 rows × 29 columns

In [ ]:

Missing Value

In [26]:

```
df.isnull().sum(axis = 0)
```

Out[26]:

```
CUST_RK                0
ternure_m              0
recency_m              0
SIN                    0
SIN_his                0
REG                    0
REG_his                0
ILP                    0
ILP_his                0
AHa                    0
AHa_his                0
AHb                    0
AHb_his                0
AHc                    0
AHc_his                0
AHd                    0
AHd_his                0
VIP_CLASS         127632
VIP                    0
WEALTH_LEVEL           0
CLIENT_MARITAL     42221
CLIENT_INCOME          0
DIGI_FLG               0
TOPCARD                0
GENDER                 0
stick_level2           0
cust_group2            0
TOTAL_AUM          11786
INSURED_DOB            0
dtype: int64
```

In [27]:

```
# 有Missing value的列數(客戶數)
sum([1 for i in df.isnull().sum(axis = 1) if i != 0])
```

Out[27]:

```
127687
```

1 客戶RK

In [15]:

```
df['CUST_RK'].value_counts(dropna=False)
```

Out[15]:

```
4094      1
158823    1
209996    1
212045    1
205902    1
         ..
240416    1
244514    1
236326    1
234279    1
2049      1
Name: CUST_RK, Length: 130487, dtype: int64
```

2 客戶戶齡 (月)

In [138]:

```
df['ternure_m'].value_counts(dropna=False)
# 可以做個圖看戶齡分配
```

Out[138]:

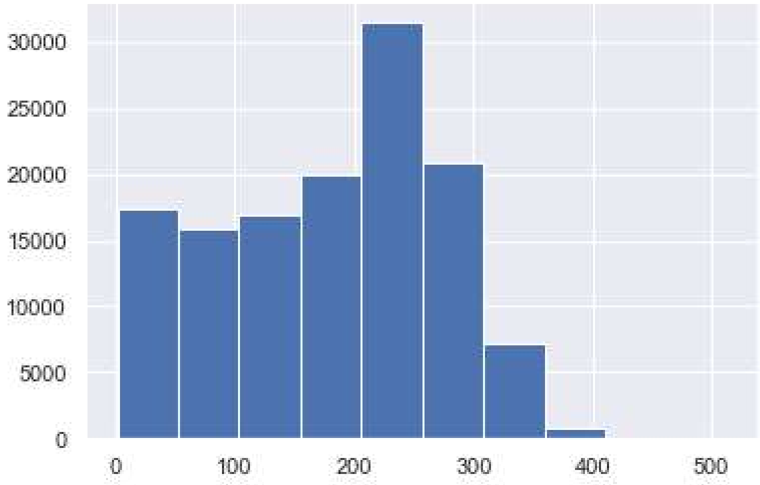
```
229    2463
181    2000
205    1760
13     1330
169    1170
       ...
452       1
430       1
486       1
420       1
450       1
Name: ternure_m, Length: 454, dtype: int64
```

In [222]:

```
sns.set()
plt.hist(df['ternure_m'])
```

Out[222]:

```
(array([1.7345e+04, 1.5848e+04, 1.6969e+04, 1.9921e+04, 3.1444e+04,
        2.0831e+04, 7.2290e+03, 8.1900e+02, 6.9000e+01, 1.2000e+01]),
 array([  1. ,  52.2, 103.4, 154.6, 205.8, 257. , 308.2, 359.4, 410.6,
        461.8, 513. ]),
 <a list of 10 Patch objects>)
```

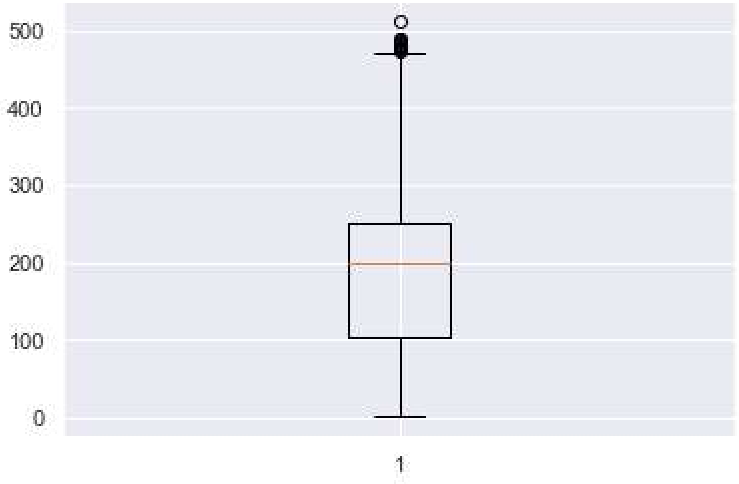


In [135]:

```
plt.boxplot(df['ternure_m'])
df['ternure_m'].describe()
```

Out[135]:

```
count    130487.000000
mean        179.262432
std          92.741422
min           1.000000
25%         102.000000
50%         199.000000
75%         250.000000
max         513.000000
Name: ternure_m, dtype: float64
```



4 現在、過去持有保單

In [85]:

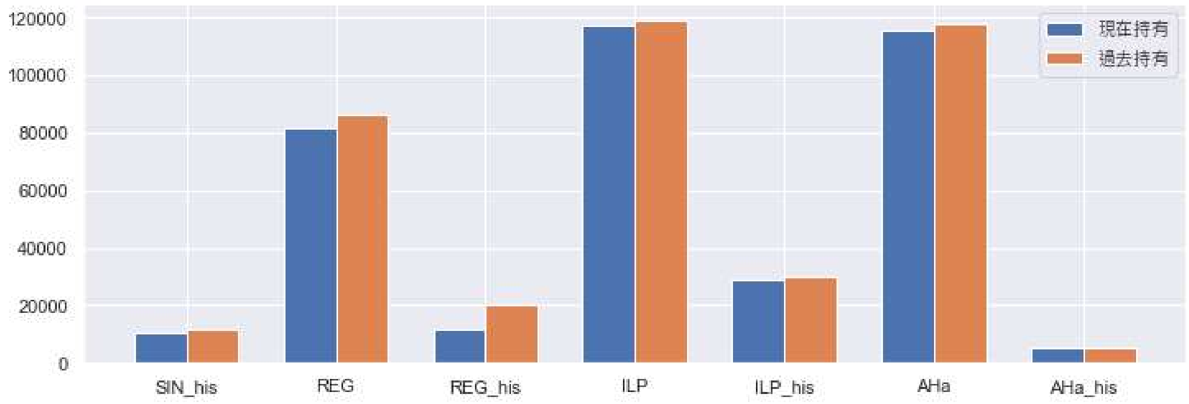
```
policy1 = ['SIN', 'REG', 'ILP', 'AHa',
          'AHb', 'AHc','AHd', ]
policy1_num = []
for item in policy1:
    policy1_num.append(df[item].value_counts().sort_index().values[1])

policy2 = [ x+"_his" for x in policy1  ]
policy2_num = []
for item in policy2:
    policy2_num.append(df[item].value_counts().sort_index().values[1])
```

In [104]:

```
sns.set()

plt.figure(figsize=(12,4))
x = np.arange(len(policy1))
width = 0.35
plt.bar(x - width/2, policy1_num, width, label='現在持有')
plt.bar(x + width/2, policy2_num, width, label='過去持有')
plt.gca().set_xticklabels(policy)
plt.rcParams['font.family']= ['Microsoft JhengHei']
plt.legend()
```



5 VIP等級

In [91]:

```
df['VIP_CLASS'].value_counts(dropna=False)
```

Out[91]:

```
NaN    127632
V05      1789
V04       932
V03        88
V02        27
V01        19
Name: VIP_CLASS, dtype: int64
```

In [92]:

```
df['VIP'].value_counts(dropna=False)
```

Out[92]:

```
0    127632
1      2855
Name: VIP, dtype: int64
```

6 財富等級

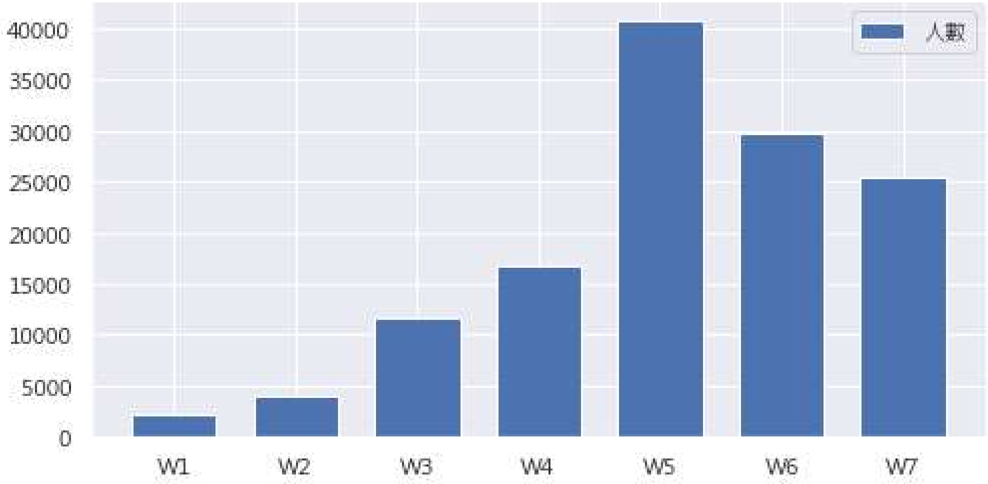
W1最高-->W7最低

In [97]:

```
wealth = df['WEALTH_LEVEL'].value_counts(dropna=False).sort_index()
```

In [179]:

```
sns.set()
plt.figure(figsize=(8,4))
width = 0.7
plt.rcParams['font.family']= ['Microsoft JhengHei']
plt.bar(wealth.index , wealth.values, width, label='人數')
plt.legend()
```



7 婚姻狀況

In [128]:

```
df['CLIENT_MARITAL'].value_counts(dropna=False)
```

Out[128]:

```
M      50244
NaN    42221
S      38022
Name: CLIENT_MARITAL, dtype: int64
```

8 客戶年收入

In [130]:

```
df['CLIENT_INCOME'].value_counts(dropna=False).sort_index()
```

Out[130]:

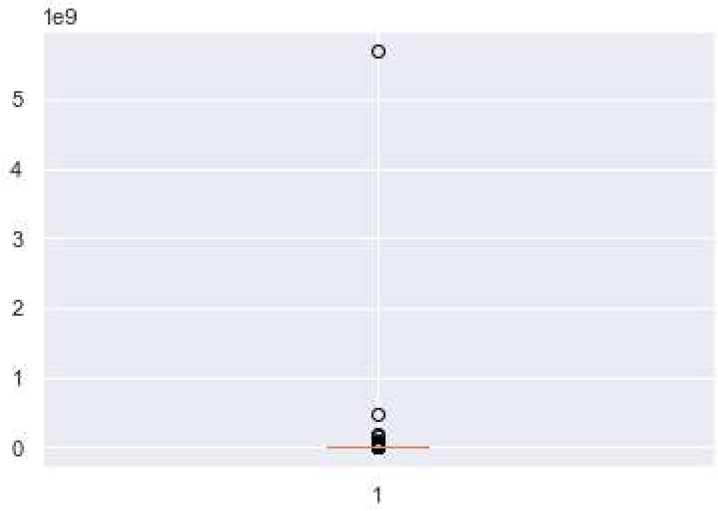
```
0             36186
9500              1
19000             9
28500             8
33250             1
              ...  
95000000          2
152000000         1
180500000         1
475000000         1
5700000000        1
Name: CLIENT_INCOME, Length: 267, dtype: int64
```

In [136]:

```
plt.boxplot(df['CLIENT_INCOME'])
df['CLIENT_INCOME'].describe()
# 57億 ?
# 0 是真的沒有收入還是沒資料 ?
```

Out[136]:

```
count    1.304870e+05
mean     7.574329e+05
std      1.589034e+07
min      0.000000e+00
25%      0.000000e+00
50%      5.700000e+05
75%      9.500000e+05
max      5.700000e+09
Name: CLIENT_INCOME, dtype: float64
```

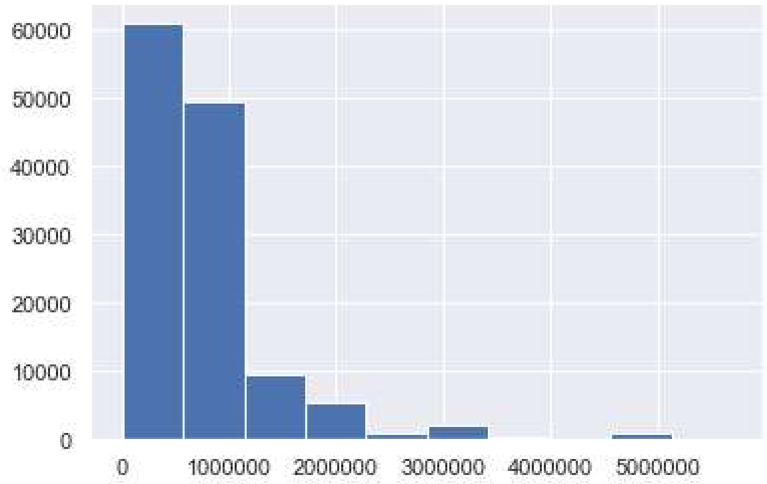


In [202]:

```
import copy
tmp = copy.deepcopy(df['CLIENT_INCOME'])
plt.hist(sorted(tmp)[:int(0.995*len(tmp))])
# 去掉最後0.5%的年收入分配
```

Out[202]:

```
(array([6.0870e+04, 4.9354e+04, 9.4550e+03, 5.5020e+03, 9.3100e+02,
        2.2500e+03, 3.8000e+02, 5.1000e+01, 9.3800e+02, 1.0300e+02]),
 array([      0.,  570000., 1140000., 1710000., 2280000., 2850000.,
        3420000., 3990000., 4560000., 5130000., 5700000.]),
 <a list of 10 Patch objects>)
```



9 數位客戶

1:數位客戶 0:非數位客戶

In [162]:

```
df['DIGI_FLG'].value_counts(dropna=False)
```

Out[162]:

```
0    120091
1     10396
Name: DIGI_FLG, dtype: int64
```

10 頂級卡

In [168]:

```
df['TOPCARD'].value_counts(dropna=False)
```

Out[168]:

```
0    128761
1      1726
Name: TOPCARD, dtype: int64
```

11 性別

1:女 0:男

In [169]:

```
df['GENDER'].value_counts(dropna=False)
```

Out[169]:

```
1    70831
0    59656
Name: GENDER, dtype: int64
```

12 忠誠度

S01最高-->S10最低

In [185]:

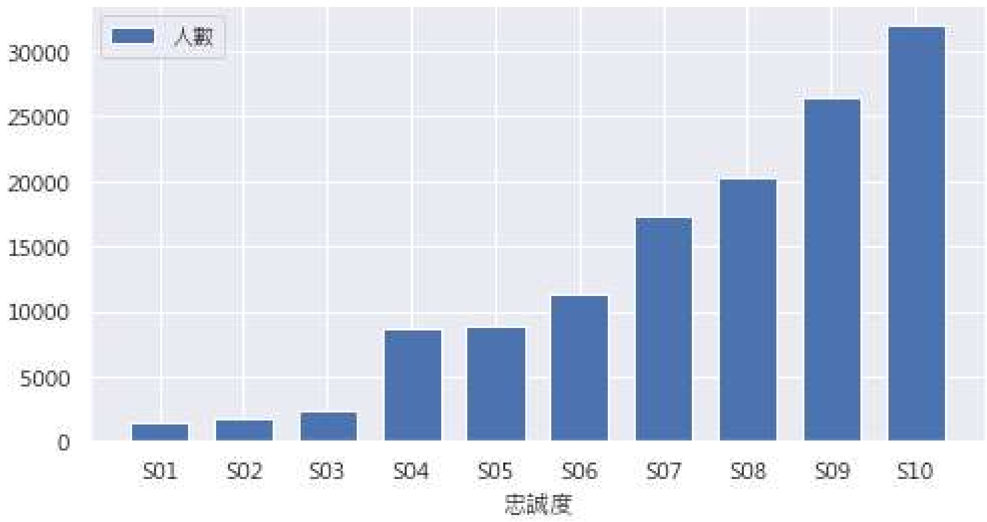
```
stick = df['stick_level2'].value_counts(dropna=False).sort_index()
stick
```

Out[185]:

```
S01     1396
S02     1761
S03     2355
S04     8700
S05     8799
S06    11298
S07    17386
S08    20320
S09    26499
S10    31973
Name: stick_level2, dtype: int64
```

```
In [181]:
```

```
sns.set()
plt.figure(figsize=(8,4))
width = 0.7
plt.rcParams['font.family']= ['Microsoft JhengHei']
plt.bar(stick.index, stick.values, width, label='人數')
plt.legend()
plt.xlabel('忠誠度')
```



13 客戶分群

G0最高-->G4最低

```
In [183]:
```

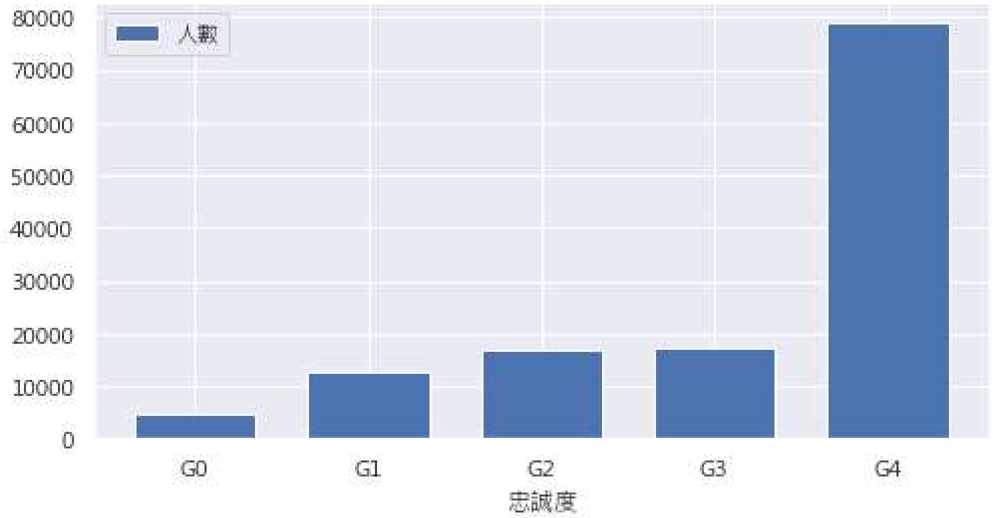
```
group = df['cust_group2'].value_counts(dropna=False).sort_index()
group
```

```
Out[183]:

G0     4745
G1    12647
G2    16917
G3    17292
G4    78886
Name: cust_group2, dtype: int64
```

In [184]:

```
plt.figure(figsize=(8,4))
width = 0.7
plt.rcParams['font.family']= ['Microsoft JhengHei']
plt.bar(group.index, group.values, width, label='人數')
plt.legend()
plt.xlabel('忠誠度')
sns.set()
```



14 總資產

In [190]:

```
df['TOTAL_AUM'].value_counts(dropna=False).sort_index()
```

Out[190]:

```
5.700000e+02        2
6.070500e+02        1
6.270000e+02        3
6.355500e+02        2
6.441000e+02        1
                ...
8.381121e+07        1
8.781562e+07        1
2.126112e+08        1
4.223700e+08        1
NaN             11786
Name: TOTAL_AUM, Length: 62882, dtype: int64
```

In [189]:

```
df['TOTAL_AUM'].describe()
```

Out[189]:

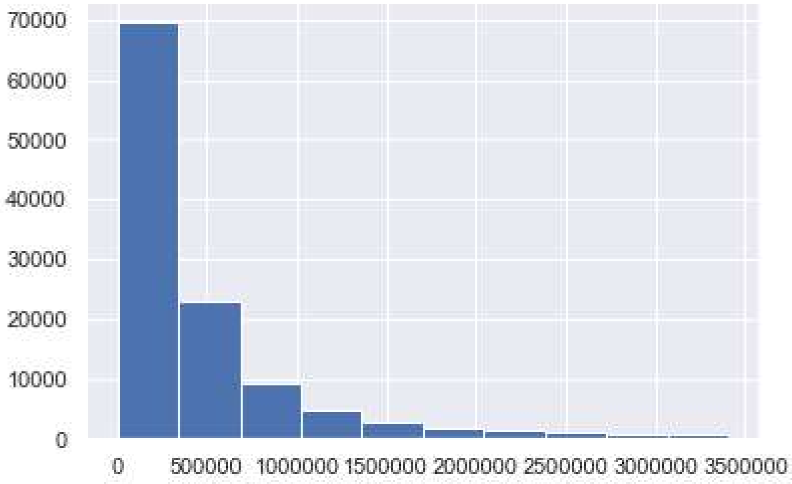
```
count    1.187010e+05
mean     6.592805e+05
std      2.175334e+06
min      5.700000e+02
25%      1.033695e+05
50%      2.618770e+05
75%      5.973980e+05
max      4.223700e+08
Name: TOTAL_AUM, dtype: float64
```

In [214]:

```
import copy
tmp2 = copy.deepcopy(df['TOTAL_AUM']).dropna()
plt.hist(sorted(tmp2)[:int(0.97*len(tmp2))])
# 去掉最後 3% 的總資產分配
```

Out[214]:

```
(array([69524., 23087.,  9115.,  4785.,  2773.,  1936.,  1451.,  1016.,
          835.,   617.]),
 array([5.700000e+02, 3.411070e+05, 6.816440e+05, 1.022181e+06,
        1.362718e+06, 1.703255e+06, 2.043792e+06, 2.384329e+06,
        2.724866e+06, 3.065403e+06, 3.405940e+06]),
 <a list of 10 Patch objects>)
```



In [ ]:

```

```