

套件匯入

In [1]:

```
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import calendar
import seaborn as sns
import datetime as dt
```

理賠資料

將兩張理賠表匯入

In [2]:

```
df1 = pd.read_excel("CLAIM_ACCT_FIN_1.xlsx")
df2 = pd.read_excel("CLAIM_ACCT_FIN_2.xlsx")
```

將兩張理賠表合併

In [3]:

```
df2.index = [120000+i for i in range(len(df2))]
frames = [df1, df2]
df = pd.concat(frames)
```

In [4]:

df

Out[4]:

	INJURED_RK	Claim_RK	Policy_RK	BundleSubtype2	illness_code	illness_desc	Diagno
0	14	4034	72224	5.N疾病醫療	A09	感染性胃腸炎及大腸炎	01.傳
1	14	62576	72224	4.C重大疾病	N19	腎衰竭	14.減
2	14	90970	72224	5.N疾病醫療	H26.9	白內障	
3	14	230243	72224	1.D身故給付	V99	運輸意外事故	20.勞
4	24	157920	91033	5.N疾病醫療	R19.7	腹瀉,未特定	18.症別
...	
234423	249733	30121	166605	6.N意外醫療	Y99.8	其他外因狀態	
234424	249733	86463	166605	5.N疾病醫療	D33	大腦及中樞神經系統其他部位之良性腫瘤	
234425	249733	109718	166605	6.N意外醫療	Z04.2	來院接受對工作意外後之檢查及觀察	21.影響
234426	249733	204841	166605	6.N意外醫療	V23	摩拖車騎士與汽車,小貨車或箱型車碰撞受傷	20.勞
234427	249795	189792	97319	5.N疾病醫療	K05	齒齦炎及牙周疾病	

234428 rows × 13 columns

在新的理賠表中做運算，產生新的資料欄

In [5]:

```
df3 = df.set_index(['POLICY_HOLDER_RK','Policy_RK']).sort_values(['POLICY_HOLDER_RK','Policy_RK','claim_settle_dt'])
df3['計數'] = 1
df3['累積理賠金額'] = df3.groupby('Policy_RK')['REIMBURSED_YR_TW'].transform('cumsum')
df3['初次理賠時間'] = df3.groupby('Policy_RK')['claim_settle_dt'].transform(min)
df3['累積理賠次數'] = df3.groupby('Policy_RK')['計數'].transform('cumsum')
df3 = df3.drop(columns=['計數'])
df3['結案後120天'] = df3['claim_settle_dt'] + dt.timedelta(days=120)
df3['結案後180天'] = df3['claim_settle_dt'] + dt.timedelta(days=180)
df3['結案後360天'] = df3['claim_settle_dt'] + dt.timedelta(days=360)
```

將處理完的理賠檔存檔

In [6]:

```
df3.to_excel('ready_CLAIM_ACCT_FIN.xlsx')
```

In [213]:

```
# df3 = df3.loc[~df3.index.duplicated(keep='last')]
# df3 = df3.groupby(level=df3.index.names)
# df3 = df3.last()
# df3
```

查看某位要保人的理賠資訊

In [7]:

```
df3.loc[130950].tail()
```

Out[7]:

	INJURED_RK	Claim_RK	BundleSubtype2	illness_code	illness_desc	DiagnosisCode_
Policy_RK						
105429	130950	160458	5.N疾病醫療	C57	其他及未明示 女性生殖器官 之惡性腫瘤	0
105429	130950	169555	5.N疾病醫療	C57	其他及未明示 女性生殖器官 之惡性腫瘤	0
105429	130950	178279	5.N疾病醫療	C57	其他及未明示 女性生殖器官 之惡性腫瘤	0
105429	130950	187467	5.N疾病醫療	C57	其他及未明示 女性生殖器官 之惡性腫瘤	0
105429	130950	222326	5.N疾病醫療	C57	其他及未明示 女性生殖器官 之惡性腫瘤	0

再購資料

In [8]:

```
df4 = pd.read_excel("COV_ACCT_FIN.xlsx")
```

In [9]:

```
df4.sort_values(['POLICY HOLDER_RK', 'Policy_RK', 'EFFECTIVE_DT'])
```

Out[9]:

	INSURED_RK	Policy_RK	RIDER_CD	payment_period	EFFECTIVE_DT	SHORT_NAME	p
13965	26347	56714	0	月繳	2017-10-29	PAR	
12107	23125	179172	1	月繳	2018-11-24	CAB	
12108	23125	179173	0	月繳	2018-11-24	NPBBR	
12109	23125	179173	0	月繳	2018-11-24	NAI	
12110	23125	179173	0	月繳	2018-11-24	HIR	
...	
134471	251945	177625	1	躉繳保費	2018-08-31	BVA	
79401	134583	90488	0	年繳	2017-08-31	N1TR	
79402	134583	90488	1	年繳	2017-08-31	NPHI	
79555	134869	90491	0	年繳	2017-09-08	N1TR	
79556	134869	90491	1	年繳	2017-09-08	NPHI	

134472 rows × 11 columns



理賠再購合併

要vs要 合併

In [10]:

```
df3['再購(120天)'] = 0
for i in range(len(df3.index)):
    time_list = df4[df4['POLICY_HOLDER_RK']==df3.index[i][0]]['EFFECTIVE_DT'].unique()
    for time in time_list:
        if (df3.iloc[i,6] < time) and (time < df3.iloc[i,14]):
            df3.iloc[i,17] = 1
            break
    if i % 10000 == 0:
        print("第{}筆資料已標記!!".format(i))
```

第0筆資料已標記!!
第10000筆資料已標記!!
第20000筆資料已標記!!
第30000筆資料已標記!!
第40000筆資料已標記!!
第50000筆資料已標記!!
第60000筆資料已標記!!
第70000筆資料已標記!!
第80000筆資料已標記!!
第90000筆資料已標記!!
第100000筆資料已標記!!
第110000筆資料已標記!!
第120000筆資料已標記!!
第130000筆資料已標記!!
第140000筆資料已標記!!
第150000筆資料已標記!!
第160000筆資料已標記!!
第170000筆資料已標記!!
第180000筆資料已標記!!
第190000筆資料已標記!!
第200000筆資料已標記!!
第210000筆資料已標記!!
第220000筆資料已標記!!
第230000筆資料已標記!!

In [11]:

```
df3['再購(180天)'] = 0
for i in range(len(df3.index)):
    time_list = df4[df4['POLICY_HOLDER_RK']==df3.index[i][0]]['EFFECTIVE_DT'].unique()
    for time in time_list:
        if (df3.iloc[i,6] < time) and (time < df3.iloc[i,15]):
            df3.iloc[i,18] = 1
            break
    if i % 10000 == 0:
        print("第{}筆資料已標記!!".format(i))
```

第0筆資料已標記!!
第10000筆資料已標記!!
第20000筆資料已標記!!
第30000筆資料已標記!!
第40000筆資料已標記!!
第50000筆資料已標記!!
第60000筆資料已標記!!
第70000筆資料已標記!!
第80000筆資料已標記!!
第90000筆資料已標記!!
第100000筆資料已標記!!
第110000筆資料已標記!!
第120000筆資料已標記!!
第130000筆資料已標記!!
第140000筆資料已標記!!
第150000筆資料已標記!!
第160000筆資料已標記!!
第170000筆資料已標記!!
第180000筆資料已標記!!
第190000筆資料已標記!!
第200000筆資料已標記!!
第210000筆資料已標記!!
第220000筆資料已標記!!
第230000筆資料已標記!!

In [12]:

```
df3['再購(360天)'] = 0
for i in range(len(df3.index)):
    time_list = df4[df4['POLICY_HOLDER_RK']==df3.index[i][0]]['EFFECTIVE_DT'].unique()
    for time in time_list:
        if (df3.iloc[i,6] < time) and (time < df3.iloc[i,16]):
            df3.iloc[i,19] = 1
            break
    if i % 10000 == 0:
        print("第{}筆資料已標記!!".format(i))
```

第0筆資料已標記!!
第10000筆資料已標記!!
第20000筆資料已標記!!
第30000筆資料已標記!!
第40000筆資料已標記!!
第50000筆資料已標記!!
第60000筆資料已標記!!
第70000筆資料已標記!!
第80000筆資料已標記!!
第90000筆資料已標記!!
第100000筆資料已標記!!
第110000筆資料已標記!!
第120000筆資料已標記!!
第130000筆資料已標記!!
第140000筆資料已標記!!
第150000筆資料已標記!!
第160000筆資料已標記!!
第170000筆資料已標記!!
第180000筆資料已標記!!
第190000筆資料已標記!!
第200000筆資料已標記!!
第210000筆資料已標記!!
第220000筆資料已標記!!
第230000筆資料已標記!!

理賠再購合併檔案輸出

In [13]:

```
df3.to_excel('理賠再購合併.xlsx')
```

In [14]:

df3

Out[14]:

		INJURED_RK	Claim_RK	BundleSubtype2	illness_code	illness
POLICY HOLDER_RK	Policy_RK					
3	49082	4324	104018	5.N疾病醫療	D36	其他及部位良
4	56714	26347	90482	6.N意外醫療	W18	其他滑倒或絆
	56714	26347	184848	6.N意外醫療	W18	其他滑倒或絆
5	84973	93160	183589	5.N疾病醫療	D24	乳房良
6	65407	94922	175992	5.N疾病醫療	L03.90	蜂窩
...	
251945	102196	123213	112437	6.N意外醫療	V23	摩拖車、汽車，或箱型
251947	16872	86041	179452	6.N意外醫療	Y99.8	其他外
251949	100554	161848	113170	6.N意外醫療	V99	運輸意
251951	96970	134869	174868	5.N疾病醫療	D36	其他及部位良
251954	74765	156320	136040	4.C重大疾病	I25.1	自體的脈粥樣

234428 rows × 20 columns



理賠再購檔.客戶屬性檔合併

In [16]:

```
df4 = pd.read_excel('CUST_PROPERTY_FIN_1.xlsx')
df5 = pd.read_excel('CUST_PROPERTY_FIN_2.xlsx')
```

In [17]:

```
df5.index = [65000+i for i in range(len(df5))]
frames = [df4,df5]
df6 = pd.concat(frames)
```

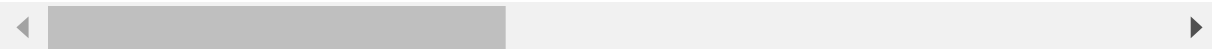
In [18]:

df6

Out[18]:

	CUST_RK	ternure_m	recency_m	SIN	SIN_his	REG	REG_his	ILP	ILP_his	AHa	...
0	24	201	201	0	0	1	1	0	0	0	...
1	152	175	175	0	0	1	1	0	0	1	...
2	155	227	227	0	0	1	1	0	0	1	...
3	158	165	165	0	0	0	0	0	0	1	...
4	204	181	181	0	0	1	1	0	0	1	...
...
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	...

130487 rows × 29 columns



In [19]:

```
df7 = df3.merge(right=df6, how='left', left_on='POLICY HOLDER_RK',right_on='CUST_RK')
```

In [21]:

```
df7.index= df3.index
```

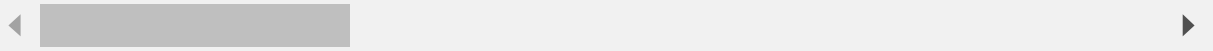
In [22]:

df7

Out[22]:

		INJURED_RK	Claim_RK	BundleSubtype2	illness_code	illness
POLICY HOLDER_RK	Policy_RK					
3	49082	4324	104018	5.N疾病醫療	D36	其他及部位良
4	56714	26347	90482	6.N意外醫療	W18	其他滑倒或絆
	56714	26347	184848	6.N意外醫療	W18	其他滑倒或絆
5	84973	93160	183589	5.N疾病醫療	D24	乳房良
6	65407	94922	175992	5.N疾病醫療	L03.90	蜂窩
...	
251945	102196	123213	112437	6.N意外醫療	V23	摩托車、汽車、或箱型
251947	16872	86041	179452	6.N意外醫療	Y99.8	其他外
251949	100554	161848	113170	6.N意外醫療	V99	運輸意
251951	96970	134869	174868	5.N疾病醫療	D36	其他及部位良
251954	74765	156320	136040	4.C重大疾病	I25.1	自體的脈粥樣

234428 rows × 49 columns



In [23]:

```
df8 = df7[['INJURED_RK', 'Claim_RK', 'BundleSubtype2', 'illness_code',
            'illness_desc', 'DiagnosisCode_DESC', 'claim_settle_dt',
            'REIMBURSED_YR_TW', 'INSURED_RK', 'MATURITY_BENEFICIARY_RK',
            'DEATH_BENEFICIARY_RK', '累積理賠金額', '初次理賠時間', '累積理賠次數', '結案後120天',
            '結案後180天', '結案後360天', 'CUST_RK',
            'ternure_m', 'recency_m', 'SIN', 'SIN_his', 'REG', 'REG_his', 'ILP',
            'ILP_his', 'AHa', 'AHa_his', 'AHb', 'AHb_his', 'AHc', 'AHc_his', 'AHd',
            'AHd_his', 'VIP_CLASS', 'VIP', 'WEALTH_LEVEL', 'CLIENT_MARITAL',
            'CLIENT_INCOME', 'DIGI_FLG', 'TOPCARD', 'GENDER', 'stick_level2',
            'cust_group2', 'TOTAL_AUM', 'INSURED_DOB', '再購(120天)', '再購(180天)', '再購(360天)']]
```

In [24]:

df8

Out[24]:

		INJURED_RK	Claim_RK	BundleSubtype2	illness_code	illness
POLICY HOLDER_RK	Policy_RK					
3	49082	4324	104018	5.N疾病醫療	D36	其他及部位良
4	56714	26347	90482	6.N意外醫療	W18	其他滑倒或絆
	56714	26347	184848	6.N意外醫療	W18	其他滑倒或絆
5	84973	93160	183589	5.N疾病醫療	D24	乳房良
6	65407	94922	175992	5.N疾病醫療	L03.90	蜂窩
...	
251945	102196	123213	112437	6.N意外醫療	V23	摩拖車、汽車，或箱型
251947	16872	86041	179452	6.N意外醫療	Y99.8	其他外
251949	100554	161848	113170	6.N意外醫療	V99	運輸意
251951	96970	134869	174868	5.N疾病醫療	D36	其他及部位良
251954	74765	156320	136040	4.C重大疾病	I25.1	自體的脈粥樣

234428 rows × 49 columns

In [25]:

df8.to_excel('理賠再購屬性合併.xlsx')