

# 南山理賠服務及理賠再購預測

財金二 施詠瀚

# PTT Insurance Crawl

Using keyword “南山” to crawl data from ptt

Three categories : Author, Title, URL

	author	title	url
0	basloilu	[意外/醫療] 32歲女 新保單檢視/南山	/bbs/Insurance/M.1556985859.A.07F.html
1	zack2808	[心得] 三商人壽 南山人壽 理賠超快速	/bbs/Insurance/M.1556867843.A.A05.html
2	WHungYun	[險種] 南山乙式車險請教	/bbs/Insurance/M.1556842393.A.5FC.html
3	libby2	[險種] 57歲女 新保單規劃/南山	/bbs/Insurance/M.1556678712.A.551.html
4	linalain15	[險種] 31歲女 既有保單檢視/南山	/bbs/Insurance/M.1556269857.A.219.html

# PTT Insurance Crawl

Using URL & Title to collect data

[險種] 意外險/醫療險 32女 南山新保單檢視

商品名稱	[英文代號]	保額/計畫	期間	保費
南山人壽樂健康祥定期健康保險	10TDD	60 萬	10年	2274
南山人壽新人身意外傷害保險附約	PAR	200 萬	1年	2340
南山人壽新傷害醫療保險金附加條款	AMN	3 萬	1年	810
南山人壽意外傷害醫療日額給付附加條款	DHI	1000 元	1年	540
南山人壽住院費用給付保險附約	HIR	1000 元	1年	1930
住院費用給付保險附約居家療養附加條款	HR	500 元	1年	940
南山人壽好醫靠住院醫療健康保險附約甲型	NHS	1 份	1年	5400

# PTT Insurance Crawl

利用自定義字典手動新增各個保險合約後計算詞頻

```
alldata = dict(allcounter.most_common())  
print(allcounter.most_common(250))
```

## Result :

南山人壽住院醫療保險附約 : 319

PAR / 南山人壽新人身意外傷害保險附約 : 276

HIR / 南山人壽住院醫療保險附約 : 256

NPHI / 南山人壽新終身醫療保險 : 224

AMN / 南山人壽新傷害醫療保險金附加條款 : 217

# PTT Insurance Crawl

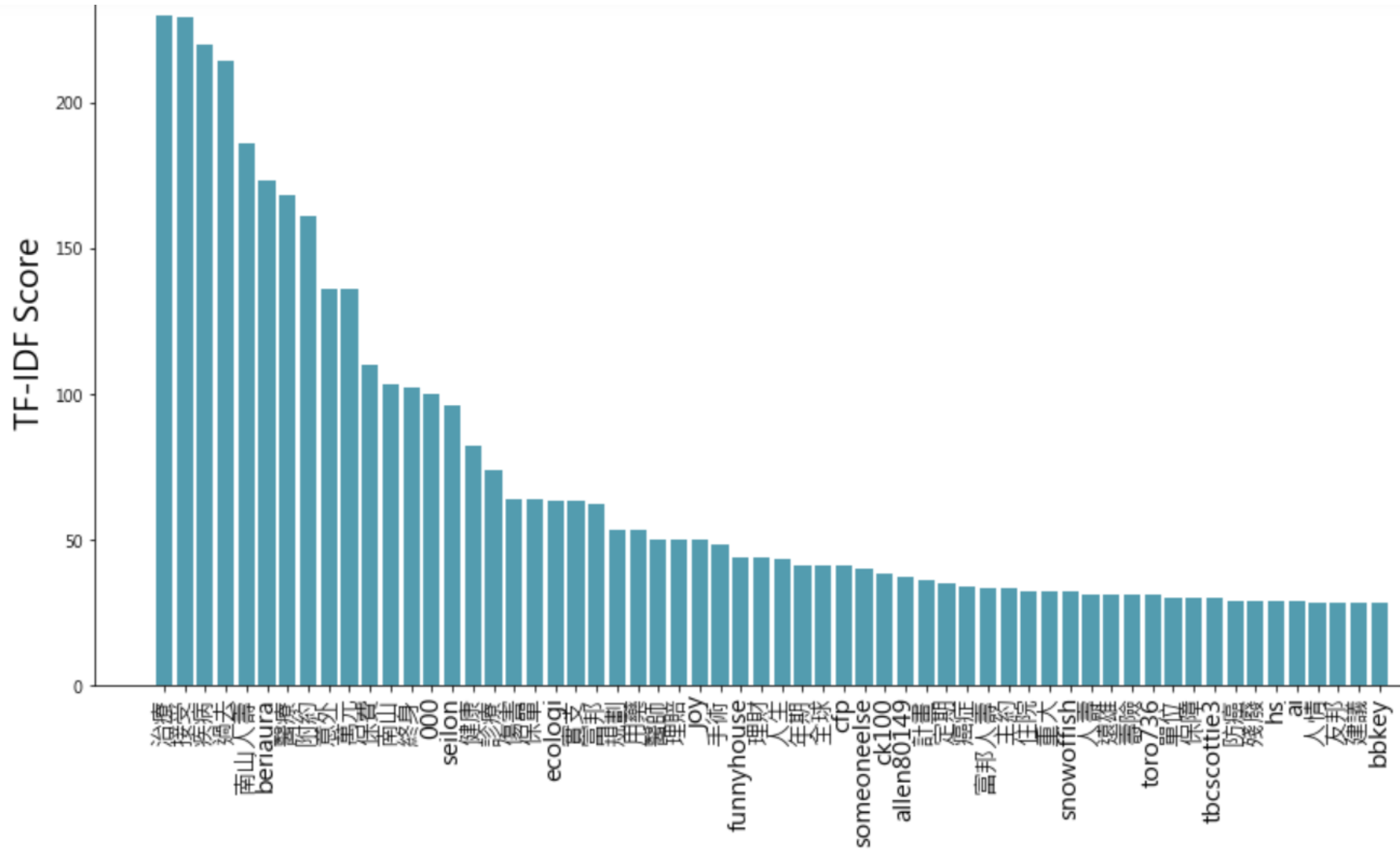
## Using Token to train Word2Vec Model

```
model2 = word2vec.Word2Vec(token, min_count=5, size=300, sg=1, hs=1, negative=5, iter=30, window  
wordlist = ['南山', '南山人壽', '理賠', '業務', '人情', '保險業']  
indexes = model2.wv.most_similar_cosmul(i)
```

```
['時間', '險種', '新舊', '保單', '健檢時間', '南山人壽', '國泰', '富邦', '三商', '男保單']  
['南山人壽住院醫療保險附約', '不分紅', '南山人壽新人身意外傷害保險附約', '新福氣', '護您', '20DDC', 'HSER',  
'南山人壽傷害保險附約', '康祥', '南山人壽新終身醫療保險']  
['融通', '糾紛', '自負額', '爭議', '嚴格', '速度', '要件', '刁難', '項', '申請']  
['開給', '業績', '傭金', '阿姨', '留言', '給業務', '嫌', '表示', '拒絕', '方便']  
['壓力', '口袋', '基', '家保險', '權益', '悲劇', '一輩子', 'genius111', '無誤', 'B2']  
['好友', '資格', '同學', '信任', '朋友', '傭金', '從事', '當業務', '薪資', '害怕']
```

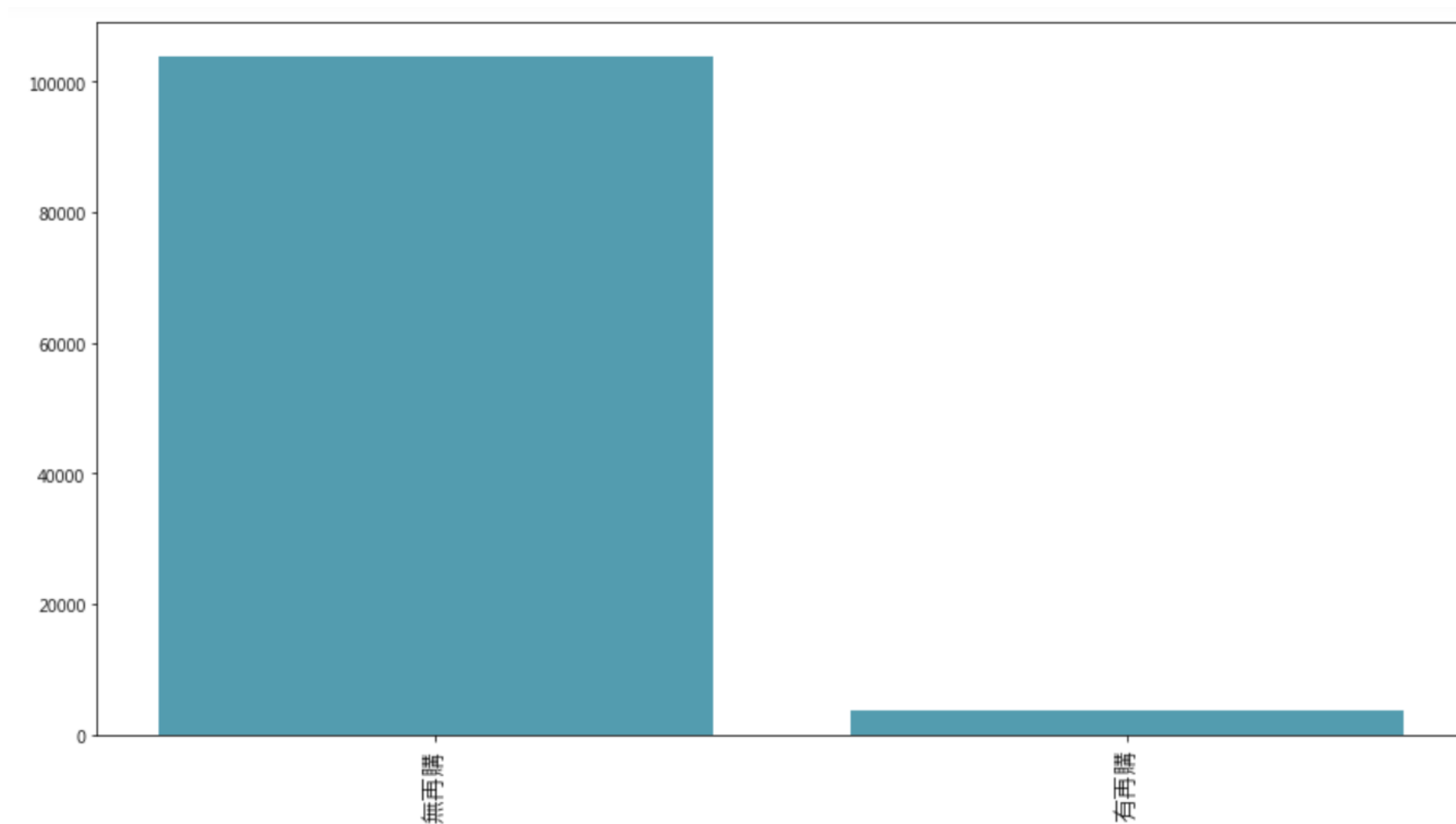
# PTT Insurance Crawl

## Using TF-IDF Score to get every article's top 35 features



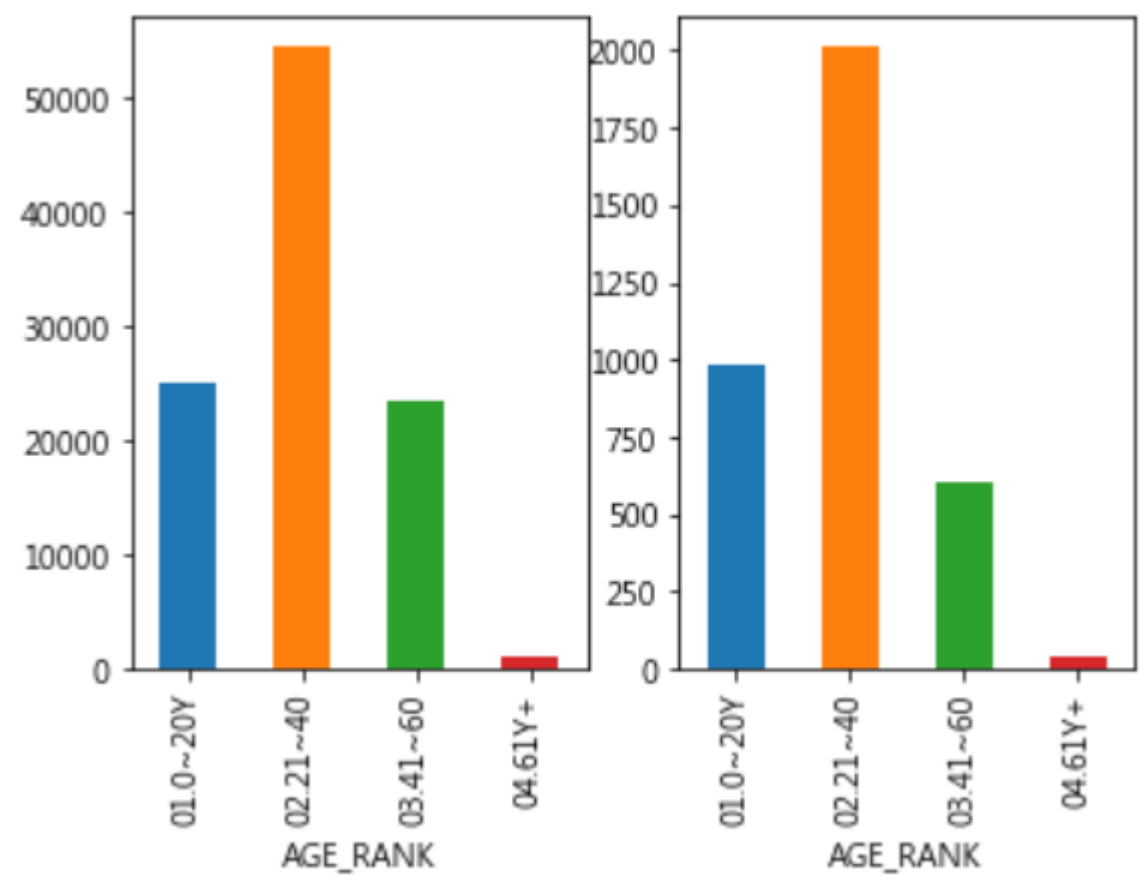
# 南山理賠服務及理賠再購資料探索

是否有再購人數

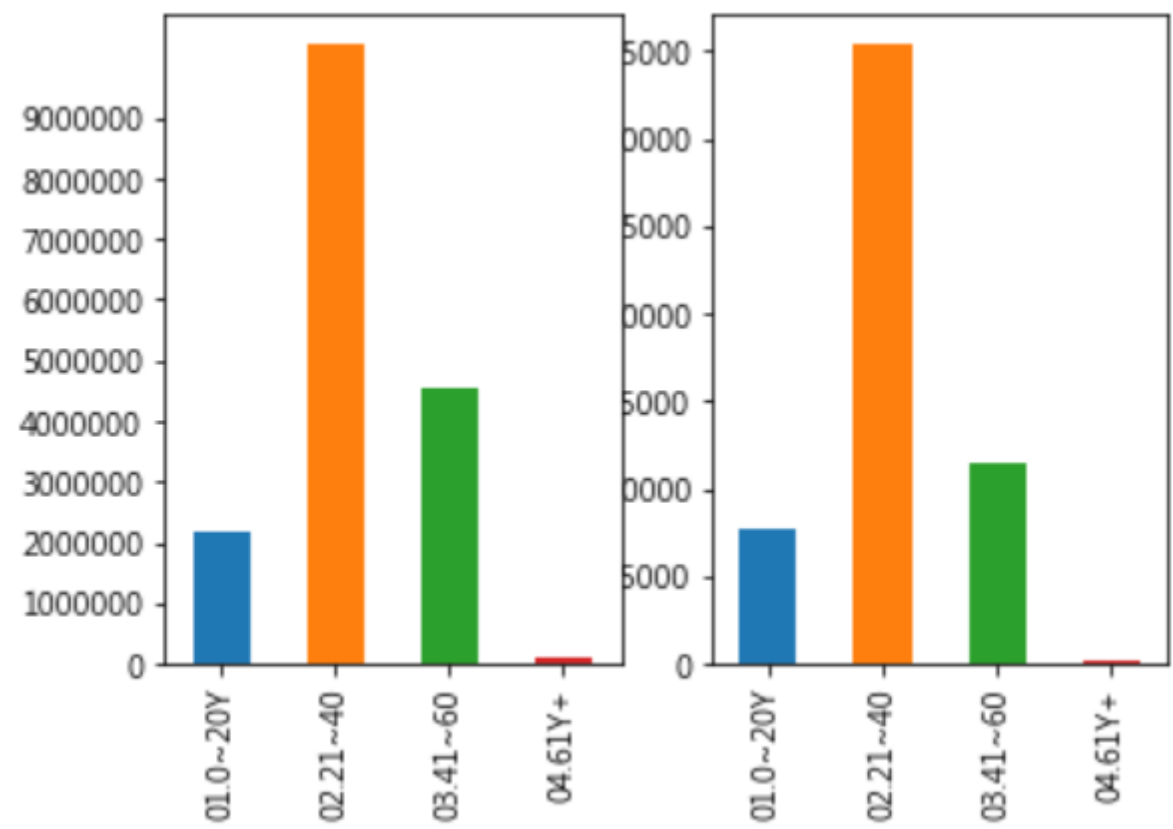


# 南山理賠服務及理賠再購資料探索

有/無再購年齡層分布



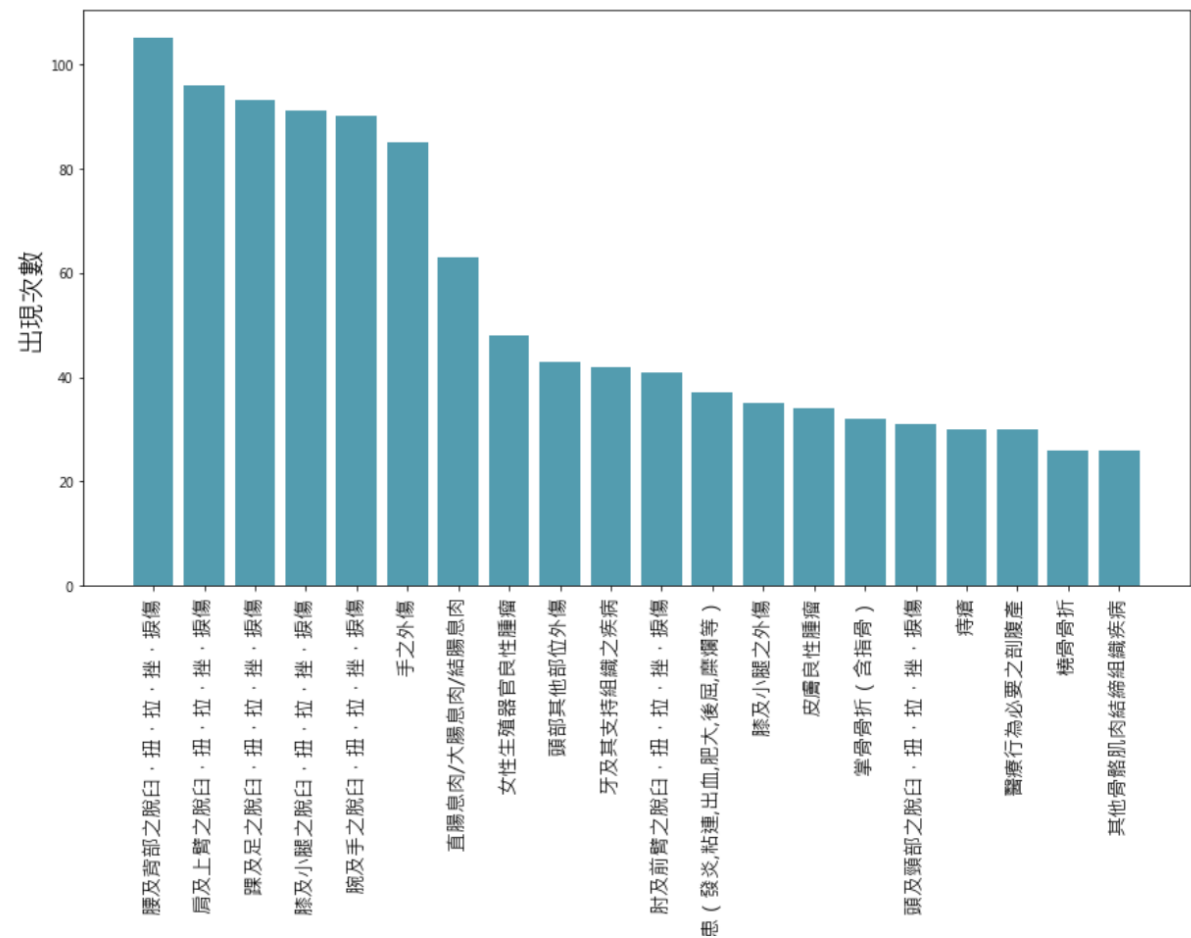
有/無再購五年內總理賠金差異



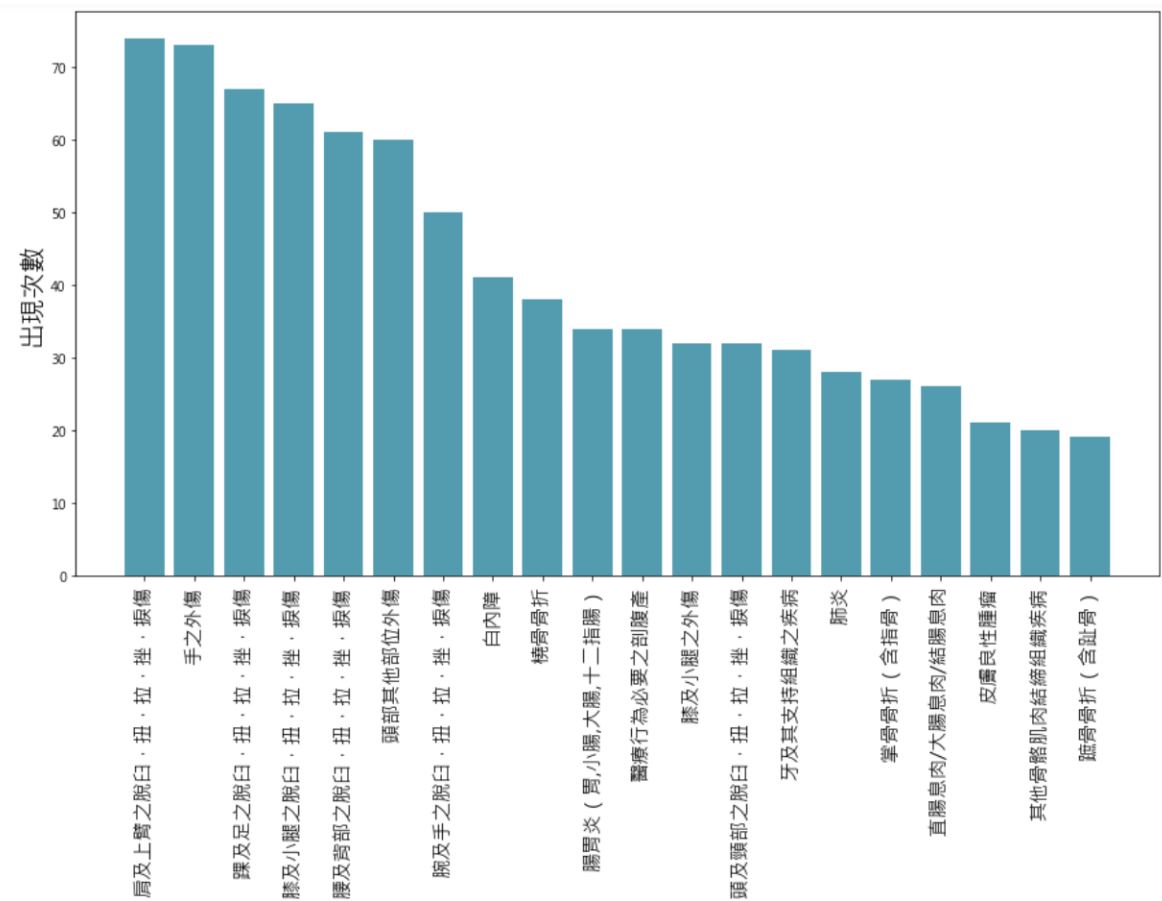


# 南山理賠服務及理賠再購資料探索

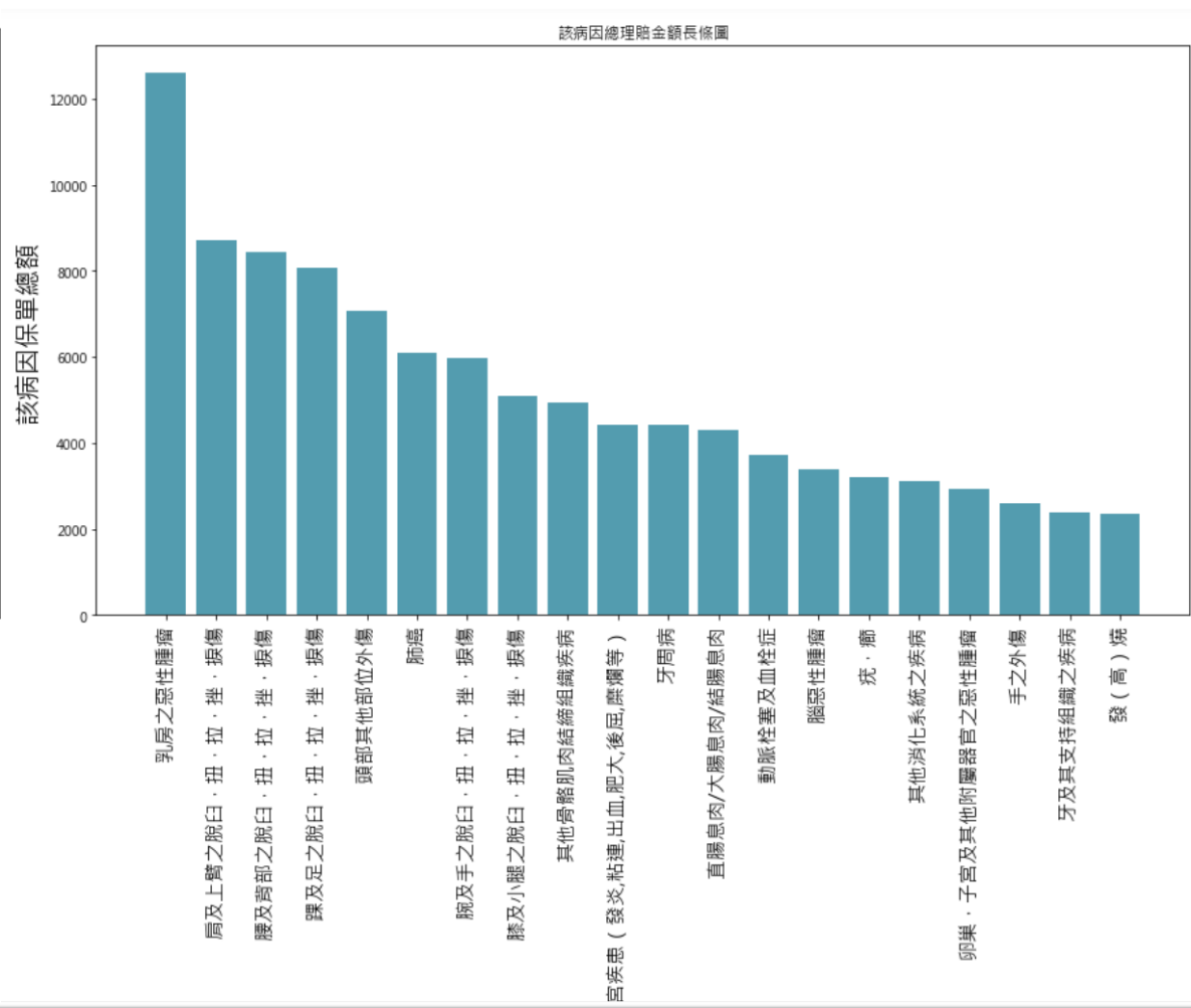
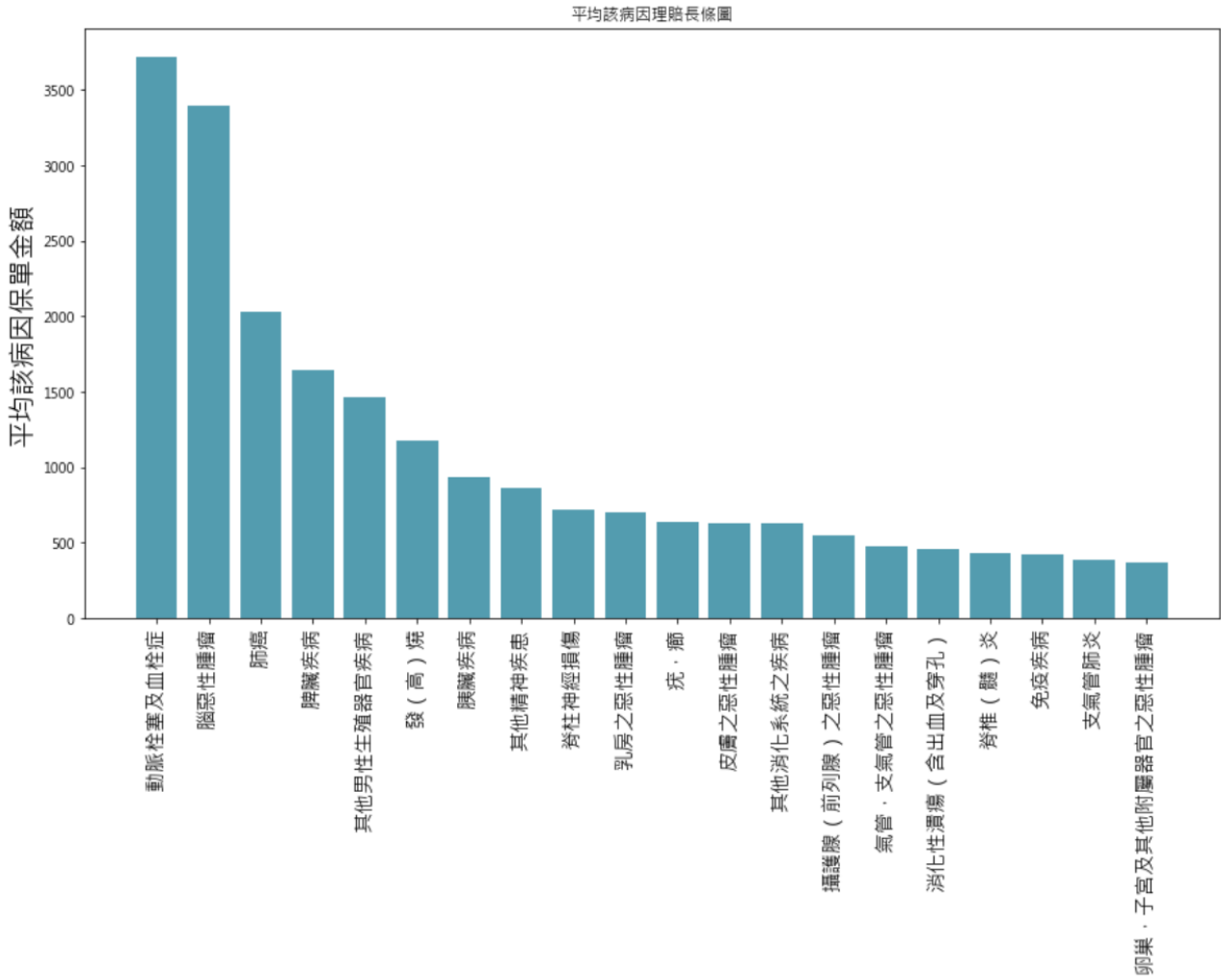
20~40前20病因



20~40以外前20病因



# 南山理賠服務及理賠再購資料探索



# 南山理賠服務及理賠再購資料探索

```
topincomedf['Income'] = + topincomedf['Sum_afyp_K'] - topincomedf['REIMB_AMT_5Y_K']
```

已購買保單總年化保費(單位:千)

五年內總理賠金額(單位:千)

## 日常型

illness_code_desc	Income
踝及足之脫臼，扭，拉，挫，擦傷	48484
直腸息肉/大腸息肉/結腸息肉	41960
皮膚良性腫瘤	34172
手之外傷	33431
肩及上臂之脫臼，扭，拉，挫，擦傷	25732
頭部其他部位外傷	25364

## 重症型

疔，癰	-2244
其他消化系統之疾病	-2424
動脈栓塞及血栓症	-2739
腦惡性腫瘤	-3399
乳房之惡性腫瘤	-5158
肺癌	-5971



# 南山理賠服務及理賠再購資料探索

## 地區分類

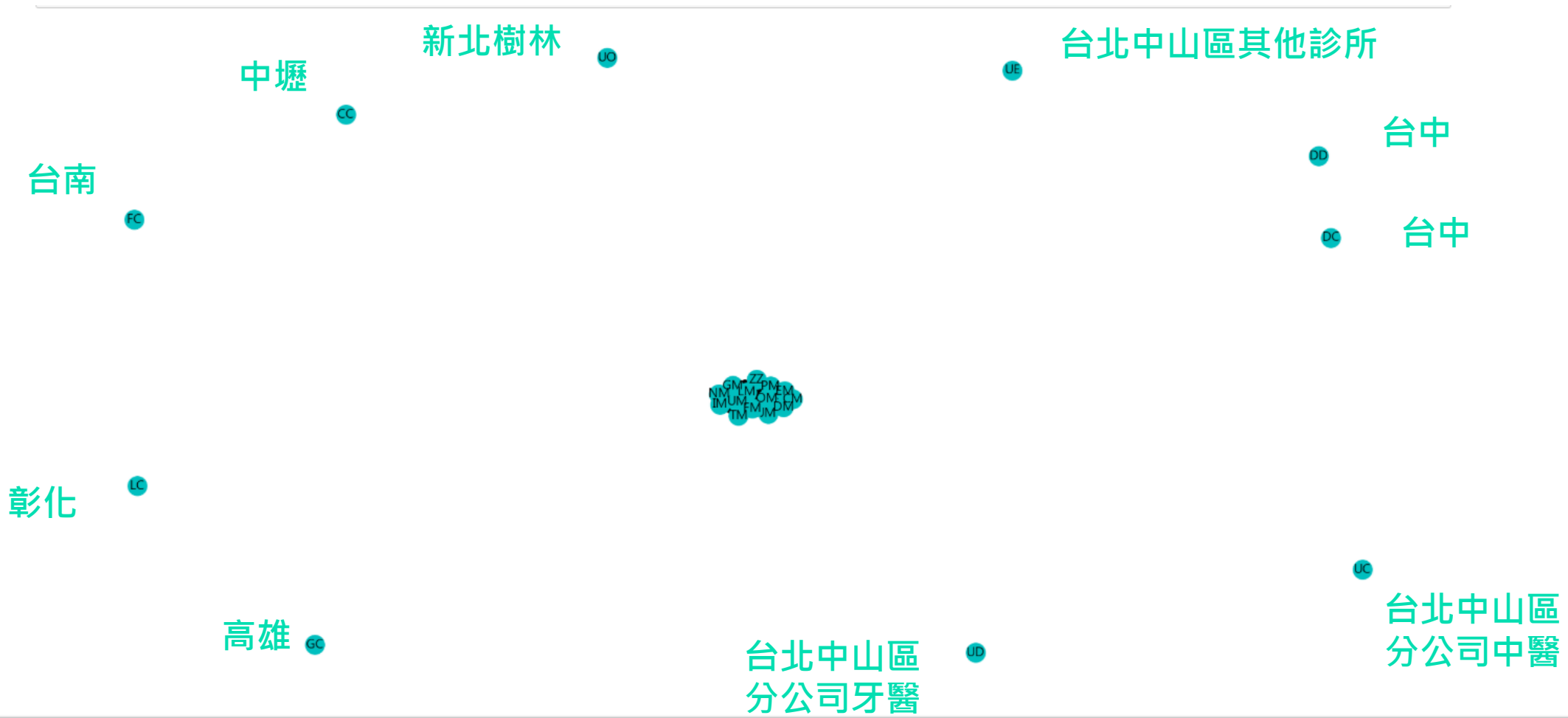
每個地區傷病frequency做Apriori

```
minsup = 0.2  
whole = apriori(graph100, min_support = minsup, use_colnames=True)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(CM)	(DM)	0.704600	0.864407	0.687651	0.975945	1.129034	0.078590	5.636804
1	(EM)	(CM)	0.733656	0.704600	0.634383	0.864686	1.227201	0.117448	2.183074
2	(CM)	(EM)	0.704600	0.733656	0.634383	0.900344	1.227201	0.117448	2.672623
3	(FM)	(CM)	0.702179	0.704600	0.607748	0.865517	1.228380	0.112992	2.196561
4	(CM)	(FM)	0.704600	0.702179	0.607748	0.862543	1.228380	0.112992	2.166646

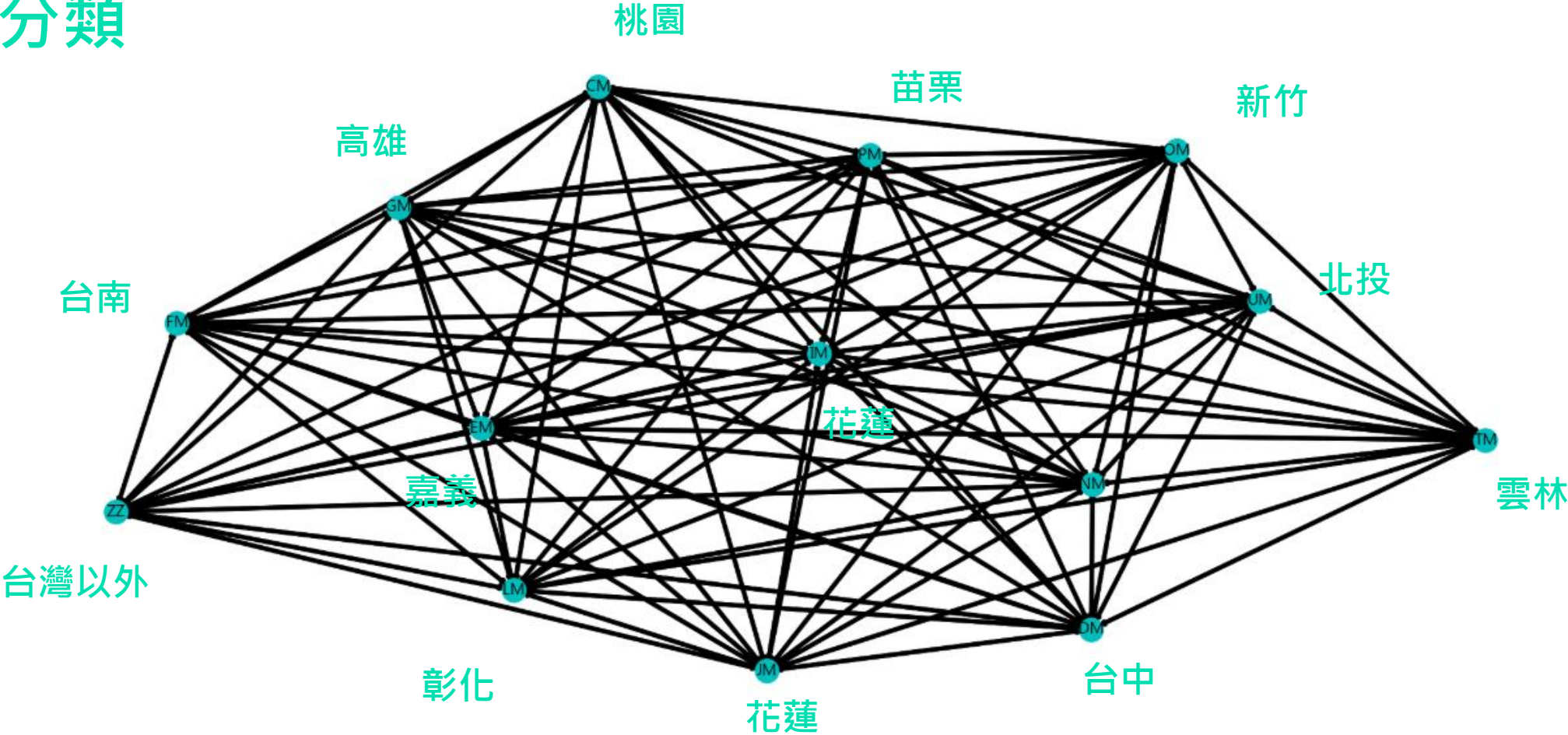
# 南山理賠服務及理賠再購資料探索

## 地區分類



# 南山理賠服務及理賠再購資料探索

## 地區分類



# 南山理賠服務及理賠再購資料探索

## Decision Tree

先將非數值資料轉換成數值資料

	TENURE	REIMB_AMT_K	CountNo	REIMB_AMT_5Y_K	RECENCY_YEAR	CountCovNo	Sum_afyp_K	illness_code	SEX	AGE_CATAGORY	REBUY
0	16	7	0.0	0.0	16	1.0	18.0	363	0	1	0.0
1	16	18	3.0	71.0	16	2.0	23.0	146	1	2	0.0
2	24	5	0.0	0.0	24	3.0	44.0	158	1	1	0.0
3	28	0	5.0	32.0	28	3.0	27.0	257	1	1	0.0
4	28	2	7.0	133.0	27	4.0	44.0	377	1	1	0.0
5	19	25	1.0	1.0	19	5.0	57.0	202	1	2	0.0
6	11	12	0.0	0.0	11	4.0	26.0	50	1	2	0.0
7	31	2	0.0	0.0	31	3.0	70.0	146	1	1	0.0
8	29	6	3.0	23.0	29	3.0	17.0	70	1	1	0.0
9	15	4	0.0	0.0	15	3.0	3.0	249	1	2	0.0
10	24	52	5.0	157.0	24	8.0	2308.0	166	1	1	0.0

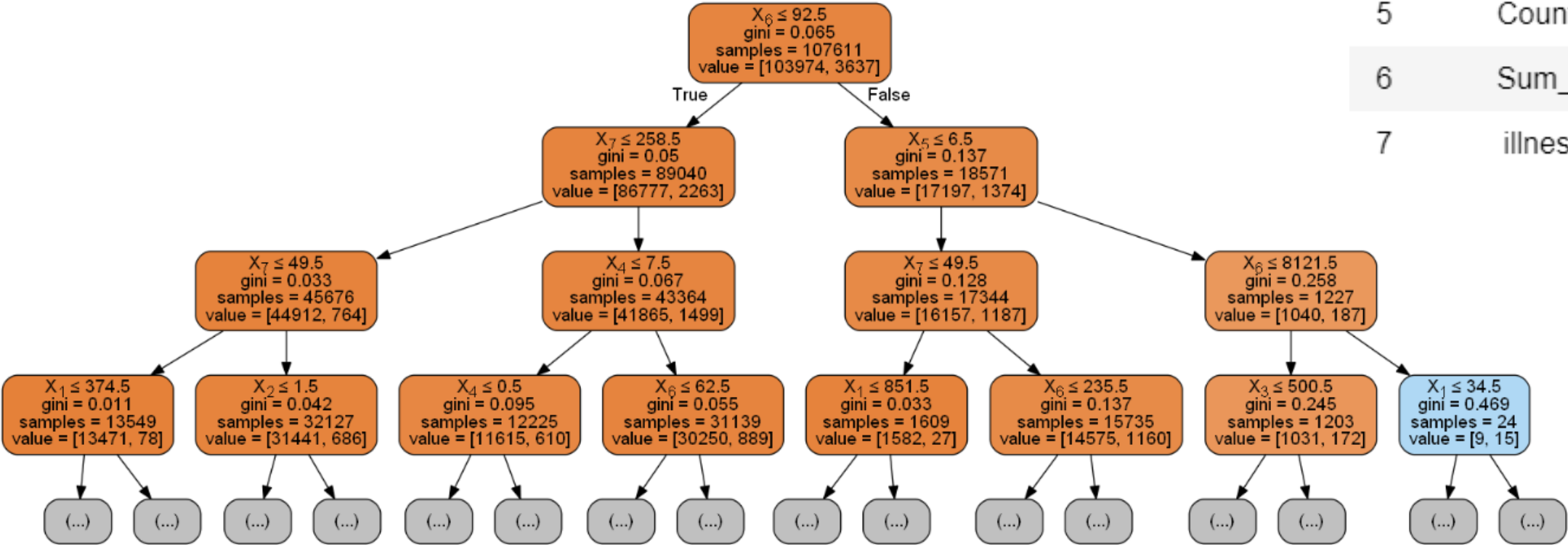


# 南山理賠服務及理賠再購資料探索

## Decision Tree

### 主要特徵

4	REGENCY_YEAR
5	CountCovNo
6	Sum_afyp_K
7	illness_code



# 南山理賠服務及理賠再購資料探索

## Decision Tree

```
[[29904 1330]
 [ 973   77]]
      precision    recall  f1-score   support

     0.0         0.97     0.96     0.96     31234
     1.0         0.05     0.07     0.06      1050

 avg / total         0.94     0.93     0.93     32284
```

Accuracy : 92.8664353859497

### 預測問題

- 再購與沒有再購的比利差太多，機器在預測時可以很簡單的預測到正確答案。

# 南山理賠服務及理賠再購資料探索

## 小結:

理賠服務可以朝現在的發展型態前進，持續推出兩種主流商品

- 符合現在主流趨勢是日常型保險服務
- 推廣雖利潤為負卻可以推廣企業形象的重症型商品
- 主要客群：20~40歲的族群

地區差異 是否可以差異化提供保險商品

- 除了序號單一的資料被排除在外，其他縣市並無太大差距

預測多項條件預測是否再購

- 應使用更多模型進行參數調整等等問題
- 資料有無再購比例差距過大