職能發展學院112年下期Python應用實戰班期末專題

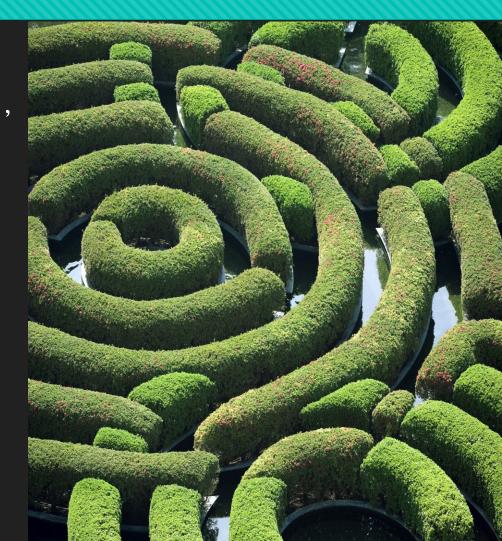
信用卡消費樣態

大綱

- 專案說明
- ○專案時程
- ○資料應用流程
 - > 蒐集資料
 - > 資料預處理
 - > 資料分析
 - > 資料解釋
- ○影片演示

專案說明

- 感於近年來受疫情影響,國人消費形態及支付方式改變,網路交易及外送盛行,希望藉由國人消費樣態分析作為發卡及收單機構推廣特約商店及行銷推廣之參考,促使其提供之牛肉能更貼近國人使用信用卡消費之需求,並預估下月份之信用卡消費金額。
- 資料來源:財團法人聯合信用卡處理中心 OPEN API
- 專案範圍:2014年1月~2023年9月
- 運用工具:Python JupyterNotebook dash



專案時程



- ○尋找組員
- ○專案小組成立

- 資料搜集
- 專案討論

- 資料清理、分析
- 程式撰寫

- 影片錄製
- 簡報準備
- 專案完成

資料應用流程



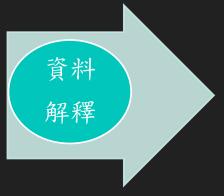
擷取公開網站 資料



資料清理 檢視資料之<u>缺</u> 失值、異常值 及重覆值



使用統計方法 和視覺化工具 來瞭解資料的 分佈、相關性 和模式。



解釋分析結果準備報告檔案

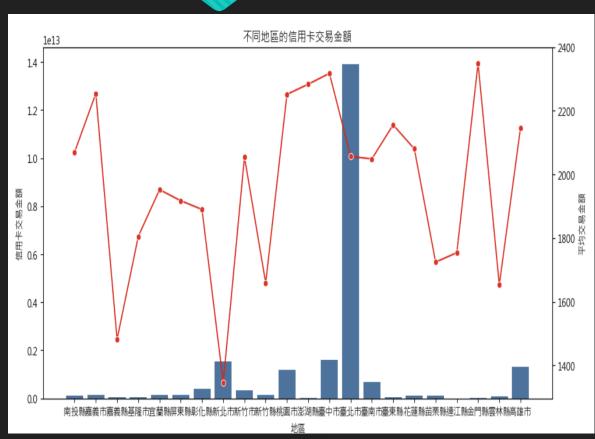
資料應用流程一蒐集資料

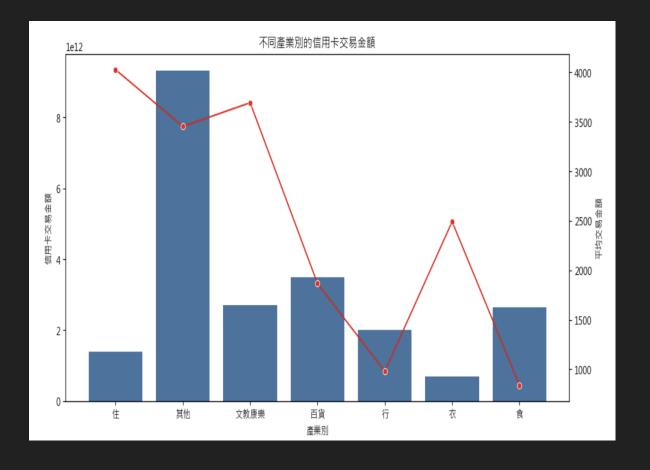
資料下載來源:聯合信用卡處理中心 OPEN API

```
#---兩性X各年齡層消費---#
for A in industry:
   for B in area:
       for C in sex:
           age url = f"https://bas.nccc.com.tw/nccc-nop/OpenAPI/C11/GenderAgeGroup/{B}/{A}/{C}
           response age = requests.request("GET", age url)
           if len(response age.text) == 0:
               continue
           folder_path = "./datasource/age/"
           file_name = f"age{B}_{A}_{C}.csv"
           file_path = os.path.join(folder_path, file_name)
           with open(file path, "wb") as file:
               file.write(response_age.content)
print("年齡層消費資料讀取成功")
```

資料應用流程--資料預處理(1)

觀察下列長條圖依地區別之消費樣態, 六都交易金額佔總交易筆數92.41%, 產業別屬其他類佔總交易筆數24.30%, 因其他類無明確項目, 故僅就六都及食、衣、住、行、百貨及文教康樂類進行分析。



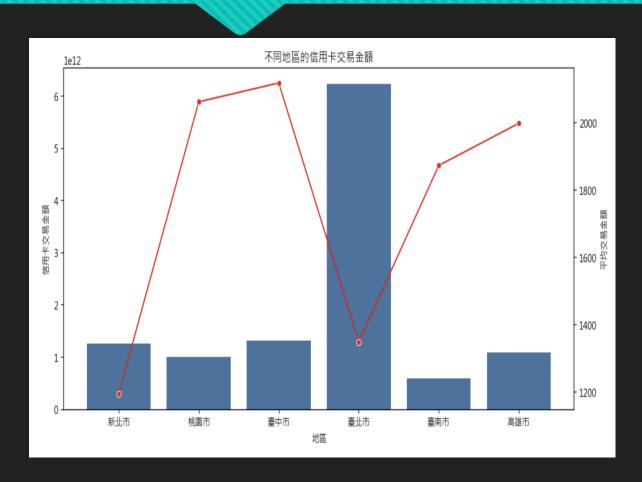


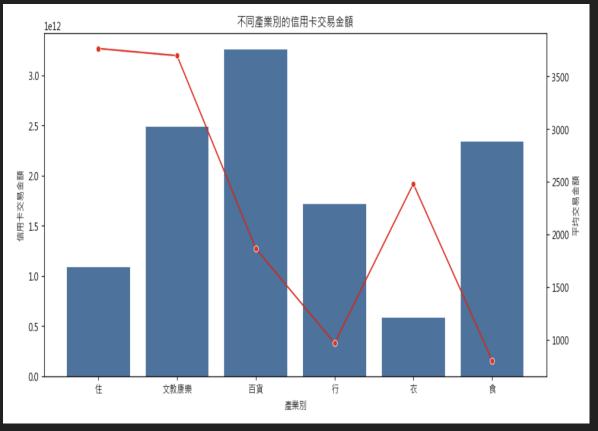
六都(不含其他類)佔總交易筆數: 68.71%

六都佔總交易筆數: 92.41% 其他類佔總交易筆數:24.30%

資料應用流程--資料預處理(2)

六都及食、衣、住、行、百貨及文教康樂類資料





資料應用流程--資料預處理(3)

○ 資料處理前

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500192 entries, 0 to 500191
Data columns (total 7 columns):
```

	•	•	
#	Column	Non-Null Count [Otype
0	年月	500192 non-null	int64
1	地區	500192 non-null	int64
2	產業別	500192 non-null	object
3	性別	500192 non-null	int64
4	年齡層	500192 non-null	object
5	信用卡交易筆數	500192 non-n	ull int64
6	信用卡交易金額[[新台幣] 500192 nor	n-null int64
d±v.n.	oc. in+64/E)	object(2)	

dtypes: int64(5), object(2)

memory usage: 26.7+ MB

○ 資料處理後

```
<class 'pandas.core.frame.DataFrame'>
Index: 117936 entries, 0 to 137591
Data columns (total 8 columns):
```

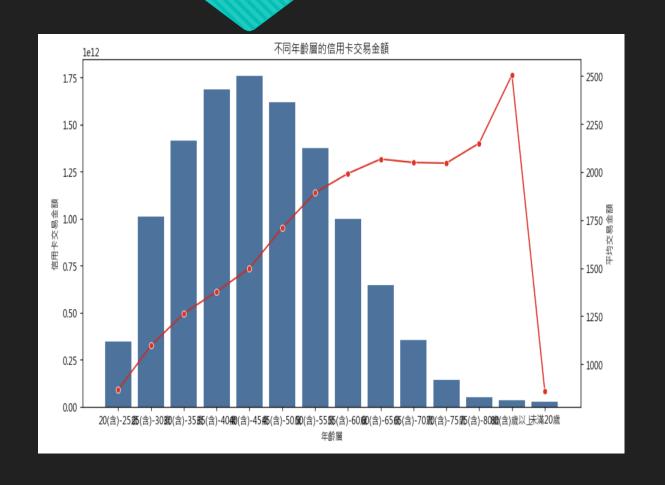
#	Column	Non-Null Count Dtype
0	年	117936 non-null int64
1	月	117936 non-null int64
2	地區	117936 non-null object
3	產業別	117936 non-null object
4	性別	117936 non-null object
5	年齢層	117936 non-null object
6	信用卡交易筆數	117936 non-null int64
7	信用卡交易金額	[新台幣] 117936 non-null int6

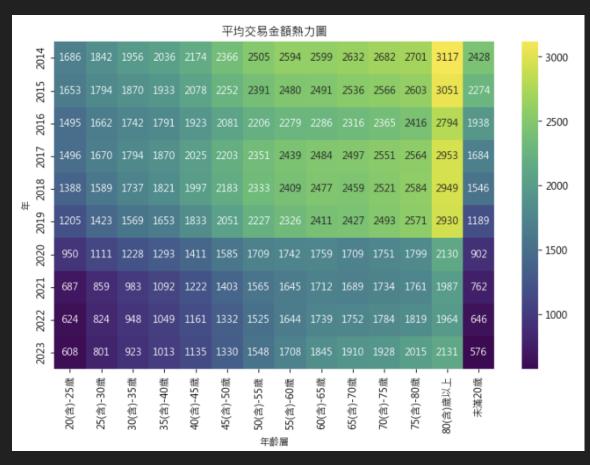
dtypes: int64(4), object(4)

memory usage: 8.1+ MB

資料應用流程--資料分析(1)

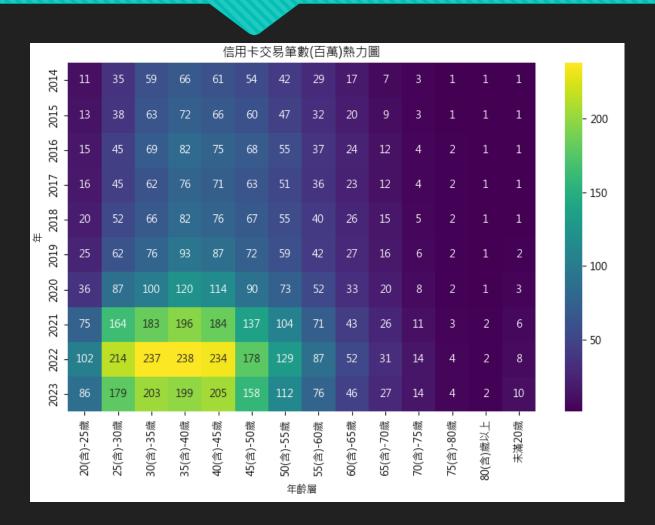
依年齡層觀察:信用卡消費集中在35歲至55歲之間,每筆平均消費金額與年齡成正比

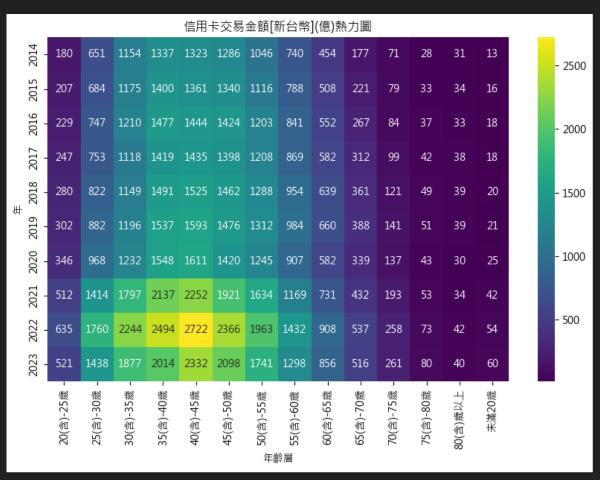




資料應用流程--資料分析(2)

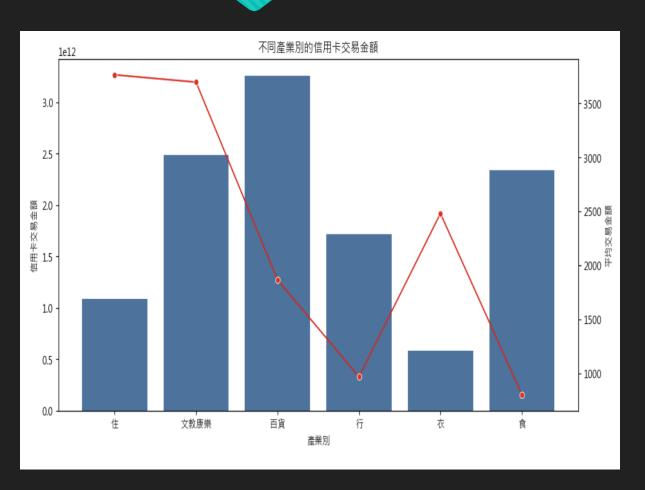
依年齡層觀察:疫情期間30歲至45歲消費金額與筆數大幅增加

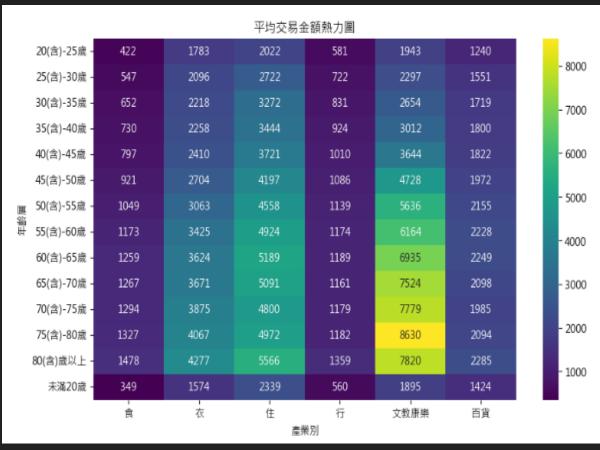




資料應用流程--資料分析(3)

依年龄層、產業別觀察:消費集中在百貨業,每筆平均消費金額以住及文教康樂較高

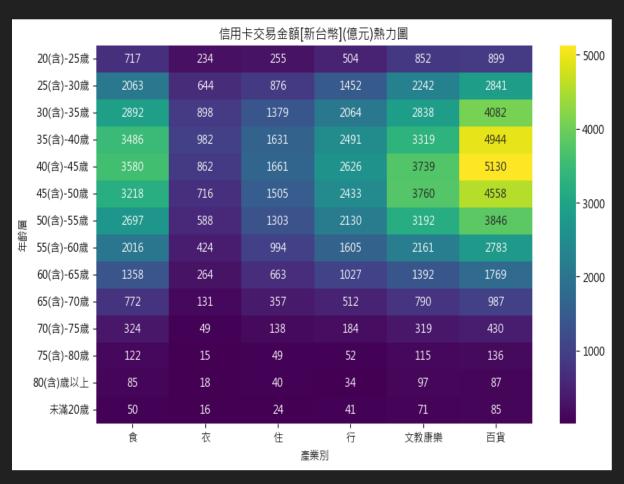




資料應用流程--資料分析(4)

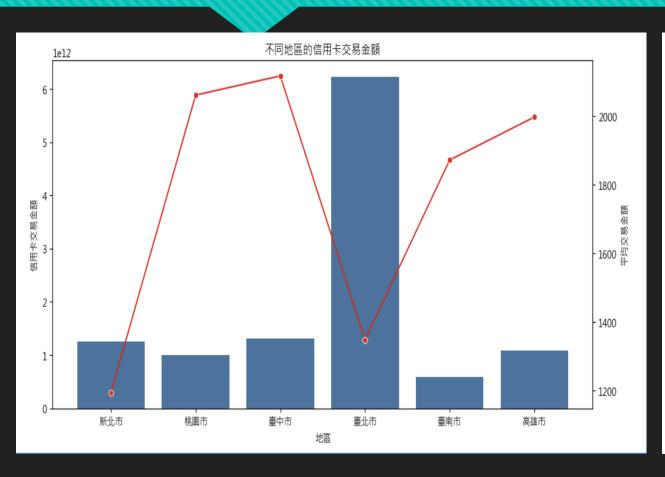
依年齡層、產業別觀察: 30歲至45歲之百貨業交易金額較高,食的交易筆數較多

	信用卡交易筆數(百萬)熱力圖									
	20(含)-25歳 -	170	13	13	87	44	73			
	25(含)-30歳 -	377	31	32	201	98	183			
	30(含)-35歳 -	444	40	42	248	107	237		- 400	
	35(含)-40歳 -	478	44	47	270	110	275			
	40(含)-45歳 -	449	36	45	260	103	282			
	45(含)-50歳 -	350	26	36	224	80	231		- 300	
興	50(含)-55歳 -	257	19	29	187	57	178			
年齡層	55(含)-60歳 -	172	12	20	137	35	125			
	60(含)-65歳 -	108	7	13	86	20	79		- 200	
	65(含)-70歳 -	61	4	7	44	10	47			
	70(含)-75歳 -	25	1	3	16	4	22			
	75(含)-80歳 -	9	0	1	4	1	7		- 100	
	80(含)歳以上 -	6	0	1	2	1	4			
	未滿20歳 -	14	1	1	7	4	6		_ 0	
		食	衣	住	· 行	文教康樂 文教康樂	_ 百貨	-	0	
				產業	巻別					



資料應用流程--資料分析(5)

依年齡層、地區觀察:信用卡消費主要集中在台北市,每筆平均消費金額以台中市及桃園市較高



平均交易金額熱力圖									2750
20(含	含)-25歳 -	618	1398	1401	764	1227	1387		- 2750
25(倉	含)-30歳 -	799	1705	1763	960	1576	1725		- 2500
30(倉	含)-35歳 -	959	1844	1952	1114	1736	1854		
35(≘	含)-40歳 -	1106	1919	1998	1223	1788	1873		- 2250
40(含	含)-45歳 -	1243	2038	2120	1335	1874	1966		2000
45(含	含)-50歳 -	1395	2229	2289	1556	1999	2128		- 2000
運 50(含	含)-55歳 -	1514	2389	2418	1763	2125	2240		- 1750
Ⅲ 50(含	含)-60歳 -	1593	2459	2502	1883	2188	2291		
60(含	含)-65歳 -	1657	2461	2534	2000	2185	2285		- 1500
65(含	含)-70歳 -	1620	2365	2384	2059	2069	2132		- 1250
70(含	含)-75歳 -	1595	2322	2299	2106	1968	2058		1230
75(含	含)-80歳 -	1664	2363	2300	2255	2037	2063		- 1000
80(含)歲以上 -	1880	2764	2670	2597	2410	2581		
未	滿20歳 -	1162	1201	1207	705	1103	1239		- 750
		新北市	桃園市	臺中市	臺北市	臺南市	高雄市	_	
地區									

資料應用流程--資料分析(6)

依年齡層、地區觀察:台北市信用卡交易金額高,交易筆數遠大於其他地區

信用卡交易筆數									
	20(含)-25歳 -	56	19	29	260	13	22		- 700
	25(含)-30歳 -	127	48	67	598	32	50		700
	30(含)-35歳 -	153	65	83	710	42	67		- 600
	35(含)-40歳 -	166	78	95	753	49	81		
	40(含)-45歳 -	159	78	94	710	49	83		- 500
	45(含)-50歳 -	127	66	83	554	42	74		
唖	50(含)-55歳 -	100	53	68	411	34	62		- 400
年齡層	55(含)-60歳 -	72	36	47	277	24	45		
	60(含)-65歳 -	46	22	28	172	15	30		- 300
	65(含)-70歳 -	26	12	16	95	8	17		
	70(含)-75歳 -	11	4	6	39	3	7		- 200
	75(含)-80歳 -	3	1	2	13	1	2		
	80(含)歳以上 -	2	1	1	8	1	1		- 100
	未滿20歳 -	5	1	2	22	1	2		
		新北市	桃園市	- 臺中市	臺北市	臺南市	高雄市	•	_
	地區								



資料應用流程--資料分析(7)

使用多項式廻歸模型預測2023年10~12月三個月六都交易金額

相關性矩陣:

年月 信用卡交易金額[新台幣] 信用卡交易筆數

年月 1.000000 0.834079 0.874661

信用卡交易金額[新台幣] 0.834079 1.000000 0.971233

信用卡交易筆數 0.874661 0.971233 1.000000

R-squared value: 0.9086274900883068

均方差: 0.006444418593204797

預測信用卡金額:

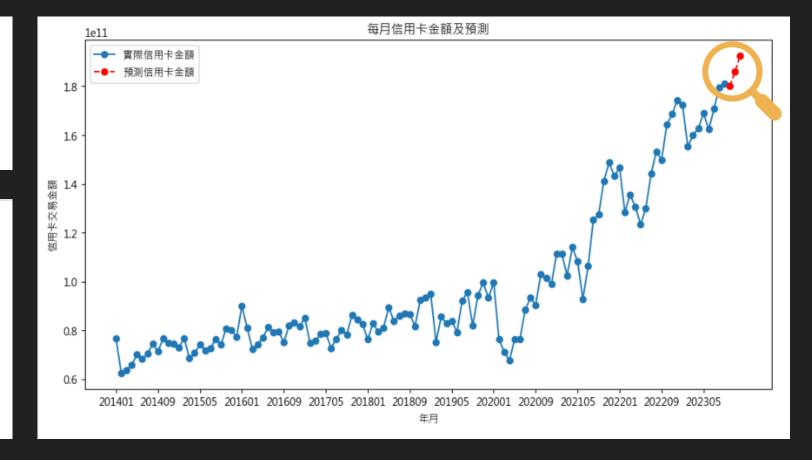
年份 月份

預測信用卡金額

0 2023 10 1.799901e+11

1 2023 11 1.859726e+11

2 2023 12 1.926103e+11



資料應用流程--資料解釋

- 依年齡層資料顯示信用卡消費金額集中於35歲至55歲之間,每筆平均消費金額與年齡成正 比,其中以住及文教康樂兩類別單筆交易金額較高,尤其年齡80歲以上之文教康樂類單筆 交易金額最高,各類別在疫情期間單筆金額有下降,2023年有逐步回升。
- 依產業別資料顯示信用卡交易金額以百貨業最高,其次是文教康樂、食等,每筆平均消費 金額以住及文教康樂較高,顯示國人願意花較多的錢獲得較高品質的生活。
- 依地區的資料顯示信用卡消費高度集中在六都,尤其是台北市,每筆平均消費金額以台中市最高,以新北市最低。
- ♀↑預估2023年10~12月三個月六都(不含其他類)交易金額應可達5,586億元。

影片演示

網址連結:https://dash-ml-knot.onrender.com



THANK YOUR