# 我們與巧克力的距離

-- 宏亞食品

臺北大學

林芳義、黃亮臻、謝沛璇、孫曉涵

### 動機

從眾效應



分析社群媒體、 論壇輿情資料



口味預測







分析 消費者資料



集中差異化



提升消費者對產品的價值







#### 流程

Step 1

文字探勘- 找出討論度高的話題

Step 2

宏亞食品於pchome、momo平台 產品銷售資料分析

Step 3

由2020年發票資料,找出口味與其他因素關係

## Step 1 文字探勘

輿情資料

大數軟體-社群媒體輿情資料

發票資料

睿點行動-發票存摺\_資策會數據\_invosdata

# 輿情資料







2021年8月論壇資料

2021年7、8月IG資料

2021年7、8月 YOUTUBE資料

# 發票資料



2020年發票資料

#### 發票資料-分性別

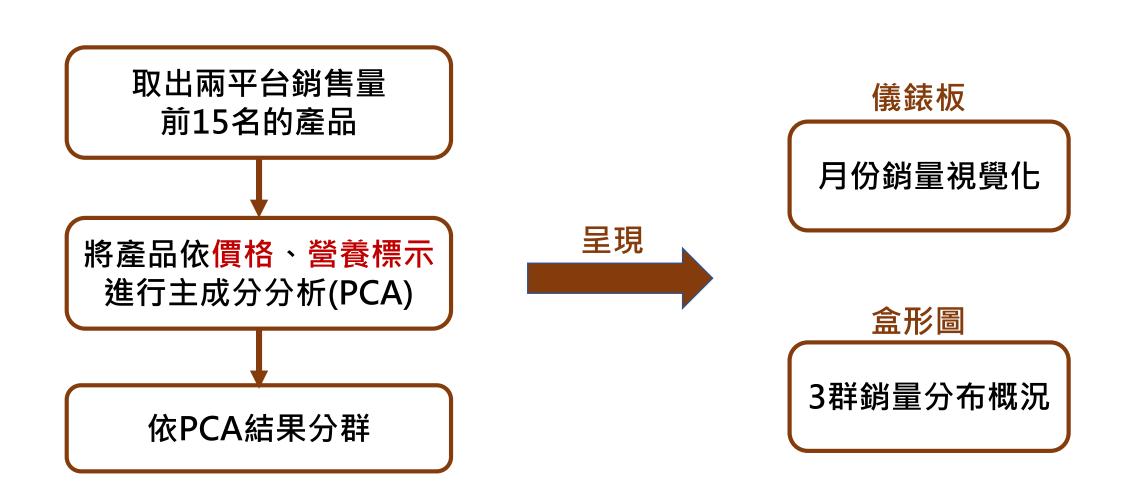




2020年發票資料-女性

2020年發票資料-男性

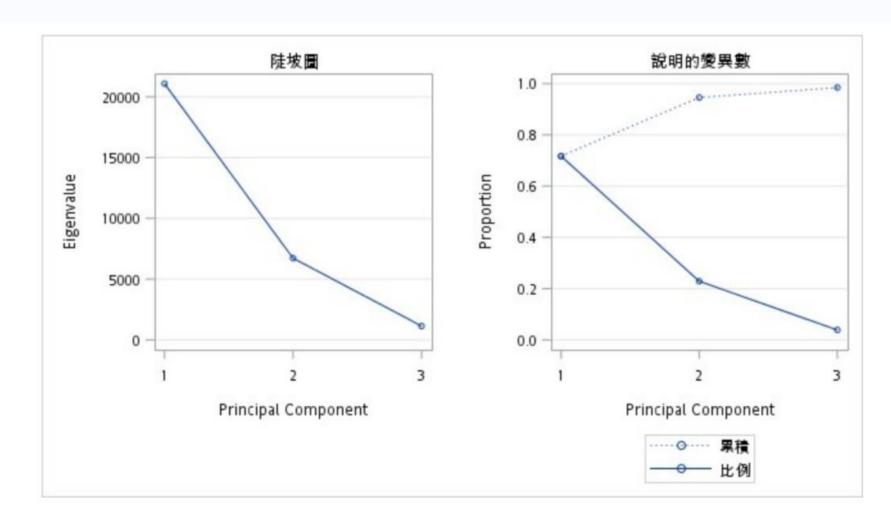
### Step 2 pchome、momo平台產品銷售資料分析



# PCA降維-報表

#### 主成分分析

#### PRINCOMP 程序



#### PCA降維-報表

3個成分共解釋了整體98.47%的變異



主成分1 高淨重&中鈉含量

主成分2 高鈉含量&低淨重



主成分3 高熱量&低價

	共變異數矩陣的特徵值						
	特徵值 差異 比例 累						
1	21075.6624	14355.0508	0.7174	0.7174			
2	6720.6117	5587.8795	0.2288	0.9461			
3	1132.7321		0.0386	0.9847			

特徵向量				
	PRIN1	PRIN2	PRIN3	
價格	0.087032	030673	169061	
淨重	0.964255	214658	0.123085	
每單位重	008034	010891	149978	
單位總數	0.048864	0.024144	0.114712	
熱量#100g	082185	0.140010	0.901533	
蛋白質#100g	0.003719	016118	007655	
脂肪#100g	017912	0.023437	0.154675	
飽和脂肪#100g	016077	0.026512	0.093517	
反式脂肪#100g	0.000134	0.000337	000204	
碳水化合物#100g	0.015980	003810	111795	
糖 <b>#100g</b>	0.001542	052483	212603	
鈉 <b>#100g</b>	0.229304	0.963533	132017	

#### CLUSTER分群-報表

 群集平均值

 群集 PRIN1 PRIN2 PRIN3

 1 -0.165223727 0.903679148 -0.097470108

 2 0.401144230 -0.686390816 -0.378248005

 3 -1.161928321 -0.806199787 2.235293509

Cluster1 (prin2/主成分2)

高鈉含量&低淨重

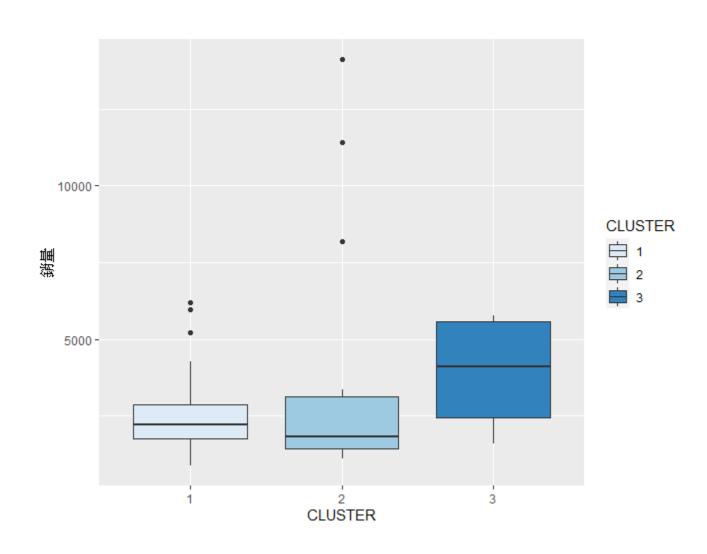
Cluster2 (prin1 /主成分1)

高淨重&中鈉含量

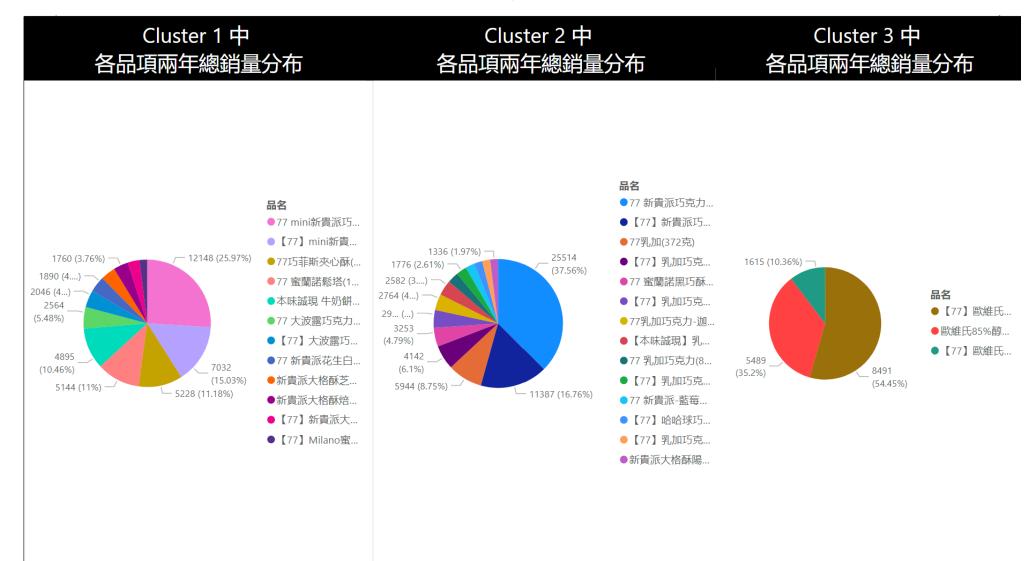
Cluster3 (prin3 /主成分3)

高熱量&低價

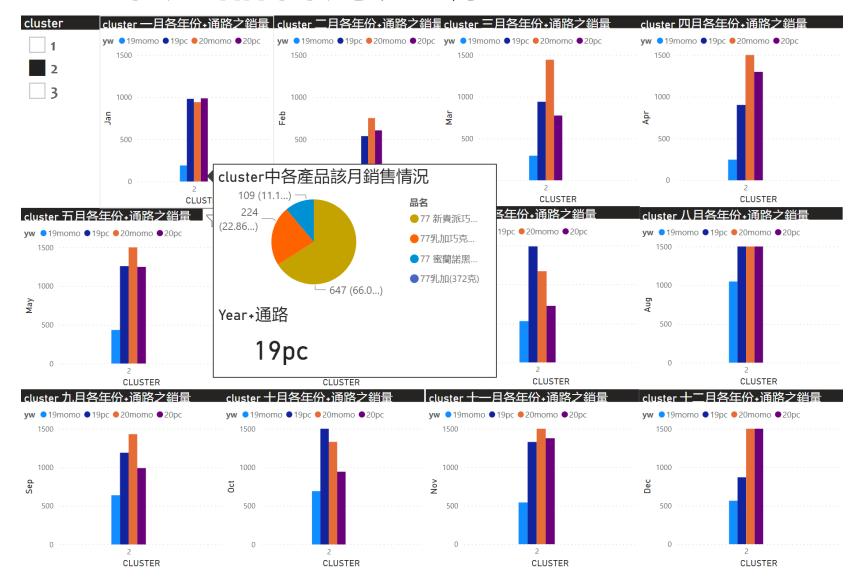
### CLUSTER 銷量分布概況



#### CLUSTER 中各品項之銷量



#### CLUSTER 中產品各月之銷量



#### 判別分析

Linear Discriminant Function for CLUSTER					
變數	1	2	3		
常數	-4777	-4370	-3713		
價格	-1.97543	-1.76463	-1.78248		
葷素類別	-78.82635	-71.91608	-30.16966		
淨重	-0.29235	-0.27804	-0.25441		
每單位重	14.26886	13.84617	11.13934		
單位總數	-6.95618	-6.04762	-4.89727		
熱量#100g	-30.18549	-30.79640	-19.63914		
蛋白質#100g	138.76962	140.53835	113.88529		
脂肪#100g	386.06271	386.60154	279.21412		
飽和脂肪#100g	-51.01806	-48.78954	-38.77933		
反式脂肪#100g	-5.70280	-38.70967	179.37107		
碳水化合物#100g	219.71984	217.76935	159.95189		
糖#100g	22.48140	21.28142	15.13517		
鈉#100g	4.16947	3.85565	2.65924		

#### 判別分析結果

#### DISCRIM 程序

校準資料的分類摘要: WORK.SORTTEMPTABLESORTED 使用下列方式的交叉驗證摘要 線性判別函數

分類為 CLUSTER 的觀測值數目和百分比						
From CLUSTER	1	2	3	總計		
1	17	1	0	18		
	94.44	5.56	0.00	100.00		
2	1	18	0	19		
	5.26	94.74	0.00	100.00		
3	0	0	4	4		
•	0.00	0.00	100.00	100.00		
總計	18	19	4	41		
₩ <u>0</u> -□	43.90	46.34	9.76	100.00		
直車	0.43902	0.46341	0.09756			
争刖						

CLUSTER 的誤差計數估計值						
	1	2	3	總計		
				0.0488	)	
事前	0.4390	0.4634	0.0976			

判錯的機率

# Step 3 由2020年發票資料,找出關係

性別與口味

年齡與口味

地區與口味

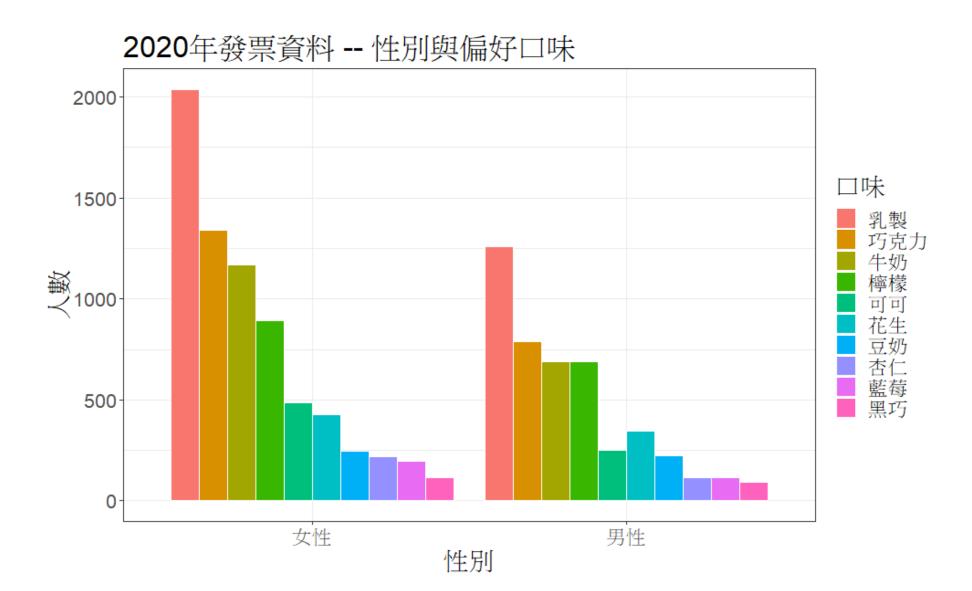
### 1. 性別與口味的關係

統計值	DF	值	機率
卡方	9	60.1614	<.0001
概度比卡方	9	59.7559	<.0001
Mantel-	1	11.1871	0.0008
Haenszel 卡方			
Phi 係數		0.0717	
列聯係數		0.0715	
Cramer V		0.0717	

p值小於0.05

性別與口味有相關!

### 1. 性別與口味的關係



### 2. 年龄層與口味的關係

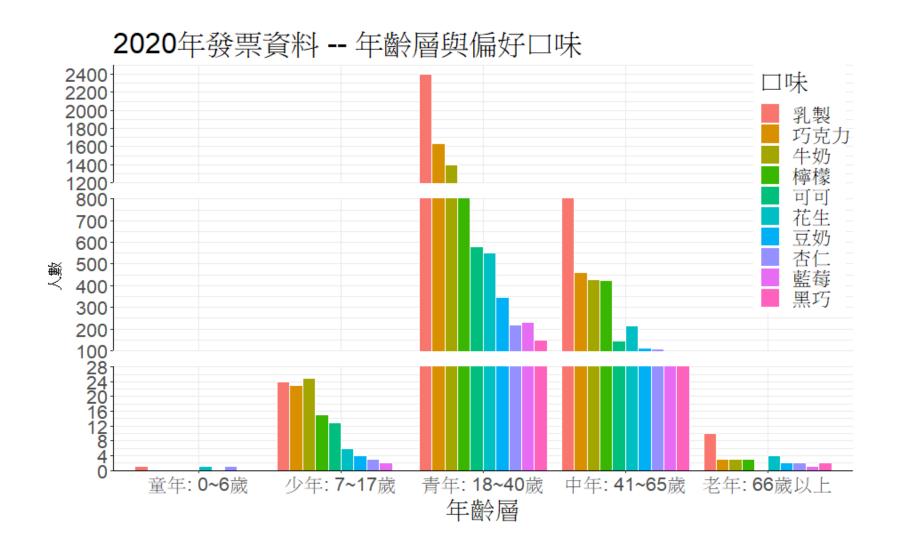
統計值	DF	值	機率
卡方	36	86.4175	<.0001
概度比卡方	36	80.7896	<.0001
Mantel-	1	0.9616	0.3268
Haenszel 卡方			
Phi 係數		0.0860	
列聯係數		0.0857	
Cramer V		0.0430	

WARNING: 44% of the cells have expected counts less than 5.

Chi-Square may not be a valid test.

細格小於5 卡方檢定結果可能 不進確

### 2. 年龄層與口味的關係



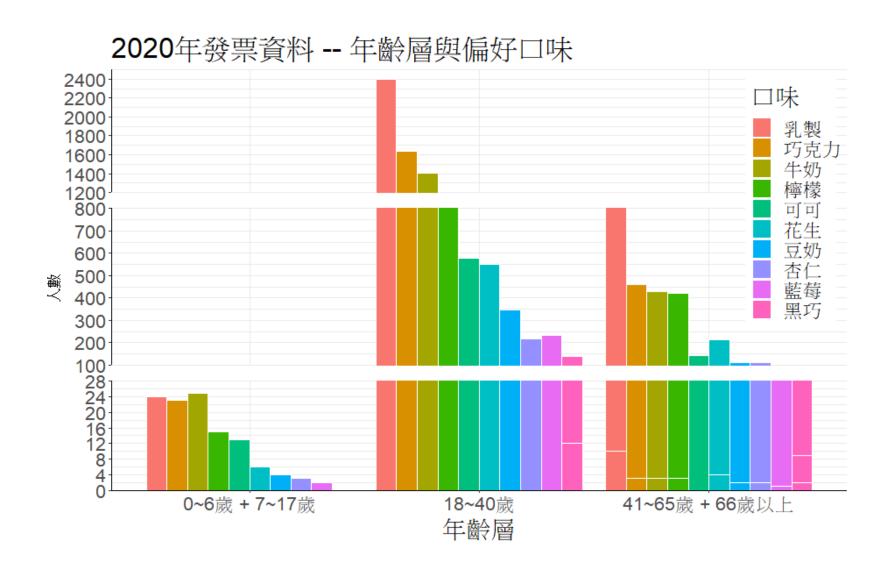
## 2. 年龄層與口味的關係 (合併後)

統計值	DF	值	機率
卡方	18	60.7837	<.0001
概度比卡方	18	62.2169	<.0001
Mantel-	1	0.9706	0.3245
Haenszel 卡方			
Phi 係數		0.0721	
列聯係數		0.0719	
Cramer V		0.0510	

p值小於0.05

年龄層與口味有相關!

# 2. 年龄層與口味的關係 (合併後)



# 3. 地區與口味的關係 (見Power BI)



### 3. 地區與口味的關係

統計值	DF	值	機率
卡方	250	566.9678	<.0001
概度比卡方	250	286.8340	0.0546
Mantel-	1	3.4034	0.0651
Haenszel 卡方			
Phi 係數		0.2202	
列聯係數		0.2150	
Cramer V		0.0696	

WARNING: 41% of the cells have expected counts less than 5.

Chi-Square may not be a valid test.

細格小於5 卡方檢定結果可能 不進確

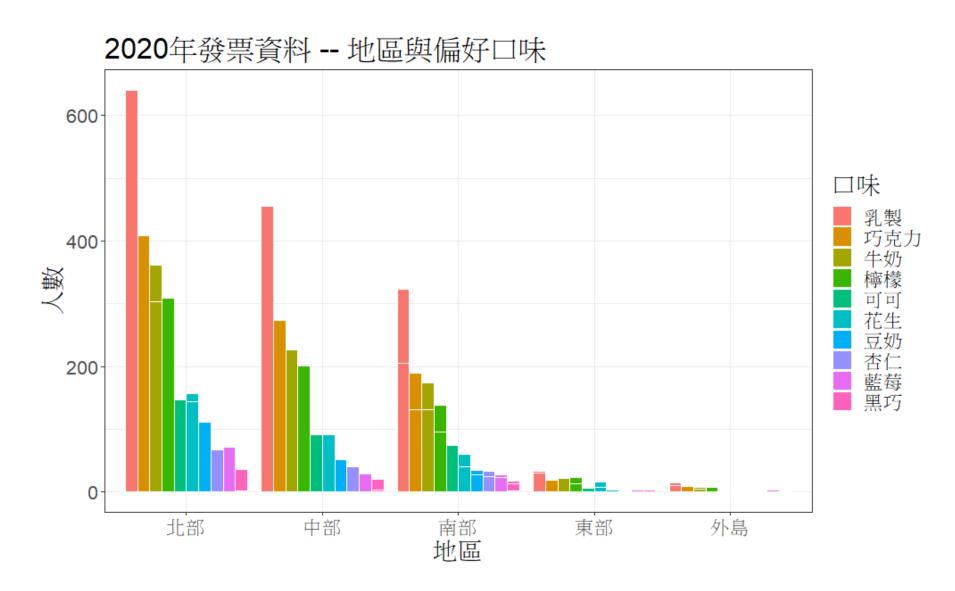
# 3. 地區與口味的關係 (合併區域)

統計值	DF	值	機率
卡方	40	67.8233	0.0039
概度比卡方	40	69.7388	0.0025
Mantel-	1	0.2630	0.6081
Haenszel 卡方			
Phi 係數		0.0762	
列聯係數		0.0759	
Cramer V		0.0381	

p值小於0.05

地區與口味有相關!

# 3. 地區與口味的關係 (合併區域)



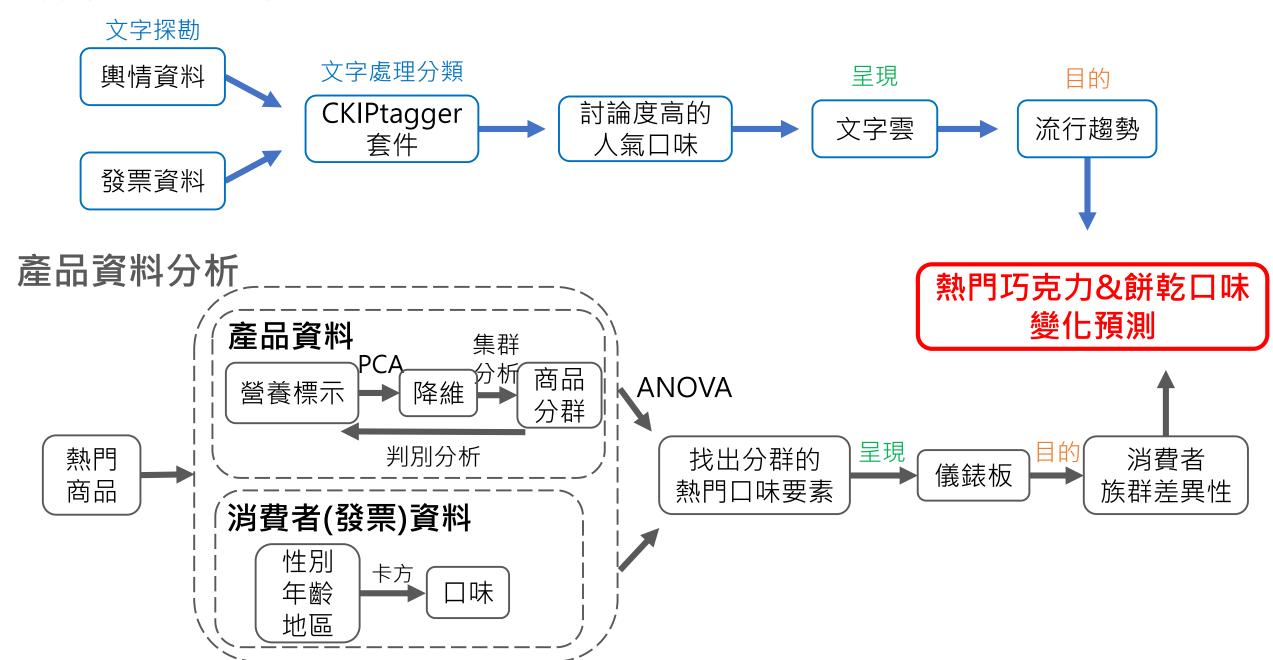
### 結論

- 1. 性別、年齡、地區都與口味有相關性,可以鎖定客群
- 2. 高熱量、低淨重、低價類產品的銷量普遍較高
- 3. 乳製、巧克力、牛奶類產品在各客群中皆占大宗
- 4. 紅茶、咖啡、奶茶、綠茶皆是消費者較常購買的商品
- 5. 可以針對不同cluster的旺季,推出新口味

#### 技術面的突破與創新

	產業現況	提案
方法	以基本資訊繪製圖表	輿情分析(文字探勘)、
		統計檢定
呈現方式	統計圖表	文字雲、統計推論、
		儀錶板
優點	1.了解現階段的銷售狀	1.預測未來熱門新口味
	況	2.能針對不同群進行調
		整。
		3.儀錶板之互動式圖表
		多元豐富
<b>缶夬黑占</b>	1.圖表功能單一	1.文字分類類別不夠精
	2.僅能了解現在的銷售	確
	狀況,較難預測與應	2.需要多種資料
	變未來趨勢。	

#### 輿情與民眾消費行為分析



#### (i) CLUSTER、品名、口味對照表

1	CLUSTER	品名	口味
2	1	【77】Milano蜜蘭諾千層鬆塔12入	無類別
3	1	【77】mini新貴派巧克力-花生294g(迷你)	花生
4	1	【77】大波露巧克力40g	巧克力
5	1	【本味誠現】乳酥餅136g(8入裝)	乳製
6	1	【77】mini新貴派巧克力-花生294g(迷你)	花生
7	1	【77】新貴派大格酥(烤焙花生口味20入)	花生
8	1	【本味誠現】乳酥餅136g(8入裝)	乳製
9	1	77 mini新貴派巧克力-花生(294g)	花生
10	1	77 大波露巧克力(40g)	巧克力
11	1	77 新貴派花生白巧風味(118g)	花生
12	1	77 蜜蘭諾鬆塔(12入)	無類別
13	1	本味誠現 牛奶餅(12入)	牛奶
14	1	77 mini新貴派巧克力-花生(294g)	花生
15	1	77 蜜蘭諾鬆塔(12入)	無類別
16	1	77巧菲斯夾心酥(300g)	無類別
17	1	本味誠現 牛奶餅(12入)	牛奶
18	1	新貴派大格酥芝麻豆奶324g	豆奶
19	1	新貴派大格酥焙烤花牛 20包入	花牛

20	2	【77】乳加巧克力-8入	乳製
21	2	【77】乳加巧克力372g(迷你)	乳製
22	2	【77】新貴派巧克力-花生18入	花生
23	2	【77】乳加巧克力-8入	乳製
24	2	【77】乳加巧克力-牛奶杏仁147g(迷你)	杏仁
25	2	【77】乳加巧克力-黑巧杏仁147g(迷你)	杏仁
26	2	【77】乳加巧克力372g(迷你)	乳製
27	2	【77】哈哈球巧克力324g	巧克力
28	2	【77】新貴派巧克力-花生18入	花生
29	2	77 新貴派巧克力-花生口味(18人)	花生
30	2	77 蜜蘭諾黑巧酥(16入)	黑巧
31	2	77乳加(372克)	乳製
32	2	77乳加巧克力-迦納可可黑巧杏仁(147g)	可可
33	2	77 乳加巧克力(8人)	乳製
34	2	77 新貴派-藍莓口味(18人)	藍莓
35	2	77 新貴派巧克力-花生口味(18人)	花生
36	2	77 蜜蘭諾黑巧酥(16人)	黑巧
37	2	77乳加(372克)	乳製
38	2	新貴派大格酥陽光檸檬324g	檸檬
39	3	【77】歐維氏85%醇黑巧克力	醇黑
40	3	【77】歐維氏70%醇黑巧克力(低糖)	醇黑
41	3	【77】歐維氏85%醇黑巧克力	醇黑
42	3	歐維氏85%醇黑巧克力(44g)	醇黑

#### (ii) 地區 (Area)卡方檢定

1	巧克力
2	花生
3	乳製
4	黑巧
5	牛奶
6	杏仁
8	可可
12	檸檬
15	豆奶
17	藍莓

 表格分析
 表格分析

 結果
 結果

 FREQ 程序
 FREQ 程序

Table of □味 by 地區								
	地區							
		中部	北部	外島	東部	南部	總計	
口味								
1	次數	383	1220	19	39	465	2126	
	預期	369.45	1217.1	16.909	42.182	480.36		
	儲存格卡方	0.4966	0.007	0.2586	0.24	0.4914		
	百分比	3.28	10.43	0.16	0.33	3.98	18.18	
	列百分比	18.02	57.38	0.89	1.83	21.87		
	欄百分比	18.85	18.23	20.43	16.81	17.60		
2	次數	130	472	1	24	147	774	
	預期	134.51	443.1	6.156	15.357	174.88		
	儲存格卡方	0.1509	1.8851	4.3184	4.8645	4.4456		
	百分比	1.11	4.04	0.01	0.21	1.26	6.62	
	列百分比	16.80	60.98	0.13	3.10	18.99		
	欄百分比	6.40	7.05	1.08	10.34	5.56		
3	次數	613	1854	28	64	739	3298	
	預期	573.12	1888	26.231	65.435	745.17		
	儲存格卡方	2.7745	0.6136	0.1194	0.0315	0.0511		
	百分比	5.24	15.86	0.24	0.55	6.32	28.20	
	列百分比	18.59	56.22	0.85	1.94	22.41		
	欄百分比	30.17	27.70	30.11	27.59	27.97		
4	次數	37	116	1	5	47	206	
	預期	35.799	117.93	1.6384	4.0872	46.545		
	儲存格卡方	0.0403	0.0316	0.2488	0.2038	0.0044		
	百分比	0.32	0.99	0.01	0.04	0.40	1.76	
	列百分比	17.96	56.31	0.49	2.43	22.82		
	欄百分比	1.82	1.73	1.08	2.16	1.78		
5	次數	309	1025	16	43	465	1858	
	預期	322.88	1063.7	14.778	36.864	419.81		
	儲存格卡方	0.5968	1.4056	0.1011	1.0212	4.8645		
	百分比	2.64	8.77	0.14	0.37	3.98	15.89	
	列百分比	16.63	55.17	0.86	2.31	25.03		
	欄百分比	15.21	15.31	17.20	18.53	17.60		
6	次數	58	190	3	3	81	335	
	預期	58.216	191.78	2.6644	6.6467	75.692		
	儲存格卡方	0.0008	0.0165	0.0423	2.0008	0.3722		
	百分比	0.50	1.62	0.03	0.03	0.69	2.86	
	列百分比	17.31	56.72	0.90	0.90	24.18		
	欄百分比	2.85	2.84	3.23	1.29	3.07		
8	次數	115	421	3	8	189	736	
	預期	127.9	421.34	5.8538	14.603	166.3		
	儲存格卡方	1.3014	0.0003	1.3912	2.9856	3.0994		
	百分比	0.98	3.60	0.03	0.07	1.62	6.29	

	Table of □味 by 地區								
				地區					
		中部	北部	外島	東部	南部	總計		
8	列百分比	15.63	57.20	0.41	1.09	25.68			
	欄百分比	5.66	6.29	3.23	3.45	7.15			
12	次數	274	908	14	37	346	1579		
	預期		903.94	12.559	31.329	356.77			
	儲存格卡方	0.0006	0.0182	0.1654	1.0266	0.3252			
	百分比	2.34	7.77	0.12	0.32	2.96	13.50		
	列百分比	17.35	57.50	0.89	2.34	21.91			
	欄百分比	13.48	13.56	15.05	15.95	13.10			
15	次數	68	298	3	4	93	466		
	預期	80.981	266.78	3.7063	9.2459	105.29			
	儲存格卡方	2.0808	3.6547	0.1346	2.9764	1.4349			
	百分比	0.58	2.55	0.03	0.03	0.80	3.99		
	列百分比	14.59	63.95	0.64	0.86	19.96			
	欄百分比	3.35	4.45	3.23	1.72	3.52			
17	次數	45	190	5	5	70	315		
	預期	54.74	180.33	2.5053	6.2499	71.173			
	儲存格卡方	1.7332	0.5184	2.484	0.25	0.0193			
	百分比	0.38	1.62	0.04	0.04	0.60	2.69		
	列百分比	14.29	60.32	1.59	1.59	22.22			
	欄百分比	2.21	2.84	5.38	2.16	2.65			
總計	次數	2032	6694	93	232	2642	11693		
	百分比	17.38	57.25	0.80	1.98	22.59	100.00		

口味\*地區之表格的統計值

統計值	DF	信	機率
卡方		57.2990	155.4
既度比卡方	36	60.6126	0.0063
Mantel-Haenszel 卡方	1	0.2973	0.5856
Phi 係數		0.0700	
列聯係數		0.0698	
Cramer's V		0.0350	

樣本大小 = 11693

#### (iii) 性別 (Sex)卡方檢定

1	巧克力
2	花生
3	乳製
4	黑巧
5	牛奶
6	杏仁
8	可可
12	檸檬
15	豆奶
17	藍莓

表格分析	
結果	
FREQ 程序	

Table of □味 by 性別					
		性	別		
		女性	男性	總計	
口味		, (j.12	J 0   1.11		
1	次數	1338	788	2126	
	預期	1295.5	830.55		
	儲存格卡方	1.3973	2.1794		
	百分比	11.44	6.74	18.18	
	列百分比	62.94	37.06		
	欄百分比	18.78	17.25		
2	次數	429	345	774	
	預期	471.63	302.37		
	儲存格卡方	3.853	6.0097		
	百分比	3.67	2.95	6.62	
	列百分比	55.43	44.57		
	欄百分比	6.02	7.55		
3	次數	2038	1260	3298	
	預期	2009.6	1288.4		
	儲存格卡方	0.4014	0.626		
	百分比	17.43	10.78	28.20	
	列百分比	61.80	38.20		
	欄百分比	28.60	27.58		
4	次數	113	93	206	
	預期	125.52	80.476		
	儲存格卡方	1.2495	1.949		
	百分比	0.97	0.80	1.76	
	列百分比	54.85	45.15		
	欄百分比	1.59	2.04		
5	次數	1168	690	1858	
	預期	1132.2	725.85		
	儲存格卡方	1.1351	1.7705		
	百分比	9.99	5.90	15.89	
	列百分比	62.86	37.14		
	欄百分比	16.39	15.11		
6	次數	219	116	335	
	預期	204.13	130.87		
	儲存格卡方	1.0834	1.6899		
	百分比	1.87	0.99	2.86	
	列百分比	65.37	34.63		
	欄百分比	3.07	2.54		
8	次數	487	249	736	
	預期	448.47	287.53		
	儲存格卡方	3.3097	5.1623		
	百分比	4.16	2.13	6.29	

	Tubic of	口味by		
		性		
			男性	總計
8	列百分比	66.17	33.83	
	欄百分比	6.84	5.45	
12	次數	891	688	1579
	預期	962.15	616.85	
	儲存格卡方	5.2609	8.2058	
	百分比	7.62	5.88	13.50
	列百分比	56.43	43.57	
	欄百分比	12.51	15.06	
15	次數	244	222	466
	預期	283.95	182.05	
	儲存格卡方	5.6212	8.7678	
	百分比	2.09	1.90	3.99
	列百分比	52.36	47.64	
	欄百分比	3.42	4.86	
17	次數	198	117	315
	預期	191.94	123.06	
	儲存格卡方	0.1912	0.2983	
	百分比	1.69	1.00	2.69
	列百分比	62.86	37.14	
	欄百分比	2.78	2.56	
總計	次數	7125	4568	11693
	百分比	60.93	39.07	100.00

表格分析 結果 FREQ 程序

口味\*性別之表格的統計值

統計值	DF	值	機率
卡方	9	60.1614	<.0001
概度比卡方	9	59.7559	<.0001
Mantel-Haenszel 卡方	1	11.1871	0.0008
Phi 係數		0.0717	
列聯係數		0.0715	
Cramer's V		0.0717	

樣本大小 = 11693

#### (iv) 年齡 (Age)卡方檢定

1	巧克力
2	花生
3	乳製
4	黑巧
5	牛奶
6	杏仁
8	可可
12	檸檬
15	豆奶
17	藍莓

表格分析 結果 FREQ 程序

Table of □味 by 年齡層								
			年齢層					
		2	3	4	總計			
口味								
1	次數	23	1638	465	2126			
	預期	21.455	1573.8	530.73				
	儲存格卡方	0.1113	2.6174	8.1399				
	百分比	0.20	14.01	3.98	18.18			
	列百分比	1.08	77.05	21.87				
	欄百分比	19.49	18.92	15.93				
2	次數	7	550	217	774			
	預期	7.8108	572.97	193.22				
	儲存格卡方	0.0842	0.9209	2.927				
	百分比	0.06	4.70	1.86	6.62			
	列百分比	0.90	71.06	28.04				
	欄百分比	5.93	6.35	7.43				
3	次數	25	2398	875	3298			
	預期	33.282	2441.4	823.3				
	儲存格卡方	2.0608	0.7721	3.2464				
	百分比	0.21	20.51	7.48	28.20			
	列百分比	0.76	72.71	26.53				
	欄百分比	21.19	27.70	29.98				
4	次數	0	151	55	206			
	預期	2.0789	152.5	51.425				
	儲存格卡方	2.0789	0.0147	0.2485				
	百分比	0.00	1.29	0.47	1.76			
	列百分比	0.00	73.30	26.70				
	欄百分比	0.00	1.74	1.88				
5	次數	25	1402	431	1858			
	預期	18.75	1375.4	463.82				
	儲存格卡方	2.0833	0.5135	2.323				
	百分比	0.21	11.99	3.69	15.89			
	列百分比	1.35	75.46	23.20				
	欄百分比	21.19	16.20	14.77				
6	次數	4	219	112	335			
	預期	3.3807	247.99	83.628				
	儲存格卡方	0.1135	3.3892	9.6254				
	百分比	0.03	1.87	0.96	2.86			
	列百分比	1.19	65.37	33.43				
	欄百分比	3.39	2.53	3.84				
8	次數	13	580	143	736			
	預期	7.4273	544.84	183.73				
	儲存格卡方	4.1811	2.2689	9.0302				
	百分比	0.11	4.96	1.22	6.29			
	日カル	0.11	4.90	1.22	0.29			

表格分析 結果

FREQ 程序

	列百分比	2	年齡層			
	列百分比	2				
	列百分比					
	ᄁᄓᄆᄭᅜ	1 77	78.80	19.43	總計	
)	欄百分比	11.02		4.90		
12	次數	15		426	1579	
	預期		1168.9		1010	
信	存格卡方			2.5693		
[[1]	百分比	0.13		3.64	13.50	
7	列百分比		72.07		. 5.00	
	欄百分比	12.71		14.59		
15	次數	4	347	115	466	
	預期	4.7026	344.97	116.33		
信	香格卡方	0.105	0.012			
	百分比	0.03	2.97	0.98	3.99	
- 2	列百分比	0.86	74.46	24.68		
	欄百分比	3.39	4.01	3.94		
17	次數	2	233	80	315	
	預期	3.1788	233.19	78.636		
信	皆存格卡方	0.4372	0.0001	0.0237		
	百分比	0.02	1.99	0.68	2.69	
3	列百分比	0.63	73.97	25.40		
7	欄百分比	1.69	2.69	2.74		
總計	次數	118			11693	
	百分比	1.01	74.03	24.96	100.00	

口味 \* 年齡層 之表格的統計值

統計值	DF	值	機率
卡方	18	60.7837	<.0001
概度比卡方	18	62.2169	<.0001
Mantel-Haenszel 卡方	1	0.9706	0.3245
Phi 係數		0.0721	
列聯係數		0.0719	
Cramer's V		0.0510	

樣本大小 = 11693

由 SAS 系統 ('Local', X64\_10HOME) 於 2021年11月14日 1:22:35 AM 產生

(v) Pchome 年月份銷售量



(vi) Momo 年月份銷售量

