

大數據資料科學家養成班 第三次專案討論

Costco好市多商品經驗老實說 2.0

第五組:林佑霖

許伊寧

呂沅蓁

中華民國111年05月

簡報大綱

- 壹、前情提要
 - (1)研究動機
 - (2)熱門度定義
 - (3)資料來源
 - (3)程式開發流程圖
 - (4)1.0 問題盤點
- 貳、2.0功能優化介紹
- 參、成果展示

- 肆、延伸應用(POWBERBI)
 - (1)資料寫入
 - (2)報表更新
- 伍、現行問題&未來展望
 - (1)綜合分析
 - (2)未來展望

壹、前情提要

- (1)研究動機
- (2)熱門度定義
- (3)資料來源
- (3)程式開發流程圖
- (4)1.0 問題盤點

研究動機

從自身需求出發

通路選擇:COSTCO、全聯、購物網站



別人有買,我也要



錢要花在刀口上,精準消費



節省比價時間



重疊區域

熱門度的定義

FB 社團

討論度

商品

熱門

推薦度

商家

推薦

購買度

資料來源(FB公開社團)

選擇原因

目標資料

社團



COSTCO 好市多 商品消費心得分享區

公開社團・52 萬位成員・一天 6 則貼文

加入社團



8 位朋友是成員



好市多商品資訊分享

公開社團·11 萬位成員·一個月8 則貼文 #Costco #好市多 #商品資訊分享 #特價資訊 #新品分享

加入社團

1 位朋友是成員



Costco好市多 商品經驗老實說

公開社團 · 188 萬位成員 · 一天 10 則以上貼文

加入社團

參考價值:活躍度最高

剛才挺好市多線上,發現飛利浦這台萬用鍋有特價了! 之前吃過朋友用這台燉的滷肉,肉很嫩湯汁也很濃,當時看他開鍋直接 用來炒菜也覺得很方便,感覺可以燉可以炒真的頗萬用,被燒到很想買

- 一台XD
- 一直蹲等終於被我等到優惠,也上來分享分享

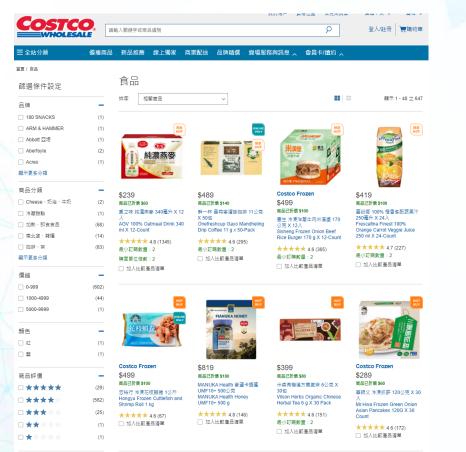
買了一罐松露醬,家裡除了我沒有一個人受得了那個味道,明明就很香 😥 😥 😥 看大家都說很容易壞掉,搞得我三餐卯起來加,松露蛋餅,松露臭豆腐,松露蛋炒

現在已經升級到松露炒米粉了😭 😭 😭

跪求各種好吃的方法!!!

技術考量:公開社團在爬蟲的可行性較高

資料來源(COSTCO官網)



1 選擇品項:飲食

2 販賣範圍:全通路 賣場限定

程式開發流程圖

前端 後端 採用語法

使用者操作介面 點擊觸發 Tkinter

獲取FB社團貼文

爬蟲

匯入MONGODB

前處理 FB 社團貼文

讀取DB

斷詞

次數統計

獲取COSTCO商品資料*2份

爬蟲

匯入MONGODB

前處理 COSTCO商品資料*2份

轉成DATAFRAME*2

模糊比對 前處理的衍生資料*2

產生討論度最高商品資料各10筆

使用者操作介面 輸出討論度最高商品*20筆

Tkinter

1.0 問題盤點



手刀購買去。

全通路TOP10

		商品連結1
	商品名稱 商品價格	商品連結2
1	新東陽績緻豬肉鬆 270公克 \$165	
2	得意的一天 三益調和油 3.3公升 \$239	商品連結3
3	桂格黃金麩片燕麥片 1.7公斤 \$249	商品連結4
4		IN HHAZINH T
5	黑橋牌 蒜味迷你腸 500公克×2入×6包 \$2,549	商品連結5
6	桂格有機燕麥飲 300毫升 X 12入 \$329	# Di#4+c
7	Chosen Foods 酪梨油 1公升 \$389	商品連結6
8	Kewpie 美奶滋 1公斤 \$167	商品連結7
9	Starbucks 春季限定咖啡豆 1.13公斤 \$699	
10) MANUKA Health 麥蘆卡蜂蜜UMF10+500公克 \$819	商品連結8
		# Di#4+0

手刀購買去c

查場限定TOP10

	商品名稱	商品連結
1	PRIMA TASTE 新加坡叻沙225公克X 3	商品連結
2	萬家香甲等醬油 3750 毫升	商品連結
3	五月花妙用綿柔藝巧抽取衛生紙 130抽 X 36包	商品連結
4 5	應頭牌PVC布紋封箱膠帶 4.8公分X 20公尺X 6 3M SCOTCH膠帶六入組3/4吋X1100吋	商品連結
6	3M SCOTCH雙面棉紙膠帶12入	商品連結
7	TWS精製砂糖(二砂)2公斤	
	Kirkland Signature 科克蘭 全脂鮮乳 1.89公升 X 2瓶	商品連結
9	HP #932/#933墨水組合包1黑3彩XL TWS結晶冰糖2公斤	商品連結
10	7 IVV Sカロ 自由/小 付着 2 2公/)	商品連結

爬蟲相關

商品連結10

商品連結10

- ◆ FB社團爬取的貼文內容 不完整
- ◆ FB社團爬蟲的連接不穩定
- ◆ 商品資料會出現非食品

- ◆ GUI的版面配置不夠精確
- ◆ (字形、字體、位置)
- ◆ 未提供商品連結



貳、2.0 功能優化介紹

2.0 功能優化介紹



2.0 功能優化介紹



全通路 討論度TOP 3

2 (買什麼好呢)

商品名稱	商品價格	商品連結
Kirkland Signature 科克蘭有機蘇門答臘咖啡豆 907公克	\$469	https://www.costco.com.tw/Water-Beverages/Kirkland-Signature
老食粹 有機枸杞 600公克	\$489	https://www.costco.com.tw/Food/Pantry-Dry-Goods/Dried-Groc
健司 繼三薯脆片 400公克	\$295	https://www.costco.com.tw/Food/Snacks/Nuts-Chips-Rice-Crack

資料排版整齊

連結至訂購頁面

賣場獨家 討論度TOP 3 3 買什麼好呢

商品名稱		商品連結
雀巢 咖啡伴侶原味罐裝 1.5公斤	https://www.costco.com.tv	/Warehouse-Only/Food-Beverages/Water-Beverages/NESTLE-Co
法國可鬆12入	https://www.costco.com.tv	//Warehouse-Only/Food-Beverages/Dry-Grocery/FRENCH-CROIS
TWS特級砂糖2公斤	https://www.costco.com.tv	/Warehouse-Only/Food-Beverages/Dry-Grocery/TWS-GRANULA

2.0 功能優化介紹

```
driver.get("https://www.costco.com.tw/c/CLWH?q=:relevance:category:WH1205:category:WH1201:category:WH0806:category:WH1206
298
299
        #讓頁面自動往下滑
        driver.execute script("window,scrollTo(0,document.body.scrollHeight)")
300
301
        time.sleep(5)
302
        soup = BeautifulSoup(driver.page source, "html.parser")
303
        for commodity2 in soup.find all("ul",class ="product-listing product-grid")[0] find all("li",class ="product-list-item g
304
            #商品名稱
305
                                                    使用webdriver
            a=commodity2.find_all("a",class_="lister-
            #網址
307
            ##9<sup>21</sup>c="https://www.costco.com.tw/"+commodity2 提升爬蟲資料準確度
              print(a)
309
              print(c)
310
              print('--
311
312
            #新增單筆資料
313
314
            st={"商品名稱":a,
315
                "網址":c
316
317
            result=col.insert one(st)
```

參、成果展示



肆、延伸應用(POWERBI)

- (1)資料寫入
- (2)報表更新

資料寫入

```
#寫入資料
df_extract=df_extract.drop(['matches','merge_key'],axis=1)
with pd.ExcelWriter('G:\\我的雲端硬碟\\project3.xlsx') as writer:
df_extract.to_excel(writer,sheet_name="文字雲切分",index=False)
final.to_excel(writer,sheet_name="CostcoTOP10",index=False)
final2.to_excel(writer,sheet_name="Costco&FTOP10",index=False)
```

更新資料於Google Drive 電腦版

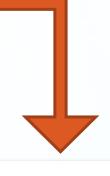


Costco 最新食品TOP10推薦排行榜

COSTCO .com.tw







		結構 / イ	事曆	關聯性	計算		
	000	× ✓					
	-	商品名稱 ▼	商品價格 ▼		網址	¥	
	\blacksquare	商品名稱	商品價格	網址		E	ă,
	誯	Starbucks 派克市場咖啡豆 1.13公斤	\$699	https://www.	costco.com.tw/Food/Drinks/Coffee-Tea/S	itar h	ntt
	48	雀巢 可可早餐脆片 500公克 X 2入	\$329	https://www.	costco.com.tw/Food/Drinks/Powdered-D	rinl h	ntt
		家樂氏 Special K 草莓早餐脆片 1.2公斤	\$389	https://www.	costco.com.tw/Food/Drinks/Powdered-D	rinl h	ntt
		Evian 天然礦泉水 1500毫升 X 12入	\$659	https://www.	costco.com.tw/Food/Drinks/Water-Bever	age h	ntt
		愛之味 純濃燕麥 340毫升 X 12入	\$299	https://www.	costco.com.tw/Food/Drinks/Water-Bever	age h	ntt
		關山一級御用米 9公斤	\$635	https://www.	costco.com.tw/Food/Pantry-Dry-Goods/F	Rice h	ntt
		紅龍冷凍牛肉捲 220 公克 X 6 入	\$459	https://www.	costco.com.tw/Food/Frozen-Foods/Prepa	arec h	ntt
		ARM & HAMMER 小蘇打粉 6.12公斤	\$299	https://www.	costco.com.tw/Food/Pantry-Dry-Goods/E	Baki h	ntt
		喜年來 原味蛋捲 72公克 X 20入	\$449	https://www.	costco.com.tw/Food/Snacks/Cookies-Dri	ed-I h	ntt
		老食粹 有機枸杞 600公克	\$489	https://www.	costco.com.tw/Food/Pantry-Dry-Goods/I	Orie h	ntt

圖片網址
https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/https://www.costco.com.tw//medias/sys_master/images/

固片網址





Simple Image
Vincent Faigt

****(13)



Costco 最新熱門討論字 ♦





word	→ num
好市	18
顯示	16
請問	14
大家	11
今天	10
因為	10
這個	10
可以	9
Costco	8
使用	8
最近	8
特價	8
發現	8
一下	7
分享	7

sites SOCIAI famous dia digital content networ information like

Word Cloud 🌣 Microsoft Corporation

****(135)

報表更新

下列項目的設定:專案3-V2

檢視資料集 🛭

上次重新整理成功: Sun May 29 2022 13:52:59 GMT+0800 (GMT+08:00) 重新整理記錄

▲資料集描述

描述此資料集的內容。

用 捨棄

▲閘道連接

若要使用資料閘道,請確定電腦已上線,而且資料來源已新增至<u>管理閘道</u>。如果您使用內部部署資料閘道 (標準模式),請選取對應的資料來源,然後按一下 [套用]。

使用內部部署或 VNet 資料閩道



開的

聞道 部門 連絡人資訊 狀態 動作

● F

Personal Gateway

套用

捨棄

▲資料來源認證

project3-1.xlsx <u>編輯認證 以譜系檢視顯示</u> ^[2] project3.xlsx <u>編輯認證 以譜系檢視顯示</u> ^[3]



伍、現行問題&未來展望

- (1)綜合分析
- (2)未來展望

綜合分析

現行問題

- ◆ FB社團爬取的貼文內容不完整
- ◆ FB社團爬蟲的連接不穩定
- ◆ 目前無法自動爬蟲,需固定時間手動操作

未來展望

- ◆ 標籤分群(生鮮、限時優惠)
- ◆ 熱門排行榜(生活用品、家電)
- ◆豐富社群來源(Ptt、Dcard、Instagram)
- ◆ 比較通路價格(家樂福、全國電子)

未來展望





簡報結束 謝謝聆聽