

商品搜尋推薦算法

報告者:林桓鈺

報告架構

1

資料尋找與處理

2

參考資料處理

3

篩選資料的爬蟲
與處理

4

商品推薦算法的生成
與優化

資料尋找與處理

主題動機: 希望藉由此次報告，能熟悉商品推薦優先方法的運作方式並搭配數據資料，優化算法的功能性。

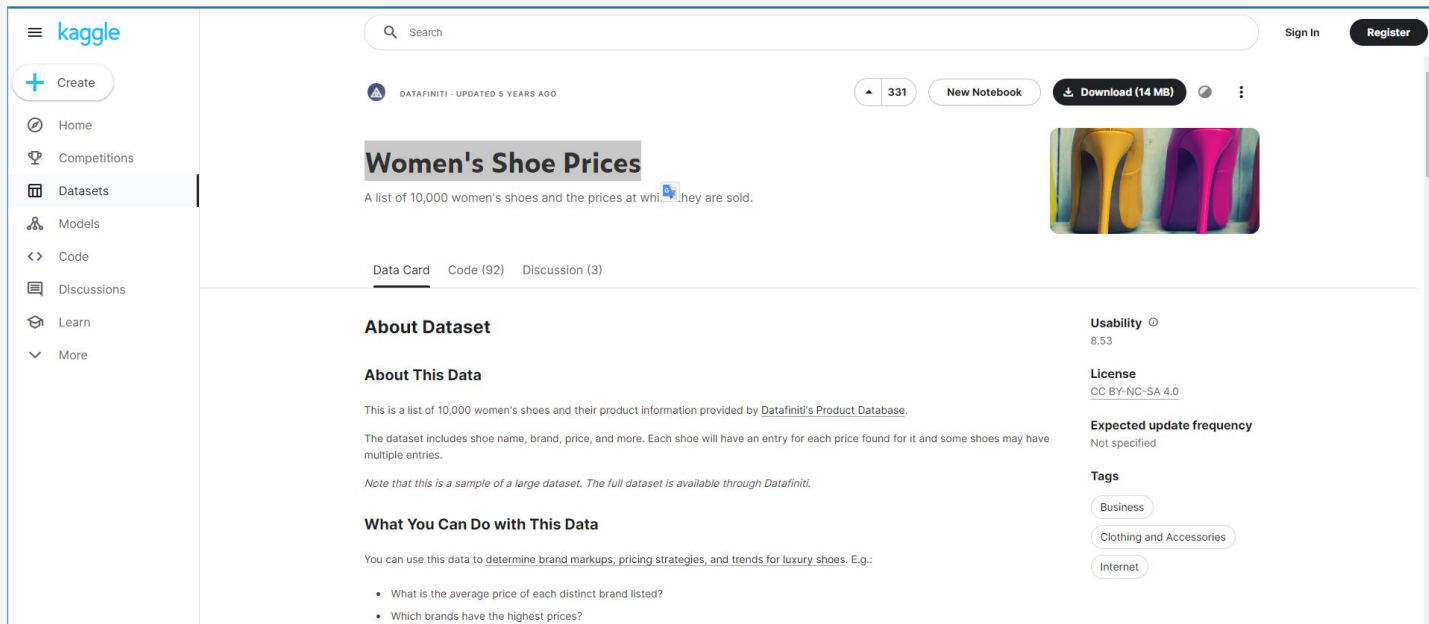
The screenshot displays the momo購物網 (momo shopping website) interface. At the top, there's a navigation bar with links like '首頁', '登入', '註冊', '會員中心', etc. Below this, the 'momo 購物網' logo is visible. A search bar is prominently featured, with a dropdown menu open for the search term '室內拖鞋'. The dropdown menu lists various sub-categories and the number of items available for each:

Category	Item Count
室內拖鞋	約16831件商品
拼字校正	
使用滑鼠左鍵開啟 (Alt+向下鍵)	
台灣製室內拖鞋	約1797件商品
室內拖	約3308件商品
兒童室內拖鞋	約615件商品
保暖室內拖鞋	約363件商品
止滑室內拖鞋	約1065件商品
防滑室內拖	約519件商品
兒童室內拖	約504件商品
防滑室內拖鞋	約1065件商品
厚底室內拖鞋	約3039件商品
室內拖鞋 保暖	約363件商品
男室內拖鞋	約3882件商品
毛毛室內拖鞋	約180件商品
亞麻室內拖鞋	約117件商品
絨毛室內拖鞋	約86件商品

On the right side of the search bar, there's a promotional banner for '年貨節' (New Year Goods Festival) with a '搶mo幣' (Grab mo coins) offer. Below the search bar, there's a horizontal navigation bar with categories like '3C', '家電', '美妆個護', etc. On the left side, there's a '會員福利特區' (Member Benefits Special Zone) with various offers. On the right side, there's a '購物車' (Shopping Cart) and a '簽到賺mo幣' (Check-in to earn mo coins) button. At the bottom, there's a '好康活動大集合' (Great collection of good deals) section.

資料尋找與處理

資料來源: kaggle 網站的Women's Shoe Prices 資料集。
年份:2023全年



The screenshot displays the Kaggle website interface for the 'Women's Shoe Prices' dataset. The left sidebar contains navigation links: Home, Competitions, Datasets (highlighted), Models, Code, Discussions, Learn, and More. The main content area features a search bar, a 'Sign In' button, and a 'Register' button. Below these, the dataset title 'Women's Shoe Prices' is prominently displayed, followed by a brief description: 'A list of 10,000 women's shoes and the prices at which they are sold.' A thumbnail image of two high-heeled shoes is shown. The dataset is attributed to 'DATAFINITI - UPDATED 5 YEARS AGO'. Below the title, there are tabs for 'Data Card' (selected), 'Code (92)', and 'Discussion (3)'. The 'About Dataset' section includes a description of the data, its source (Datafiniti's Product Database), and a note that it is a sample of a larger dataset. The 'What You Can Do with This Data' section provides examples of potential analyses, such as determining brand markups and pricing strategies. On the right side, there are sections for 'Usability' (8.53), 'License' (CC BY-NC-SA 4.0), 'Expected update frequency' (Not specified), and 'Tags' (Business, Clothing and Accessories, Internet).

Kaggle

Search

Sign In Register

DATAFINITI - UPDATED 5 YEARS AGO

331 New Notebook Download (14 MB)

Women's Shoe Prices

A list of 10,000 women's shoes and the prices at which they are sold.

Data Card Code (92) Discussion (3)

About Dataset

About This Data

This is a list of 10,000 women's shoes and their product information provided by Datafiniti's Product Database.

The dataset includes shoe name, brand, price, and more. Each shoe will have an entry for each price found for it and some shoes may have multiple entries.

Note that this is a sample of a large dataset. The full dataset is available through Datafiniti.

What You Can Do with This Data

You can use this data to determine brand markups, pricing strategies, and trends for luxury shoes. E.g.:

- What is the average price of each distinct brand listed?
- Which brands have the highest prices?

Usability 8.53

License CC BY-NC-SA 4.0

Expected update frequency Not specified

Tags

- Business
- Clothing and Accessories
- Internet

參考資料處理

資料資訊: **Women's Shoe Prices** 資料集的內的相關所帶的變數。

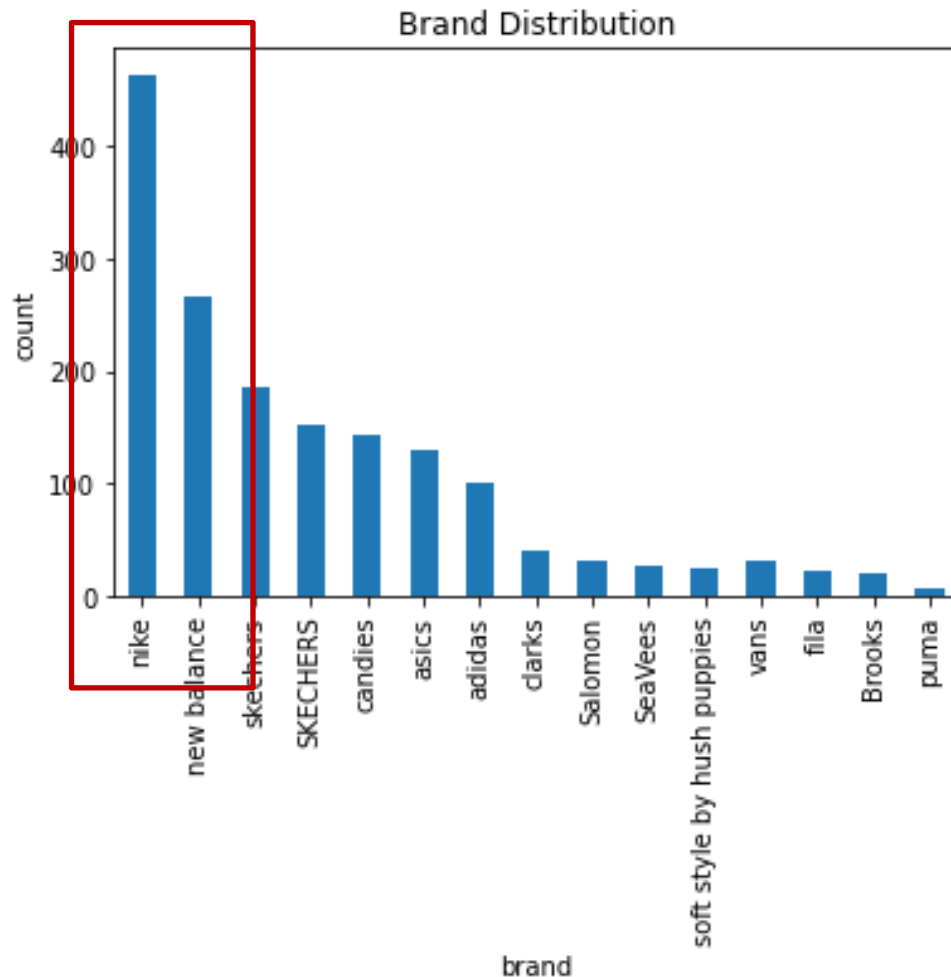
```
'''  
id - 產品編號  
brand - 品牌名稱  
colors- 產品顏色  
imageURLs - 產品網址連結  
prices.amountMax - 產品最高價格  
prices.amountMin - 產品最低價格  
prices.color - 銷售價格對應的商品顏色  
prices.currency - 售價幣值  
prices.isSale - 產品是否打折  
prices.merchants - 產品銷售的商家  
'''
```

資料的處理:

處理缺失值。

變數名稱在命名(命名成更簡便的名稱)。

參考資料處理

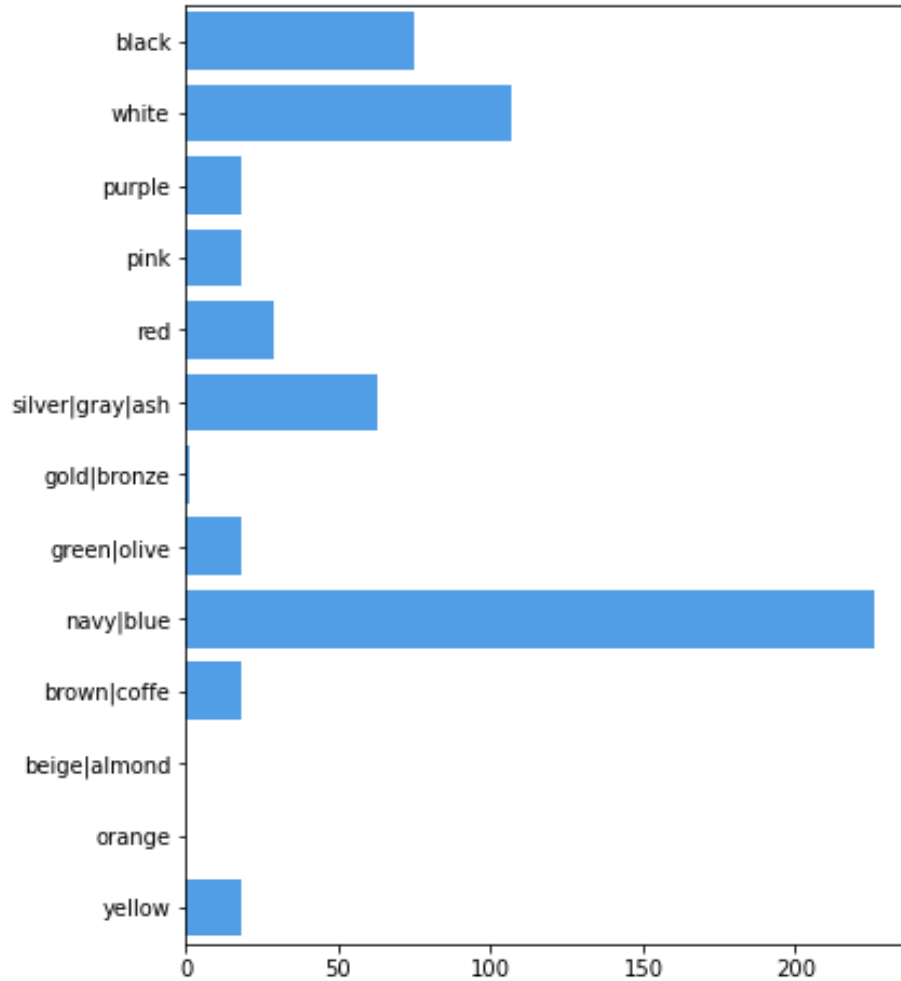


賣出商品品牌的分布圖：

可以看出在全年女鞋的銷量之中 **Nike** 為大宗超過第二名至少200多雙的量。

將統計結果作為加權依據

參考資料處理



賣出商品顏色的分布圖：

可以看出在全年女鞋的銷量以 藍色 為大家最喜歡色系。

將統計結果作為加權依據

篩選資料與爬蟲處理

爬蟲工具: 藉由python selenium工具，爬取pchome、momo等購物網站。



篩選資料與爬蟲處理

商品本身的顏色: 在資料處理過程中寫入商品本身的顏色。
使用python colorthief套件。

```
# -*- coding: utf-8 -*-
from colorthief import ColorThief
import requests
#rgb轉換函數 將rgb數值轉換為十六進制顏色
def rgb_to_hex(rgb):
    return "#{:02x}{:02x}{:02x}".format(rgb[0], rgb[1], rgb[2])
#獲取圖片中顏色並輸出RGB
def color(d):
    # 實際圖片的 URL
    new_image_url = f"{d}"

    # 使用 requests 下載圖片
    response = requests.get(new_image_url)

    # 將圖片保存到本地文件
    with open('temp_image.png', 'wb') as f:
        f.write(response.content)

    # 使用 ColorThief 獲取主要顏色
    color_thief = ColorThief('temp_image.png')
    dominant_color = color_thief.get_color(quality=1)

    #想獲取調色板中的顏色
    palette = color_thief.get_palette(color_count=6)
    # 將 dominant_color 和 palette 包裝在元組中進行返回
    return dominant_color, palette

# 測試函數
```

將rgb數值轉換成16進位制
並運用色碼表參考其顏色。

篩選資料與爬蟲處理

將資料輸出成檔案: 將資料輸出成csv檔。

[illegible]

商品推薦算法的生成與優化

加權搜尋引擎的基本原理：

關鍵字搜尋引擎：

定義： 關鍵字搜尋引擎是一種允許使用者通過輸入關鍵字來檢索相關信息或資源的系統。這些關鍵字可以是使用者認為與他們感興趣的主題或問題相關的詞彙。

原理： 關鍵字搜尋引擎使用不同的算法和方法來匹配使用者輸入的關鍵字與數據庫中的內容，以返回與關鍵字相關的結果。

Personalized Search (個性化搜索)： 考慮用戶的個人特徵、歷史行為等，為每個用戶提供定制的搜索結果，以提高搜索的精確性。

商品推薦算法的生成與優化

Apriori 算法的基本原理：某個詞類重複出現的程度。

1. **頻繁項集 (Frequent Itemsets)**：Apriori 算法的核心思想是使用 "先驗性原則"，即如果一個項集是頻繁的，那麼它的所有子集也應該是頻繁的。一個項集是指數據集中的一個物品的集合。頻繁項集是指在數據集中出現頻率高於某一預定閾值（通常由用戶指定）的項集。
2. **支持度 (Support)**：支持度是指在整個數據集中一個項集出現的次數。Apriori 算法設置一個支持度閾值，只保留那些支持度高於閾值的項集。
3. **候選項集 (Candidate Itemsets)**：算法首先生成所有可能的單個項目（1-項集），然後檢查每個項目的支持度。根據支持度閾值，只有那些高於閾值的項目被保留，形成候選項集。
4. **連接步驟 (Join)**：接下來，Apriori 算法進行連接步驟，將剩餘的項目連接成更大的項集，稱為 k -項集。這裡 k 是項集的大小。
5. **剪枝步驟 (Prune)**：然後進行剪枝步驟，剔除那些不滿足 "先驗性原則" 的項集。即，如果一個 k -項集的子集不是 $(k-1)$ -項集，則將其剔除。
6. **重複連接和剪枝**：不斷重複連接和剪枝步驟，生成 $k+1$ -項集，直到無法再生成更大的項集。最後，所有保留的項集都是頻繁項集。
7. **關聯規則產生**：最後，可以從找到的頻繁項集中生成關聯規則。這些規則描述項集之間的相關性，例如 "如果購買 A，則更有可能購買 B"。

商品推薦算法的生成與優化

生成搜尋引擎：

CONVERSE CTAS MOVE 女鞋 低筒 高筒 厚底 帆布鞋 休閒鞋

0	康威士 Converse All Star 男鞋 女鞋 M9622C	藍	https://24h.pchome.com.tw/prod/QCAL2F-A90...
1	Converse 休閒鞋 All Star 低筒 穿搭 男女鞋 ...	藍	https://24h.pchome.com.tw/prod/DXAX5K-A90...
2	Converse All Star OX 休閒 流行 女鞋 M9166C	藍	https://24h.pchome.com.tw/prod/QCAL2F-A90...
3	Converse Chuck 低筒 帆布 男鞋 女鞋 M9697C	藍	https://24h.pchome.com.tw/prod/QCAL2F-A90...
4	Converse Chuck Taylor 帆布鞋 男鞋 女鞋 M33...	藍	https://24h.pchome.com.tw/prod/QCAL2F-A90...
5	Converse 帆布鞋 Chuck 70 HI 男鞋 女鞋 抹茶...	藍	https://24h.pchome.com.tw/prod/DXAX5K-A90...
6	Converse 穆勒鞋 Chuck 70 Mule 男鞋 女鞋 莫...	藍	https://24h.pchome.com.tw/prod/DXAX5K-A90...
7	【CONVERSE品牌旗艦店】CTAS ALL STAR LIFT 1...	藍	https://m.momoshop.com.tw/goods.momo?i_co...
8	Converse 休閒鞋 All Star 低筒 穿搭 男女鞋 ...	黑	https://24h.pchome.com.tw/prod/DXAX5K-A90...
9	Converse 帆布鞋 Chuck 70 HI 男鞋 女鞋 抹茶...	黑	https://24h.pchome.com.tw/prod/DXAX5K-A90...
10	【CONVERSE】休閒鞋 All Star 男鞋 女鞋 低筒...	黑	https://m.momoshop.com.tw/goods.momo?i_co...
11	Converse 休閒鞋 All Star 穿搭 男女鞋 經典...	黑	https://24h.pchome.com.tw/prod/DXAX5K-A90...
12	Converse All Star Hi 基本款 休閒 男鞋 女鞋...	黑	https://24h.pchome.com.tw/prod/DXAX5K-A90...
13	Converse 休閒鞋 All Star Move 女鞋 厚底 舒...	白	https://24h.pchome.com.tw/prod/QCAL2F-A90...
14	Converse All Star Hi 基本款 帆布鞋 男女鞋 ...	白	https://24h.pchome.com.tw/prod/QCAL2F-A90...
15	Converse 休閒鞋 Chuck 70 GTX 高筒 男女鞋 ...	黑	https://24h.pchome.com.tw/prod/DXAX5K-A90...
16	【CONVERSE】CHUCK 70 1970 低筒 休閒鞋 帆布...	白	https://m.momoshop.com.tw/goods.momo?i_co...
17	【CONVERSE】Chuck Taylor All Star 男鞋 女...	黑	https://m.momoshop.com.tw/goods.momo?i_co...
18	Converse All Star 基本款 帆布鞋 男鞋 女鞋 ...	白	https://24h.pchome.com.tw/prod/QCAL2F-A90...
19	Converse All Star Hi 基本款 休閒 男鞋 女鞋...	白	https://24h.pchome.com.tw/prod/DXAX5K-A90...
20	Converse Chuck 低筒 帆布 男鞋 女鞋 M9697C	黑	https://24h.pchome.com.tw/prod/QCAL2F-A90...
21	Converse 休閒鞋 All Star Move 女鞋 厚底 舒...	黑	https://24h.pchome.com.tw/prod/QCAL2F-A90...
22	Converse All Star Hi 基本款 帆布鞋 男女鞋 ...	黑	https://24h.pchome.com.tw/prod/QCAL2F-A90...
23	Converse Chuck Taylor All Star Lift 黑 白	黑	https://24h.pchome.com.tw/prod/DXAX5K-A90...

商品推薦算法的生成與優化

頻繁項集展示：生成一個虛擬的消費紀錄。

```
dataset = [  
    ['NIKE', 'CONVERSE', 'ADIDAS'],  
    ['NIKE', 'NEW BALANCE'],  
    ['CONVERSE', 'ADIDAS'],  
    ['NIKE', 'NEW BALANCE', 'ADIDAS', 'FILA'],  
    ['CONVERSE', 'ADIDAS', 'NEW BALANCE']  
]
```

商品推薦算法的生成與優化

Apriori 算法展示：

CONVERSE CTAS MOVE 女鞋 低筒 高筒 厚底 帆布鞋 休閒鞋

【CONVERSE】CTAS CRUISE HI 高筒 休閒鞋 帆...	黑	https://m.momoshop.com.tw/goods.momo?i_co...
【CONVERSE】CTAS LIFT HI 高筒 休閒鞋 厚底...	黑	https://m.momoshop.com.tw/goods.momo?i_co...
【CONVERSE】CTAS ALL STAR 休閒鞋 帆布鞋 男...	黑	https://m.momoshop.com.tw/goods.momo?i_co...
【CONVERSE品牌旗艦店】CTAS MOVE 女鞋 低筒 ...	黑	https://m.momoshop.com.tw/goods.momo?i_co...
Converse 休閒鞋 All Star Move 女鞋 厚底 舒...	白	https://24h.pchome.com.tw/prod/QCAL2F-A90...
Converse 休閒鞋 All Star Move 女鞋 厚底 舒...	黑	https://24h.pchome.com.tw/prod/QCAL2F-A90...
【adidas 愛迪達】休閒鞋 女鞋 運動鞋 網球鞋 ...	藍	https://m.momoshop.com.tw/goods.momo?i_co...
ADIDAS 女鞋 慢跑鞋 RUNFALCON 3.0 W -HP7557	藍	https://24h.pchome.com.tw/prod/DXAX0W-A90...
【ADIDAS】ADILETTE COMFORT Slipper 男鞋 ...	藍	https://24h.pchome.com.tw/prod/DXAX2Y-A90...
【ADIDAS】GALAXY 6 W 跑步鞋 女鞋 白色-GW41...	藍	https://24h.pchome.com.tw/prod/DXAX2Y-A90...
【ADIDAS】RUNFALCON 3.0 W 跑步鞋 女鞋 白色...	藍	https://24h.pchome.com.tw/prod/DXAX2Y-A90...
ADIDAS 女鞋 男鞋 慢跑鞋 RUNFALCON 3.0 TR -...	藍	https://24h.pchome.com.tw/prod/DXAX0T-A90...
adidas 慢跑鞋 Racer TR21 女鞋 白 銀 豹紋 ...	藍	https://24h.pchome.com.tw/prod/DXAX0U-A90...
ADIDAS 女鞋 男鞋 慢跑鞋 RUNFALCON 3.0 TR -...	藍	https://24h.pchome.com.tw/prod/DXAX0T-A90...
【ADIDAS】GALAXY 6 W 跑步鞋 女鞋 白色-GW41...	白	https://24h.pchome.com.tw/prod/DXAX2Y-A90...
【ADIDAS】DURAMO 10 跑步鞋 女鞋 桃紅色-HQ4...	黑	https://24h.pchome.com.tw/prod/DXAX2Y-A90...

Thank you
