

# Marathon及Catalog之數據分析

支援海外線上零售商進入香港市場

**Team Navy**

Benjamin

Carman

Harris

# 數據分析簡介



太古資源



## 太古資源有限公司

太古資源是集團旗下在香港特別行政區、中國內地及澳門特別行政區眾多從事運動鞋、戶外鞋、便服鞋及時尚生活品牌零售及經銷業務的控股公司。該公司經營逾一百七十個零售點，包括銷售Nike、Adidas、New Balance、Puma、Converse 及 Asics產品的單一品牌專門店，以及銷售多種運動用品品牌的零售連鎖店馬拉松、GigaSports及Catalog。獨家經銷業務方面，太古資源代理多個國際知名的品牌，包括 Arena、Cath Kidston、Chevignon、Columbia、Crocs、Havaianas、Jockey、Repetto、Rockport、Speedo、Teva及UGG。

CATALOG

旗下兩大核心

鞋類零售網店

MARATHON  
SPORTS 馬拉松



# 提供數據支持的市場洞見

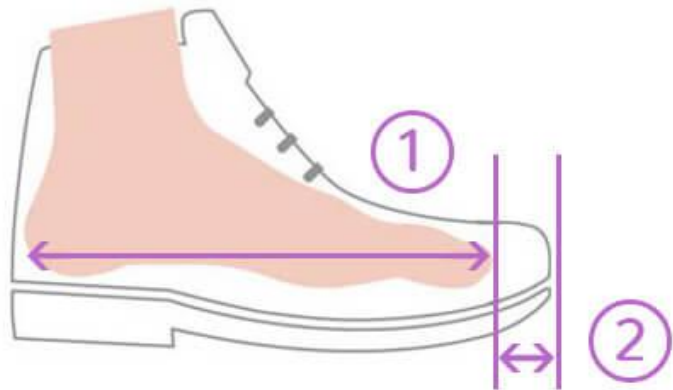
男/女鞋的價格/鞋款分別？



網店的促銷活動  
和折扣力度？



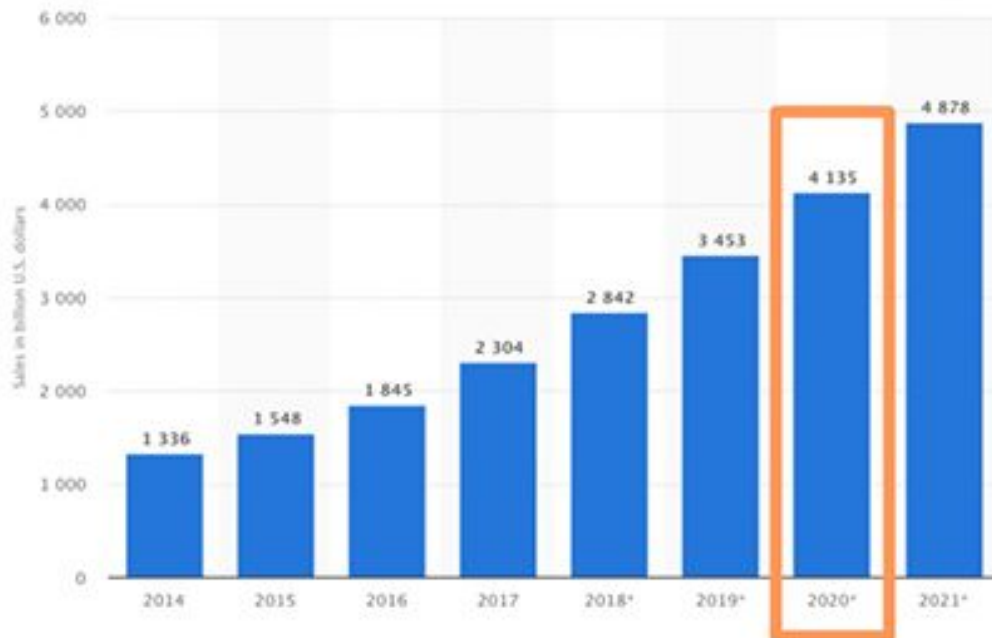
各牌子/鞋子尺碼的  
市場占有率？



## 分析數據就能解答！

# 背景與目標

2014 – 2021年全球零售業網上銷售額增長 (單位: 十億美金)



# 產業環境



- 香港作為購物天堂，擁有成熟的電商生態和獨特的消費者習慣
- 鞋類零售市場的銷售額在2025年1月達到 4,496百萬港元。
- 鞋類零售佔香港零售總額的 約5%，顯示出其在整體零售市場中的重要性
- 運動鞋和生活風格鞋需求成長顯著，疫情後市場復甦強勁

# 產業環境

成長趨勢

消費者行為



# 商業價值

消費者趨勢掌握：透過分析消費者購買行為和需求變化，預判流行趨勢，為產品開發和促銷規劃提供基礎。

市場機會挖掘：辨識潛在市場空白與突破點，推動客戶在香港市場精準拓展與銷售成長。



# 市場定位

## Marathon:

- 健康生活方式, 跑步賽事生態系統
- 針對熱衷跑步、注重健康生活的消費者
- 有較高的運動裝備消費能力
- 市場佔比相對較小但成長迅速且用戶黏性強

## catalog:

- 產品及服務展示, 零售及電商支持
- 廣泛消費者, 電商購物用戶
- 依賴零售和電商的龐大需求
- 規模龐大且數位轉型推動快速擴展

# project方向

- 促銷活動和折扣力度
- 香港男生及女生普遍鞋碼
- 牌子在marathon和catalog的市佔率
- 價格區間
- 新出的款式有什麼共同點

**解釋所獲取數據如何分析以上目標**

# 資料爬蟲架構與實現

# 資料爬蟲架構與實現

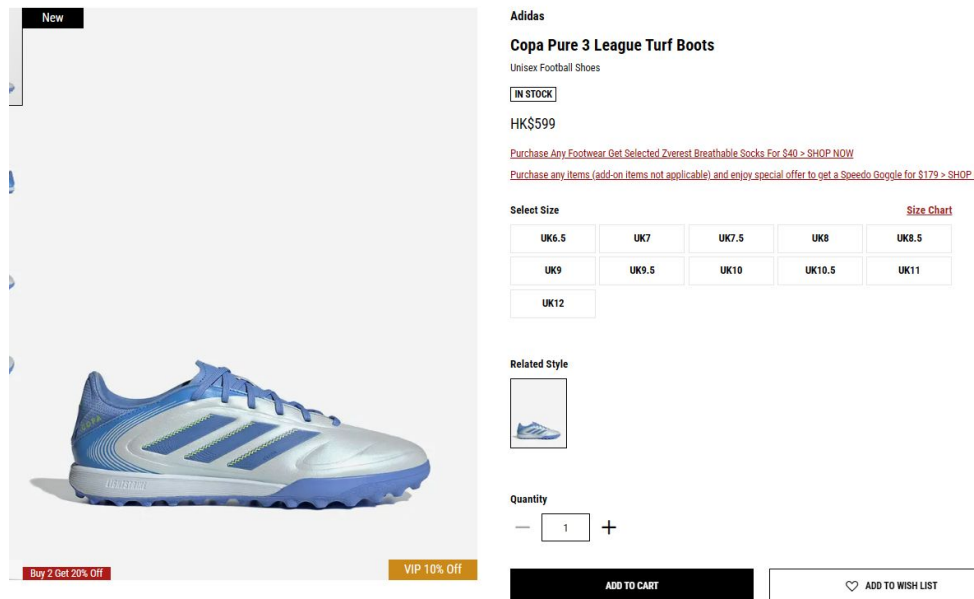
## 目標與需求



- 爬取香港運動鞋網站男女裝鞋的資料(Catalog 和 馬拉松)

- 收集的資料

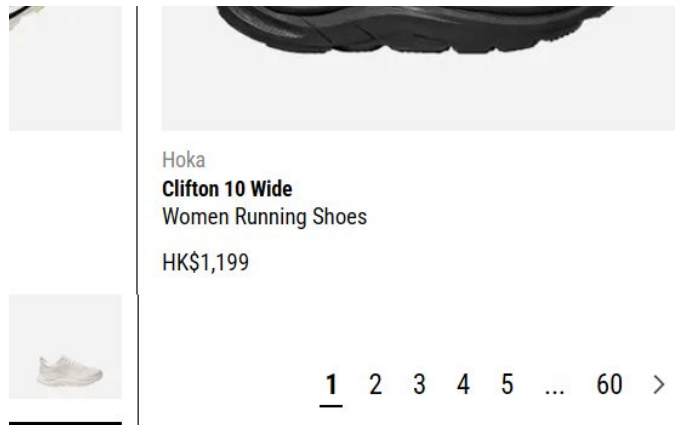
品牌、產品名稱、鞋款、  
價格、男/女裝、折扣、  
版面排序、收集時間、  
鞋碼(UK5 到 UK10)和 SKU



# 資料爬蟲架構與實現

## 目標與需求

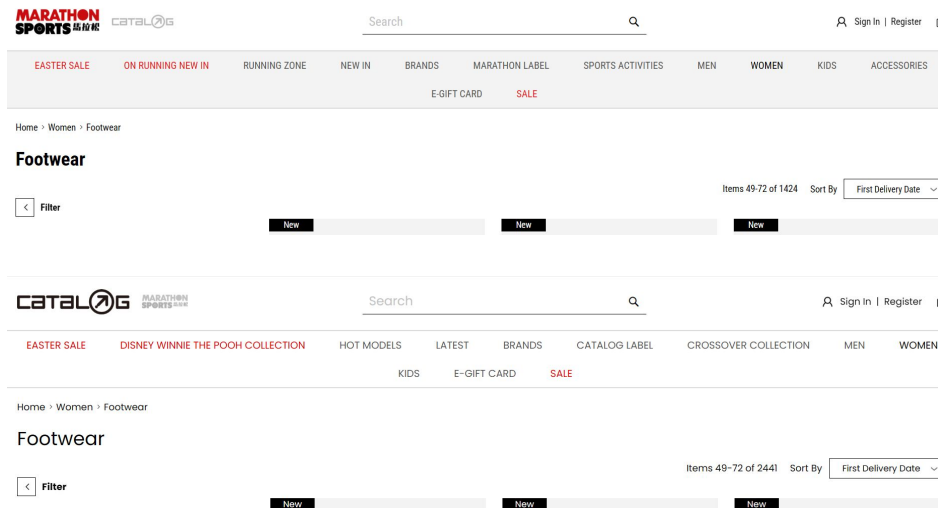
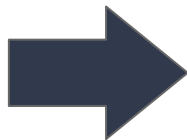
- 目標是爬取**指定頁數範圍**(例如第 1 至 25 頁), 每頁大約 24 對鞋, 兩個網頁男女裝鞋總共約 2400 對的資料
- 資料最終要整理成兩個表格
  - 1.儲存鞋的**基本資訊**
  - 2.儲存價格、折扣和鞋碼等**詳細資料**



# 資料爬蟲架構與實現

## 架構設計(1)

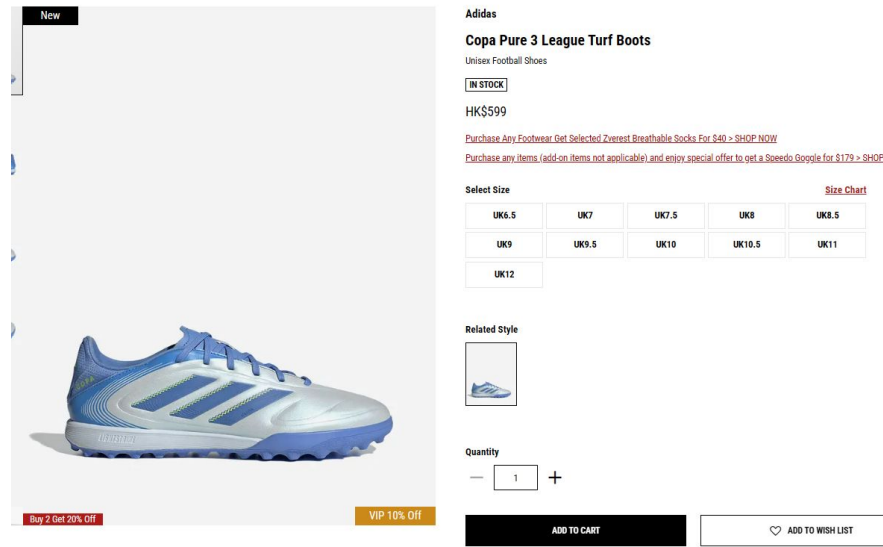
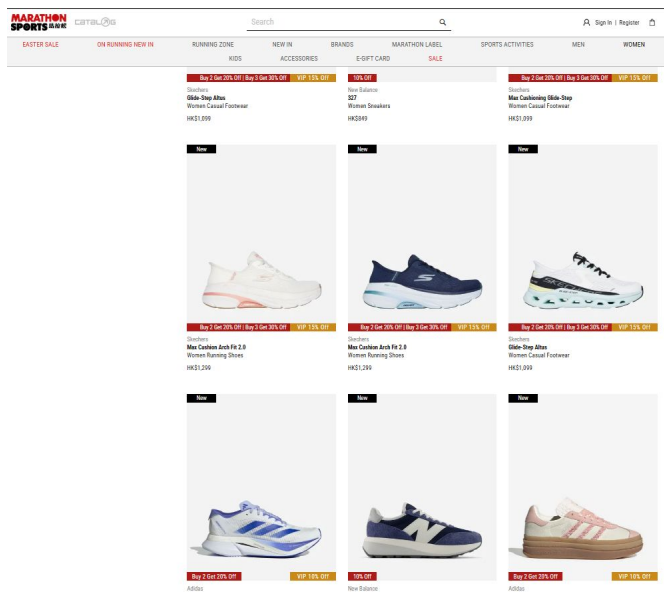
- 使用 **Python** 和 **Selenium(ChromeDriver)** 進行網頁爬蟲
- 網站動態載入，模擬瀏覽器行為



# 資料爬蟲架構與實現

## 架構設計(1)

- 主程式負責爬取**清單頁面**(每頁的產品列表), 包括品牌、名稱、價格等
- 設計多線程(最多 4 個)爬取**每個產品的內頁**, 獲取鞋碼和 SKU, 提升效率



# 資料爬蟲架構與實現

## 架構設計(2)

- 設計鞋碼轉換表(US & EU 到 UK)和折扣加權表, 方便資料處理

### | Size Chart

Adidas Men's and Women's Footwear Size Chart

UK	EU	US - MEN	US - WOMEN	HEEL-TOE (CM)
3	35.5	3.5	3.5	21.6 cm
3.5	36.0	4	5	22.1 cm
4	36.7	4.5	5.5	22.5 cm
4.5	37.3	5	6	22.9 cm

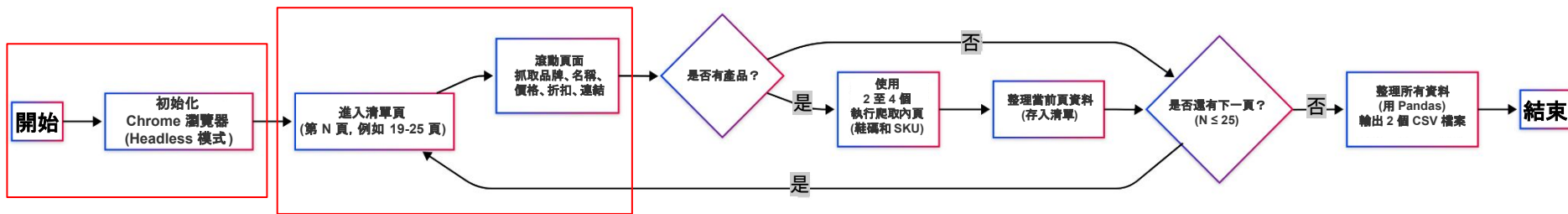
```
discount_weights = {  
    "vip 10% off": 90,  
    "vip 15% off": 85,  
    "10% off": 90,  
    "buy 1 get 10% off": 90,  
    "buy 2 get 20% off": 80,  
    "buy 1 get 15% off": 85,  
    "buy 3 get 30% off": 70,  
    "buy 2 get 10% off": 90,  
    "no discount": 100
```

- 資料儲存使用 Pandas, 將爬取結果存成 CSV 檔案, 方便後續分析





# 資料爬蟲架構與實現 程式流程(1)



## 步驟 1: 初始化

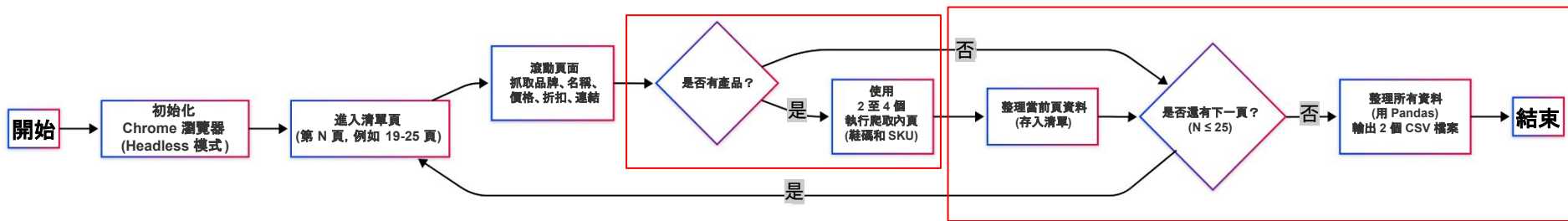
設置 Chrome 選項 (語言為英文, 後來加入 **Headless** 模式), 並打開 **Catalog** 和 **馬拉松** 的主頁面

## 步驟 2: 清單頁面爬取

從第 1 頁開始(後來改為指定範圍, 例如 19 到 25 頁), 滾動頁面載入所有產品, 抓取 **品牌、名稱、價格、折扣和產品連結**

# 資料爬蟲架構與實現

## 程式流程(2)



### 步驟 3: 內頁爬取(多線程)

對每個產品連結, 使用多執行進入內頁, 抓取鞋碼(檢查是否有貨)和 SKU

### 步驟 4: 資料整理

將所有資料存入清單, 轉成 Pandas DataFrame, 分成兩個表格輸出為 CSV 檔案

# 資料爬蟲架構與實現

## 執行細節

- 使用**等待頁面元素載入**(**WebDriverWait**), 例如產品清單和鞋碼選項, 確保資料抓取完整。
- 加入**隨機延遲**(例如 2-4 秒), 模擬人類行為, 避免被網站封鎖。
- 改進**頁面跳轉方式**, 從點擊「下一頁/頁數」按鈕改為直接訪問 URL (例如 ?p=19), 更穩定且效率更高。



← → × ⓘ marathonsports.hkstore.com/marathon\_en\_us/women/footwear?p=1

# 資料爬蟲架構與實現

## 執行結果

- 成功每日爬取 100 頁 共**2400 對鞋**的資料, 每頁 24 對鞋, 資料包括品牌、名稱、價格、折扣、鞋碼和 SKU 等等
- 輸出的 **CSV 檔案**格式大致符合要求, 並使用Excel作小修改後, 可以直接用於後續資料庫分析



# 資料爬蟲架構與實現

## 挑戰

- 挑戰 1: 資源耗盡問題

設定 4 個執行同時爬取內頁，結果在第 120 個產品時，程式報錯 (`chrome not reachable`), 表示 Chrome 崩潰  
原因是開太多瀏覽器，電腦的記憶體和 CPU 負擔太重

解決方法：

1. 更改執行緒數量從 4 減少到 2，降低資源壓力
2. 加入記憶體監控(用 `psutil` 檢查使用率)，確保不會過載

# 資料爬蟲架構與實現

## 挑戰

- 挑戰 2: 瀏覽器連線失敗

偶爾會遇到無法連接到 Chrome 的問題，特別是在爬取內頁時可能是 ChromeDriver 版本與 Chrome 不匹配，或者瀏覽器沒正確關閉

解決方法：

加入重試機制(`webdriver.Chrome()`)，最多重試 3 次，每次失敗後等待 2 秒，確保每個執行結束後都關閉瀏覽器(`driver.quit()`)。

# 資料爬蟲架構與實現

## 挑戰

- 挑戰 3: 頁面載入不穩定

有些產品頁面**載入很慢**，導致鞋碼或 SKU 抓取失敗，經常抓空白資料

解決方法：

1. 內頁延遲從 1-2 秒改為 2-4 秒，確保頁面完全載入
2. 延長等待時間(**WebDriverWait**)，從 15 秒改 20 秒，提高抓取成功率
3. 加入異常處理，即使抓取失敗也記錄為「無 SKU」或「鞋碼抓取失敗」，不影響整體流程。

# 資料爬蟲架構與實現

## 挑戰

- 挑戰 4: Headless 模式的限制

一開始用正常模式(非 Headless)測試, 方便觀察, 但發現資源消耗大且速度慢

改用 Headless 模式後, 速度提升, 但有些頁面會載入失敗

解決方法:

保留 Headless 模式, 但加入更多等待時間和重試機制, 確保穩定性。  
必要時可以切回正常模式進行除錯。



# 資料庫 Shoes Database





# 爬蟲的資訊量

Register as new member to get a HK\$100 Coupon. > Register Now

Free Home delivery for order over \$800

**MARATHON SPORTS 馬拉松 CATALOG** Search Sign In | Register

**CATALOG** Search Sign In | Register

EASTER SALE ON RUNNING NEW IN RUNNING ZONE NEW IN BRANDS MARATHON LABEL SPORTS ACTIVITIES MEN WOMEN KIDS

ACCESSORIES E-GIFT CARD SALE

DISNEY WINNIE THE POOH COLLECTION HOT MODELS LATEST BRANDS CATALOG LABEL CROSSOVER COLLECTION MEN

WOMEN KIDS E-GIFT CARD SALE

Home > Men > Footwear

ear

**Footwear**

< Filter

Category +

Product category +

Hot Model +

footwear type +

Brand +

Gender +

Activity +

Function +

Color +

Price +

Size +

Items 1-24 of 2441 Sort By First Delivery Date

**10% OFF**

**資料庫**

**PUMA**

**1 2 3**

**GORE-TEX**

**Buy 1 Get 15% Off / Buy 2 Get 20% Off**

PUMA come from: PUMA merchant

Fast-Trac NITRO™ 3 Trail Men's Running Shoes Men

HK\$999

**Buy 1 Get 15% Off / Buy 2 Get 20% Off**

PUMA come from: PUMA merchant

Fast-Trac NITRO™ 3 Trail Men's Running Shoes Men

HK\$999

**Buy 1 Get 15% Off / Buy 2 Get 20% Off**

PUMA come from: PUMA merchant

Fast-Trac NITRO™ GORE-TEX® Men's Running Shoes Men

HK\$1,099

**Buy 1 Get 15% Off / Buy 2 Get 20% Off**

PUMA come from: PUMA merchant

Fast-Trac NITRO™ 3 Trail Women's Running Shoes Women

HK\$999

**Buy 2 Get 10% Off**

PUMA come from: PUMA merchant

Fast-Trac NITRO™ GORE-TEX® Women's Running Shoes Women

HK\$1,099 HK\$659

**Buy 2 Get 10% Off**

PUMA come from: PUMA merchant

Fast-Trac NITRO™ 3 Trail Women's Running Shoes Women

HK\$999 HK\$599

兩個網站上有的數據→統一資料庫儲存→市場分析

# 解決方案 - ETL

**E**xtract: 爬蟲

**MARATHON**  
**SPORTS** 馬拉松



**CATAL**  **G**

**T**ransform: 清潔 + 整理



非時效性資訊

時效性資訊

**L**oad: 分析

分析



# 數據轉換 - Transform Data

Web Scraper

**MARATHON**  
**SPORTS** 馬拉松

**CATAL**  **G**

python

output: csv 檔案

clean

Shoes DB



資料庫

postgreSQL

input: csv 檔案

store

shoes

非時效性資訊

scrape\_data

時效性資訊

# 數據選擇

- 時效性資訊 = 根據搜集日子有機會更變

- 非時效性資訊 = 不會根據搜集日子而改變

**MARATHON SPORTS 馬拉松 CATALON**

Search

EASTER SALE ON RUNNING NEW IN RUNNING ZONE NEW IN BRANDS MARATHON LABEL SPORTS ACTIVITIES MEN WOMEN KIDS

ACCESSORIES E-GIFT CARD SALE

**網頁**

**男/女裝**

**價格**

**品牌**

**鞋名**

**鞋款**

**鞋碼**

**折扣**

**排序(第幾對鞋顯示)**

**SKU**

Skechers  
Go Walk Arch Fit 2.0  
Men Casual Footwear  
IN STOCK  
HK\$899  
Purchase Any Footwear Get Selected Zverest Breathable Socks For \$40 > SHOP NOW  
Purchase any items (add-on items not applicable) and enjoy special offer to get a Speedo Goggle for \$179 > SHOP NOW

Select Size

US2	US4	US6	US7	US8
US9	US10	US11		

Size Chart

Related Style

Quantity

— 1 +

ADD TO CART

ADD TO WISH LIST

Share SKU# 185216518NVY

f

WhatsApp

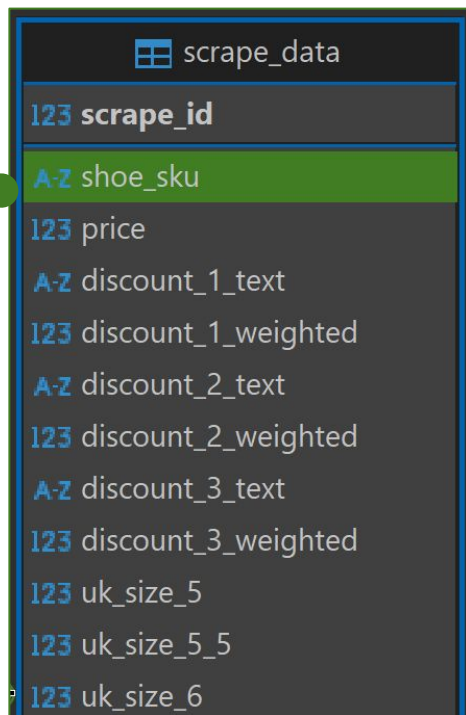
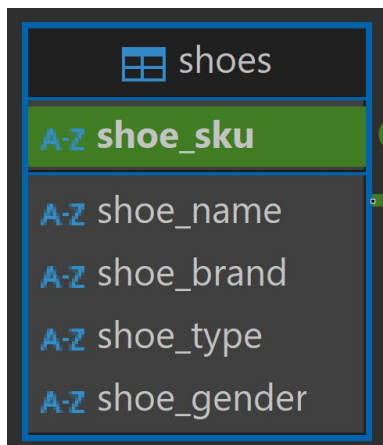
Telegram

Buy 2 Get 20% Off | Buy 3 Get 30% Off

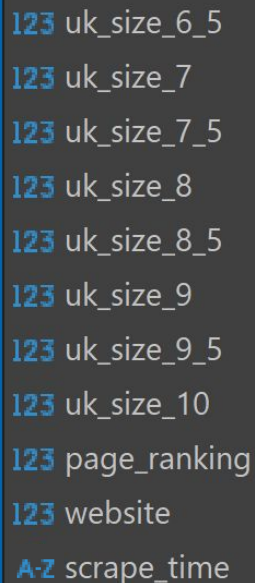
VIP 15% Off

# 資料庫設計 - Shoes DB ER Diagram

shoes  
非時效性資訊



scrape\_data  
時效性資訊





# 數據清理 - Data Cleaning 1

確保 **shoes\_sku** 在 shoes table 是**獨一無二**的

1. Create temp\_table with **SELECT DISTINCT**
2. Replace temp\_table to shoes\_table\_csv
3. Add Primary Key



```
CREATE TABLE temp_table AS  
SELECT DISTINCT ON (shoe_sku) *  
FROM "shoes_csv";
```

```
DROP TABLE "shoes_csv";
```

```
ALTER TABLE temp_table RENAME  
TO "shoes_csv";
```

```
ALTER TABLE "shoes_csv"  
ADD PRIMARY KEY (shoe_sku);
```



# 數據清理 - Data Cleaning 2

確保加進去 shoes table 的 **shoes\_sku** 不會重複

1. Insert column by column
2. Using **WHERE NOT EXISTS**



```
INSERT INTO shoes (shoe_sku, shoe_name, shoe_brand, shoe_type, shoe_gender)
SELECT shoe_sku, shoe_name, shoe_brand, shoe_type, shoe_gender
FROM "shoes_csv" s
WHERE NOT EXISTS (
    SELECT 1
    FROM shoes
    WHERE shoe_sku = s.shoe_sku);
```

# 數據清理 - Data Cleaning 3

## 確保 **shoe\_type** 在正確位置

1. Set Case: set shoe\_type...
2. when shoe\_name ILIKE then...



```
UPDATE shoes
SET shoe_type = CASE
    WHEN shoe_name ILIKE
'%Sneakers%' THEN 'Sneakers'
    WHEN shoe_name ILIKE
'%Slip-on%' THEN 'Slip-on Shoes'
    ...
    ELSE shoe_type
END
WHERE shoe_brand = 'PUMA come
from PUMA merchant'
```

# 數據清理 - Data Cleaning 4

## 替換中文字 -> 英文字 in shoes\_type

1. Set Case: set shoe\_type...
2. when shoe\_type ILIKE '中文字' then...



```
UPDATE shoes
SET shoe_type = CASE
    WHEN shoe_type ILIKE '%高筒帆布鞋%' THEN 'Canvas Shoes'
    WHEN shoe_type ILIKE '%登山鞋%' THEN 'Hiking Shoes'
    ...
    ELSE shoe_type
END
```

# 資料庫設計 - shoes Primary Key: shoe\_sku

**shoe\_sku:** 112U574RBJ

- shoe\_brand: New Balance
- shoe\_name: 574
- shoe\_type: Sneakers
- shoe\_gender: Unisex



# 資料庫設計 - scrape\_data Primary Key: scrape\_id

scrape\_id: 22041527

A. → B. → C. → D.

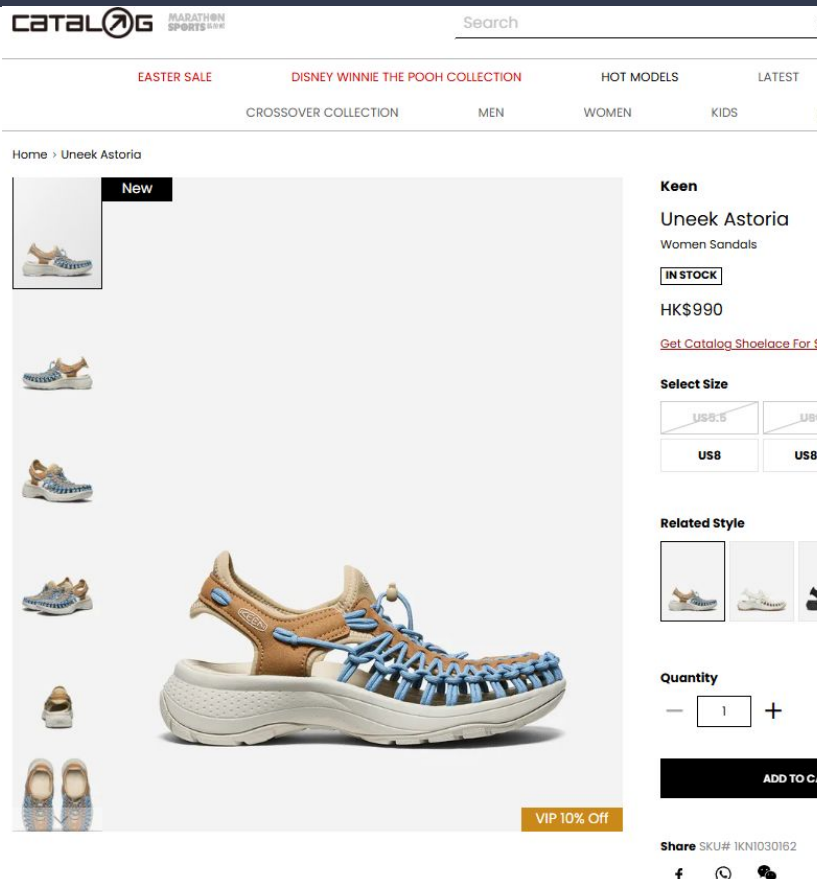
scrape\_id: 每一個爬取記錄都是獨一無二的!

A. 網店編號: 1 = marathon, 2 = catalog

B. 男女裝頁面: 1 = men, 2 = women

C. 爬蟲日期: MMDD

D. 頁面順序: 自然數, 由0開始, i.e. 0, 1, 2...



# 資料庫設計 - discount

總共最多有

3 重折扣

Skechers

**Max Cushion Premier 2.0**

Men Running Shoes

IN STOCK

HK\$1,099



discount\_1

discount\_2

discount\_3

Buy 2 Get 20% Off | Buy 3 Get 30% Off


VIP 15% Off

# 資料庫設計 - discount

- discount\_text = 折扣的描述
- discount\_weighting = 折扣的計算百分比%
- 計算鞋子折實價
- 推測網店多想賣這款鞋子

折實價 = price \* discount\_weighting

- price = 1099 (\$HKD)
- if buy 2, price = 1099\*80% (discount\_1\_weighting)
- if buy 3, price = 1099\*70% (discount\_2\_weighting)
- if VIP, price = 1099\*85% (discount\_3\_weighting)



discount\_1 discount\_2 discount\_3

Buy 2 Get 20% Off | Buy 3 Get 30% Off VIP 15% Off

**Skechers**

**Max Cushion Premier 2.0**

Men Running Shoes

**IN STOCK**

HK\$1,099



## 資料庫設計 - uk\_size

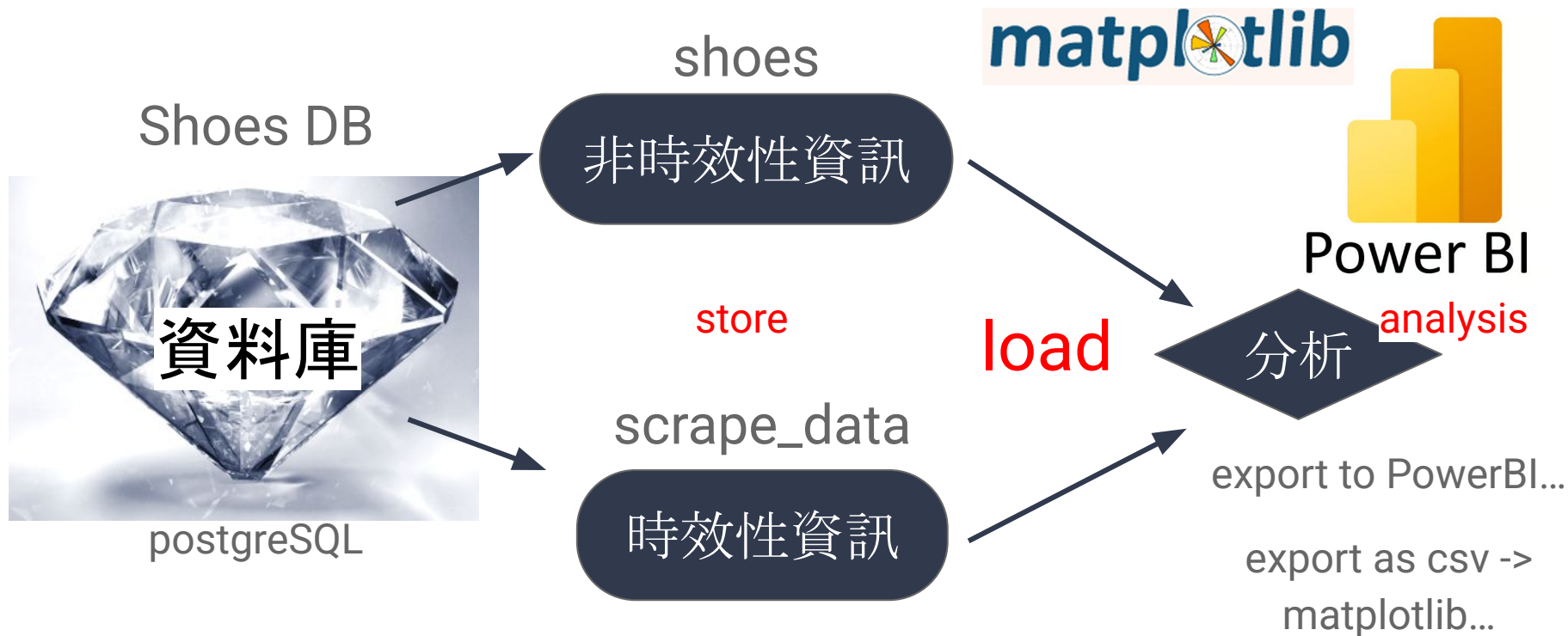
- 用uk\_size來統一計算, 方便比較
- uk\_size\_5 - uk\_size\_10, 涵蓋主流男女裝鞋子尺碼
- Categorise, 省卻位置, 用 0, 1, 2 來代表:
  - 0 = 沒有該尺碼提供
  - 1 = 有該尺碼提供但暫時缺貨
  - 2 = 有該尺碼且現在有貨
- 分析普遍受歡迎/不受歡迎的鞋碼
- 分析賣剩/熱賣/冷門/絕版鞋款

**OUT OF STOCK**

US7	UK5	EU37.5
US7.5	UK5.5	EU38
US8	UK6	EU39
US8.5	UK6.5	EU39.5
US9	UK7	EU40
US9.5	UK7.5	EU41
US10	UK8	EU41.5
US10.5	UK8.5	EU42
US11	UK9	EU42.5
US11.5	UK9.5	EU43
US12	UK10	EU44



# 數據提取 - Load



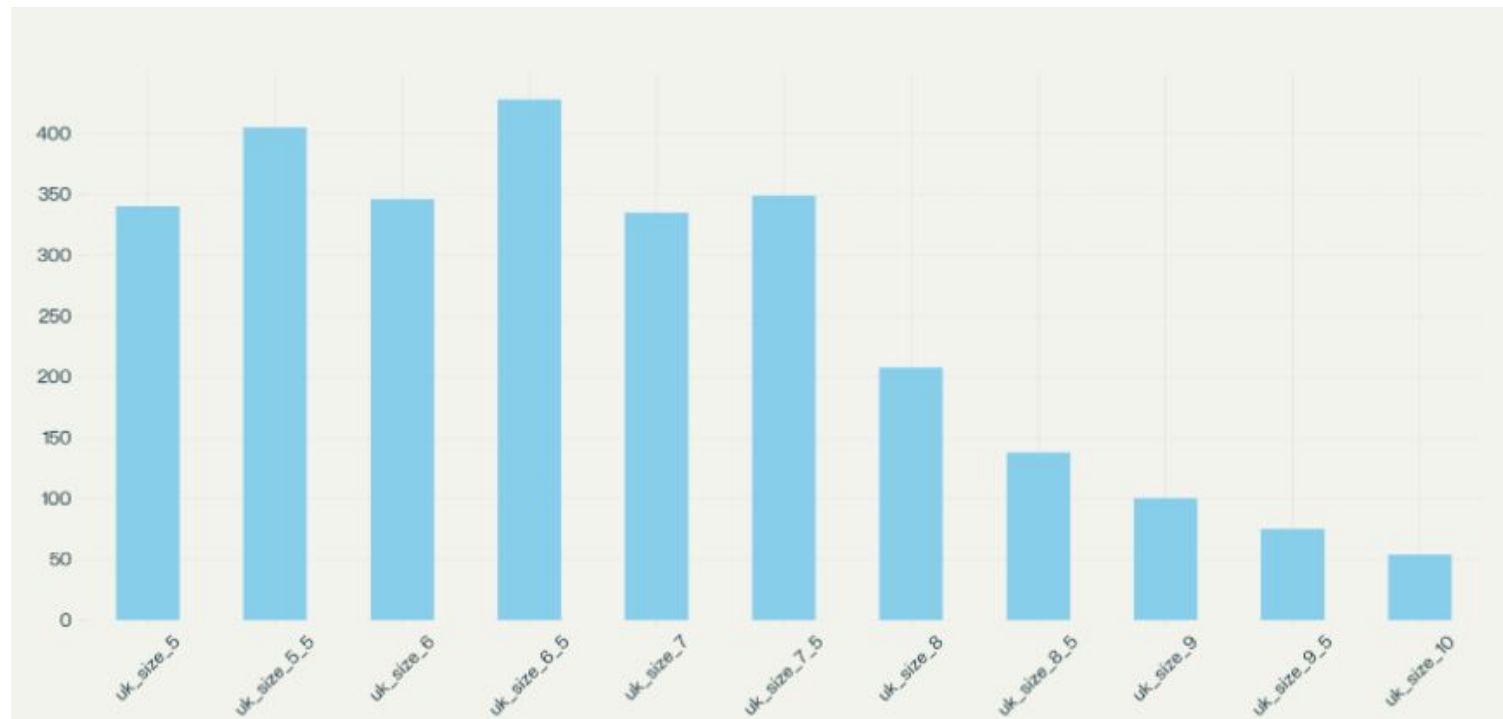
# 分析成果



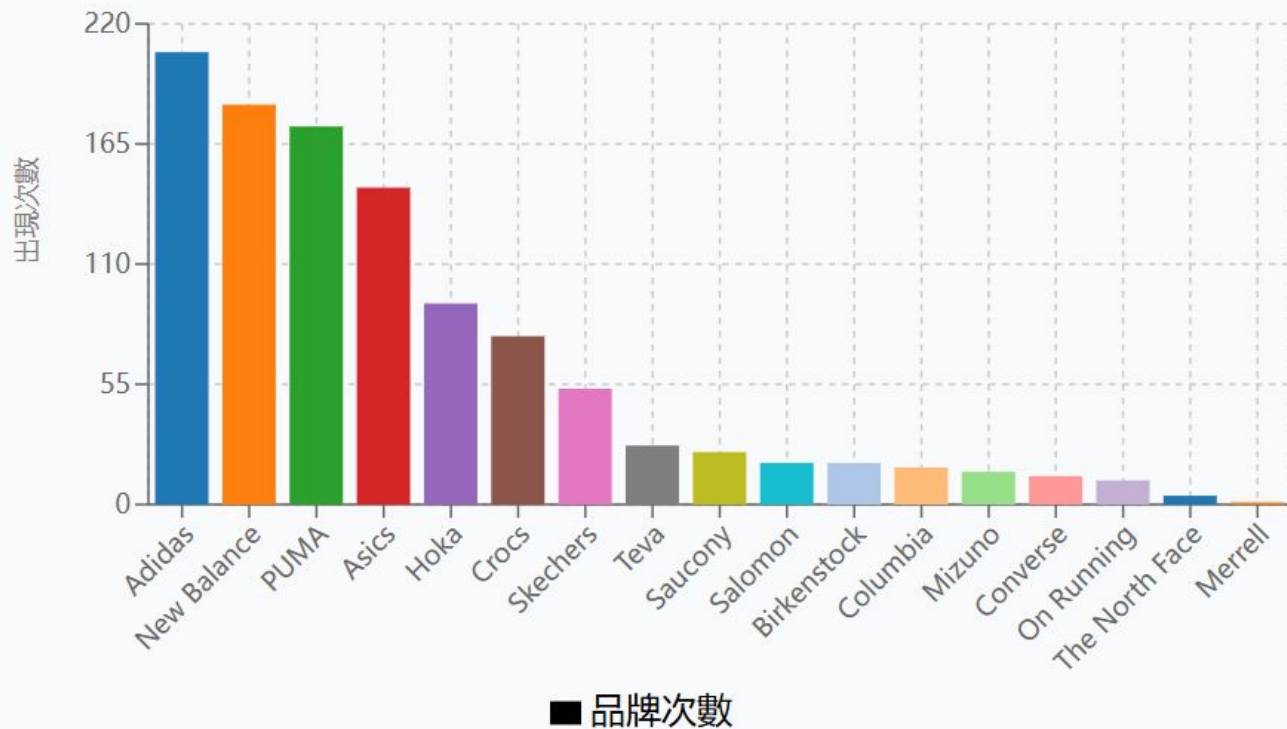
# 分析成果：香港男生普遍鞋碼



# 分析成果：女生普遍鞋碼

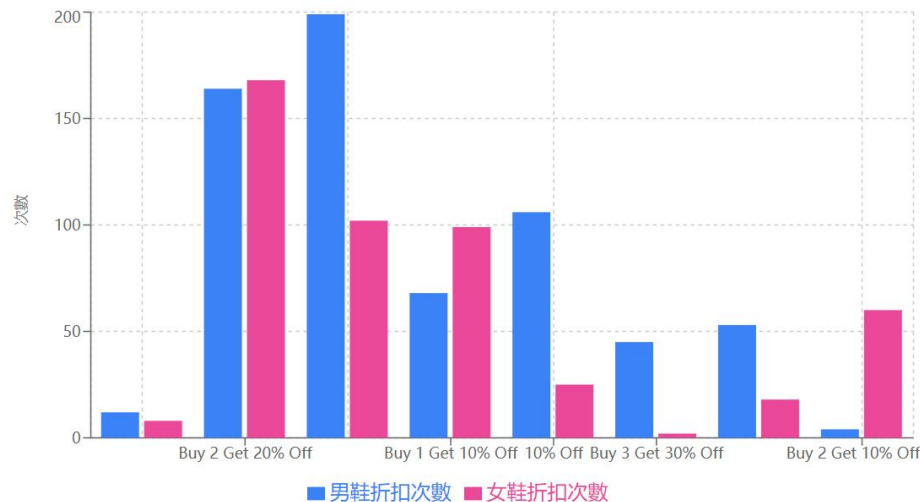


# 分析成果:各牌子在marathon和catalog的市佔率



# 分析成果:

## 促銷活動和折扣力度

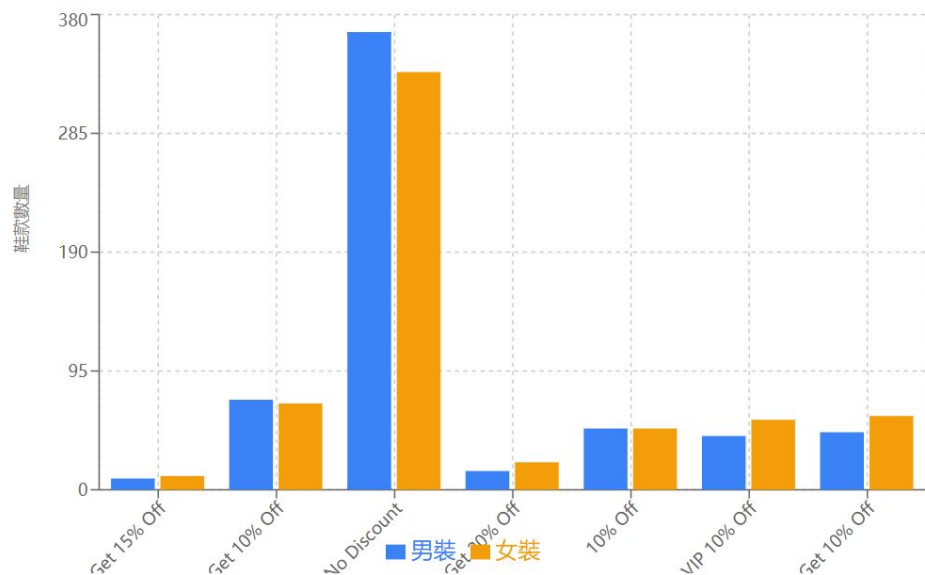


## Marathon:

- New Balance買2件8折, Skechers買3件7折, 滿HK\$1100減HK\$100
- 快閃折扣:跑鞋低至3折, 精選跑鞋7折左右
- 會員專屬優惠
  - 偏向專業跑步鞋, 運動鞋
  - 針對運動愛好者, 專業運動員等消費群

# 分析成果:

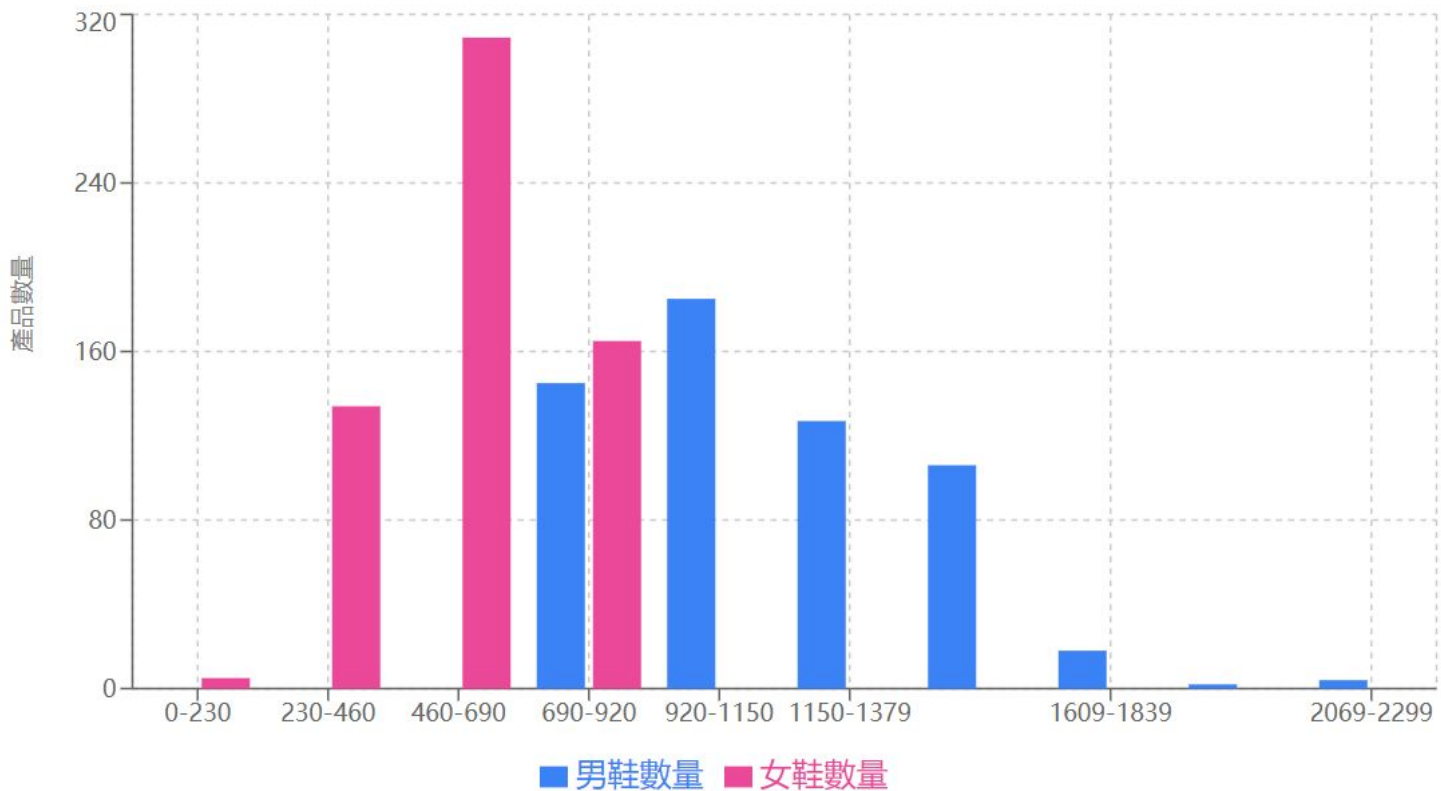
## 促銷活動和折扣力度



## Catalog:

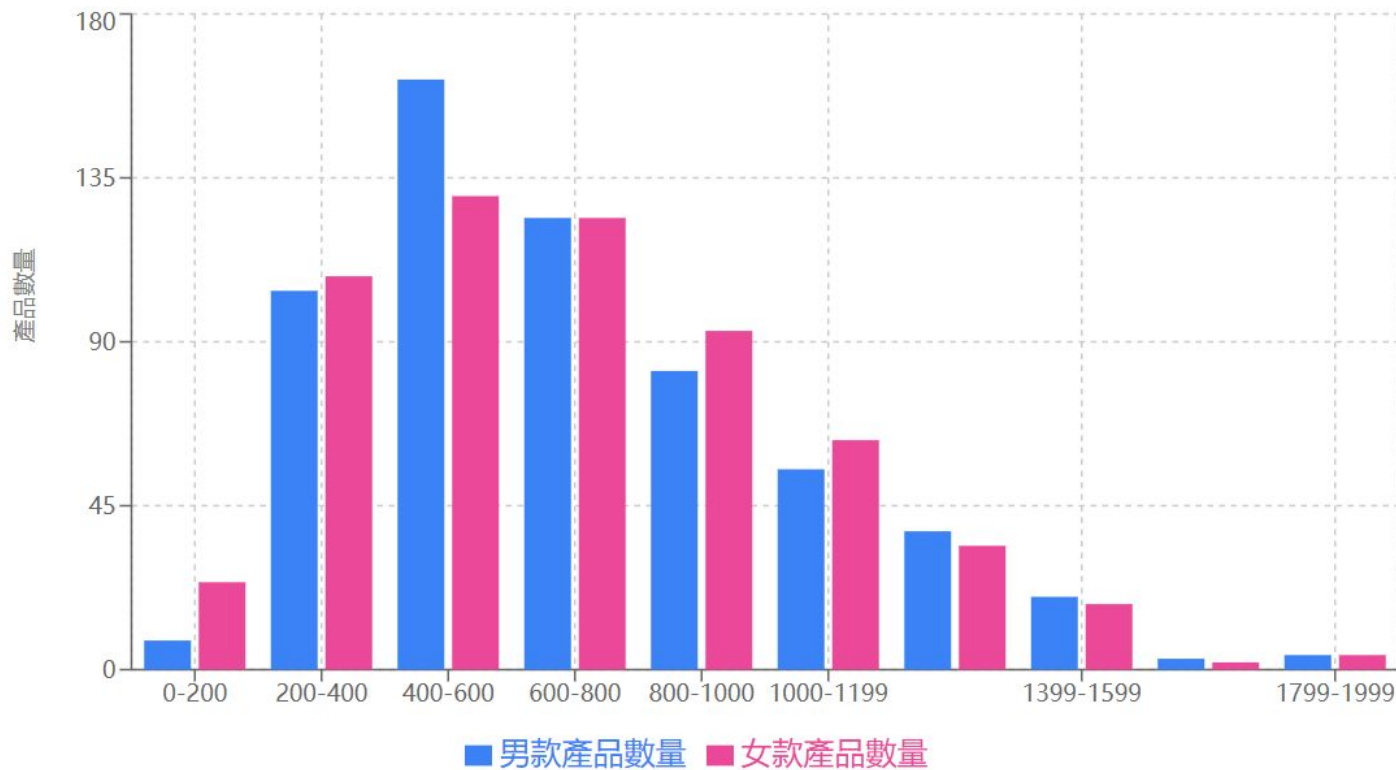
- 電商額外折扣: 限時折扣5折, 優惠碼疊加使用
- New Balance低至5折, 部分鞋款VIP會員10折
- 滿HK\$700免運費, 滿HK\$800減HK\$100
- Kuromi、Pokémon聯乘鞋款折扣40%-70%
  - 偏向潮流休閒鞋
  - 針對廣大網上年輕消費者

# 分析成果: Marathon價格區間





# 分析成果: Catalog價格區間



# 分析成果：新出的款式共同點

Marathon:

總數~1200(男女)

~700 sports shoe

~340running shoes

Catalog:

總數~1200

~750 sports shoes

~400 running shoes

可見香港本地對運動鞋有更高需求，使 catalog 的市場定位有所偏移

# 分析成果:

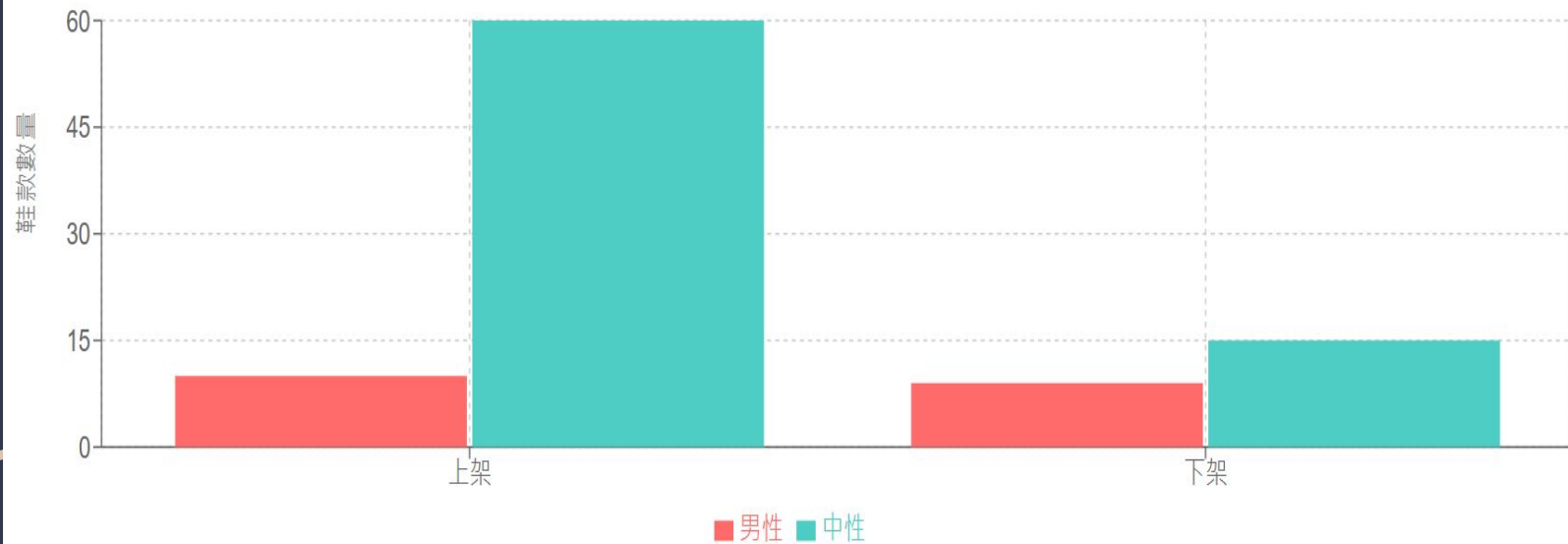
## 貨品上新落架

鞋款名稱	品牌	性別	價格	SKU	下架時間
Nike Air Zoom Spiridon Barely Volt Hot Punch	Nike	Men	\$160	SP7936-002	20250421
Nike Ja 2 Ja-Breaker PE	Nike	Men	\$170	JA7936-003	20250421
adidas Harden Vol 9 'Hell Cat'	adidas	Men	\$160	JH6483	20250421
New Balance 990v6 Made in USA (Avocado/Natural Mint)	New Balance	Unisex	\$200	U990AB6	20250421
Nike Air Max 1 'Big Head Mode'	Nike	Men	\$160	IB9898-001	20250421

鞋款名稱	品牌	性別	價格	SKU	上架時間
Nike Total 90 III SP (Safety Orange/Medium Gum Brown-White)	Nike	Men	\$100	HJ9351-800	20250415
Nike Total 90 III SP (Dynamic Yellow/Green Spark-Gum Medium Brown)	Nike	Men	\$110	HJ9351-700	20250415
adidas Superstar (Blue/White)	adidas	Unisex	\$160	JR2702	20250415
adidas Superstar (White/Black)	adidas	Unisex	\$160	JR2701	20250415
Nike LeBron 22 'Frozen Grapes'	Nike	Men	\$180	HV8454-001	20250415

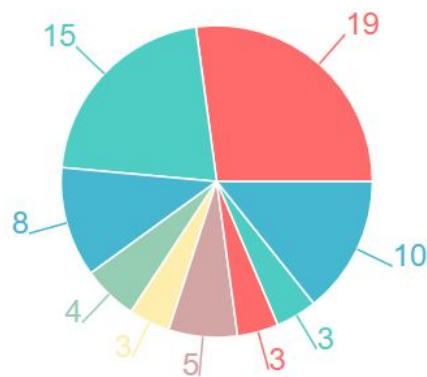
# 分析成果:

## 貨品上新落架



# 分析成果:

## 貨品上新落架



■ PUMA ■ New Balance ■ Adidas ■ Salomon ■ Crocs ■ Converse ■ Asics ■ Birkenstock ■ Hoka

# Q&A