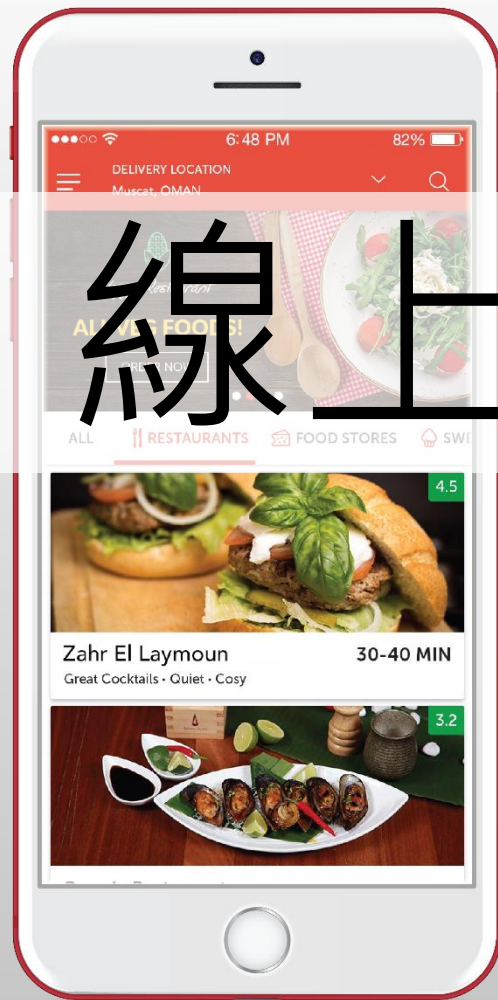


# 線上外送平台



akeed

# 交易分析與推薦系統

	組員一	組員二
姓名	林介友	楊昇達
經歷	關貿網路公司【資料分析師】 台琦環境工程【資料分析師】	海洋大學環境暨漁業科學系
專長	統計、資料科學、資料分析、資料視覺化、資料庫管理	
使用軟體	<b>MySQL, MariaDB, Python, R, SPSS, Linux</b>	<b>SQL, linux ,hadoop</b>

# 分組工作

組員	資料蒐集 匯入MariaDB	資料清理 R	分析 R	視覺化設計 R	報告製作/說明
林介友		○	○	○	○
楊昇達	○	○		○	○

# 背景說明

- 外送**APP**用戶：

近來，**外送APP**平台如雨後春筍般地增加，顧客選擇更為多元，但選擇到中意餐點的**難度與時間**增加。

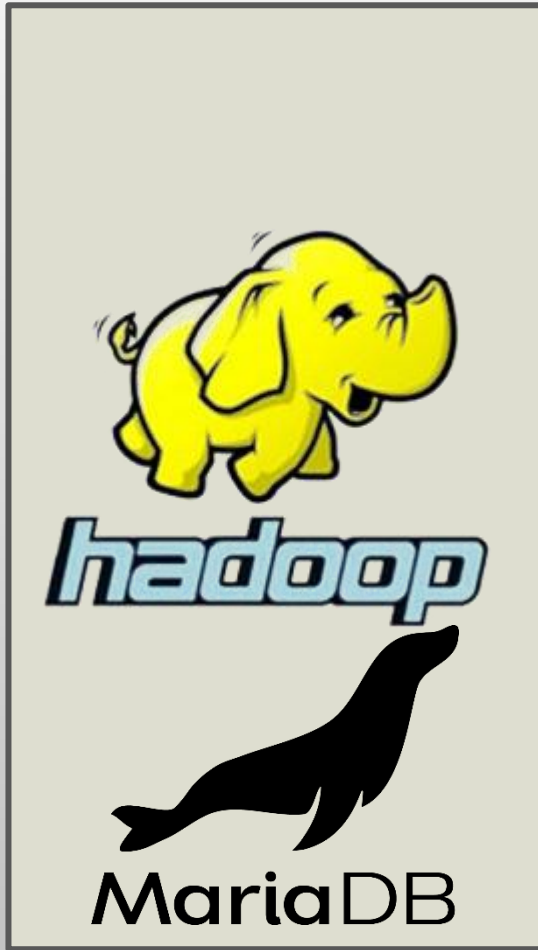
- **APP**合作餐廳

面對外送經濟的趨勢，許多餐廳欲了解**外送生意興隆的關鍵**。

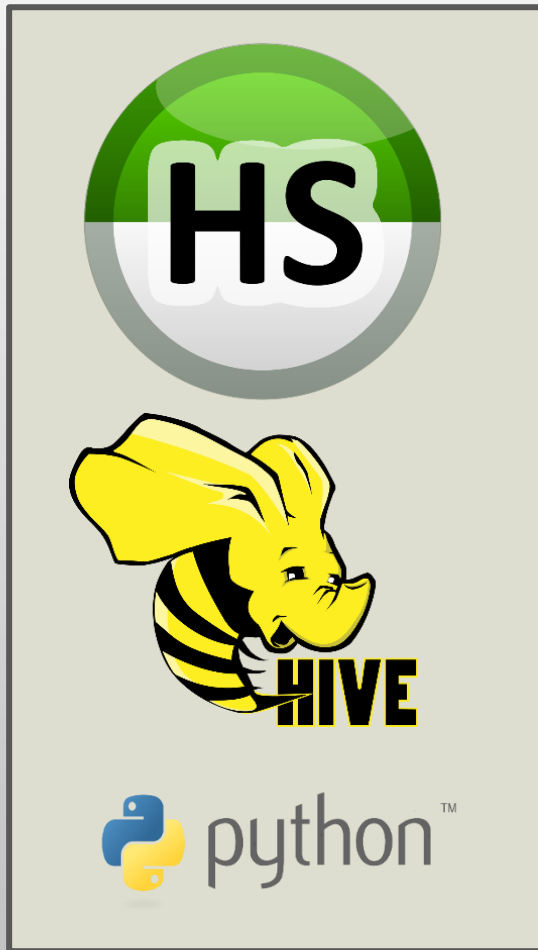
# 專題目標

1. 探討影響外送**消費金額**、餐廳**評分**的關鍵因子
2. 利用歷史訂單資料**推薦**給顧客中意的外送餐廳，使顧客更快地決定下訂單，意味著更高的商業潛力

# 系統架構



進資料庫



資料預處理



資料分析



特徵工程

# 資料說明

- 資料：**Akeed**外送**APP**用戶資料
- 資料來源：**Kaggle**網站([Restaurant Recommendation Challenge](#))
- 訂單資料：共**135,303**筆，**27**個欄位
- 餐廳資料：共**100**筆，**59**個欄位
- 顧客資料：共**34,647**筆，**8**個欄位，包含**App**客戶**ID**、性別、出生年、使用語言，和帳號當前狀態、認證狀態、建立與更新時間。

# 資料說明

訂單資料(27欄)		
顧客相關	運送相關	交易相關
顧客ID	訂單編號	購買物品數量
是否加餐廳為最愛	訂單創建日期	購買金額
是否有評價	訂單接受時間	付款方式
餐廳的評分	外送員接單時間	付款折扣碼
外送員的評分	取餐時間	折扣量
顧客地點編號	送達時間	折扣比例
顧客訂購位置	運送日期	餐廳相關
	外送距離	餐廳代碼
		備餐時間
		準備好餐點的時間



# 資料說明

餐廳資料(局部)		
位置	類型	帳號
緯度	餐廳型態	餐廳代號
經度	型態代碼	驗證代碼
服務距離	特徵標籤	餐廳帳號狀態
時間相關	排名相關	餐廳帳號認證
準備時間	餐廳評分	餐廳帳號創立日期
營業時間	等級	餐廳帳號更新日期
營業時間2	費用相關	裝置種類
	運費	
	折扣百分比	

# 資料處理步驟

- 1. 辨識欄位**
- 2. 空缺值處理**
- 3. 重複資料處理**

# 資料預處理

**排除**定義不明欄位：如經緯度、裝置種類、驗證代碼

**排除**缺值過多(**90%**以上)的欄位：如出生年份

**排除**與餐廳選擇無直接關聯的欄位：如顧客帳號建立、更新時間

**排除**重複的顧客資料

**新增**依據「訂單建立的時間」欄位，新增「星期」、「時段」欄位

**轉換**將性別中**male**、**Male**、改為**M**；**female**、**Female**、改為**F**

# 餐廳評價關鍵因子分析

**1. 資料筆數：135303筆**

**2. 使用欄位：**物品購買數量、消費金額、支付方式、有無折扣碼、折扣量、折扣比例、是否加入最愛、是否評分、餐廳評價、外送員評分、運送距離、備餐時間、位置代碼、位置類型、顧客性別、帳號狀態、帳號是否認證、**APP**使用語言、訂單建立星期、訂單建立時段，共**20**欄

# 連續變數轉為類別變數

連續變數	轉換類別	標準
購買物品數量	高、中、低	$\geq 3(Q3)$ 為高、 $\leq 1(Q1)$ 為低
購買金額	高、中、低	$\geq 18.5(Q3)$ 為高、 $\leq 8.2(Q1)$ 為低
折扣量	多、一些、無	$\geq 0.5$ 為多、 $> 0$ 為一些、 $0$ 為無
外送距離	長、中、短	$\geq 7.9(Q3)$ 為長、 $0$ 為短
備餐時間	長、中、短	$\geq 45(Q3)$ 為長、 $< 40$ 為短

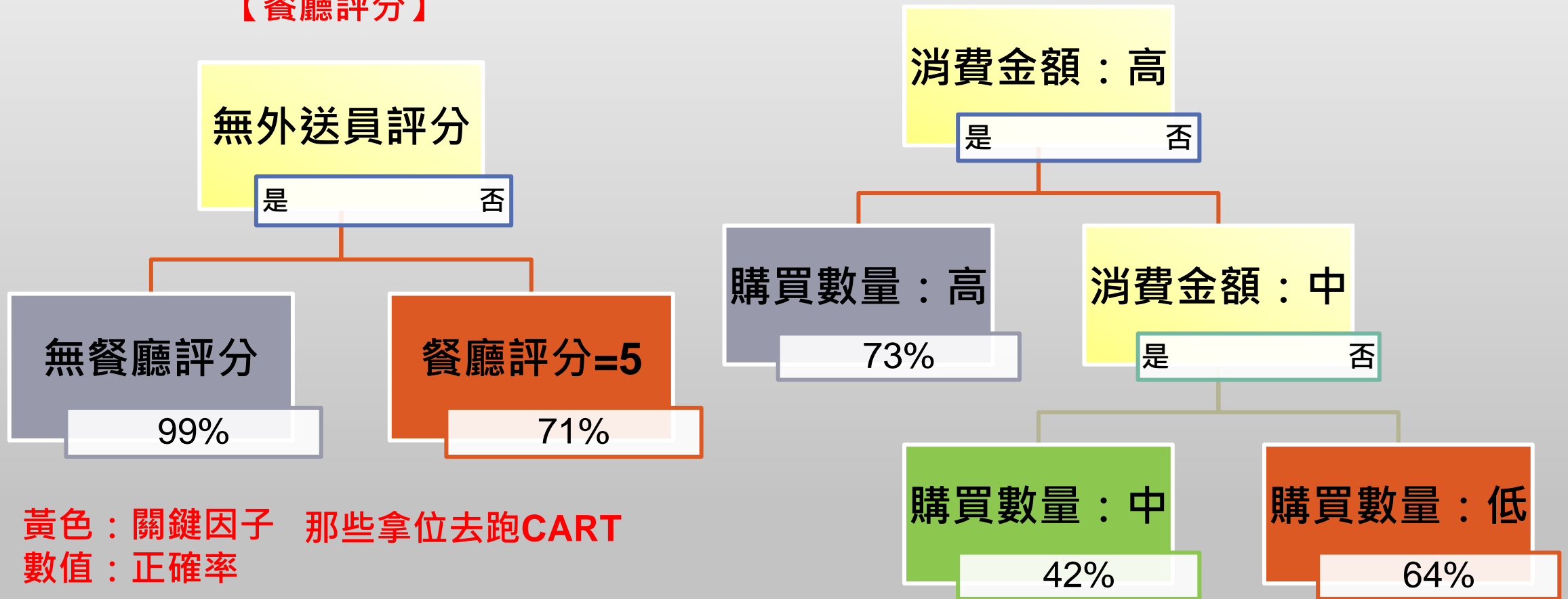
# 餐廳特徵 & 集群

1. 將「餐廳特徵標籤欄位」執行**one-hot-encoding**，新增**64**個標籤欄位
2. 為了將有稀少標籤的餐廳較易歸在同集群，將上述**64**個欄位分別除以各自欄位的總和。
3. 以**kmeans**演算法，採用歐式距離(**Euclidean distance**)對**100**家餐廳做集群，共分為**20**群，並定義為欄位**Cluster\_20**

# CART分析結果 資料分析-分類樹

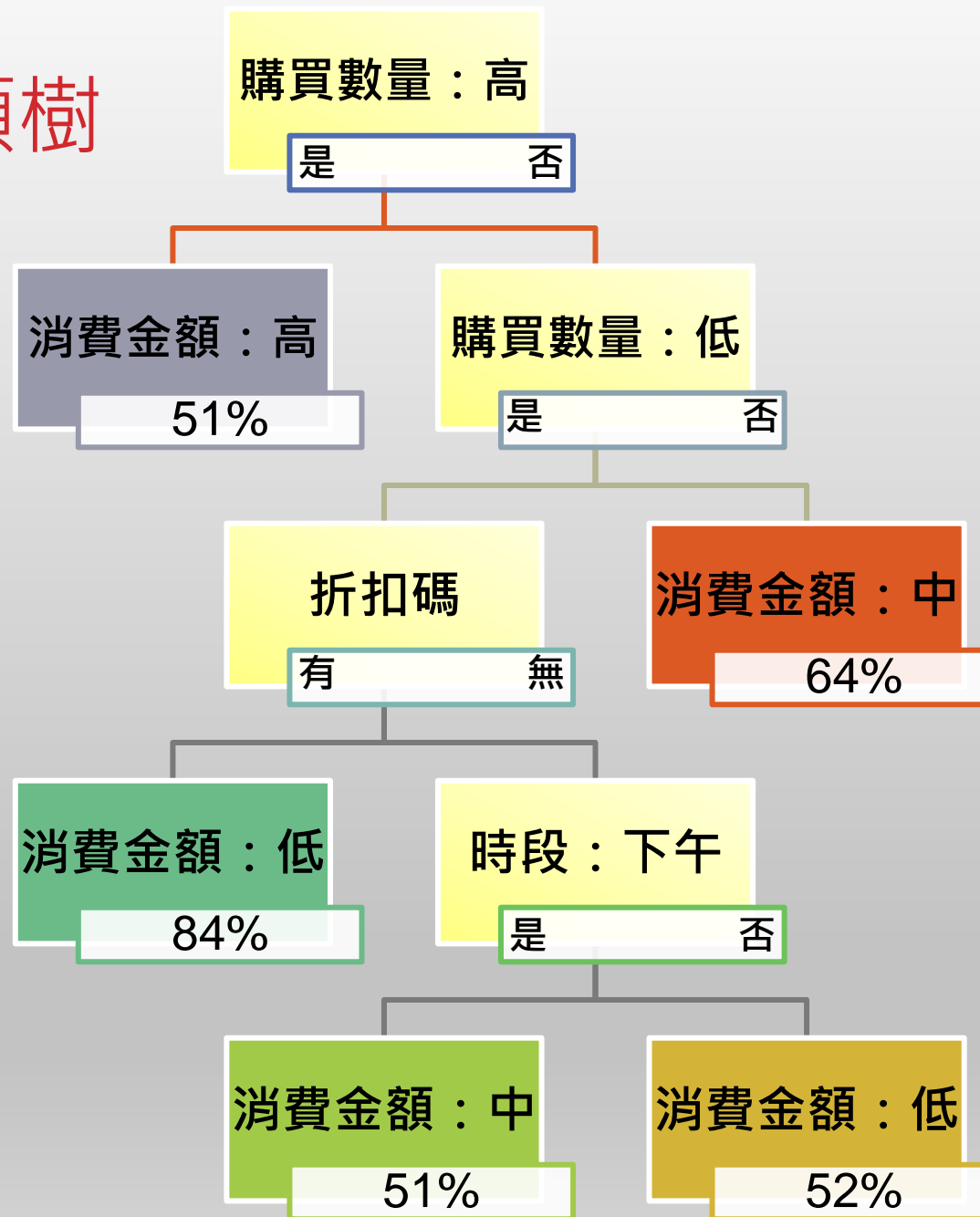
【購買數量】

【餐廳評分】



黃色：關鍵因子 那些拿位去跑CART  
數值：正確率

# 資料分析-分類樹





# 分類樹結果

	關鍵因子	關係
評分	外送員分數	正向(將0作未評分)
購買數量	購買價格	正向

- 購買數量=低、無折扣條件下：  
下午的消費金額(中)多於其他時段(低)

# 關聯規則

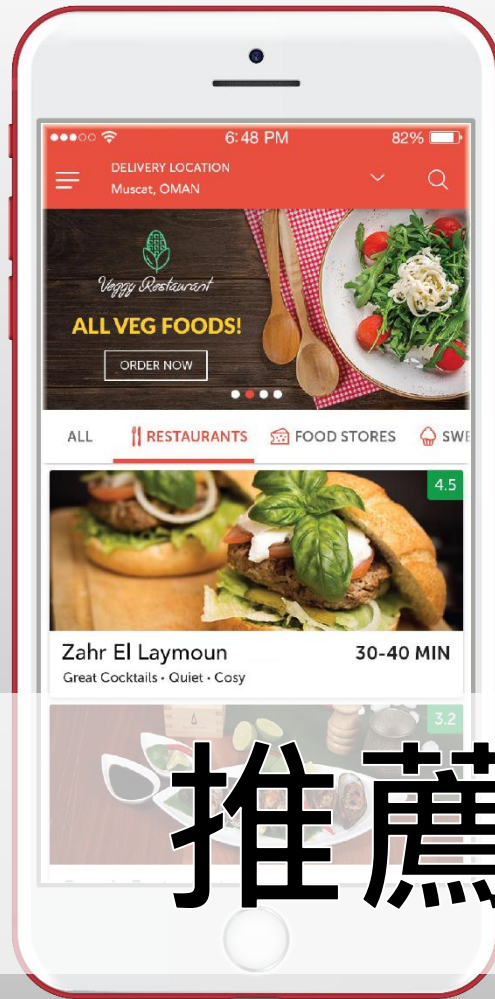
- 關聯規則可用於發現交易資料中購買產品之間的規律性，也可在資料庫中發現變數之間的有趣性關係。
- 例如，從銷售資料中發現規則 {洋蔥, 馬鈴薯} → {漢堡} 表示如果顧客一起買洋蔥和馬鈴薯，他們也有可能買漢堡的肉。
- **R package: arules**
- 參數選擇：

$$\begin{aligned} \text{support} &= P(LHS \cap RHS) \\ \text{confidence} &= \frac{P(LHS \cap RHS)}{P(LHS)} \end{aligned}$$

# 關聯規則結果

## ➤ 結果表格

規則		說明
選 <b>17</b> 號餐廳者 (美式速食類)	高評分	美式餐廳素來在外送業務已具備一定經驗， 推測外送效率上有優勢
男	低評分	可朝向男性感興趣的方向應對，如：增加 餐點份量、廣告噱頭
支付方式 <b>1</b>	低消費	由於方式 <b>1</b> 傾向造成低消費，因此鼓勵採用 多元的支付方式



akeed

# 推薦系統介紹

# 餐廳推薦系統

Option Tabs Restaurant Rank Restaurant Category About

**Gender**

Male  
 Female

**Location Type**

Home  
 Work  
 Other

**Time**

morning  
 afternoon  
 night  
 dawn

**Day**

Sunday ▼

**Payment Mode**

1 5

1 2 3 4 5

Get Ranks



# 餐廳推薦系統

Show  entries      Search:

cluster_20	Score	Pred_time
6	90	2.36
2	85	7.61
19	80	4.84
18	75	2.44
17	65	5.89
5	60	10
12	45	2.39
8	40	6.53
3	35	3.22
4	30	3.97

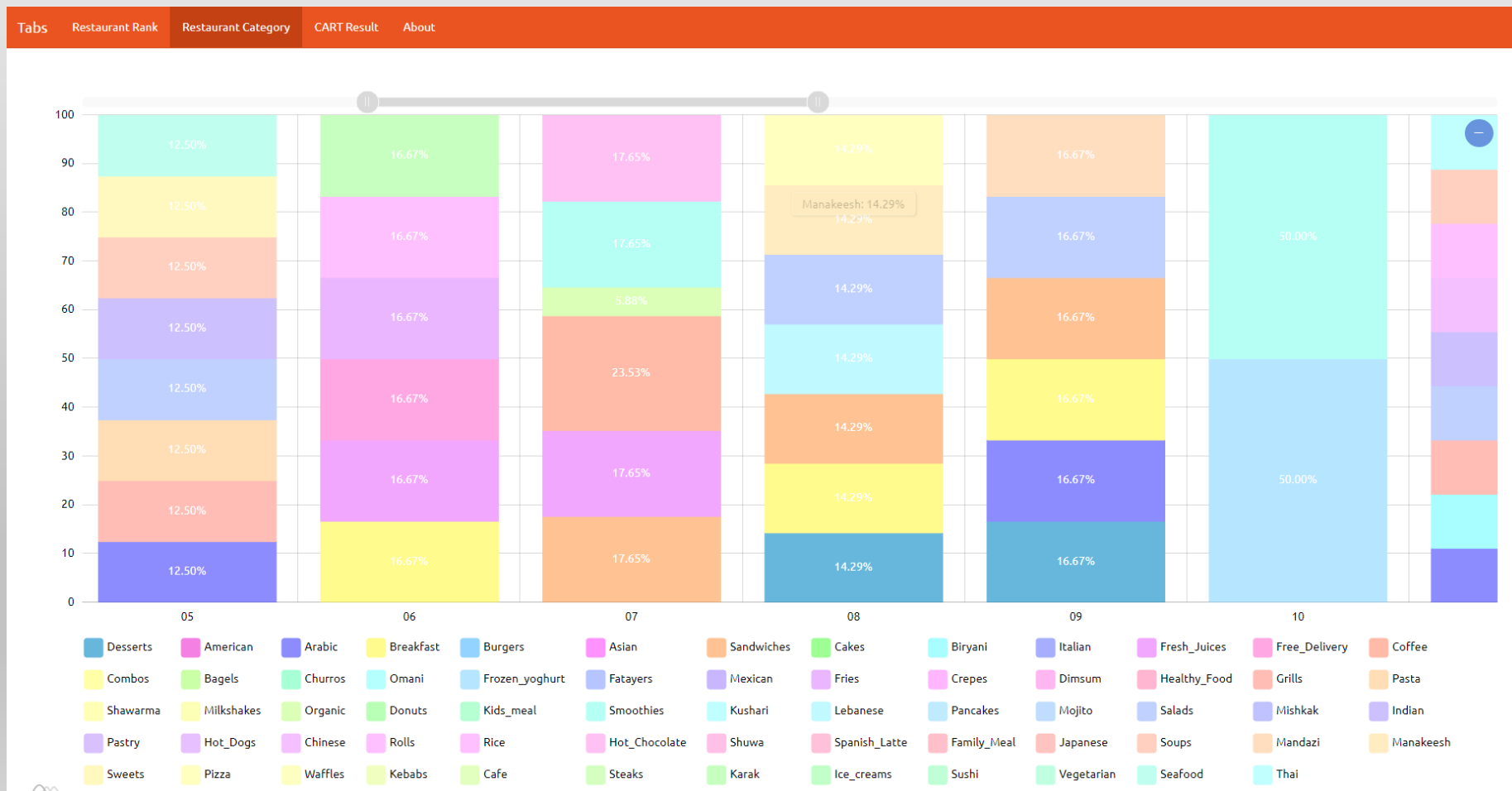
Showing 1 to 10 of 14 entries      Previous            2      Next



# 餐廳推薦系統



可在 **Restaurant Category** 分頁中查詢每個餐廳集群的特徵



# 餐廳推薦系統



可在**Restaurant Category**分頁中查詢每個餐廳集群的特徵

Show  entries Search:

vendor_tag_name	cluster_20
Arabic,Breakfast,Burgers,Desserts,Free Delivery,Grills,Lebanese,Salads,Sandwiches,Shawarma	3
Breakfast,Cakes,Crepes,Italian,Pasta,Pizzas,Salads,Sandwiches,Soups	17
Breakfast,Desserts,Free Delivery,Indian	17
Burgers,Desserts,Fries,Salads	17
Burgers	17
Desserts,Mexican	17
American,Burgers,Fries,Sandwiches	17
American,Burgers,Fries,Sandwiches	17
Breakfast,Desserts,Grills,Milkshakes,Salads,Sandwiches,Soups	3
Breakfast,Burgers,Desserts,Indian,Salads	17

Showing 1 to 10 of 100 entries Previous  2 3 4 5 ... 10 Next