

2017巨量專題報告



710433107 賴彥丞

710533115 張瑛珊

710433109 李昱廷

目錄

- 分析動機
- 公司介紹-Instacart
- 資料來源
- 資料處理
- 分析方法
- 結論與討論

分析動機



- 現今的社會形態下，資訊以及資料環繞著我們，也因為這些資料加上網路的發達，我們以大量資料來做比較分析與預測。我們這組選擇了Instacart的資料集，其中包含了客戶透過他們系統選擇在超市要購買的品項，且詳細記錄購物的時間與商品，因此我們打算透過商品的種類以及品項進行關聯性分析，將數據整理分析，並發想促銷策略，並與合作超市配合，以增進公司系統的使用率，進而培養出客戶使用忠誠度，達到未來業績的增長。

Instacart

- Instacart是2012年6月由先前於 Amazon 擔任過供應鏈工程師的 Apoorva Mehta成立的O2O新創公司，
- 透過他們的平台或是app就能夠透過網路輕鬆訂購生鮮食品
- 在客戶指定送達的時間內交貨，以維持食材的新鮮度，分為一小時和兩小時兩種方案，
- 依送達時間決定運費的收取方式，而送達時間越短當然就必須付較高的運費。
- Instacart就好比是物流業的Uber，所以運費則對他們而言是最主要的收益來源，

資料來源

- 資料出處：Kaggle - Instacart Market Basket Analysis 競賽專案
- 資料型態：共有6個資料檔(皆為csv檔案)
 1. orders.csv
 2. order_products__train.csv
 3. order_products__prior.csv
 4. products.csv
 5. aisles.csv
 6. departments.csv
- 皆為類別型變數

資料來源 - orders.csv

order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
2539329	1	prior	1	2	8	NA
2398795	1	prior	2	3	7	15
473747	1	prior	3	3	12	21
2254736	1	prior	4	4	7	29
431534	1	prior	5	4	15	28
3367565	1	prior	6	2	7	19

- Observations: 3,421,083
- Variables: 7
- Data size : 109 Mb

資料來源 - order_products__train.csv & order_products__prior.csv

order_id	product_id	add_to_cart_order	reordered
1	49302	1	1
1	11109	2	1
1	10246	3	0
1	49683	4	0
1	43633	5	1
1	13176	6	0

- Observations: 1,384,617 & 32,434,489
- Variables: 4
- Data size : 24.7 Mb & 577.6 Mb

資料來源 - product.csv

product_id	product_name	aisle_id	department_id
1	Chocolate Sandwich Cookies	61	19
2	All-Seasons Salt	104	13
3	Robust Golden Unsweetened Oolong Tea	94	7
4	Smart Ones Classic Favorites Mini Rigatoni With Vodka Cream Sauce	38	1
5	Green Chile Anytime Sauce	5	13
6	Dry Nose Oil	11	11

- Observations: 49,688
- Variables: 4
- Data size : 2.2 Mb

資料來源 - aisle.csv

aisle_id	aisle
1	prepared soups salads
2	specialty cheeses
3	energy granola bars
4	instant foods
5	marinades meat preparation
6	other

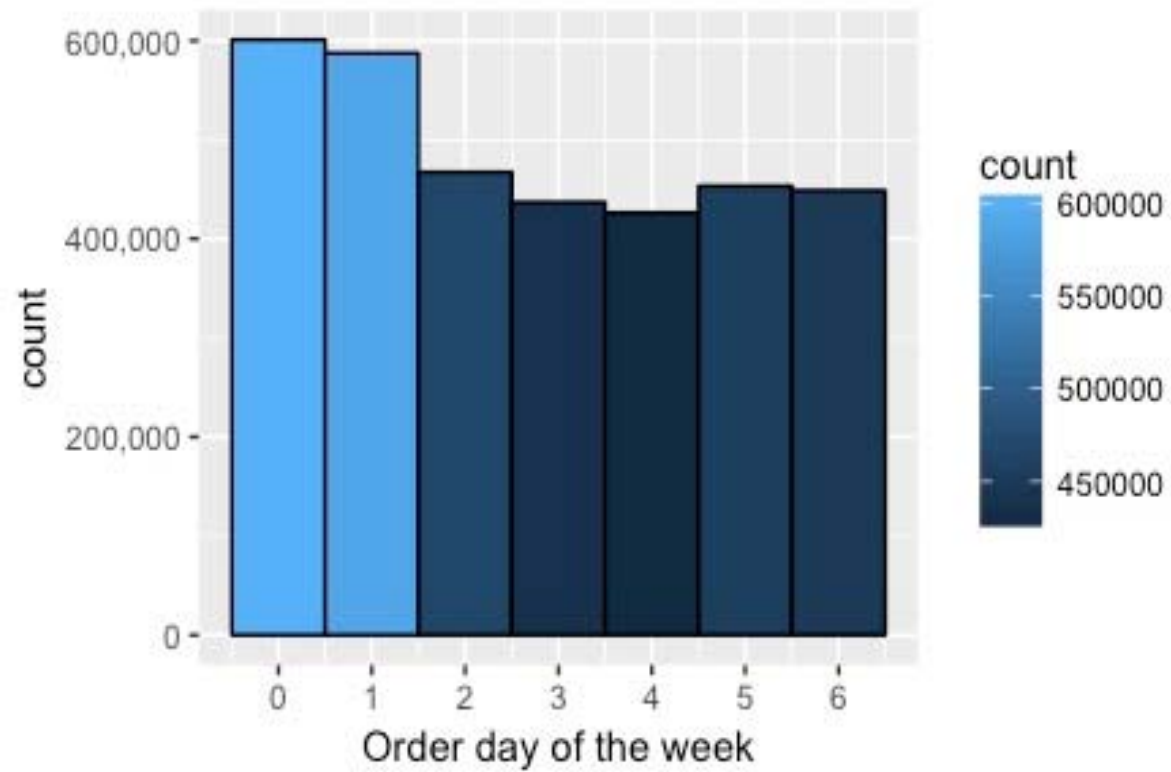
- Observations: 134
- Variables: 2
- Data size : 3 kb

資料來源 - product.csv

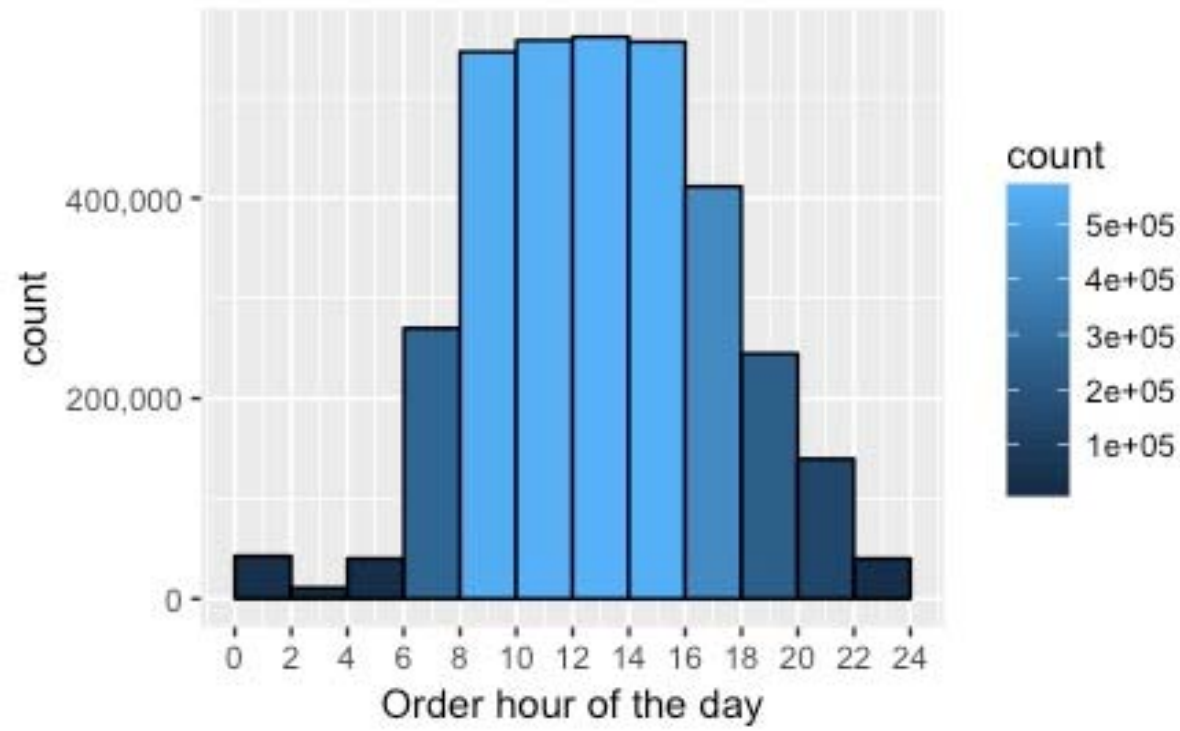
department_id	department
1	frozen
2	other
3	bakery
4	produce
5	alcohol
6	international

- Observations: 21
- Variables: 2
- Data size : 2.2 Mb

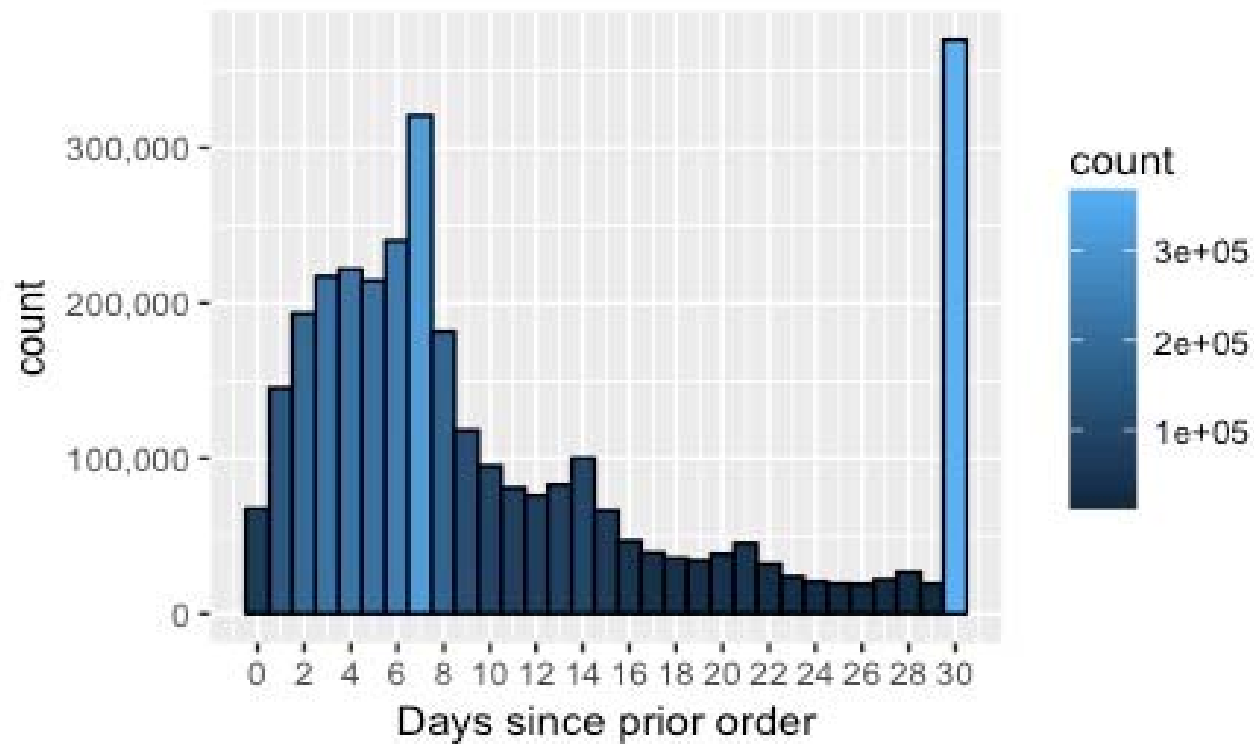
資料處理 - 當週購買日



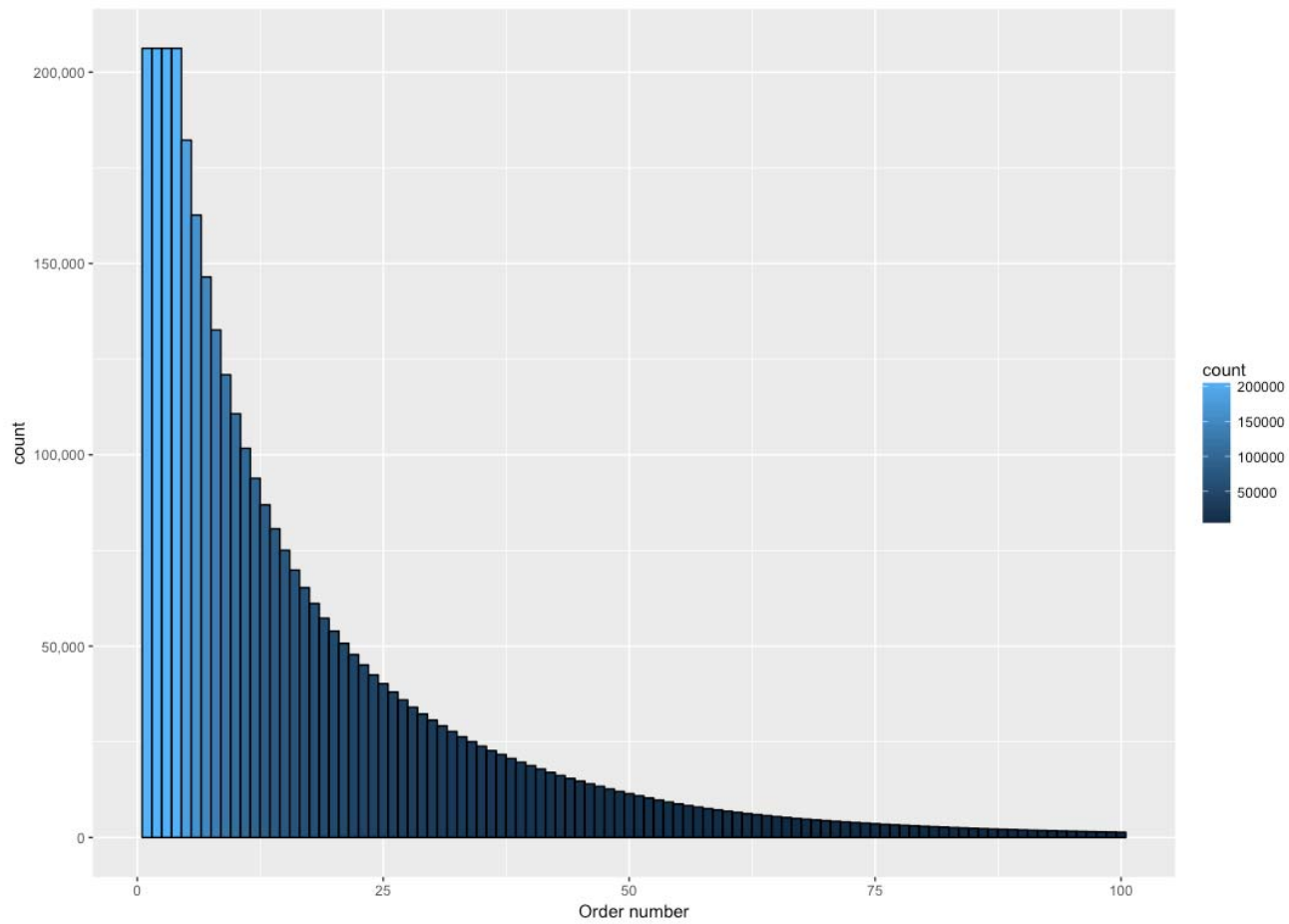
資料處理 - 當日購買時間



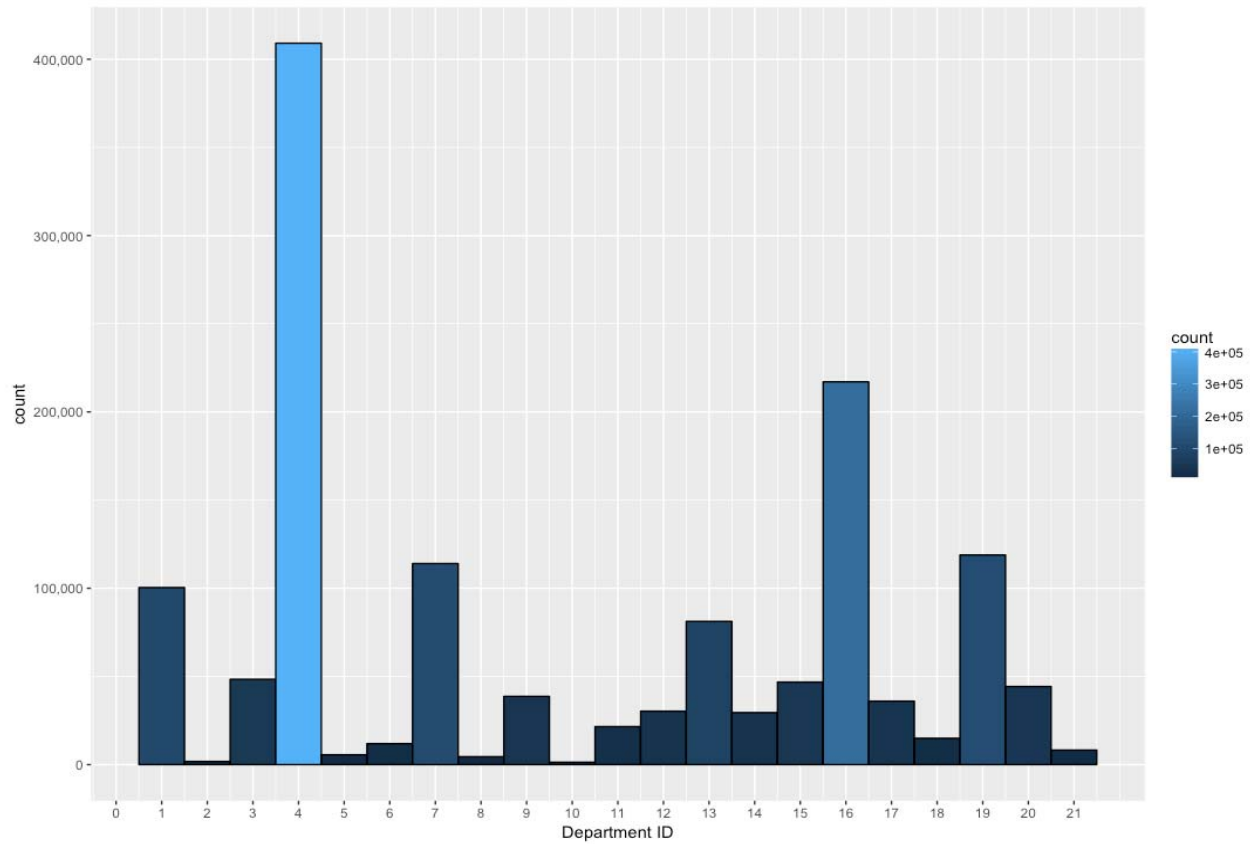
資料處理 - 距離前次購買期間



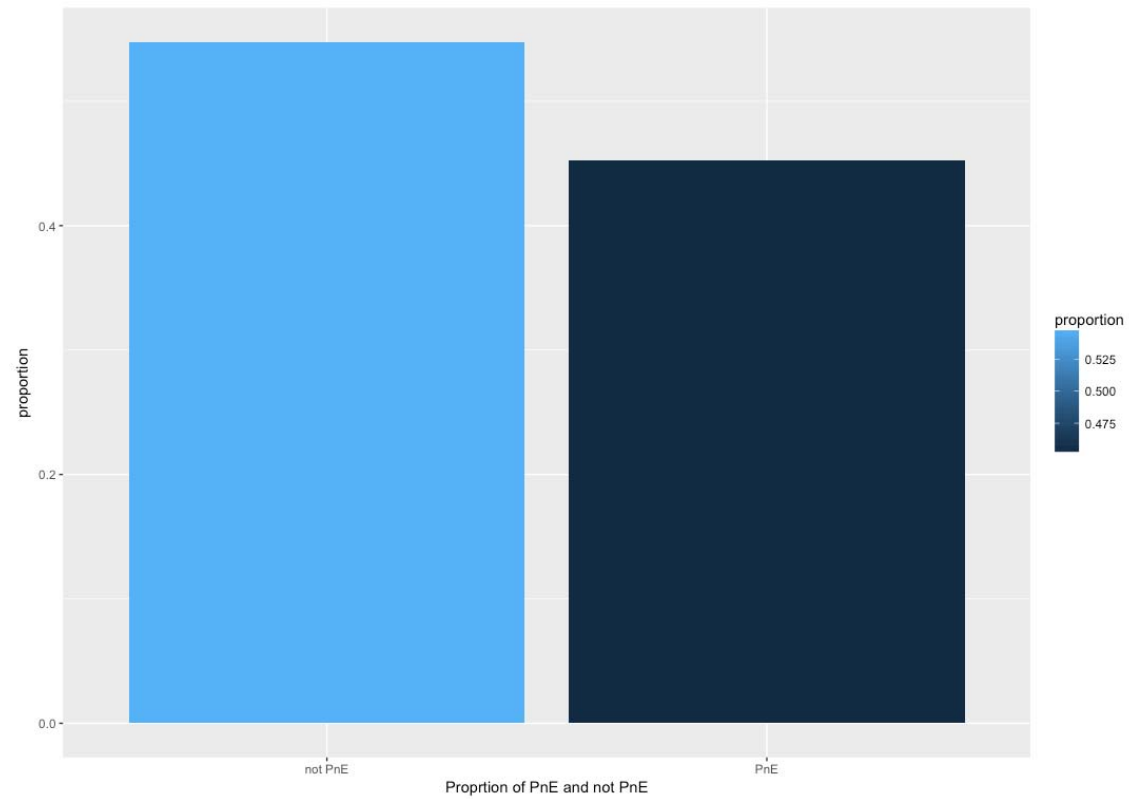
資料處理 - 訂單分佈



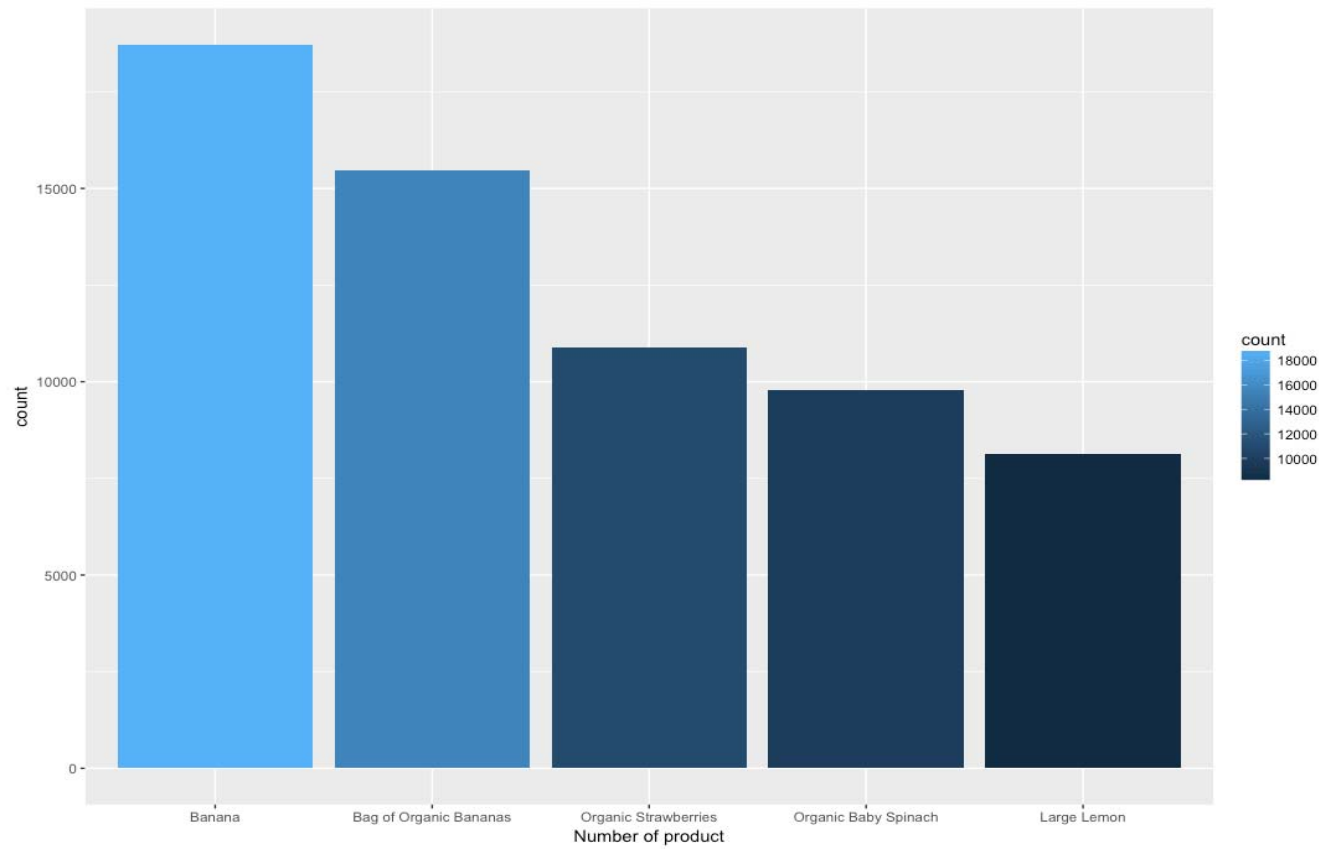
資料處理 - 商品種類



資料處理 - 商品種類比例



資料處理 - 商品細項



分析方法

- 支持度 (Support)
 - $\text{Support}(A \Rightarrow B) = P(A \cup B)$
- 信賴度 (Confidence)
 - $\text{Confident}(A \Rightarrow B) = P(B|A) = P(A \cup B)/P(A)$
- 提升度 (Lift)
 - $\text{Lift}(A \Rightarrow B) = P(B|A)/P(B)$
= $\text{confident}(A \Rightarrow B)/P(B) = P(A \cup B)/P(A)P(B)$

分析方法 – Association(Banana)

Items in LHS Group

- 1 rules: {Blueberries, Honeycrisp Apple, +1 items}
- 3 rules: {Eggo Homestyle Waffles, Grade A Large White Eggs, +4 items}
- 3 rules: {No Pulp Calcium & Vitamin D Pure Orange Juice, Cucumber Kirby, +4 items}
- 3 rules: {Feta Cheese Crumbles, Organic Red Radish, Bunch, +5 items}
- 4 rules: {Fresh Cauliflower, Roasted Turkey, +6 items}
- 1 rules: {Cherubs Heavenly Salad Tomatoes, Gala Apples}
- 5 rules: {Organic Kiwi, Organic Romaine, +10 items}
- 9 rules: {Classic Hummus Family Size, Lime, +15 items}
- 7 rules: {Green Beans, Organic Blueberry Waffles, +13 items}
- 11 rules: {Brownberry 100% 12 Grain Bread, Gluten Free Whole Grain Bread, +14 items}
- 14 rules: {Shredded Parmesan, Mild Cheddar Cheese Sticks, +22 items}
- 4 rules: {Cut & Peeled Baby Carrots, Whole Strawberries, +8 items}
- 5 rules: {Lime Sparkling Water, Original Veggie Straws, +9 items}
- 31 rules: {Blackberries, Organic Grape Tomatoes, +42 items}
- 19 rules: {Baby Spinach, Creamy Peanut Butter, +33 items}
- 1 rules: {Red Raspberries, Gala Apples}
- 2 rules: {Total 2% Lowfat Plain Greek Yogurt, Sweet Potato Yam, +2 items}
- 12 rules: {Organic Baby Arugula, Organic Gala Apples, +17 items}
- 31 rules: {Blueberries, Bunched Cilantro, +39 items}
- 11 rules: {Berry Medley, Orange Juice, +19 items}

Grouped Matrix for 177 Rules

RHS
{Banana}

Size: support
Color: lift

分析方法 – Association(Bag of Organic Banana)

Items in LHS Group

- 2 rules: {100% Whole Wheat Bread, Organic Zucchini, +4 items}
- 5 rules: {Unsweetened Almondmilk, Organic Large Green Asparagus, +10 items}
- 8 rules: {Organic Yams, Spring Water, +14 items}
- 14 rules: {Organic Lime, Frozen Organic Wild Blueberries, +18 items}
- 37 rules: {Globe Eggplant, No Pulp Calcium & Vitamin D Pure Premium 100% Pure Orange Juice, +43 items}
- 33 rules: {Dha Omega 3 Vitamin D Milk, Everything Bagels, +41 items}
- 25 rules: {Organic Extra Virgin Olive Oil, Organic Sour Cream, +33 items}
- 29 rules: {Lactose Free Half & Half, Organic Large Grade AA Brown Eggs, +38 items}
- 23 rules: {100% Recycled Paper Towels, Organic Dark Sweet Cherries, +30 items}
- 40 rules: {Italian Sparkling Mineral Water, Organic Stringles Mozzarella String Cheese, +43 items}
- 61 rules: {Organic Cheese Frozen Pizza, Bananas, Raspberries & Oats Organic Baby Food, +59 it
- 63 rules: {Basil Pesto, Instant Oatmeal Variety Pack, +51 items}
- 78 rules: {Authentic French Brioché Hamburger Buns, Organic Grade A Free Range Large Brown
- 55 rules: {Clementines, Bag, Guacamole, +50 items}
- 29 rules: {Grape White/Green Seedless, Organic Broccoli Crown, +32 items}
- 42 rules: {Baby Food Stage 2 Blueberry Pear & Purple Carrot, Organic Fancy Shredded Mozzarella
- 49 rules: {Baby Food Stage 2 Pumpkin Banana, Bag of Organic Lemons, +54 items}
- 24 rules: {Gluten Free Whole Grain Bread, Organic Frozen Mango Chunks, +36 items}
- 9 rules: {Fresh Cauliflower, Cantaloupe, +14 items}
- 49 rules: {2-Ply Right Size 100% Recycled Paper Towels, Marinara Sauce, +45 items}

Grouped Matrix for 675 Rules

Size: support
Color: lift

RHS

{Bag of Organic Bananas}

分析方法 – Association(Organic Strawberries)

Items in LHS Group

- 1 rules: {Organic Baby Spinach, Cantaloupe, +1 items}
- 1 rules: {Organic Navel Orange, Organic Baby Carrots, +1 items}
- 2 rules: {Organic Hothouse Cucumbers, Organic Snipped Green Beans, +2 items}
- 2 rules: {Organic Whole String Cheese, Organic Baby Carrots, +4 items}
- 3 rules: {Backyard Barbeque Potato Chips, YoKids Strawberry Banana/Strawberry Yogurt, -
- 3 rules: {YoKids Blueberry & Strawberry/Vanilla Yogurt, Banana, +6 items}
- 5 rules: {Organic Whole Milk Strawberry Beet Berry Yogurt Pouch, Organic Red Grapes, +9
- 6 rules: {Uncured Genoa Salami, Organic Yellow Onion, +12 items}
- 11 rules: {Single Serve Crispy Wheat Crackers, Sunny Days Strawberry Snack Bars, +18 ite
- 2 rules: {Honeycrisp Apple, Organic D'Anjou Pears, +4 items}
- 4 rules: {Original Hummus, Organic Broccoli, +9 items}
- 2 rules: {Fresh Cauliflower, Organic Kiwi, +2 items}
- 7 rules: {Baked Aged White Cheddar Rice and Corn Puffs, Cinnamon Rolls with Icing, +12 i
- 10 rules: {Marinara Sauce, Organic Blueberry Waffles, +16 items}
- 5 rules: {Organic Freeze Dried Strawberries, Organic Peeled Whole Baby Carrots, +17 items
- 14 rules: {Organic Half & Half, Organic Kale Greens, +25 items}
- 12 rules: {Organic Black Beans, Organic Cilantro, +20 items}
- 12 rules: {Mini Original Babybel Cheese, Organic Mandarins, +18 items}
- 22 rules: {Blueberry Whole Milk Yogurt Pouch, Cane Sugar, +29 items}
- 23 rules: {Organic Small Bunch Celery, 100% Whole Wheat Bread, +33 items}

Grouped Matrix for 147 Rules

RHS

{Organic Strawberries}

Size: support
Color: lift

結論與討論

- 生鮮食品與蛋奶類為最大宗（占比約45%）
- 購買時間大多集中在 8:00～ 16:00
- 重複回購的時段大約在7天和30天
- 下單訂購日約集中在六日

結論與討論

- 購買藍莓、蜜脆蘋果 → 香蕉
- 購買有機哈斯酪梨、有機檸檬、有機覆盆莓 → 袋裝有機香蕉
- 購買袋裝有機香蕉、哈密瓜、有機覆盆莓 → 有機草莓

結論與討論

- 利用會員制的方式，結合熱門及冷門商品做促銷
- 探討回購類型，提升下單率，增加運費收入
- 利用資料與超市合作