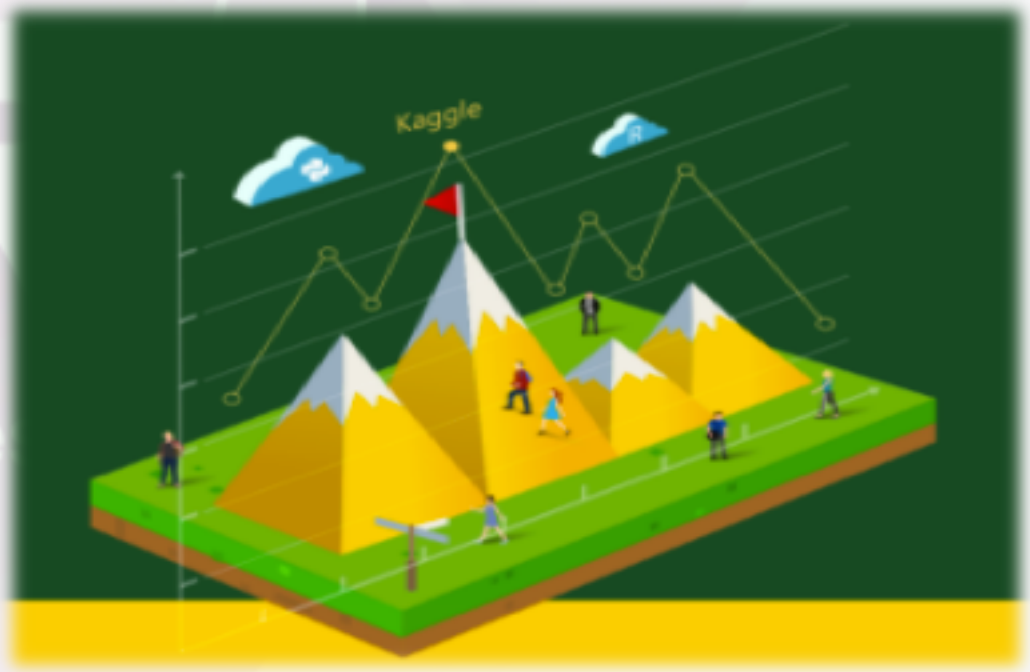
× kaggle™

R-LADIES TAIPEI KAGGLE 大賽

10 mins quick show

July 2017

妳知道 kaggle™ 嗎？



kaggle

- Website : https://www.kaggle.com
- 資料科學和機器學習競賽平台
- 目前已累積超過50萬名、遍布超過194個國家的註冊用戶
- 涵蓋電腦科學、電腦視覺、生物、醫藥

妳知道 2017.07.22-23 發生什麼事嗎？


2017.07.23

臺灣 NO1.全女性 Kaggle 黑客松競賽

當天的流程是這樣的....

#參加Kaggle松保證肉變鬆

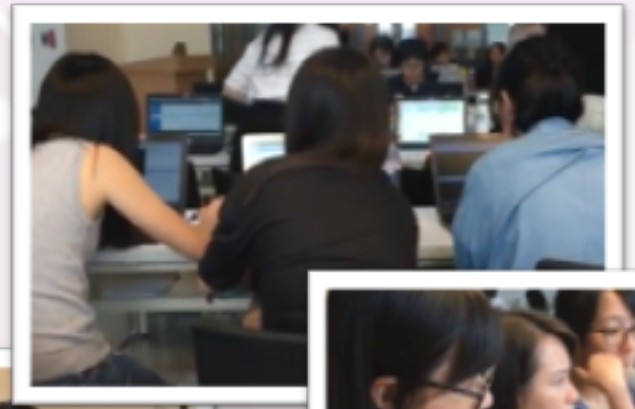
7/22 11:00

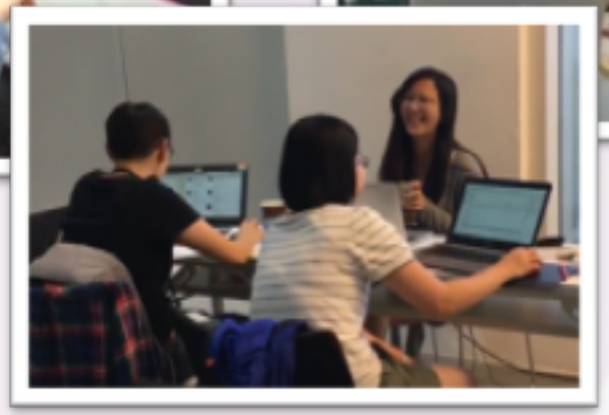
7/22 13:00

7/23 13:00

7/23 14:30

Summit





7/22 11:00

7/22 13:00

7/22 18:00

7/23 10:30

7/23 11:30

7/23 13:00

7/23 14:30

Summit

助教攤

7/22 11:00

7/22 13:00

7/22 18:00

7/23 10:30

7/23 11:30

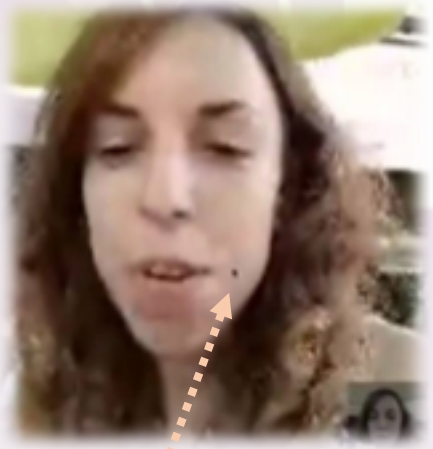
7/23 13:00

7/23 14:30

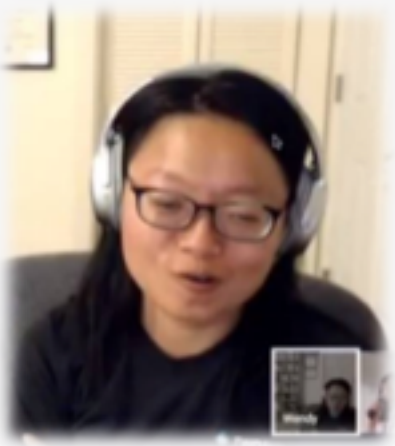
Summit

助教攤

神秘嘉賓

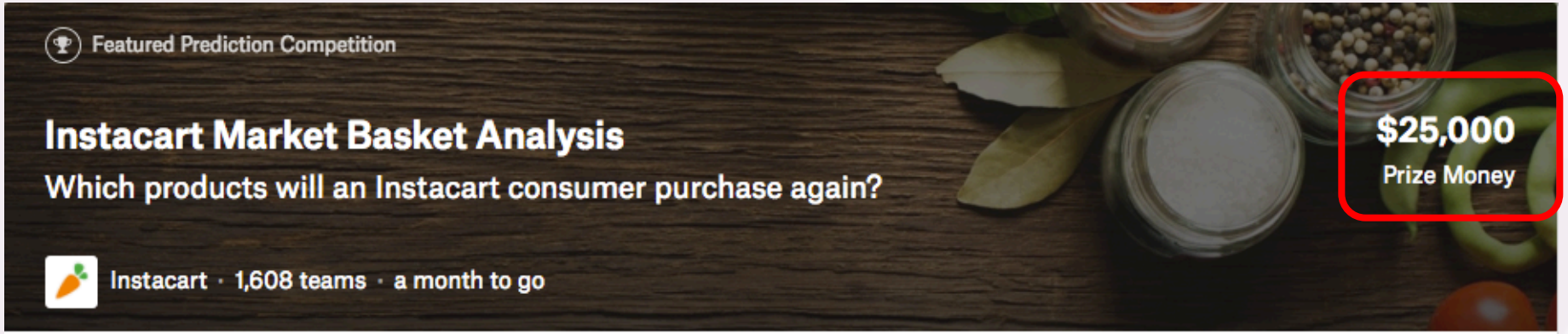


是滑鼠不是痣



這次選定的比賽是?!

Competition



Overview Data Kernels Discussion Leaderboard Rules

Overview

Description
Evaluation
Prizes
Timeline

Whether you shop from meticulously planned grocery lists or let whimsy guide your grazing, our unique food rituals define who we are. Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you.

Competition

- Website
  https://www.kaggle.com/c/instacart-market-basket-analysis
- Instacart

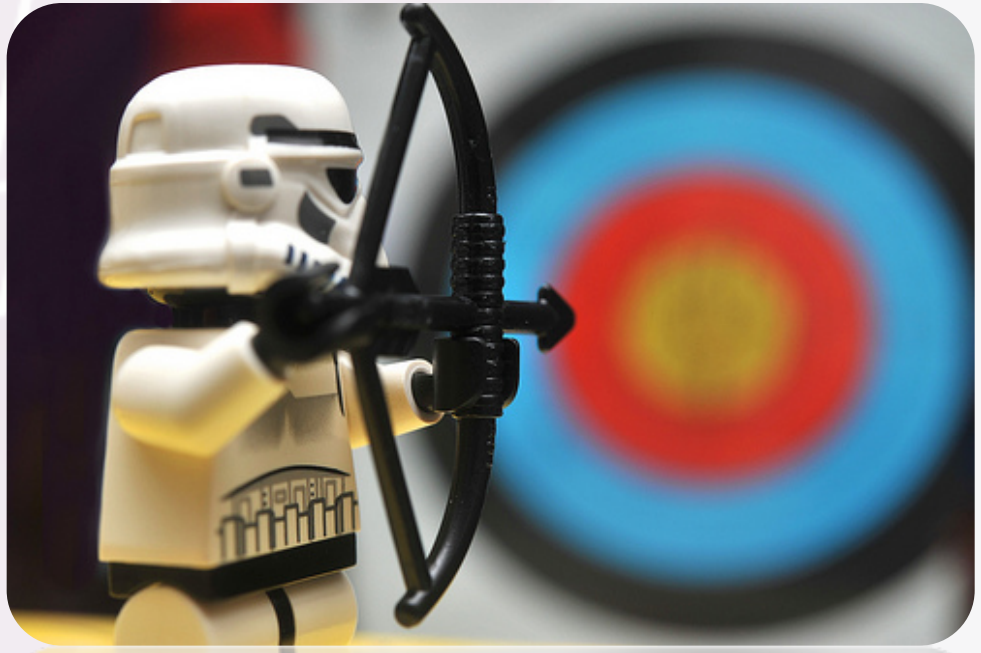
  利用募集群眾外包配送的服務，主要為配送的物品以生活必需品與雜貨為主，商品大約為 30 多萬個品項。

詳細介紹：https://www.inside.com.tw/2014/10/09/instacart

Competition

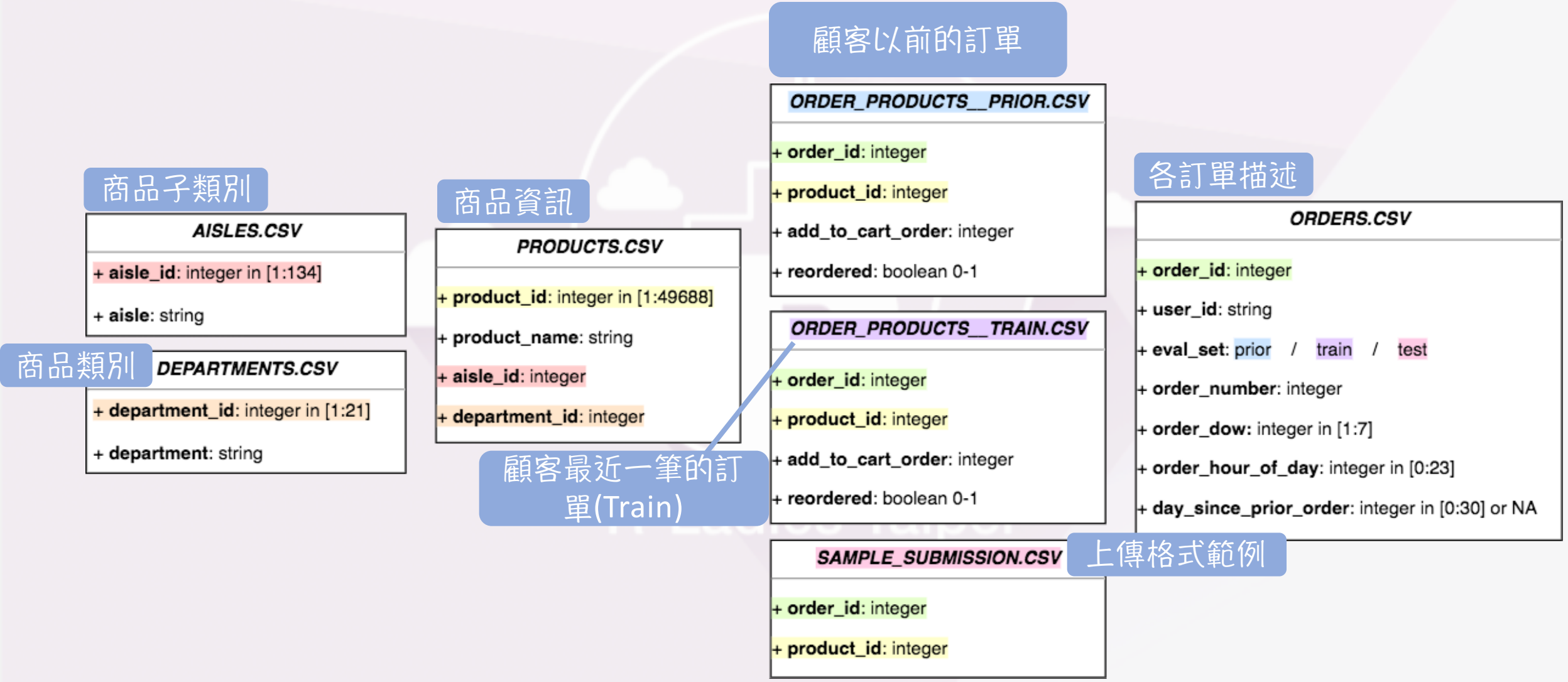
- Target

  預測顧客下一筆訂單中會有哪些商品



那比賽資料集長怎樣呢?!

About the Data -- 檔案關聯及資料屬性

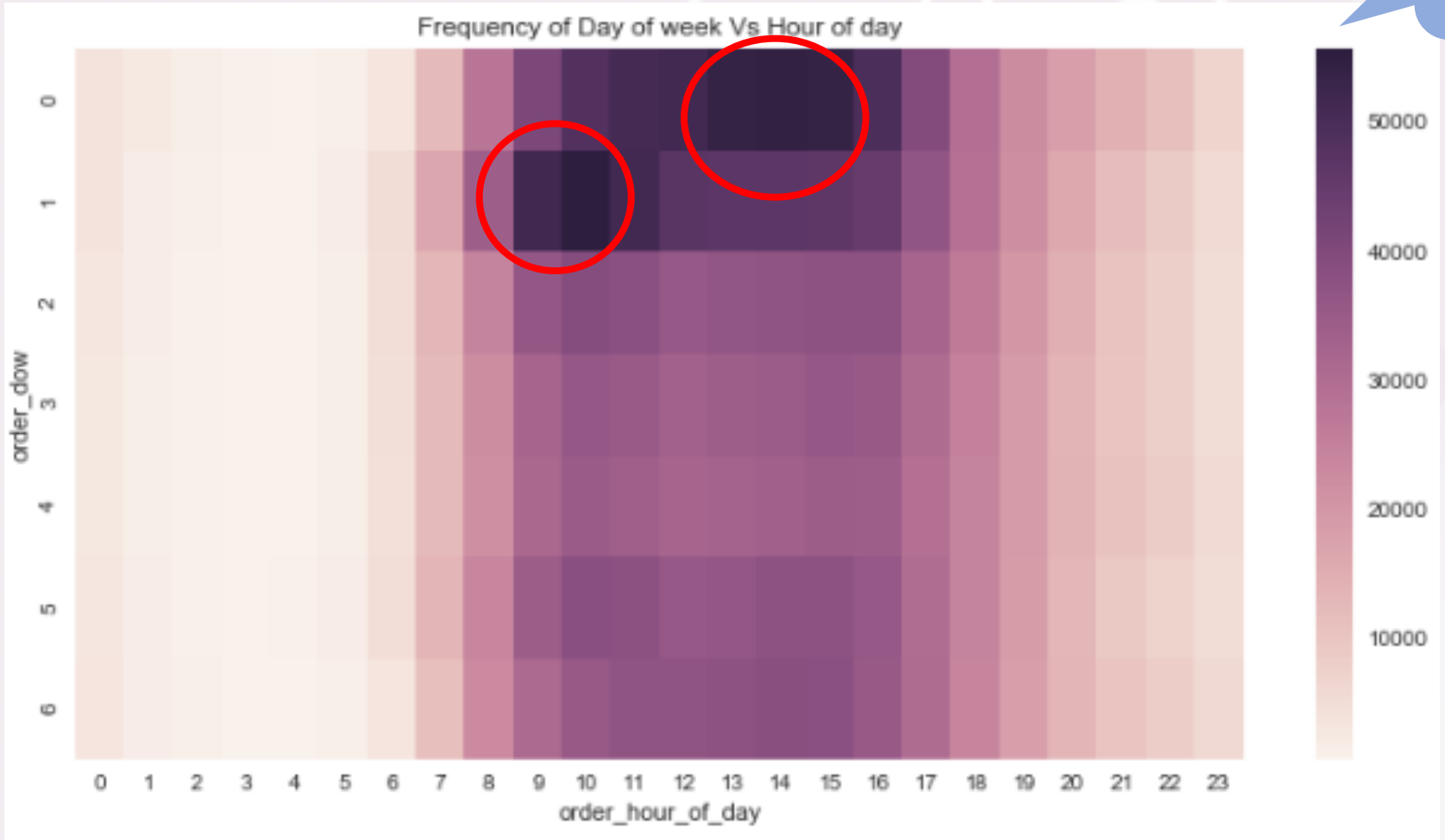


Ref：https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/33128#183176

我們發現了…

Exploratory Data Analysis

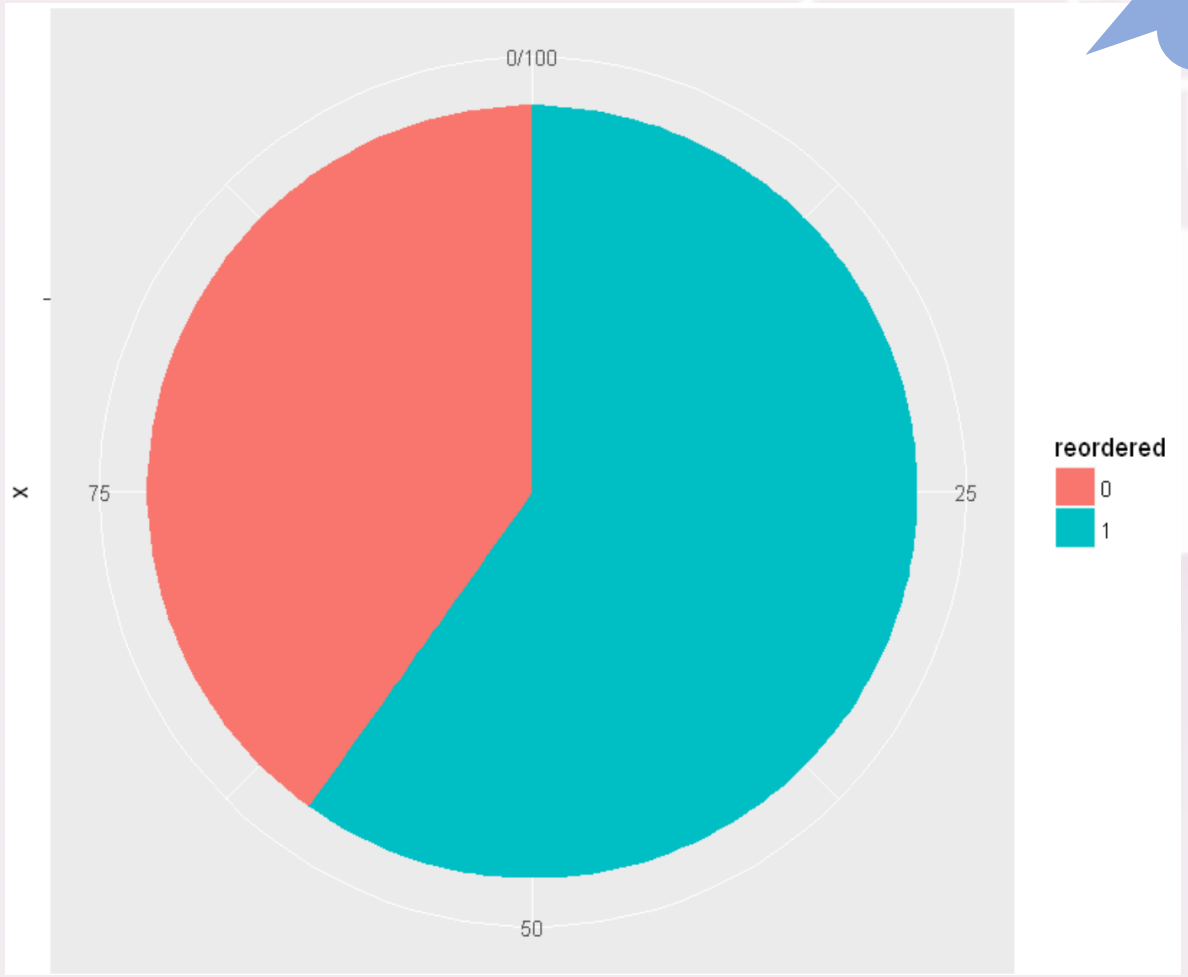
顧客大約都在何時訂購？



週六中午過後及週日早上為最密集的購買時間

Exploratory Data Analysis

商品回購率？

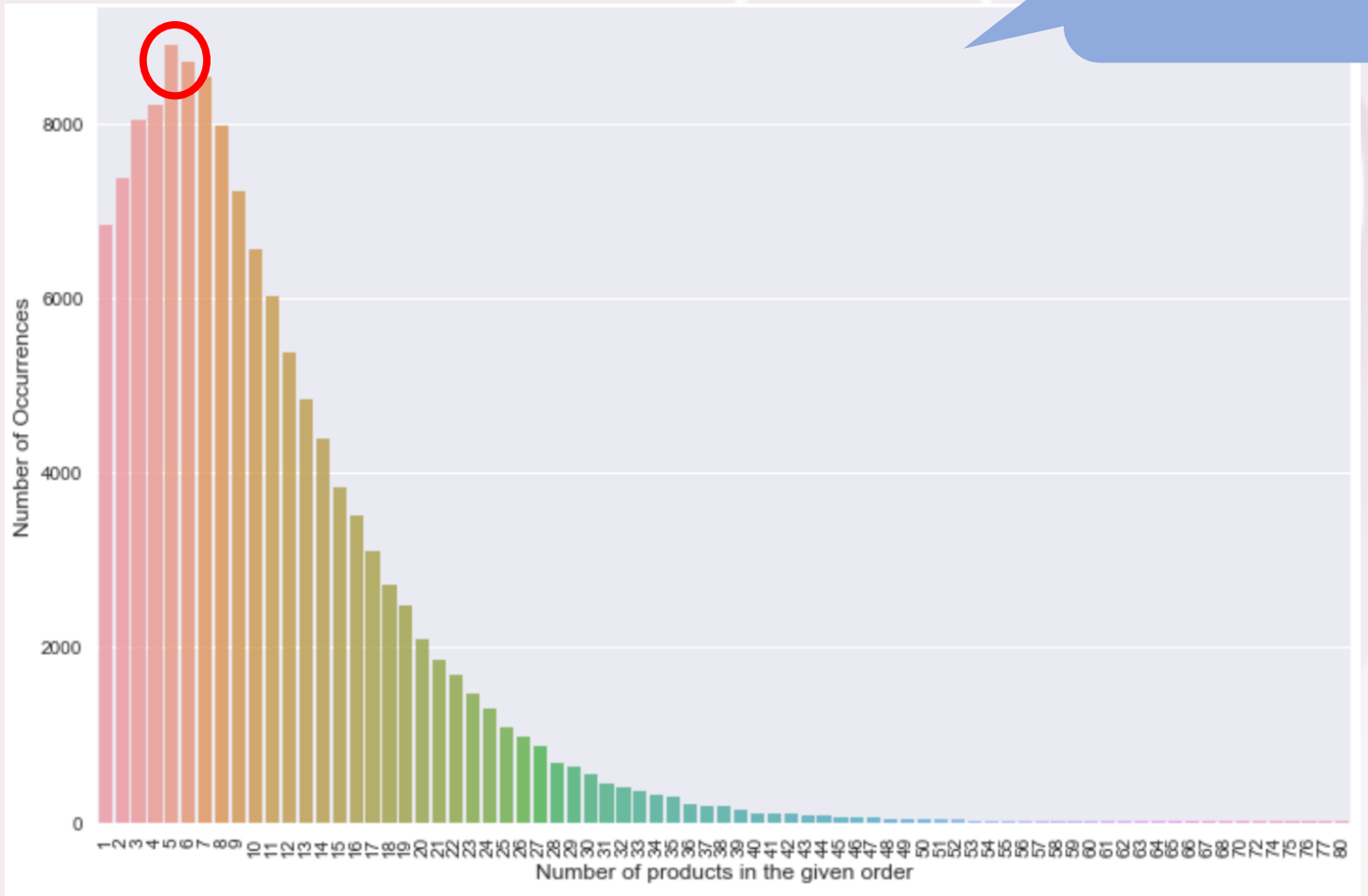


回購率大約為 59%

Exploratory Data Analysis

一筆訂單中有多少件商品? (train)

大部分訂單一次購買 **5** 件商品



如果比賽的過程中遇到問題?

上傳格式範例

sample_submission.csv	
order_id	訂單編號
product_id	商品編號

order_id	products
2774568	17668 21903 39190 47766 18599 43961 23650 24810
1528013	21903 38293
1376945	33572 28465 27959 44632 24799 34658 14947 30563 8309 13176
1356845	11520 14992 7076 28134 10863 13176
2161313	11266 196 10441 12427 37710 48142 14715 27839
1416320	5134 21903 21137 24852 17948 41950 24561

助教攤常見問題 1

該用什麼**Model**？

Order_id	Product_id	商品相關資訊	reorder	Buy or not
A	001		0	0
A	002		1	1
B	006		0	0
B	008		0	0
B	003		1	1
C	009		0	0

X

Y

助教攤常見問題 2

最後 **summit** 格式該怎麼整理?

Order_id	Product_id
A	001
A	002
B	006
B	008
B	003
C	00

order_id	products
2774568	17668 21903 39190 47766 18599 43961 23650 24810
1528013	21903 38293
1376945	33572 28465 27959 44632 24799 34658 14947 30563 8309 13176
1356845	11520 14992 7076 28134 10863 13176
1313	11266 196 10441 12427 37710 48142 14715 27839
	5134 21903 21137 24852 17948 41950 24561

助教攤常見問題 2

最後 **summit** 格式該怎麼整理?

Order_id	Product_id
A	001
A	002
B	006
B	008
B	003
C	009

order_id	products
2774568	17668 21903 39190 47766 18599 43961 23650 24810
1528013	21903 38293
1376945	33572 28465 27959 44632 24799 34658 14947 30563 8309 13176
1356845	11520 14992 7076 28134 10863 13176
2161313	11266 196 10441 12427 37710 48142 14715 27839
1416320	5134 2190 17948 41950 24561

這時候你需要!!!

助教攤常見問題 3

我的電腦跑不動...

助教攤常見問題 3

我的電腦跑不動...

這時候你需要!!!

助教攤常見問題 3

Microsoft Azure VM Sizes

A0	A1	A2	A3	A4
Shared Core (low IO)	1 x 1.6Ghz (moderate IO)	2 x 1.6Ghz (high IO)	4 x 1.6Ghz (high IO)	8 x 1.6Ghz (high IO)
768 MB memory 1 Data Disk (1TB) 1 x 500 Max IOPs	1.75 GB memory 2 Data Disks (1TB) 2 x 500 Max IOPs	3.5 GB memory 4 Data Disks (1TB) 4 x 500 Max IOPs	7.0 GB memory 8 Data Disks (1TB) 8 x 500 Max IOPs	14 GB memory 16 Data Disks (1TB) 16 x 500 Max IOPs

A5	A6	A7	A8	A9
2 x 1.6Ghz (high RAM)	4 x 1.6Ghz (high RAM)	8 x 1.6Ghz (high RAM)	8 x 2.2GHz (high compute)	16 x 2.2GHz (high compute)
14 GB memory 4 Data Disks (1TB) 4 x 500 Max IOPs	28 GB memory 8 Data Disks (1TB) 8 x 500 Max IOPs	56 GB memory 16 Data Disks (1TB) 16 x 500 Max IOPs	56GB memory 8 Data Disk (1TB) 8 x 500 Max IOPs 40 Gbps NIC	112 GB memory 16 Data Disks (1TB) 16 x 500 Max IOPs 40 Gbps NIC

Azure Pass

☜ KRISTEN

Until : 8/14

EMAIL : rladies.taipei@gmail.com

密切關注

2018
RLadies Taipei Kaggle 松