

全球超市銷售數據分析

資訊碩一 鄭智謙

電子碩一 陳立穎

電資四 莊東翰

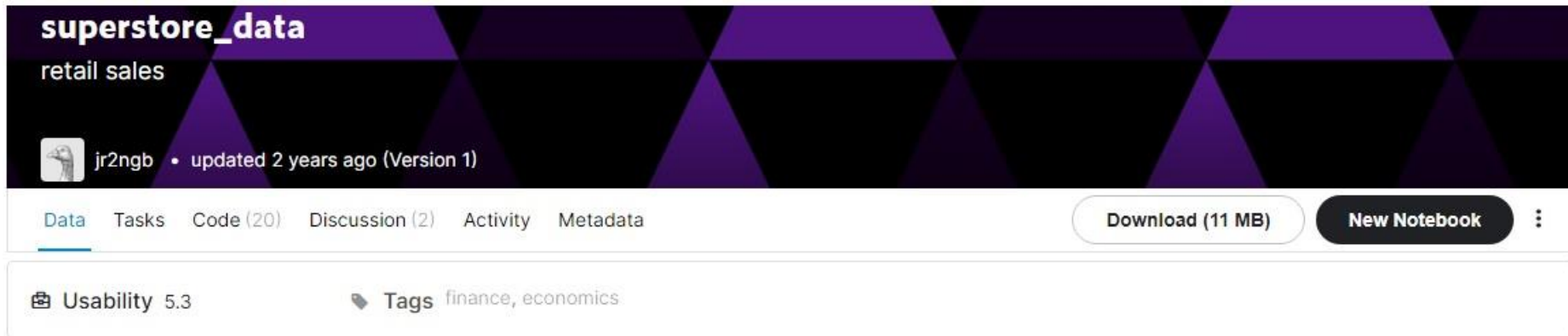
目錄

- 一、資料集來源
- 二、資料集介紹
- 三、資料前處理
- 四、資料模型與解讀

一、資料集來源

3

- 資料集名稱：superstore_dataset2011-2015
- 來源：<https://www.kaggle.com/jr2ngb/superstore-data>
- 最近更新日期：2019/01/30



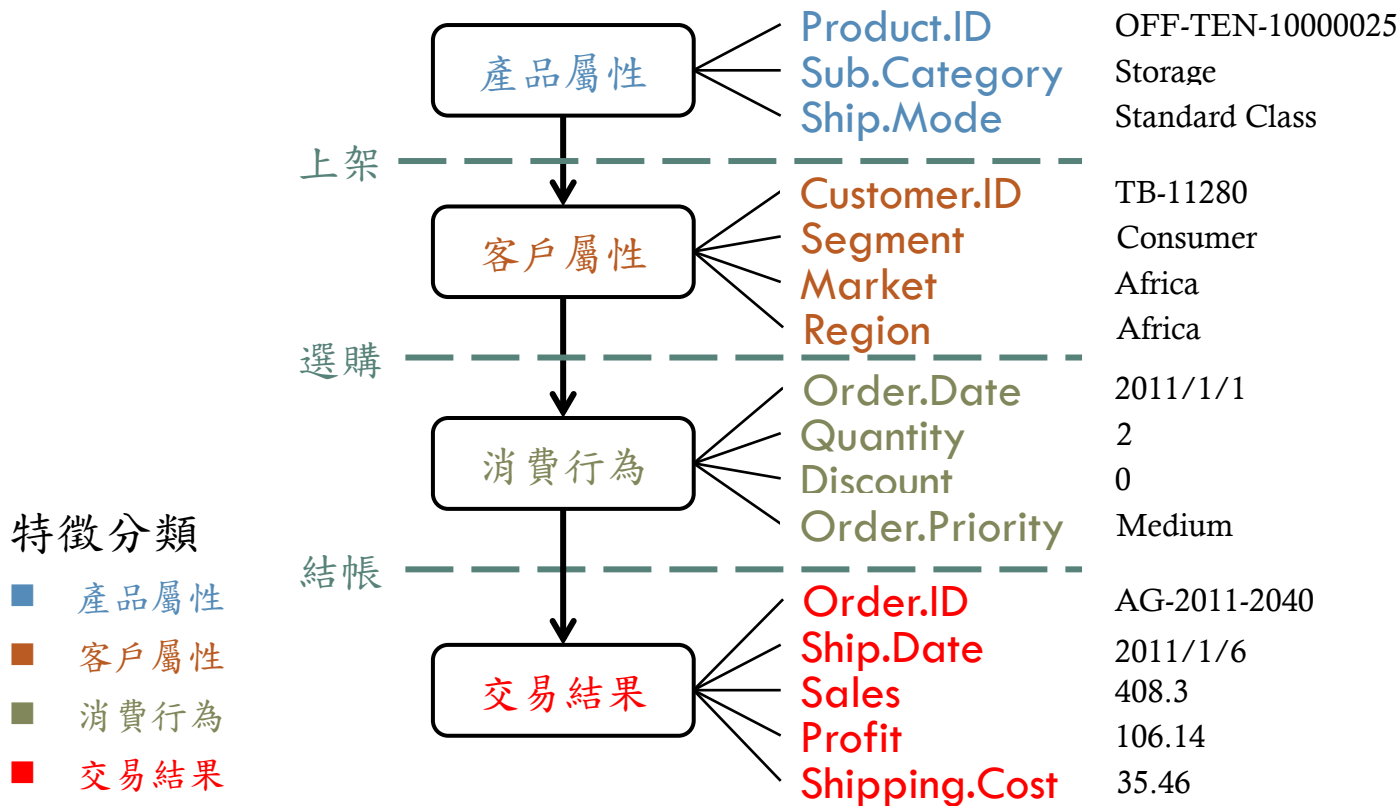
二、資料集介紹

4

- 簡介：全球超市的零售數據，時間為2011/01/01 ~ 2014/12/31
- 特徵欄位(共24項)：
Row.ID, Order.ID, Order.Date, Ship.Date, Ship.Mode, Customer.ID, Customer.Name, Segment, City, State, Country, Postal.Code, Market, Region, Product.ID, Category, Sub.Category, Product.Name, Sales, Quantity, Discount, Profit, Shipping.Cost, Order.Priority
- 原始資料量：共51263筆

二、資料集介紹

5



二、資料集介紹

6

□ 各類別的商品數量

Binders	Storage	Art	Paper	Chairs	Phones	Furnishings	Accessories	Labels	Envelopes	Supplies
6152	5059	4883	3538	3434	3357	3170	3075	2606	2435	2425
Fasteners	Bookcases	Copiers	Appliances	Machines	Tables					
2420	2411	2223	1755	1486	861					

□ 前10大客戶購買的商品類別

	Binders	Storage	Art	Paper	Chairs	Phones	Furnishings	Accessories	Labels	Envelopes
PO-18850	14	10	6	7	8	10	3	4	6	6
BE-11335	13	7	9	10	9	5	6	4	5	6
JG-15805	12	12	9	6	7	2	4	8	6	2
SW-20755	10	5	4	4	10	6	5	9	2	7
EM-13960	10	8	10	2	6	4	2	5	3	4
MY-18295	10	6	7	6	8	3	7	5	3	4
MP-17965	11	8	11	7	10	7	5	6	3	3
ZC-21910	10	8	7	8	6	5	6	6	1	6
CK-12205	6	5	10	9	5	5	7	5	5	4
AF-10870	6	3	6	5	7	6	8	5	2	8

三、資料集前處理

7

- 修復原始csv檔因編碼格式造成的資料短少問題，將原始檔轉為xlsx檔後，再轉回csv檔即可
→修正後之資料量為 51291 筆
- 統一日期格式為年/月/日

51260	32593	CA-2014-1	31-12-2014	4/1/2015
51261	47594	MO-2014-2	31-12-2014	5/1/2015
51262	8857	MX-2014-1	31-12-2014	2/1/2015
51263	6852	MX-2014-1	31-12-2014	6/1/2015
51264	36388	CA-2014-1	31-12-2014	4/1/2015
51265				

superstore_dataset2011-2015

轉檔與修改

51287	32593	CA-2014-1	2014/12/31	2015/1/4
51288	47594	MO-2014-2	2014/12/31	2015/1/5
51289	8857	MX-2014-1	2014/12/31	2015/1/2
51290	6852	MX-2014-1	2014/12/31	2015/1/6
51291	36388	CA-2014-1	2014/12/31	2015/1/4
51292				

superstore_dataset2011-2015_fix

四、資料模型與解讀

8

數據轉換：客戶數據模型

□ 客戶共有 1590位

cID	D0	Df	DD	FF	MM	TT	UU	NN	BB	DD0	FF0	MM0	BB0	Q0	Qf
AA-10315	2011-03-31	2014-12-23	19	19	13747.413	145	723.5481	7.631579	71.73684 days	(10,20]	(10,20]	(1e+04,1e+05]	(70,90]	11Q1	14Q4
AA-10375	2011-04-21	2014-12-25	23	23	5884.195	139	255.8346	6.043478	58.43478 days	(20,30]	(20,30]	(999,1e+04]	(50,70]	11Q2	14Q4
AA-10480	2011-05-04	2014-08-28	20	20	17695.590	150	884.7795	7.500000	60.60000 days	(10,20]	(10,20]	(1e+04,1e+05]	(50,70]	11Q2	14Q3

□ FF-造訪頻次，MM-消費金額，TT-購買件數，BB-回購週期，D0-首次來店日，Df-最近來店日，DD0-下單日期次數，FF0-總下單數，MM0-總獲利

四、資料模型與解讀

9

客戶價值模型

直:總下單數

橫:總獲利

	(-1e+05,-1e+04]	(-1e+04,-999]	(-999,-99]	(-99,-9]	(-9,0]	(0,9]	(9,99]	(99,999]	(999,1e+04]	(1e+04,1e+05]	Sum
(0,1]	0	0	0	0	0	1	4	4	1	0	10
(1,10]	0	0	0	0	0	0	3	203	536	0	742
(10,20]	0	0	0	0	0	0	0	1	109	46	156
(20,30]	0	0	0	0	0	0	0	0	100	430	530
(30,40]	0	0	0	0	0	0	0	0	7	144	151
(40,50]	0	0	0	0	0	0	0	0	0	1	1
(50,1e+05]	0	0	0	0	0	0	0	0	0	0	0
Sum	0	0	0	0	0	1	7	208	753	621	1590

- 消費99-999的客戶佔比13%(=208/1590)，消費999-9999的客戶佔比47 %(=753/1590)，消費9999-99999的客戶佔比39 %(=621/1590)。
- 有可能因為訂單都是量販的關係，所以單筆消費金額都偏高。一般能負擔量販的都是企業非普通家庭。
- 應該與電商合作增加更多普通民眾會購買的商品(數量少，物美價廉)，並鼓勵民眾滿指定消費額運費下降及享有優惠，讓一般民眾更願意使用此物流系統。

四、資料模型與解讀

10

客戶駐留模型

- 主要營收都是靠老客戶支撐,自從11Q4開始就大幅減少新客戶的加入,到了13Q1減少更多,甚至到了14Q1幾乎沒有新客戶加入。
- 應想辦法做促銷活動讓新客戶加入、打廣告增加曝光度或增加企業正面形象。
- 要是有更多訂單是關於衣服、飾品等,應能縮短消費週期,且不要量販銷售,讓普羅大眾下單意願增加。

直:初次下單日期
橫:最近下單日期

	11Q3	12Q1	12Q2	12Q4	13Q1	13Q2	13Q3	13Q4	14Q1	14Q2	14Q3	14Q4	Sum
11Q1	0	0	1	0	1	1	3	3	9	15	45	434	512
11Q2	0	0	0	1	2	0	6	6	13	20	40	320	408
11Q3	1	1	0	0	3	1	7	6	11	15	33	175	253
11Q4	0	0	0	1	0	2	2	5	4	9	22	91	136
12Q1	0	0	0	0	1	0	0	2	5	4	15	30	57
12Q2	0	0	0	0	1	2	0	2	2	5	19	25	56
12Q3	0	0	0	1	0	0	2	2	2	6	18	27	58
12Q4	0	0	0	2	0	1	0	2	0	5	10	19	39
13Q1	0	0	0	0	0	0	1	1	1	1	7	8	19
13Q2	0	0	0	0	0	1	0	0	0	1	6	7	15
13Q3	0	0	0	0	0	0	1	1	1	3	4	5	15
13Q4	0	0	0	0	0	0	0	3	0	1	0	3	7
14Q1	0	0	0	0	0	0	0	0	0	0	1	5	6
14Q2	0	0	0	0	0	0	0	0	0	1	2	4	7
14Q3	0	0	0	0	0	0	0	0	0	0	1	1	2
Sum	1	1	1	5	8	8	22	33	48	86	223	1154	1590

四、資料模型與解讀

11

客戶流失模型

	(-1,30]	(30,50]	(50,70]	(70,90]	(90,110]	(110,130]	(130,150]	(150,170]	(170,190]	(190,210]	(210,230]	(230,250]	(250,300]	(300,600]
11Q3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12Q1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
12Q2	0	0	0	0	0	1	0	0	0	0	0	0	0	0
12Q4	2	1	0	0	0	0	0	0	1	0	0	0	0	1
13Q1	0	0	0	2	2	0	0	1	3	0	0	0	0	0
13Q2	1	0	0	0	2	3	1	0	0	0	1	0	0	0
13Q3	1	0	1	0	2	2	4	3	2	3	2	0	0	2
13Q4	3	1	0	2	3	4	3	2	2	4	2	2	1	4
14Q1	0	1	0	1	3	4	6	5	6	7	3	5	2	5
14Q2	1	0	2	5	6	8	16	14	7	5	7	6	7	2
14Q3	3	15	14	8	22	27	31	28	25	12	10	5	9	14
14Q4	0	354	368	61	41	63	58	66	38	36	11	17	22	19

直:最近下單日期
橫:回購週期

- 大部分購買週期為30-70日，這代表客戶約1至2個月內會下單一次，老客戶直到現在都還會支持。
- 如果電商量販的物品是久久需要補充一次的物品，則久久才會有一次下單。要是能有更多訂單是有關於民生用品，這樣才能增加下單量。
- 應觀察國際趨勢，注意各國缺乏哪方面的物品，並對症下藥，讓該區域的客戶增加下單意願。