Order Recommendation System Improvement for Instacart

Instacart Overview

Industry Background

Online grocery shopping is changing the whole supermarket industry. Same-day grocery delivery industry is on trend and is predicted to worth $18 billion value by the end of 2018. Niche online grocery services are competing over conventional grocery business with its advantages such as convenience and selection. According to an online survey, 25% of millennials said they would pay a premium for same-day delivery. Grocery giants such as Amazon, Costco and Target are offering same-day grocery delivery service.

What is Instacart?

Instacart is an American company that operates as a same-day grocery on-demand delivery service and has its headquarter at San Francisco. Instacart's service is mainly provided through online order by an app and Instacart personal shoppers simply go to a store and purchase the ordered items at a chosen retail. As of March 2017, Instacart services 36 markets, composed of 1,200 cities in 25 states including California, Arizona and New York. In the same month it was valued at approximately $3.4 billion dollars and was considered "one of the most valuable companies in the America."

Instacart's Business Model

Instacart Customers are given a decision to shop from any store that is situated nearby. Personal shoppers who are found nearby the store will receive the order and starts collecting items as mentioned in the order. Shoppers are paid hourly rate while customers are also allowed to tip the shoppers. Shoppers pays the bill through Instacart's prepaid debit card which is accepted at the store. The shopper then goes to deliver the groceries to the customer.

Instacart's Revenue Model

- Delivery charges: Delivery charges are Instacart's fundamental profits. The charges may be higher if the customer requires speedier delivery or in busier circumstances. There are two options: Order above $35 entices a fixed delivery charge of $3.99 to $5.99; orders less than $35 are priced $7.99 to $9.99.
- Membership fee: Instacart also extends an annual membership scheme priced at $99 and is termed "Instacart Express". Users can opt for the membership to get groceries delivered at no cost for the entire one year.
- Mark-up charges: Products sold through the Instacart platform, some are priced as the same as store while in some cases priced 15% higher than the retail price. The revenue generated through these mark-up fees goes to the Instacart wallet.
- First Free Delivery: Instacart hypnotizes customers with a free delivery options for users who chooses to use the Instacart facility app for the first time, so pretty much a smart way to solve the "cold start" problem and provide more incentive for new app users to place an order.

Dataset

The dataset published by Instacart itself, called "The Instacart Online Grocery Shopping Dataset 2017", is a relational set of files describing customers' orders over time. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, between 4 and 100 of their orders are provided, with the sequence of products purchased in each order.

The **"aisles" table** (Figure 1 left) shows all aisles and their unique aisle id.
The "department" table (Figure 1 right) shows the departments and their unique department id.

	A	B	C	D
1	aisle_id	aisle		
2		1 prepared soups salads		
3		2 specialty cheeses		
4		3 energy granola bars		
5		4 instant foods		
6		5 marinades meat preparation		
7		6 other		
8		7 packaged meat		
9		8 bakery desserts		
10		9 pasta sauce		
11		10 kitchen supplies		
12		11 cold flu allergy		
13		12 fresh pasta		
14		13 prepared meals		
15		14 tofu meat alternatives		

	A	B	C
1	department_id	department	
2		1 frozen	
3		2 other	
4		3 bakery	
5		4 produce	
6		5 alcohol	
7		6 international	
8		7 beverages	
9		8 pets	
10		9 dry goods pasta	
11		10 bulk	
12		11 personal care	
13		12 meat seafood	
14		13 pantry	
15		14 breakfast	

(Figure 1)

The **"products" table** (Figure 2) shows the products and the aisles and departments they belong to. It can be joined with aisles and departments table via aisle_id and department_id.

	A	B	C	D
1	product_id	product_name	aisle_id	department_id
2		1 Chocolate Sandwich Cookies	61	19
3		2 All-Seasons Salt	104	13
4		3 Robust Golden Unsweetened Oolong Tea	94	7
5		4 Smart Ones Classic Favorites Mini Rigatoni With Vodka Cream Sauce	38	1
6		5 Green Chile Anytime Sauce	5	13
7		6 Dry Nose Oil	11	11
8		7 Pure Coconut Water With Orange	98	7
9		8 Cut Russet Potatoes Steam N' Mash	116	1
10		9 Light Strawberry Blueberry Yogurt	120	16
11		10 Sparkling Orange Juice & Prickly Pear Beverage	115	7
12		11 Peach Mango Juice	31	7
13		12 Chocolate Fudge Layer Cake	119	1
14		13 Saline Nasal Mist	11	11
15		14 Fresh Scent Dishwasher Cleaner	74	17

(Figure 2)

The **"order_products_"** tables contains two sub-tables, which are **order_products_prior** table and **order_products_train** table. They are the most essential tables we are using for our analysis. They specify which products were purchased in each order and the add-to-cart order of each product purchased. They also contain a binary attribute 'reordered', with 1 indicates the customer has a previous order that contains the product. These tables can be connected with products table via product_id and with orders table via order_id.

Order_products_prior:

	A	B	C	D
1	order_id	product_id	add_to_cart_order	reordered
2	2	33120	1	1
3	2	28985	2	1
4	2	9327	3	0
5	2	45918	4	1
6	2	30035	5	0
7	2	17794	6	1
8	2	40141	7	1
9	2	1819	8	1
10	2	43668	9	0
11	3	33754	1	1
12	3	24838	2	1
13	3	17704	3	1
14	3	21903	4	1
15	3	17668	5	1

Order_products_train:

	A	B	C	D
1	order_id	product_id	add_to_cart_order	reordered
2	1	49302	1	1
3	1	11109	2	1
4	1	10246	3	0
5	1	49683	4	0
6	1	43633	5	1
7	1	13176	6	0
8	1	47209	7	0
9	1	22035	8	1
10	36	39612	1	0
11	36	19660	2	1
12	36	49235	3	0
13	36	43086	4	1
14	36	46620	5	1
15	36	34497	6	1

(figure 3)

The **"orders" table**, which contains 3,421,083 purchasing records, tells to which evaluation set (prior, train, test) an order belongs. Plus the day and time an order was placed, which may help us determine when is the best time to send the recommendation. It is another essential table for our analysis.

	A	B	C	D	E	F	G
1	order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
2	2539329	1	prior	1	2	8	
3	2398795	1	prior	2	3	7	15
4	473747	1	prior	3	3	12	21
5	2254736	1	prior	4	4	7	29
6	431534	1	prior	5	4	15	28
7	3367565	1	prior	6	2	7	19
8	550135	1	prior	7	1	9	20
9	3108588	1	prior	8	1	14	14
10	2295261	1	prior	9	1	16	0
11	2550362	1	prior	10	4	8	30
12	1187899	1	train	11	4	8	14
13	2168274	2	prior	1	2	11	
14	1501582	2	prior	2	5	10	10
15	1901567	2	prior	3	1	10	3

(Figure 4)

Questions

Customer satisfaction is a major influence factor of customer retention rate. Most of Instacart users are busy people who do not have time going to grocery store on their own. By predicting which products an user will buy again and recommend these products to users during their shopping process, Instacart may offer better convenience to its users. Therefore, customer buying activities and satisfaction rate are both likely to increase, which not only drives to a higher revenue, but also more loyal customers.

In this project, the core question is: Can we increase active customers and customer buying activities by improving the order recommendation system based on their previous orders?

To know when and what should we recommend to each customer, we need to glean some insights on customers' order placing behaviors. Therefore, we can break down the core question into the following sub-questions:

1) When should we send our recommendation?
2) How many items should our recommendation contain?
3) What products should we recommend?
4) Bonus: output your prediction on test dataset and show which kind of products would mostly reordered.