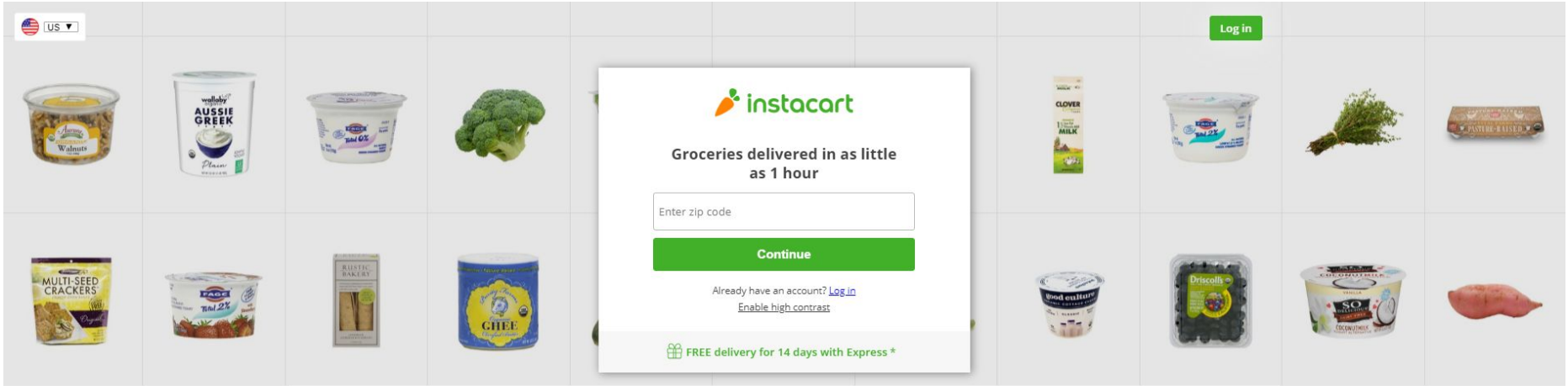
Instacart 고객분석

2018년 12월 19일 | 김수연 송민지 임동원

목 차

01 의뢰인_ Instacart



Products you love

Find 1,000's of products from the stores you already shop at.

Same-day delivery

We make deliveries in cities like Los Angeles, Miami, New York City, Chicago, Austin, Washington D.C, Houston, Atlanta and many more.

Save time & money

Find exclusive deals on popular products — delivered to your front door!

02 데이터 종류

데이터 파일	데이터 사이즈
orders.csv	3.4m 행, 206k 사용자
products.csv	50k 행
aisles.csv	134 행
depts.csv	21 행
prior.csv	3.2m 행

02 데이터 종류

1) orders	
order_id	주문 식별자
user_id	고객 식별자
eval_set	데이터셋 종류
order_number	이 사용자의 주문 순서 번호 (1 = first, n = nth)
order_dow	주문이 접수 된 요일
order_hour_of_day	주문이 접수 된 시간
days_since_prior	마지막 주문 이후 일, 30 일로 제한

```
orders.head().T
```

	0	1	2	3	4
order_id	2539329	2398795	473747	2254736	431534
user_id	1	1	1	1	1
eval_set	prior	prior	prior	prior	prior
order_number	1	2	3	4	5
order_dow	2	3	3	4	4
order_hour_of_day	8	7	12	7	15
days_since_prior_order	NaN	15	21	29	28
vip	NaN	NaN	NaN	NaN	NaN

02 데이터 종류

2) products	
product_id	제품 식별자
product_name	제품 이름
aisle_id	외래 키
department_id	외래 키

	product_id	product_name	aisle_id	department_id
0	1	Chocolate Sandwich Cookies	61	19
1	2	All-Seasons Salt	104	13
2	3	Robust Golden Unsweetened Oolong Tea	94	7
3	4	Smart Ones Classic Favorites Mini Rigatoni Wit...	38	1
4	5	Green Chile Anytime Sauce	5	13

02 데이터 종류

3) aisle	
aisle_id	통로 식별자
aisle	통로 이름

	aisle_id	aisle
0	1	prepared soups salads
1	2	specialty cheeses
2	3	energy granola bars
3	4	instant foods
4	5	marinades meat preparation

4) depts	
department_id	부서 식별자
department	부서명

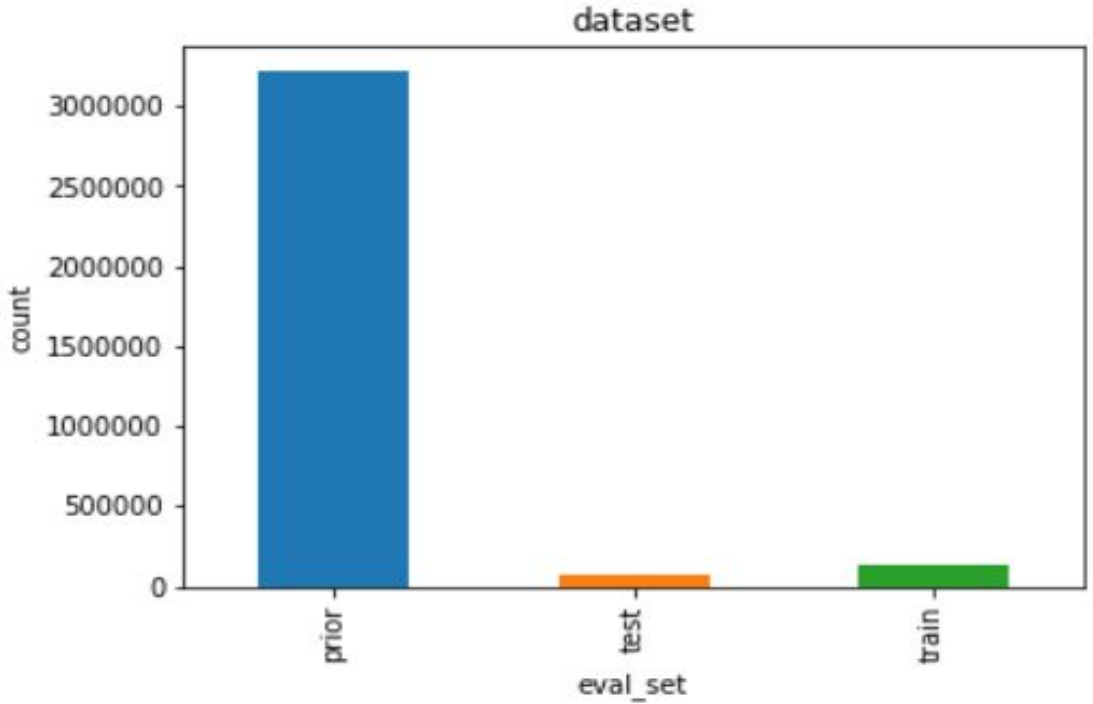
	department_id	department
0	1	frozen
1	2	other
2	3	bakery
3	4	produce
4	5	alcohol

02 데이터 종류

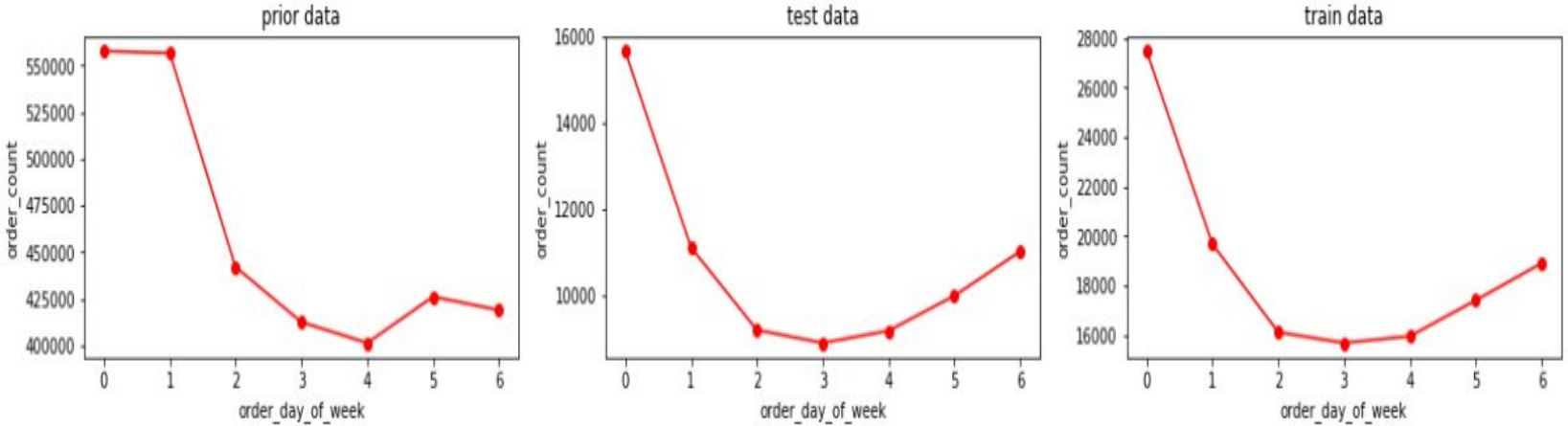
5) prior	
order_id	외래 키
product_id	외래 키
add_to_cart_order	각 제품이 장바구니에 추가 된 순서
reordered	이 제품이 과거에 이 사용자에 의해 주문 되었으면 1, 그렇지 않으면 0

	order_id	product_id	add_to_cart_order	reordered
0	2	33120	1	1
1	2	28985	2	1
2	2	9327	3	0
3	2	45918	4	1
4	2	30035	5	0

03 데이터 분석 - 데이터 셋 개수 확인

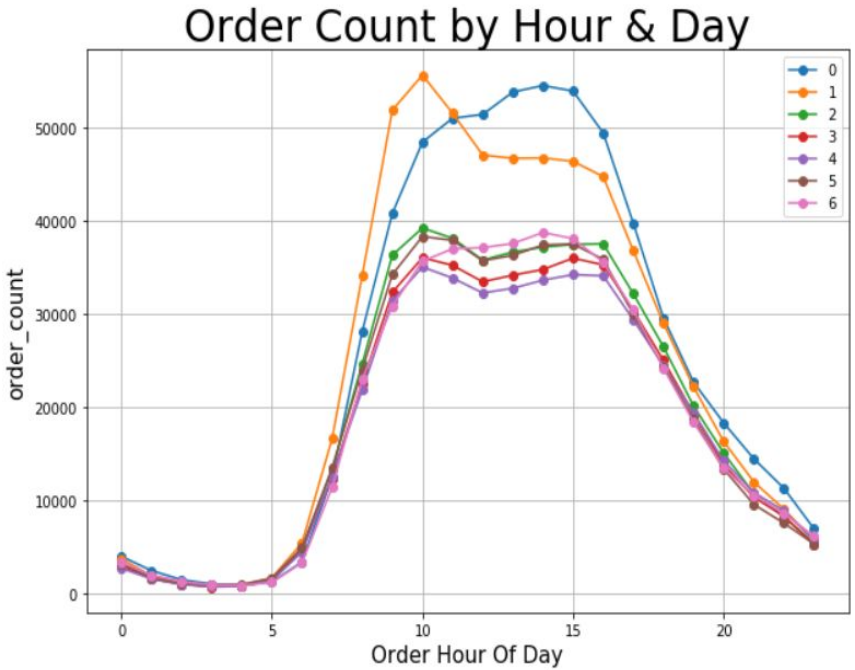


03 데이터 분석 - 각 데이터셋(prior, test, train) 별로 day_of_week에 따른 주문량 확인



- prior data의 그래프가 test data와 train data의 분포와 다르게 그려지는 것 알 수 있다.
- 이것은 전체 데이터에서 prior data와 test data, train data를 구분할 때의 오류일 수 있을 것이다.

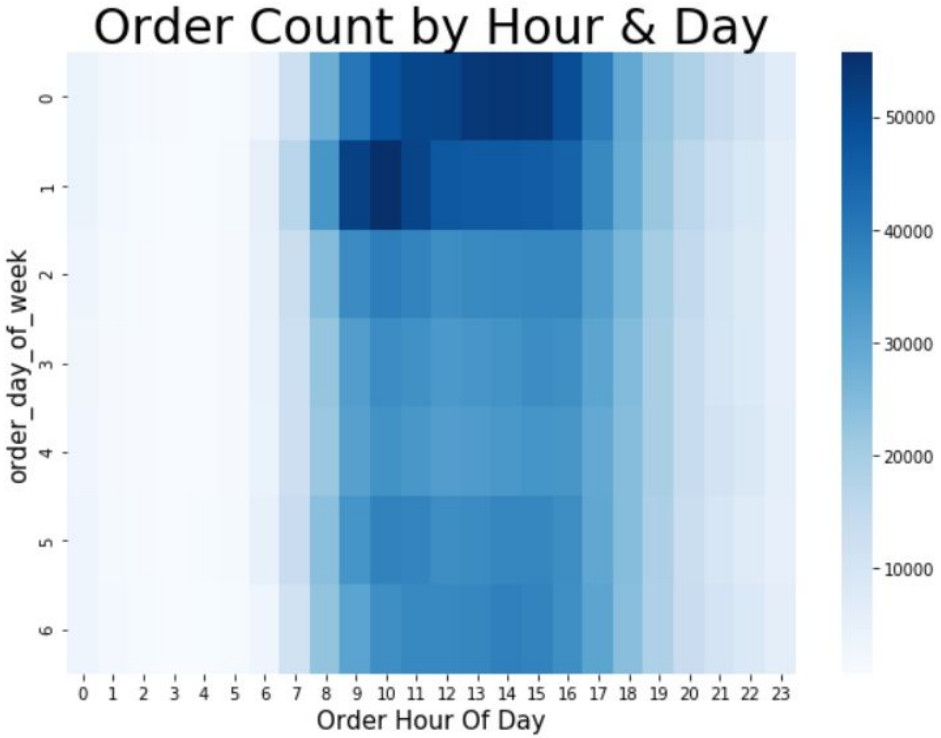
03 데이터 분석 - day_of_week, 시간대별 주문량 비교로 day_of_week의 0~6의 요일 파악



■ 주말
- 0과 1에서 주문량이 확연히 많음 → 0과 1은 주말을 나타냄 (0 : 토요일, 1: 일요일)
- 토요일은 3시를 기점으로 주문량이 많아
- 일요일은 10시를 기점으로 주문량이 많아짐

■ 평일
- 비슷한 패턴을 보이는데, 다른 평일은 오전 10시를 기점으로 주문량이 하락하다 다시 증가하는 추세
- 반면, 금요일의 경우 계속 증가하는 추세를 보임

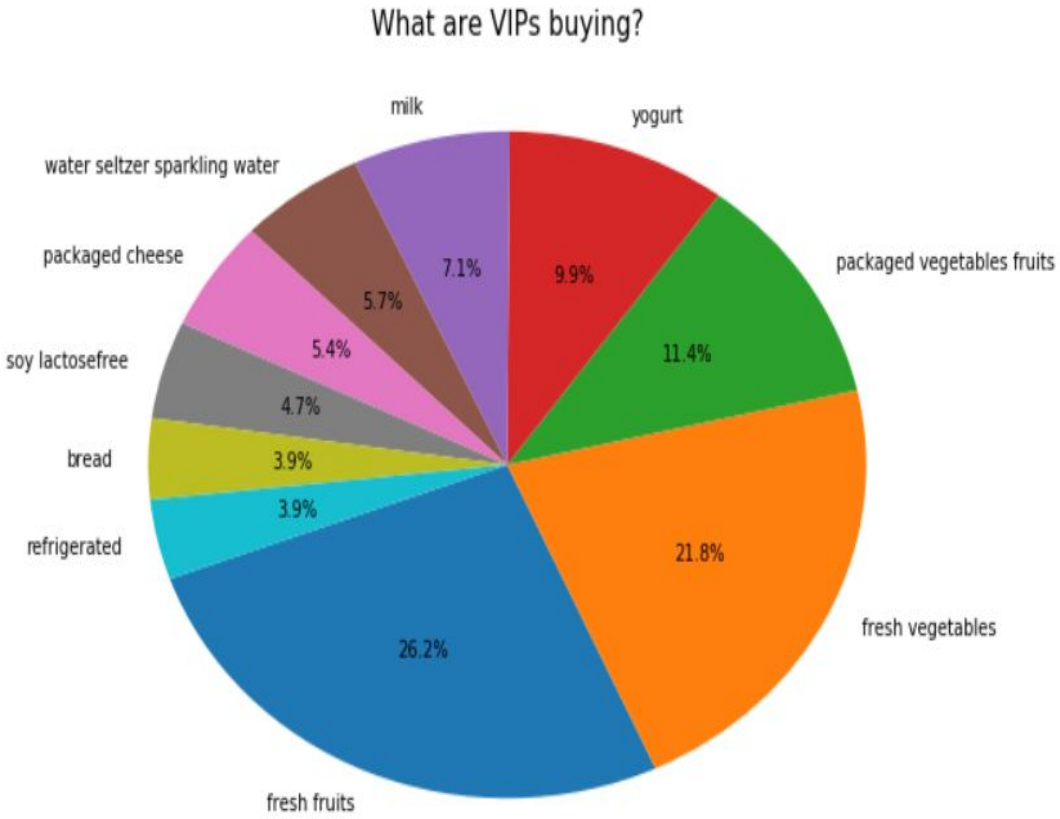
03 데이터 분석 - 시간, 요일별 주문량 Heatmap



주말 열두시 전후 2시간에 주문량이 몰림

03 데이터 분석 - 총 주문 횟수가 같은 사람들이 몇 명인지 확인

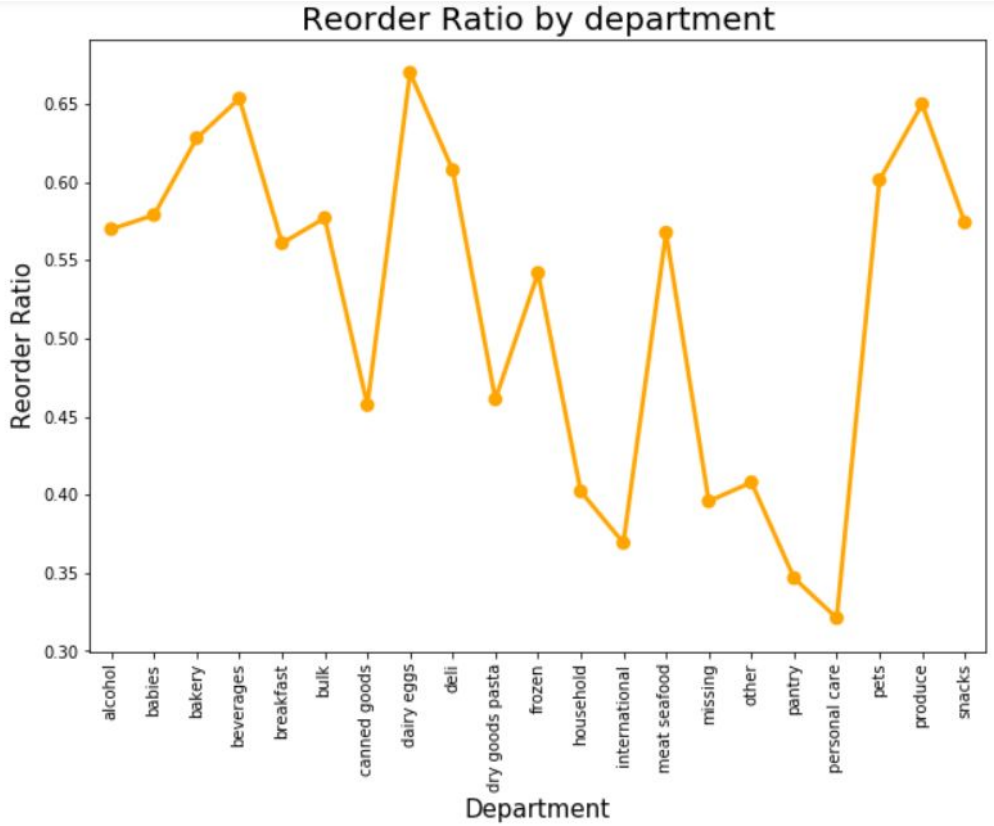
03 데이터 분석 - VIP가 구매하는 제품의 품목 비율



과일, 야채 품목은 VIP 구매 제품의 절반 이상을 차지

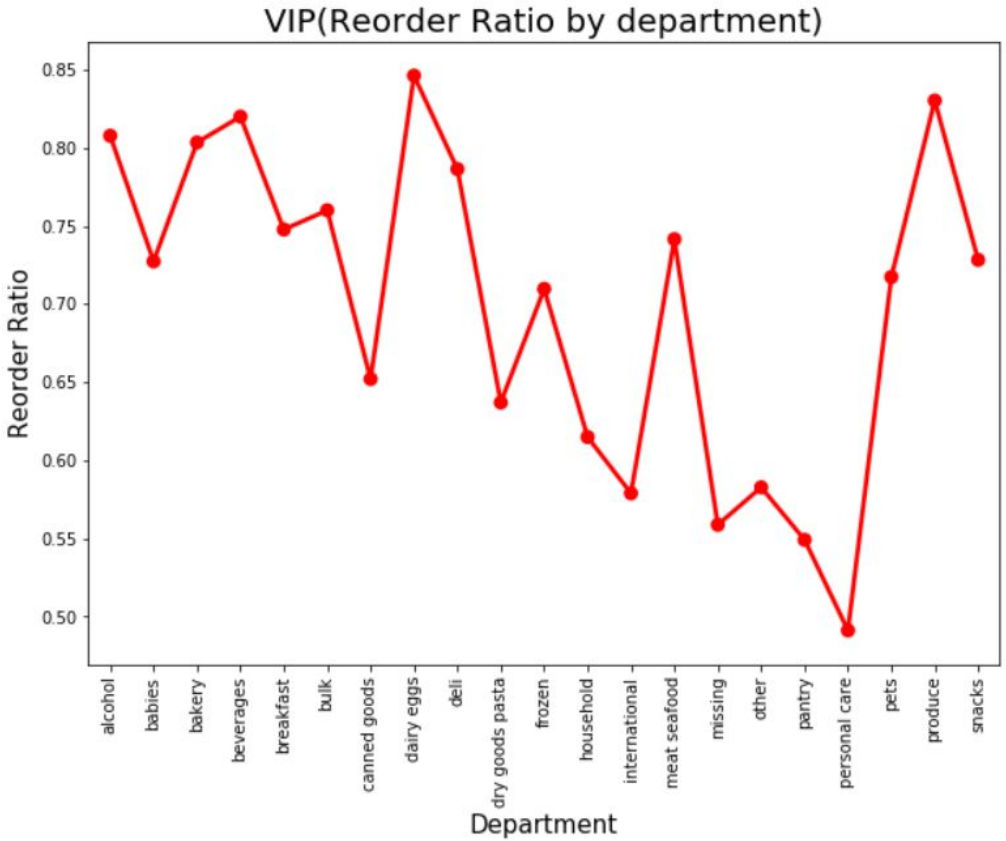
과일, 야채, 유제품 순으로 많이 구매

03 데이터 분석 - 전체고객 재구매비율이 높은 물품



- VIP 고객의 재구매비율 그래프와 매우 흡사
- 유제품, 음료, 농산물 순으로 높음
- 해외 제품, pantry , 개인 생활 용품이 제일 낮음

03 데이터 분석 - VIP 고객 재구매비율이 높은 물품



- 전체 고객의 재구매비율 그래프와 매우 흡사
- 유제품, 농산물, 음료 순으로 높음
- missing, pantry, 개인 생활 용품이 제일 낮음