

CP1_PROJECT

Instacart Data를 활용한 고객 세분화 및 재구매 강화 분석

AIB_15 박준영

Content 1

프로젝트 개요

Content 2

프로젝트 구성

Content 3

프로젝트 수행 절차 및 방법

Content 4

프로젝트 수행 결과

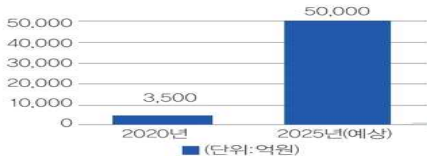
Content 5

자체 평가 의견

CONTENT 01 프로젝트 개요

퀵커머스 시장의 성장

- 퀵커머스란?
: 퀵(Quick)과 커머스(Commerce)의 합성어로 일반적으로 2시간 이내의 배송을 의미
- 2020년 3,500억 원 -> 2025년 5조 원대로 성장 예상, 유통물류 기업과 스타트업 등 다양한 업체 진출
- (코로나 이전) 맞벌이 부부, 1인 가구, 거동불편한 고객 등 직접 장보기 불편한 그룹이 예측
- (코로나 이후) 물리적으로 장보기 한계 → **많은 고객층 유입** → **고객의 성향에 따른 재분류가 필요**
- 생활유형이 아닌 주문 상품 군에 따라 새로운 고객그룹이 설정될 것이다



Instacart

1) 개요

- 온라인 기반 농작물 및 식료품 배송 서비스이다.
- 기존에 점포를 소유하고 있는 식품 회사들과 관계를 맺어가고 있으며 이에 따라 인스타카트 사용자들이 점포 가격에 따라 구입

2) 고객층

- 맞벌이 부부, 1인 가구, 거동불편 등등
- 코로나 이후 고객층이 늘어남 → 고객의 성향에 따른 분류가 필요

3) 특징 및 강점

- 3 무 → `물류창고x, 재고x, 트럭x` → 비용감소
- 다양한 소매업체와 제휴 및 협력 → 경쟁이 아닌 `win-win`관계
- 대형 마트의 독과점적 구조와 대비



문제 설정

- (코로나 이전) 맞벌이 부부, 1인 가구, 거동불편한 고객 등 직접 장보기 불편한 그룹이 예측
- (코로나 이후) 물리적으로 장보기 한계 → **많은 고객층 유입**
→ **고객의 성향에 따른 재분류가 필요**
- 생활유형이 아닌 주문 상품 군에 따라 새로운 고객그룹이 설정될 것이다

프로젝트 목표

- Instacart 식료품 주문 데이터를 분석하여 구매 패턴 파악
- 데이터를 바탕으로 비슷한 특성을 가진 고객 그룹 세분화
- 그룹별 특성을 활용하여 특성별 재구매 상품 추천
- 소형 소매업체에서의 활용방안 제안

CONTENT 02 프로젝트 구성

프로젝트 참여자	담당 업무
박준영	<ul style="list-style-type: none">▶ 1인 프로젝트▶ 데이터 정제 및 분석▶ 탐색적 데이터 분석, 시각화▶ PCA, KMeans Clustering

01

데이터
정제 및 분석

02

탐색적
데이터 분석,
시각화

03

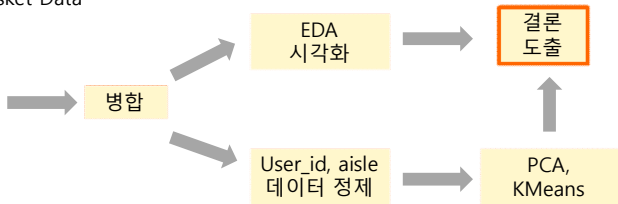
PCA,
KMeans
Clustering

CONTENT 03 프로젝트 수행 절차 및 방법

프로젝트 과정

Instacart Market Basket Data

- aisles.csv
- departments.csv
- order_products__prior.csv
- order_products__train.csv
- orders.csv
- products.csv



CONTENT 04 프로젝트 수행 결과

1) DATA 탐색 - 병합

Instacart Market Basket Data

aisles.csv

departments.csv

order_products__prior.csv

order_products__train.csv

orders.csv

products.csv

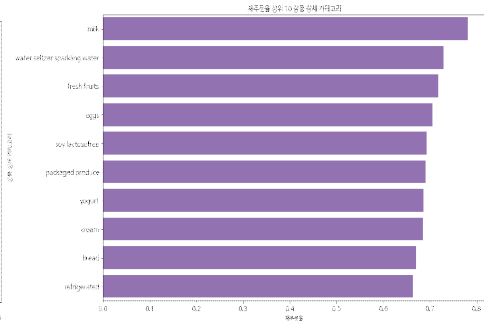
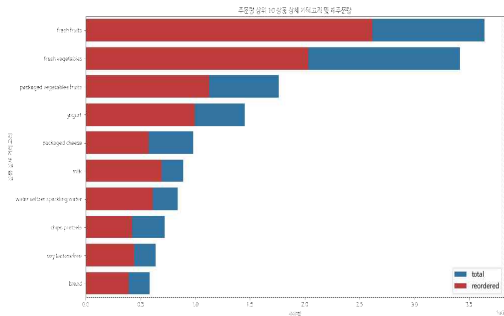


merge 함수를 이용하여
하나의 데이터로 병합

order_id	product_id	add_to_cart_order	reordered	product_name	aisle_id	department_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order	aisle	department
0	2	33120	1	1 Organic Egg Whites	06	16	202279	prior	3	5	9	8.0	eggs	dairy eggs
1	26	33120	5	0 Organic Egg Whites	06	16	153404	prior	2	0	16	7.0	eggs	dairy eggs
2	120	33120	13	0 Organic Egg Whites	06	16	23750	prior	11	6	8	10.0	eggs	dairy eggs
3	327	33120	5	1 Organic Egg Whites	06	16	58707	prior	21	6	9	8.0	eggs	dairy eggs
4	390	33120	28	1 Organic Egg Whites	06	16	166654	prior	48	0	12	9.0	eggs	dairy eggs

CONTENT 04 프로젝트 수행 결과

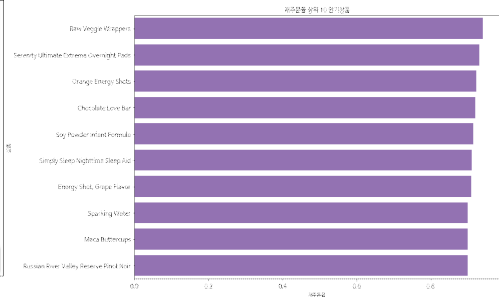
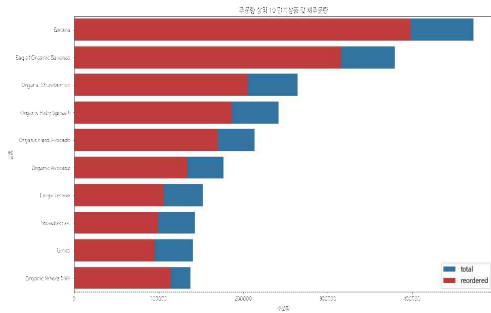
1) DATA 탐색 - 상위 판매 카테고리



- 신선 과일과 채소, 유제품의 카테고리가 상위 주문
- 우유, 탄산수, 신선과일, 계란 등의 카테고리가 상위 재주문

CONTENT 04 프로젝트 수행 결과

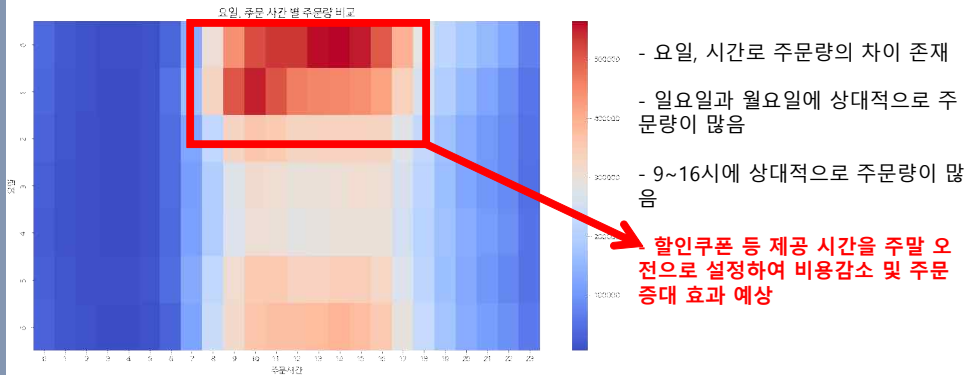
1) DATA 탐색 - 상위 판매 상품



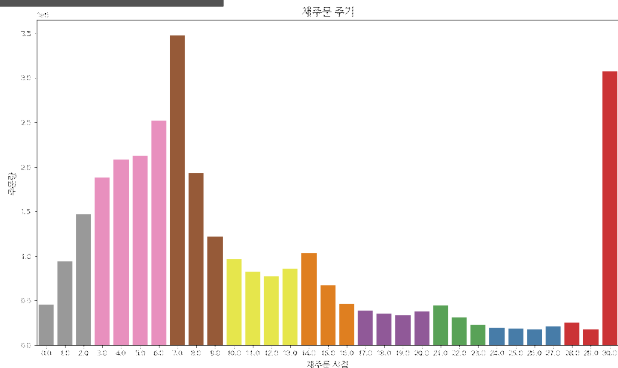
- 바나나, 딸기, 아보카도 등 과일이 상위 주문
- 간편식, 생리대, 에너지음료 등이 상위 재주문

CONTENT 04 프로젝트 수행 결과

1) DATA 탐색 - 요일, 주문시간 별 주문량 비교



1) DATA 탐색 - 재주문 주기



- 정기적으로 필요한 재주문 상품의 경우 1주, 1달 주기로 재주문

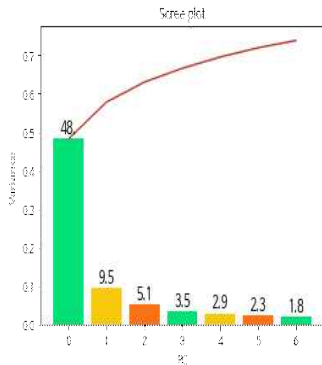
- 1주, 1달 주기로 푸시알림 발송을 고려

CONTENT 04 프로젝트 수행 결과

2) PCA

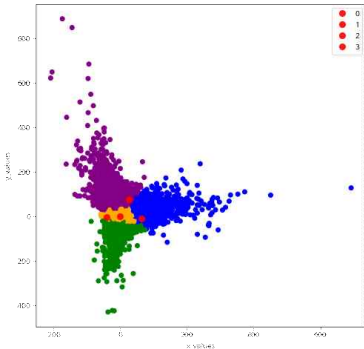
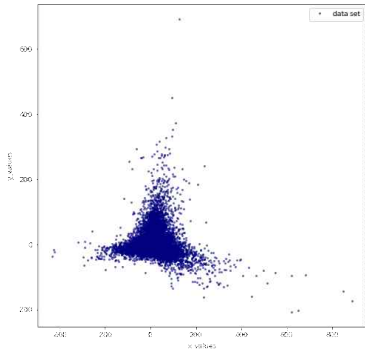
- clustering을 통한 고객 분류 이전에 특성에 대한 차원 축소
- 분류 기준은 aisles(상품 상세 카테고리)이 적절할 것으로 판단

index	0	1	2	3	4	5	6
0	-24.215659	2.429427	-2.466370	-0.145681	0.269055	-1.432755	2.142516
1	6.463208	36.751116	8.382553	15.097530	-6.920934	-0.978339	6.012066
2	-7.990302	2.404383	-11.030064	0.672239	-0.442365	-2.823118	-6.283449
3	-27.991129	-0.755823	-1.921732	2.091888	-0.288228	0.926198	0.827523
4	-19.896394	-2.637225	0.533229	3.679231	0.612828	-1.623953	-3.935466
...
206204	-24.200168	1.369802	1.325348	4.935903	0.318349	0.004891	0.398322
206205	-4.837135	-0.152500	5.263276	-21.059406	0.728541	-13.415968	13.338668
206206	8.611433	7.701299	7.952402	-5.906237	1.718246	-3.266852	-5.629643
206207	84.086220	20.418734	8.054107	-15.558599	9.035234	-37.417508	-16.398797
206208	-13.953456	6.646218	-5.230303	-0.979936	1.539632	-5.098273	0.899366



3) KMeans Clustering

- 주문 데이터 기준으로 고객 분류를 위해 KMeans Clustering 사용
- 4개의 클러스터로 그룹 분류



3) KMeans Clustering

group 1

aisle	
fresh fruits	12.996203
fresh vegetables	11.259972
packaged vegetables fruits	6.531537
yogurt	4.840170
packaged cheese	3.754606

group 2

aisle	
fresh fruits	84.542404
yogurt	63.034564
packaged vegetables fruits	28.158321
water seltzer sparkling water	25.796156
fresh vegetables	22.910808

group 3

aisle	
fresh vegetables	96.875269
fresh fruits	51.391756
packaged vegetables fruits	27.902778
fresh herbs	11.309857
packaged cheese	10.643459

group 4

aisle	
baby food formula	90.031453
fresh fruits	72.334056
fresh vegetables	50.059111
packaged vegetables fruits	34.557484
yogurt	33.242950

- 그룹 1,2,3은 차이는 있으나 비슷한 구매 패턴이 보임.

- 그룹 4의 경우 **아기 분유**에 대한 구매가 최대로 나타나는 특징이 보임.

- 그룹 4는 아기가 있는 가정으로 예상하고 그룹 4에 아기 용품을 추천 할 수 있음.

4) 소형 소매업체에서의 활용방안



오늘의 추천 집밥

레시피 식재료 구매 가능한 마트로 연결됩니다.



열정의 삼계탕



파리고추 매추리알 조림



지도 검색



주식장표



홈



주변내역



최근본

동네 소매점 연계 장보기 어플
슈핑

- 대형 마트들이 각자의 쿼커머스 서비스를 진행하기에 진입장벽이 높다.

- 인근 다른 소매점과 연계 → 하나의 소매점이 아닌 다수의 소매점 쇼핑 가능

- 슈핑, 로마켓 등 동네 거점 소매점과 연계한 회사에 적용

- 쿠폰, 할인권 등 기존에 제공하던 서비스를 고객그룹 특성에 맞추어 주말 오전, 낮 시간에 제공하여서 비용을 줄인다면 대형마트의 독과점 구조 속에서 Instacart와 같이 소형 소매점들과 협력하여 수익을 창출할 수 있을 것이라 판단



- 소형 소매점에서도 활용할 수 있는 고객 세분화와 추천 분석 제공
- KMeans Clustering을 이용한 것에 대한 성취는 있음
-> 그룹 1,2,3에 대한 명확한 차이는 부족
- XGBoost와 같은 분류 모델링 진행한다면 더 나은 분석이 될 것
- 소형 소매업체에서의 더욱 실질적인 방안 모색이 필요함

도메인에 대한 공부와 고민 특히 좋은 문제 설정을 위한 논리적 사고가 더욱 필요함

노션 페이지

<https://www.notion.so/cp1-Instacart-4ee66bd1c0b04ce49328a18c1cbb086e>

감사합니다

AIB_15 박준영