Instacart 데이터 분석 프로젝트

AI 15기 김수영

프로젝트 목차

프로젝트 개요

프로젝트 목표 :

Instacart 데이터 분석을 통한 주요 상품 파악

고객 구매 패턴 분석에 따른 마케팅 전략 제시

Instacart란?

인스타카트는 세계최대의 온라인 기반 식료품 배송업체입니다.

홈페이지나 앱을 통해 주문을 하면 대신 장을 보고 몇 시간 내로 배송을 해주는 서비스를 제공하고 있습니다.

Instacart 데이터는?

인스타카트의 데이터 사이언스 팀에서

오픈소스로 배포한 익명화된 고객 주문 데이터 입니다.

총 206,209명 고객의 약 330만 건의 식료품 주문 정보를 담고 있으며,

총 주문 상품의 개수는 3380만 개에 달합니다.

전체 상품 종류는 49,688개이며,

21종의 대분류(department)와 134종의 소분류(aisle, 진열대)로 구분되어 있습니다.

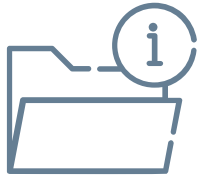
프로젝트 구성

훈련생	담당 업무
김수영 (개인프로젝트)	▶ 데이터 전처리 ▶ 문제 정의 ▶ 데이터 분석 및 시각화 ▶ 비즈니스 인사이트 도출

프로젝트 수행 절차 및 방법

문제 정의

해결해야 할
문제를 객관적이고
구체적으로
정의합니다.



데이터 수집 및 전처리

데이터를 수집하고,
수집한 데이터를
분석 용도에 맞게
Pandas모듈을
사용하여
가공합니다.

데이터 시각화 & 분석

Plotly 라이브러리를
활용하여 데이터를
시각화하고,
데이터를 분석하여 문제
정의 단계에서 정의한
문제들을 해결합니다.

비즈니스 인사이트 도출

분석 결과를 바탕으로
비즈니스 인사이트를
도출합니다.

프로젝트 수행 결과

문제 정의

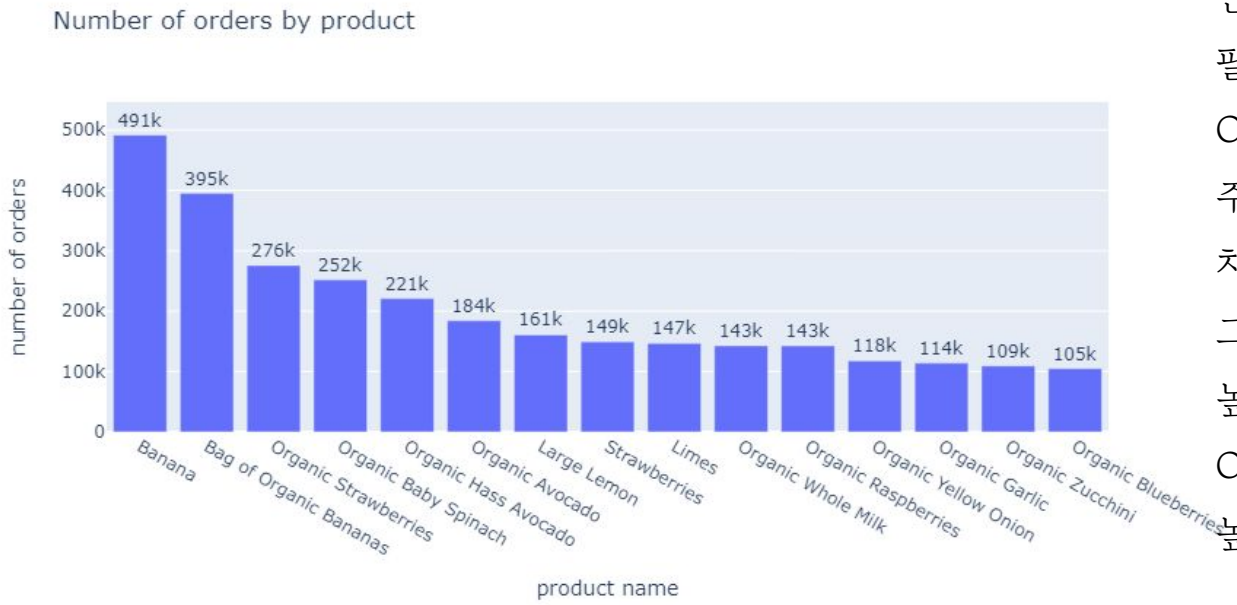
저는 이번 프로젝트에서 아래 4가지의 문제를 해결하여 비즈니스 인사이트를 도출하고자 합니다.

1. 고객이 가장 많이 주문한 상품/상품분류는 무엇일까?
2. 시간대마다 주로 주문하는 상품/상품분류가 다를까?
3. 재주문율이 가장 높은 상품/상품분류는 무엇일까?
4. 고객이 장바구니에 가장 먼저 담는 상품/상품분류는 무엇일까?

프로젝트 수행 결과

데이터 시각화 & 분석

1. 고객이 가장 많이 주문한 상품/상품분류는 무엇일까?



단일 품목으로는 바나나가 가장 많이 팔렸습니다. Banana 상품과 Bag of Organic Bananas 상품을 합치면 전체 주문(33.5M)의 2.6% 정도의 높은 비중을 차지합니다.

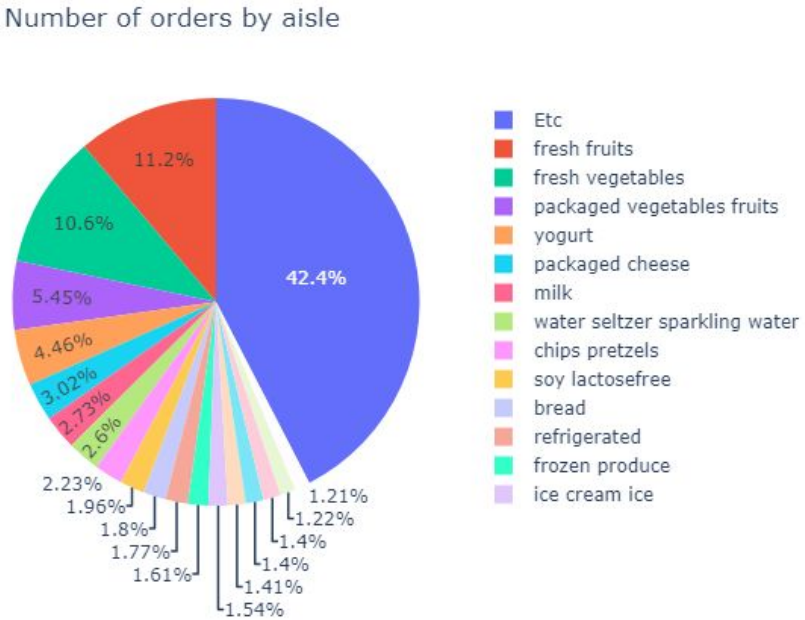
그 외에도 전반적으로 과일과 야채종류가 높은 주문량을 나타내고 있습니다.

Organic(유기농) 상품들의 순위가 높은것을 확인 할 수 있습니다.

프로젝트 수행 결과

데이터 시각화 & 분석

1. 고객이 가장 많이 주문한 상품/상품분류는 무엇일까?



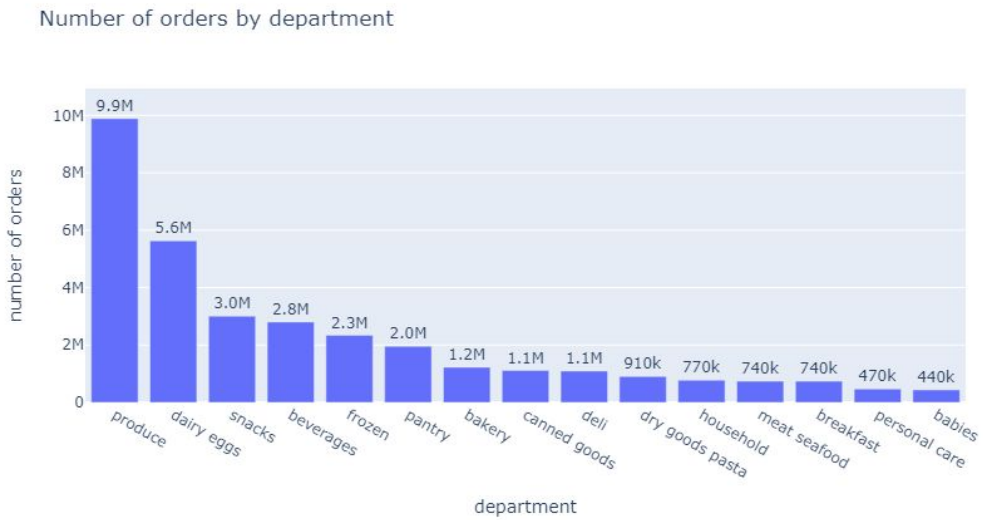
소분류(aisle)별 주문량에서는 신선 과일과 신선 채소가 높은 비중을 차지하였습니다.

3번째 많이 주문한 소분류도 포장된 채소, 과일이므로 1~3위를 합쳐 전체 주문(33.5M)의 25% 가량을 야채와 과일이 차지하는 것을 확인 할 수 있습니다.

프로젝트 수행 결과

데이터 시각화 & 분석

1. 고객이 가장 많이 주문한 상품/상품분류는 무엇일까?

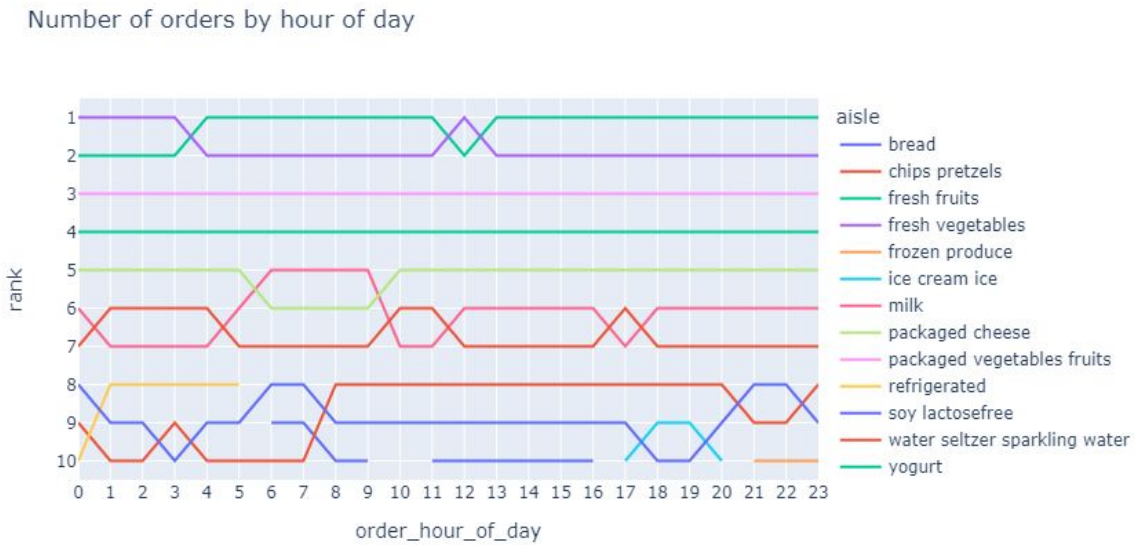


대분류(department)별 주문량을 살펴보면 produce(농작물) 분류가 가장 높고, 그 뒤로 유제품&계란, 스낵, 음료, 냉동식품 순으로 높은 비중을 차지했습니다.

프로젝트 수행 결과

데이터 시각화 & 분석

2. 시간대마다 주로 주문하는 상품/상품분류가 다를까?



1~4위는 거의 변동이 없습니다.

과일의 주문량이 가장 많지만 낮 12시에는 채소의 주문량이 많아집니다.

아침 6~9시에 우유의 주문량이 늘어나는 것을 확인할 수 있습니다.

21시에서 24시까지는 냉동제품이, 0시에서 5시까지는 냉장제품 진열대의 주문량이 평소보다 높습니다.

6~9시, 11~16시에 빵의 주문량이 늘어납니다.

프로젝트 수행 결과

데이터 시각화 & 분석

3. 재주문율이 가장 높은 상품/상품분류는 무엇일까?

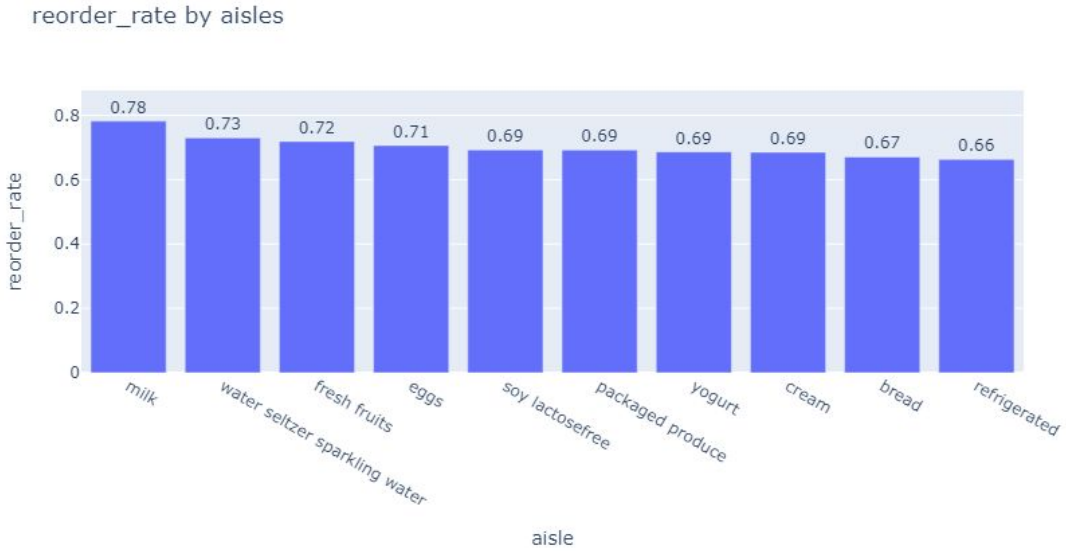
product_name	order_count	reorder_rate
Half And Half Ultra Pasteurized	2995	0.861436
Whole Organic Omega 3 Milk	9410	0.859830
Organic Lactose Free Whole Milk	8742	0.859186
Organic Homogenized Whole Milk	4095	0.858120
Ultra-Purified Water	1524	0.856955
Milk, Organic, Vitamin D	20770	0.854742
Organic Reduced Fat Milk	36869	0.851501
Goat Milk	5353	0.850177
Banana	491291	0.845051
Organic Lowfat 1% Milk	15352	0.841193

상품별 재주문율을 데이터프레임으로 분석한 결과,
다양한 우유 제품들의 재주문율이 높게 나온것을 확인할 수 있었습니다.
대부분의 고객들이 원래 마시던 우유제품를 재주문 하는 경우가 많기 때문이라고 생각됩니다.

프로젝트 수행 결과

데이터 시각화 & 분석

3. 재주문율이 가장 높은 상품/상품분류는 무엇일까?



소분류 별로 재주문율을 확인한 결과,
역시나 우유류의 재주문율이 눈에 띄게
높은것을 확인 할 수 있었습니다.

그렇다면 고객들은 재주문율이 높은
상품들을 평균적으로 몇 일마다
재구매할까 한번 확인해보았습니다.

프로젝트 수행 결과

데이터 시각화 & 분석

3. 재주문율이 가장 높은 상품/상품분류는 무엇일까?

```
#1위 우유
tmp = order_products_df.query('(aisle=="milk") and (days_since_prior_order>0)')
tmp.days_since_prior_order.mean()
```

```
10.784601852270894
```

```
#2위 water seltzer sparkling water
tmp = order_products_df.query('(aisle=="water seltzer sparkling water") and (days_since_prior_order>0)')
tmp.days_since_prior_order.mean()
```

```
11.590479639564434
```

```
#3위 fresh fruits
tmp = order_products_df.query('(aisle=="fresh fruits") and (days_since_prior_order>0)')
tmp.days_since_prior_order.mean()
```

```
10.906245099637745
```

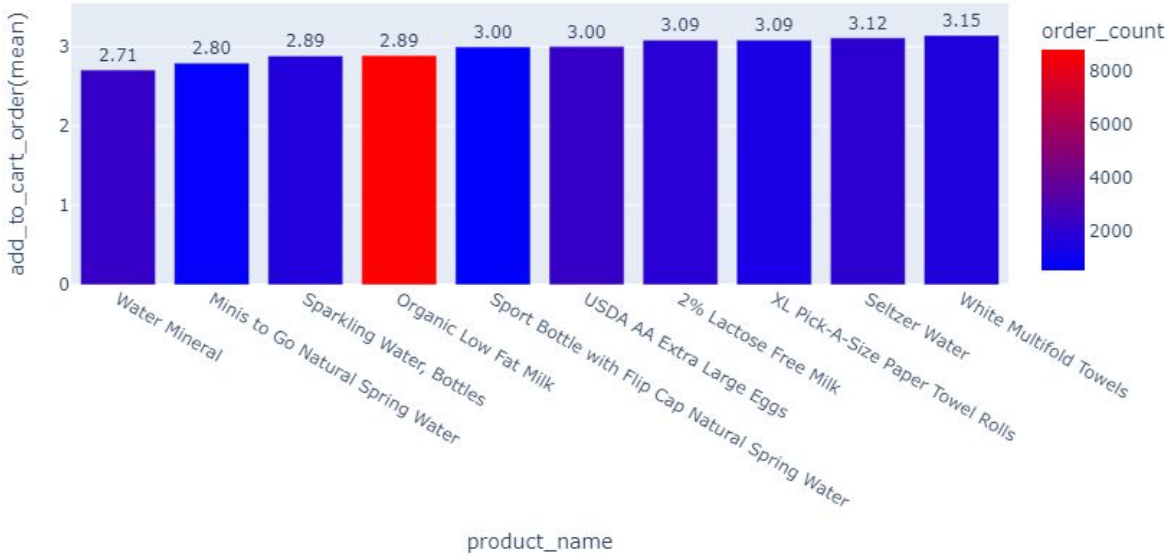
우유, 물, 과일 모두 평균적으로
구매한지 10~11일 정도 후에 재주문을
하는것으로 나타났습니다.

우유, 물, 과일을 구매한지 7일이 넘은
고객들에게 이전에 구매한것과 동일한
제품을 추천한다면 주문하거나
장바구니에 담을 가능성이 높을것이라
기대할 수 있습니다.

프로젝트 수행 결과

데이터 시각화 & 분석

4. 고객이 장바구니에 가장 먼저 담는 상품/상품분류는 무엇일까?



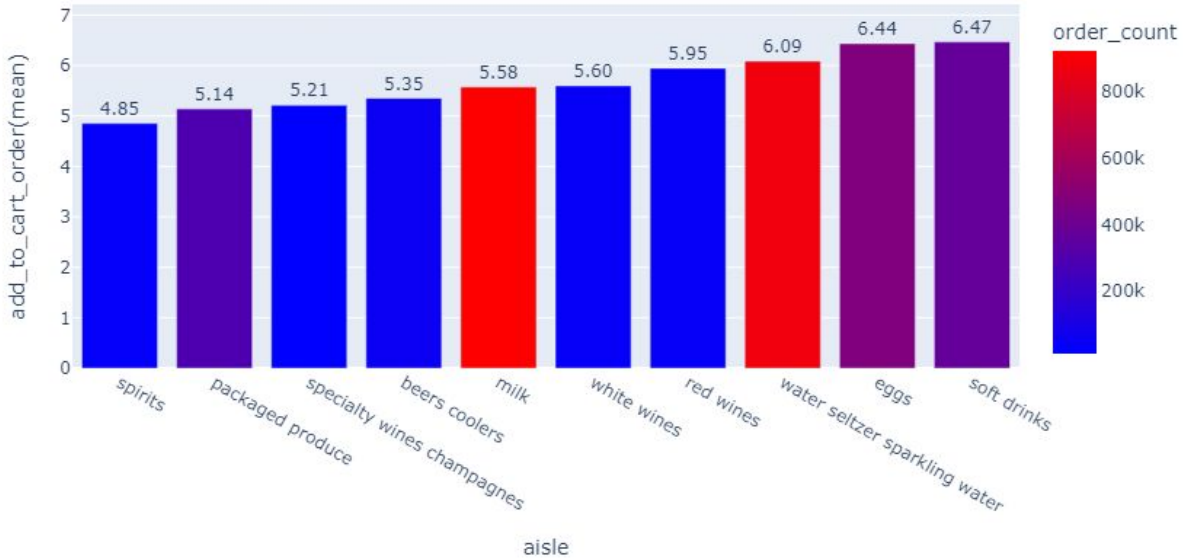
인스타카트 고객들은 장바구니에 먼저 넣는 상품은 물 제품들이 많은것을 확인할 수 있었습니다.

물 외에도 우유나 콜라 등의 음료류가 빠른 장바구니 순서를 나타내고 있습니다.

프로젝트 수행 결과

데이터 시각화 & 분석

4. 고객이 장바구니에 가장 먼저 담는 상품/상품분류는 무엇일까?



소분류별 장바구니 평균 순서는 증류주, 포장제품, 샴페인 순으로 상품별과는 조금 다른 결과가 나왔습니다.

하지만 상품별과 비슷하게 증류주, 맥주, 샴페인, 우유, 물, 두유 각종 드링크 등 음료류가 장바구니에 담기는 순번이 빠른 경향을 보이는 것을 확인 할 수 있었습니다.

프로젝트 수행 결과

비즈니스 인사이트 도출 & 마케팅 전략 제시

주문량이 많은 상품 4종은 시간에 관계없이 언제나 인기가 좋습니다.
fresh fruits, fresh vegetables, packaged vegetables fruits, yogurt 4분류의 상품은 shop페이지 상단에 계속 위치시켜 접근성을 높여야 합니다.

시간대마다 많이 주문하는 상품의 종류가 달라집니다.
12시에는 채소, 6~9시에 우유, 21시~24시 냉동제품, 0시~5시 냉장식품, 6~9시, 11~16시에 빵의 주문량이 증가합니다. 해당시간에 인스타카트의 shop페이지에 해당 aisle의 상품들을 더 노출시키면 클릭률을 높일 수 있습니다.

프로젝트 수행 결과

비즈니스 인사이트 도출 & 마케팅 전략 제시

우유, 물, 과일은 같은 제품의 재주문율이 높은 상품들입니다.
우유, 물, 과일을 구매한지 7일이 넘은 고객들에게 이전에 구매한 것과
동일한 제품을 추천한다면 다른 제품에 비해 주문할 가능성이 더 높습니다.

인스타카트의 고객들은 주류와 음료를 장바구니에 빠른 순서로 담는 경향이 있습니다.
현재 장바구니가 비어있는 고객들에게 주류/음료를 추천하여 클릭률을 높일 수 있습니다.

자체 평가 의견

아쉬웠던 점
전체적인 프로젝트 완성도가 부족하다고 느꼈습니다.
이커머스에 대한 도메인 지식이 더 많았다면 좀 더 심층적이고 전문적인 분석을 할 수 있었을 텐데 그 점이 아쉽습니다.
프로젝트 기간중에 아파서 시간을 온전히 활용할 수 없었던 점도 아쉽습니다.

만족스러웠던 점
Plotly라는 새로운 시각화 라이브러리에 대해 공부할 수 있어서 좋았습니다.

감사합니다