



# 분석 보고서 (김예진)

## 🔗 분석 목적

Instacart Market Basket Analysis

Which products will an Instacart consumer purchase again?

[k https://www.kaggle.com/c/instacart-market-basket-analysis](https://www.kaggle.com/c/instacart-market-basket-analysis)



- United States와 Canada에서 온라인 기반 농작물 배송 서비스 하는 업체의 거래 내역을 분석하여 주문 패턴의 경향과 주문 특징을 파악

이를 바탕으로 마케팅 전략을 수립하여 효과적으로 상품을 판매하고 이익 극대화를 목적으로 함

## 🔗 데이터 소개

### (1) Orders.csv

- 주문 정보, 마지막 주문으로부터 얼마나 걸렸는지에 대한 정보
- 3421083 rows , 7 columns 로 구성

Aa column 명	≡ 기능 설명	▼ type	▼ NaN값 여부
<u>order_id</u>	주문 ID	int64	X

Aa column 명	기능 설명	type	NaN값 여부
<u>user_id</u>	유저 ID	int64	X
<u>eval_set</u>	prior : 과거 구매한 주문내역인 경우 train , test : 최근 구매한 주문내역인 경우	object	X
<u>order_number</u>	구매 순서 ( 몇번째 구매인지)	int64	X
<u>order_dow</u>	구매 요일 (0: Sunday ~ 6: Saturday)	int64	X
<u>order_hour_of_day</u>	구매 시간	int64	X
<u>days_since_prior_order</u>	마지막 구매일로부터 걸린 시간(단위 : 일, NA : 첫구매 )	float64	206209개

## (2) order\_products\_\_prior.csv

- orders.csv 파일의 eval\_set 칼럼값이 prior 인 주문내역에 대한 데이터
- 어떤 상품을 구매했고, 상품을 장바구니에 담은 순서, 재구매한 상품인지에 대한 정보
- 32434489 rows , 4 columns 로 구성

Aa column 명	기능 설명	type	NaN값 여부
<u>order_id</u>	주문 ID	int64	X
<u>product_id</u>	주문한 상품ID	int64	X
<u>add_to_cart_order</u>	장바구니에 담은 순서	int64	X
<u>reordered</u>	재구매한 상품인지 아닌지 표현 (1: 재구매 , 0: 최초 구매)	int64	X

## (3) order\_products\_\_train.csv

- orders.csv 파일의 eval\_set 칼럼값이 train 인 주문내역에 대한 데이터
- 구매 상품, 장바구니에 담은 순서, 재구매한 상품인지에 대한 정보
- 1384617 rows , 4 columns 로 구성

Aa column 명	기능 설명	type	NaN값 여부
<u>order_id</u>	주문 ID	int64	X
<u>product_id</u>	주문한 상품ID	int64	X
<u>add_to_cart_order</u>	장바구니에 담은 순서	int64	X
<u>reordered</u>	재구매한 상품인지 아닌지 표현 (1: 재구매 , 0: 최초 구매)	int64	X

#### (4) products.csv

- product\_id에 대한 정보
- 49688 rows , 4 columns 로 구성

Aa column 명	기능 설명	type	NaN 값 여부
<u>product_id</u>	order_products__prior.csv, order_products__train.csv 의 product_id와 연결할 수 있는 Join Key	int64	X
<u>product_name</u>	상품 이름	object	X
<u>aisle_id</u>	상품 세부 카테고리 ID	int64	X
<u>department_id</u>	상품 카테고리 ID	int64	X

#### (5) department.csv

- 상품 카테고리ID (department\_id) 에 대한 정보
- 21 rows , 2 columns 로 구성

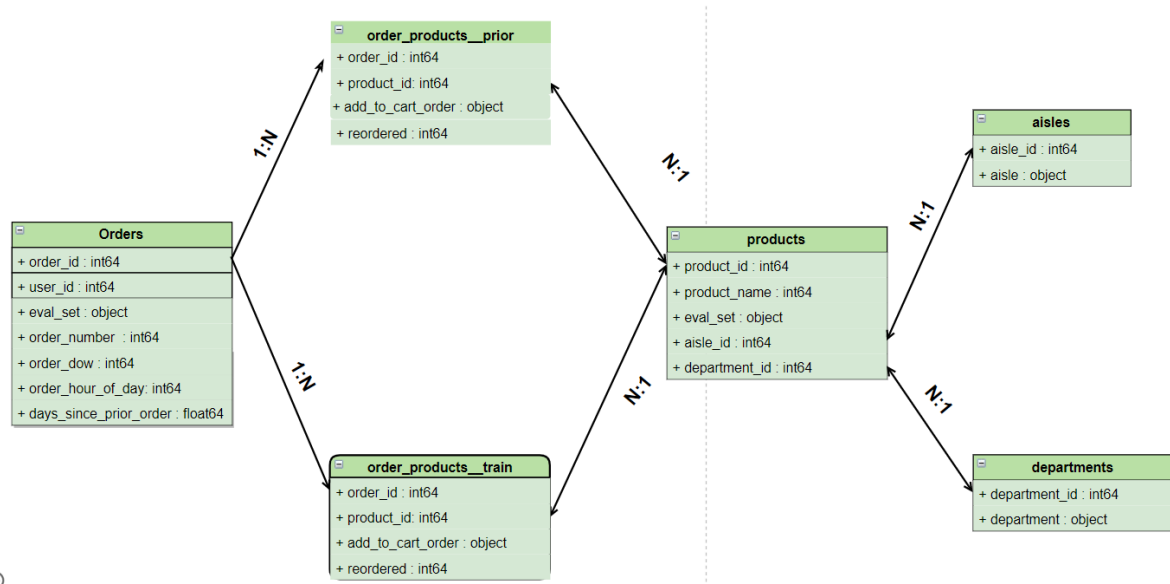
Aa column 명	기능 설명	type	NaN 값 여부
<u>product_id</u>	order_products__prior.csv, order_products__train.csv 의 product_id와 연결할 수 있는 Join Key	int64	X
<u>product_name</u>	상품 이름	object	X
<u>aisle_id</u>	상품 세부 카테고리 ID	int64	X
<u>department_id</u>	상품 카테고리 ID	int64	X

#### (6) aisles.csv

- 상품 세부 카테고리ID에 대한 정보
- 134 rows , 2 columns 로 구성

Aa column 명	기능 설명	type	NaN 값 여부
<u>aisle_id</u>	상품 세부 카테고리ID	int64	X
<u>aisle</u>	세부 카테고리 이름	object	X

## 데이터 관계



2