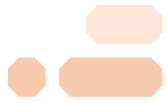


# 제품 추천 모델 구축

이커머스 사용자 행동 데이터 기반

AI 13 황한재

AI 13 한재원



# Contents

---

## 1. 프로젝트 개요

- 프로젝트 배경

## 2. 데이터 소개

- 경량화
- 전처리

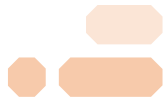
## 3. 데이터 분석 및 결과

- 코호트 분석

## 4. 추천 시스템 적용 및 결과

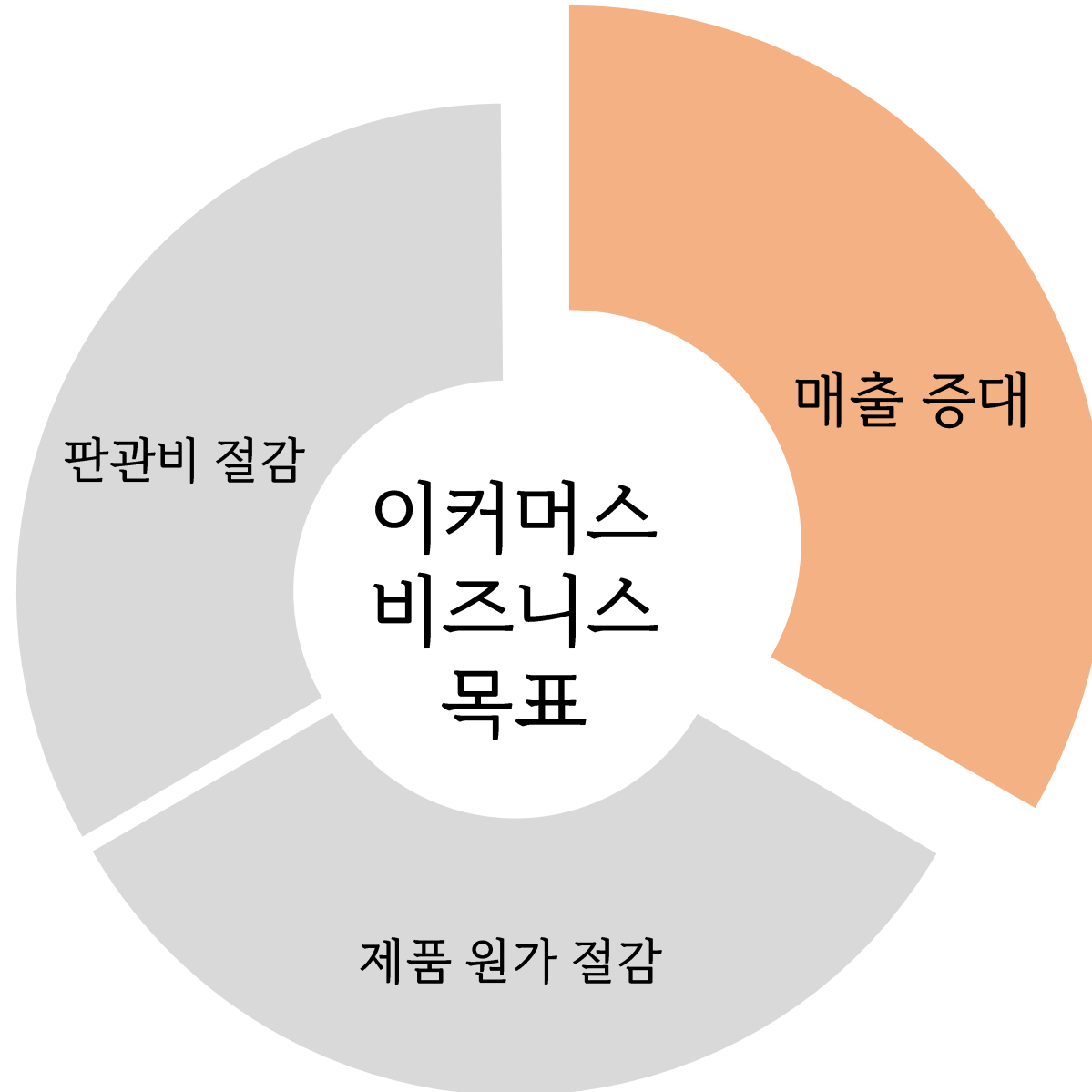
- 타당성
- 추천시스템 소개 및 선정 이유
- 성능 평가
- 결과 확인

## 5. 회고



# 프로젝트 개요

프로젝트 배경



# 데이터 소개

## 캐글

### eCommerce behavior data from multi category store

This dataset contains 285 million users' events from eCommerce website



Data Code (25) Discussion (14) Metadata

#### About Dataset

##### About

This file contain behavior data for 7 months (from October 2019 to April 2020) from a large multi-category online store.

Each row in the file represents an event. All events are related to products and users. Each event is like many-to-many relation between products and users.

Data collected by [Open CDP](#) project. Feel free to use open source customer data platform.

##### More datasets

Checkout another datasets:

1. <https://www.kaggle.com/mkechinov/ecommerce-behavior-data-from-multi-category-store> - you're reading it right now
2. <https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-electronics-store>
3. <https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop>
4. <https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-jewelry-store>
5. <https://www.kaggle.com/mkechinov/ecommerce-events-history-in-electronics-store>

How to read it

Usability ⓘ  
10.00

License  
Data files © Original Authors

Expected update frequency  
Quarterly

**event\_time** : 이벤트 발생 시간

- 2019-10-01 ~ 2019-10-31

**event\_type** : 유저 행동 유형

- view, purchase, cart

**product\_id** : 제품 ID

- 고유값 : 166744

**category\_id** : 카테고리 ID

**category\_code** : 제품이 속한 카테고리류

**brand** : 브랜드

**price** : 가격

**user\_id** : 유저 ID

- 고유값 : 3019274

**user\_session** : 유저 세션

# 데이터 소개

경량화

## 캐글

### eCommerce behavior data from multi category store

This dataset contains 285 million users' events from eCommerce website



Data Code (25) Discussion (14) Metadata

#### About Dataset

##### About

This file contain behavior data for 7 months (from October 2019 to April 2020) from a large multi-category online store.

Each row in the file represents an event. All events are related to products and users. Each event is like many-to-many relation between products and users.

Data collected by [Open CDP](#) project. Feel free to use open source customer data platform.

##### More datasets

Checkout another datasets:

1. <https://www.kaggle.com/mkechinov/ecommerce-behavior-data-from-multi-category-store> - you're reading it right now
2. <https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-electronics-store>
3. <https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop>
4. <https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-jewelry-store>
5. <https://www.kaggle.com/mkechinov/ecommerce-events-history-in-electronics-store>

How to read it

Usability ①  
10.00

License  
Data files © Original Authors

Expected update frequency  
Quarterly

행 : 약 4200만개

열 : 9개

크기 : 약 5GB csv

object  
float64  
int64



category  
float32  
int32

크기 : 약 1GB parquet



# 데이터 소개

전처리

event\_time

1. 2019-10-01 00:00:00 UTC

2. 2019-10-01 00:00:00

+4

3. 2019-10-01 04:00:00

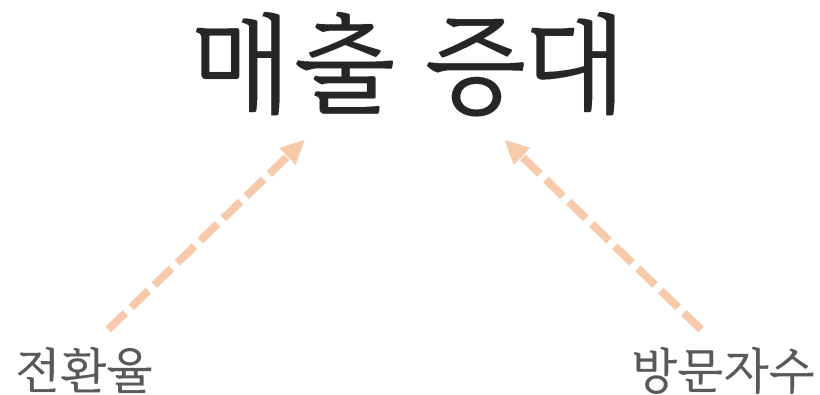
4. 2019-11-01 04:00:00

11월 이후 데이터 제거



## 데이터 분석 및 결과

전환 : 소비자가 제품을 장바구니에 담거나 구매를 한 경우

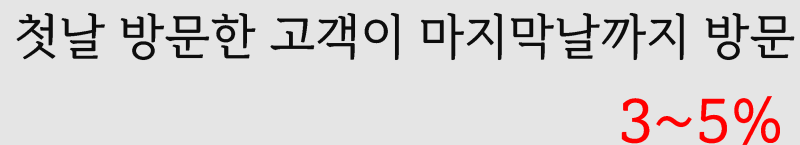


고객 세분화

전환이 일어나지 않은 사용자 VS 전환이 일어난 사용자

## Cohort Analysis

## 사이트 이용률

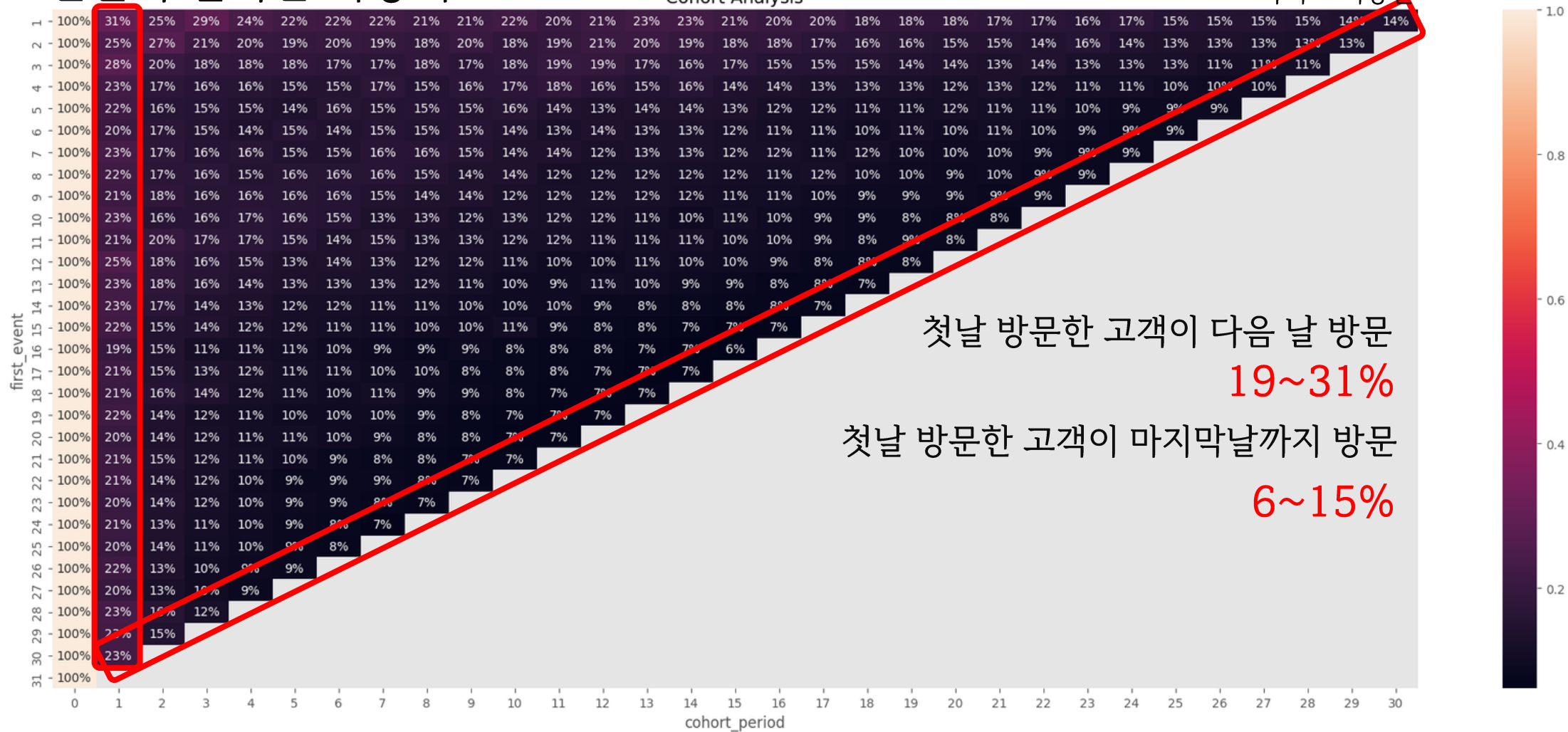


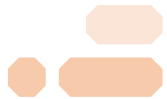


## 코호트 분석

## Cohort Analysis

## 사이트 이용률





## 추천 시스템 적용 및 결과 타당성

소비자 5명 중 3명,  
코로나19 이후  
온라인 쇼핑 시 이용하는  
“플랫폼의 개수가 늘어났다”



소비자 4명 중 1명,  
코로나19 이후  
구매여정이  
“길어지고 복잡해졌다”



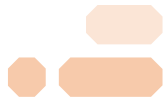
Q. (2020년 2월 이후, 코로나 이전 대비) 귀하께서는 이용하는 온라인 플랫폼(정보탐색/구매하는 사이트 등)에 변화가 있었나요? (늘어났다: 58.1%)

Q. (2020년 2월 이후, 코로나 이전 대비) 귀하께서는 온라인 쇼핑 구매 과정에 변화가 있었나요? (길어지거나 복잡해졌다: 26.2%)

Base: Total N=3,024 / Unit: %, SA

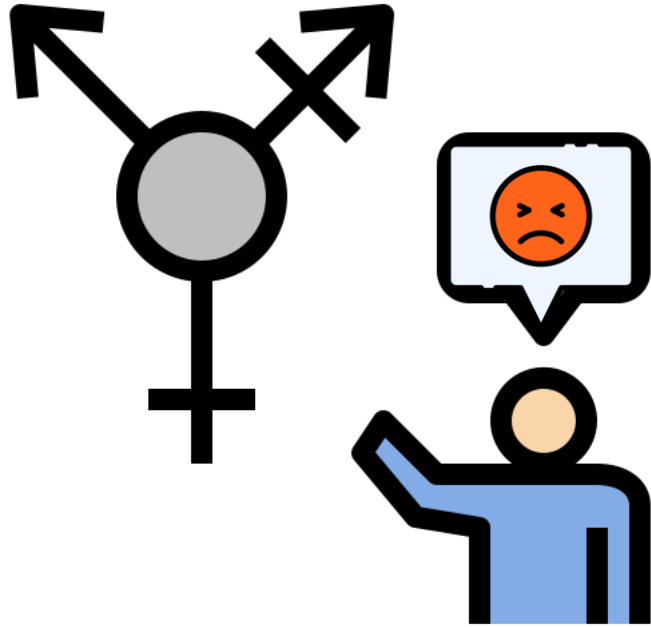
이커머스의 시장 규모 **증가**

제품 구매 시간 **지연**

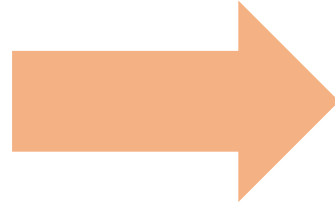


## 추천 시스템 적용 및 결과 타당성

웹 로그 분석을 통한 소비자 구매지연행동 연구 논문



대안이 너무 많음  
+  
부정적인 감정



제품 구매 지연



# 추천 시스템 적용 및 결과

추천시스템 소개 및 선정이유

## 데이터 형태

- Explicit feedback data

명시적 피드백 데이터

- 좋아요, 평점(1~5)

- Implicit feedback data

암시적 피드백 데이터

- view, cart, purchase 가 사용자가 이 제품을 선호한다! 라고 판단하기 어렵다.

사용자가 아직 보지못한 제품 추천 하기 위하여 CF 모델을 선정

협업 필터링

## Implicit 라이브러리

암시적 피드백 데이터 세트를 위한 협업필터링

Collaborative Filtering for Implicit Feedback Datasets

# 추천 시스템 적용 및 결과

성능평가

- hit@k: k개의 추천 아이템 중 사용자가 본 제품이 있다면 1 없다면 0  
10개의 추천 아이템 중 본 제품이 하나라도 있다면: 1
- precision@k : k개 추천 아이템 중 몇 개의 제품을 봤는지  
10개의 추천 아이템 중 4개를 봤다면 : 0.4

## 전환이 일어나지 않은 사용자

### 기준 모델

- hit@k : 0.27
- precision@k : 0.04

### 가중치 조절 모델

- hit@k : 0.35
- precision@k : 0.05

### 튜닝 후 모델

- hit@k : 0.37
- precision@k : 0.09

## 전환이 일어난 사용자

### 기준 모델

- hit@k : 0.47
- precision@k : 0.08

### 가중치 조절 모델

- hit@k : 0.51
- precision@k : 0.09

### 튜닝 후 모델

- hit@k : 0.53
- precision@k : 0.09

10000명의 사용자를 랜덤샘플링



## 추천 시스템 적용 및 결과 결과확인

전환이 일어나지 않은 사용자

### 사용자가 높은 비율로 본 제품

	product_id	category_code	brand	view_ratio
868	4700419	auto.accessories.videoregister	sho-me	3.066165
763	4700478	auto.accessories.videoregister	sho-me	3.025820
4726	4700557	auto.accessories.videoregister	sho-me	2.837547
4864	4700590	auto.accessories.videoregister	sho-me	2.770307
14595	5800802	electronics.audio.subwoofer	kenwood	2.649274
2019	5700788	auto.accessories.player	kenwood	2.595481
2814	5701062	auto.accessories.player	pioneer	2.474449
4495	5701086	auto.accessories.player	pioneer	2.474449
487	5700384	auto.accessories.player	pioneer	1.707907
23871	6100194	auto.accessories.radar	NaN	1.654115
1703	5700793	auto.accessories.player	kenwood	1.654115
4658	5700791	auto.accessories.player	sony	1.654115
96235	5701058	auto.accessories.player	pioneer	1.627219
6978	5801483	electronics.audio.subwoofer	pioneer	1.613771
1622	6100191	auto.accessories.radar	sho-me	1.573427
3527	5801482	electronics.audio.subwoofer	pioneer	1.546530
4922	5701063	auto.accessories.player	pioneer	1.519634
20750	5700863	auto.accessories.player	kenwood	1.519634
2015	4700497	auto.accessories.videoregister	NaN	1.385153
5047	5800140	electronics.audio.subwoofer	sony	1.344809

### 사용자에게 추천한 제품

	product_id	category_code	brand
1786	5701128	auto.accessories.player	NaN
2267	5801218	electronics.audio.subwoofer	NaN
753	5701166	auto.accessories.player	NaN
5682	5700282	auto.accessories.player	alpine
8364	5700850	auto.accessories.player	alpine
4729	5701246	auto.accessories.player	NaN
367	5700518	auto.accessories.player	NaN
4405	5701087	auto.accessories.player	jvc
2345	5701247	auto.accessories.player	NaN
1618	4700589	auto.accessories.videoregister	xiaomi



# 추천 시스템 적용 및 결과 결과확인

전환이 일어난 사용자

## 사용자가 높은 비율로 본 제품

	product_id	category_code	brand	view_ratio
1724	5100563	electronics.clocks	samsung	68.805310
635	5100718	electronics.clocks	samsung	5.309735
388	5100737	electronics.clocks	samsung	2.876106
113	1004249	electronics.smartphone	apple	2.654867
51555	5100862	electronics.clocks	NaN	1.991150
88	1005115	electronics.smartphone	apple	1.548673
1878	5100738	electronics.clocks	samsung	1.327434
556	1004250	electronics.smartphone	apple	0.884956
1104	1004777	electronics.smartphone	xiaomi	0.884956
122	5100816	NaN	xiaomi	0.663717
656	4803977	electronics.audio.headphone	samsung	0.663717
435	1005144	electronics.smartphone	apple	0.663717
56	1005105	electronics.smartphone	apple	0.663717
879	5100564	electronics.clocks	samsung	0.663717
531	1005104	electronics.smartphone	apple	0.663717
91645	5100868	electronics.clocks	NaN	0.663717
721	5100577	electronics.clocks	samsung	0.442478
24384	4600430	appliances.kitchen.dishwasher	electrolux	0.442478
17500	4600386	appliances.kitchen.dishwasher	bosch	0.442478
91556	5100867	electronics.clocks	NaN	0.442478

## 사용자에게 추천한 제품

	product_id	category_code	brand
51950	5100861	electronics.clocks	NaN
53090	5100863	electronics.clocks	NaN
124455	5100874	electronics.clocks	NaN
124393	5100875	electronics.clocks	NaN
65234	5100865	electronics.clocks	NaN
378	4803976	electronics.audio.headphone	samsung
1731	5100572	electronics.clocks	apple
137538	5100878	electronics.clocks	NaN
144	5100239	electronics.clocks	samsung
124477	5100876	electronics.clocks	NaN

### 배운점

- 이커머스 관련 도메인이 부족하여 인사이트 도출에 어려움을 느꼈지만 다양한 분석기법과 이커머스 도메인을 학습할 수 있는 좋은 기회였다.
- 비즈니스에 직접적으로 영향을 주는 추천시스템에 대하여 학습, 적용 및 개선해볼 수 있는 좋은 경험이었다.

### 보완점

- 향후에는 다양한 분석 기법과 이커머스 도메인 지식을 활용하여 다양한 분석 인사이트를 도출할 것이며, implicit 라이브러리 뿐만 아니라 다른 추천시스템을 학습 후 적용할 것입니다.





## 참고자료

- [\[링크 1\]](#)
- [\[링크 2\]](#)
- [\[웹 로그 분석을 통한 소비자 구매지연행동 연구 논문\]](#)

| 감사합니다.