

E-Commerce 데이터 분석 및 추천 시스템 구현


잠재고객 확보 및 기존고객 구매 장려

4팀 9기 정호영, 10기 박민경



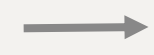
목차



 팀 구성 및
프로그램 수행 절차



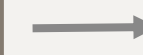
 프로젝트 주제 및 데이터 설명



 EDA 및 데이터 분석



 가설 검증



 Action plan



 추천시스템 및 성능 평가



 결론 및 회고

팀 구성 및 프로그램 수행 절차

👉 팀 구성

훈련생	담당 업무
박민경 (10기)	<ul style="list-style-type: none">▶ 데이터 분석: EDA 및 가설검정, Action plan 관련 조사 진행▶ 추천시스템: surprise 패키지 이용한 모델 & LightFM 구현 시도, word2vec 사용한 content-based 모델 구현
정호영 (9기)	<ul style="list-style-type: none">▶ 데이터 분석: EDA 및 Funnel/Cohort 분석, Action plan 관련 조사 진행▶ 추천시스템: baseline model 및 TF-IDF 사용한 content-based 모델 구현

👉 프로젝트 수행 절차

구분	기간	활동	비고
데이터 전처리	▶ 5/10(화) ~ 5/13(금)	<ul style="list-style-type: none">▶ 도메인 지식 학습▶ 데이터 전처리 및 EDA 진행▶ 가설 설정	▶ 팀별 주간보고 실시
데이터 분석	▶ 5/16(월) ~ 5/20(금)	<ul style="list-style-type: none">▶ 데이터 분석 및 가설 검정▶ Action Plan 도출	▶ 팀별 주간보고 실시
모델링	▶ 5/23(월) ~ 5/31(화)	<ul style="list-style-type: none">▶ 추천 알고리즘 학습▶ Baseline 모델 구현 및 성능 도출▶ 개선 모델 구현 및 성능 도출	▶ 팀별 주간보고 실시
정리	▶ 6/1(수) ~ 6/2(목)	▶ 코드 및 발표 자료 정리	▶ 최종 리포트 제출
총 개발기간	▶ 5/10(화) ~ 6/2(목)	-	-

주제 및 데이터 설명

👉 데이터 크기: 5.27GB (42448764, 9)

👉 목표: 2019년 10월 E-commerce 데이터 분석을 통한 매출 향상 액션 플랜 및 추천 시스템 도출

👉 컬럼 설명

- event_time : 이벤트 발생일자(년-월-일) + 발생시간(시:분:초)
- event_type : 이벤트 발생 타입
 - view / cart / purchase
- product_id : 제품번호
- brand : 브랜드명
- price : 단가
- user_id : 유저 id
- user_session : 세션 id

주제 및 데이터 설명

☞ 카자흐스탄 쇼핑몰에서 구매 상위 브랜드 중 생소한 브랜드를 가장

많이 찾아볼 수 있었음: <https://kaspi.kz/>, www.technodom.kz

- cordiant : 러시아 타이어 회사, 아마존에서는 조회 안됨
- artel : 우즈베크 전자제품 회사
- dauscher : 카자흐스탄 전자제품 회사
- respect : 러시아 신발회사
- elenberg : 러시아 전자제품회사
- stels : 러시아 자전거 회사
- polaris : 러시아 전자제품회사
- nokian : 핀란드 타이어회사, 러시아에 생산시설 있음

☞ 카자흐스탄 전자상거래 시장에서 **의류/전자제품/가전제품**이 가장 규모가 크며 성장률도 높음

“전자상거래가 가장 활발히 이뤄지는 상품 분야는 의류와 전자제품(일반 및 가전)이었다. 이 두 분야는 각각 전체 시장 규모의 24% (897억 텡게, 약 2억3000만 달러), 26% (956억 텡게, 약 2억5000만 달러)를 차지했다”

https://dream.kotra.or.kr/kotranews/cms/news/actionKotraBoardDetail.do?SITE_NO=3&MENU_ID=180&CONTENTS_NO=1&bbsSn=243&pNttSn=183671

2019년 상품 분야별 전자상거래 시장 규모
(단위: 십억 텡게, %)

분야	시장 규모	전년대비 성장률
Apparel and Footwear	89.7(약 2억3000만)	33
Beauty and Personal Care	9.2(약 2400만)	12.2
Consumer Appliances	36.7(약 9600만)	38.4
Consumer Electronics	58.9(약 1억5000만)	40.5
Consumer Health	1.6(약 400만)	13.5
Food and Drink	12.1(약 3200만)	28.8
Home Care	0.8(약 200만)	10.7
Home Improvement and Gardening	18.2(약 4800만)	28.5
Homewares and Home Furnishings	1.8(약 500만)	27.1
Media Products	6.3(약 1600만)	28.6
Personal Accessories and Eyewear	5.7(약 1500만)	12.5
Pet Care	-	-
Traditional Toys and Games	1.1(약 300만)	30.8
Video Games Hardware	-	-
Other	130.8(약 3억4000만)	21.6

주: ()의 단위는 달러 기준

자료: Euromonitor

주제 및 데이터 설명

👉 카자흐스탄
많은

👉 데이터가 카자흐스탄 자국 쇼핑몰 1위인 **Kaspi.kz**의 데이터임을, 그리고 분석 시점은 **2020년 여름**임을 가정하여 분석 및 Action Plan 도출

👉 카
모

“전

이 두 분

테게, 약 2억

대비 성장률
33
12.2
38.4
40.5
13.5
28.8
10.7
28.5
27.1
28.6
12.5
-
30.8
-
21.6

EDA 및 데이터 분석

👊 EDA를 위한 Feature engineering

- 👉 `category_code` 및 `brand` 결측치: “unknown” 으로 대체
- 👉 가격(`price`)이 0 이하인 데이터는 제거
- 👉 `user_session` 결측치는 2개로, 결측치의 앞뒤 `session id`를 토대로 유추해 채워넣음
- 👉 `event_time`에서 날짜만 추출해 `event_date` 컬럼과 월/일/요일/시간/주차에 대한 정보를 추가로 추출해 각각 `event_month`, `event_day`, `day_of_week`, `event_hour`, `event_week` 라는 컬럼 추가 생성해 feature engineering 진행

EDA 및 데이터 분석

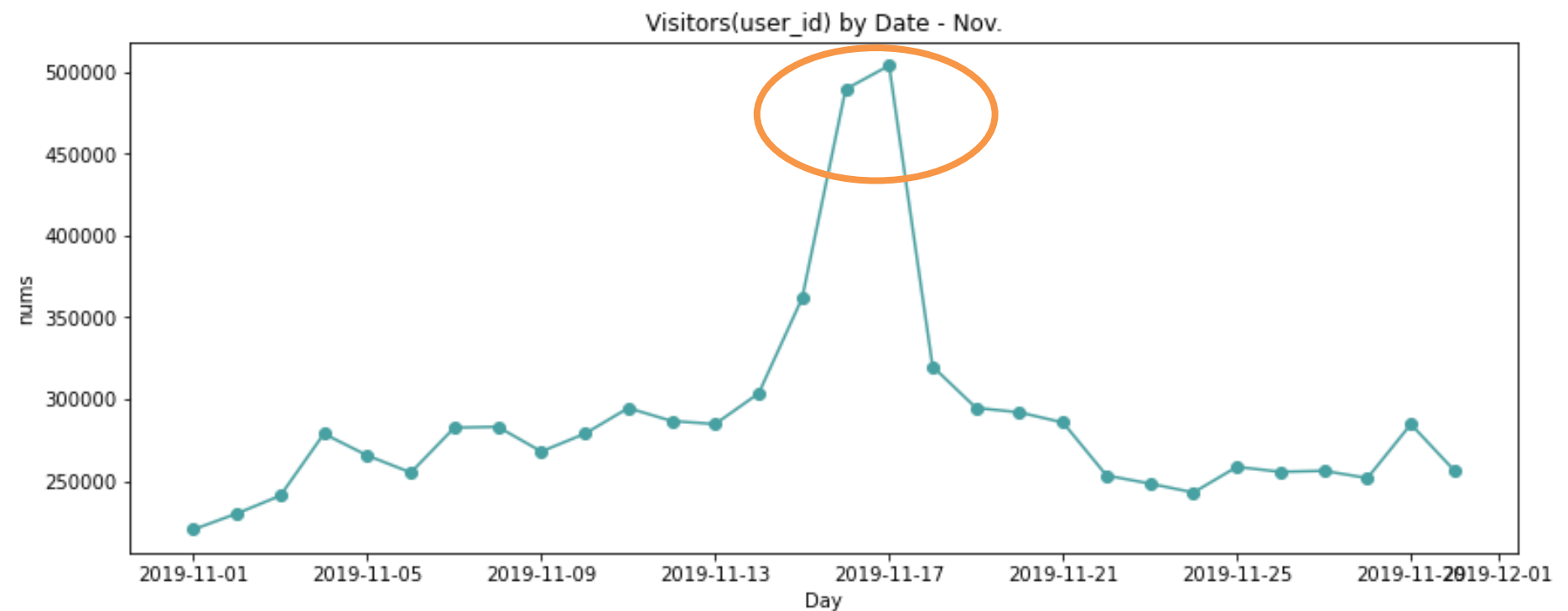
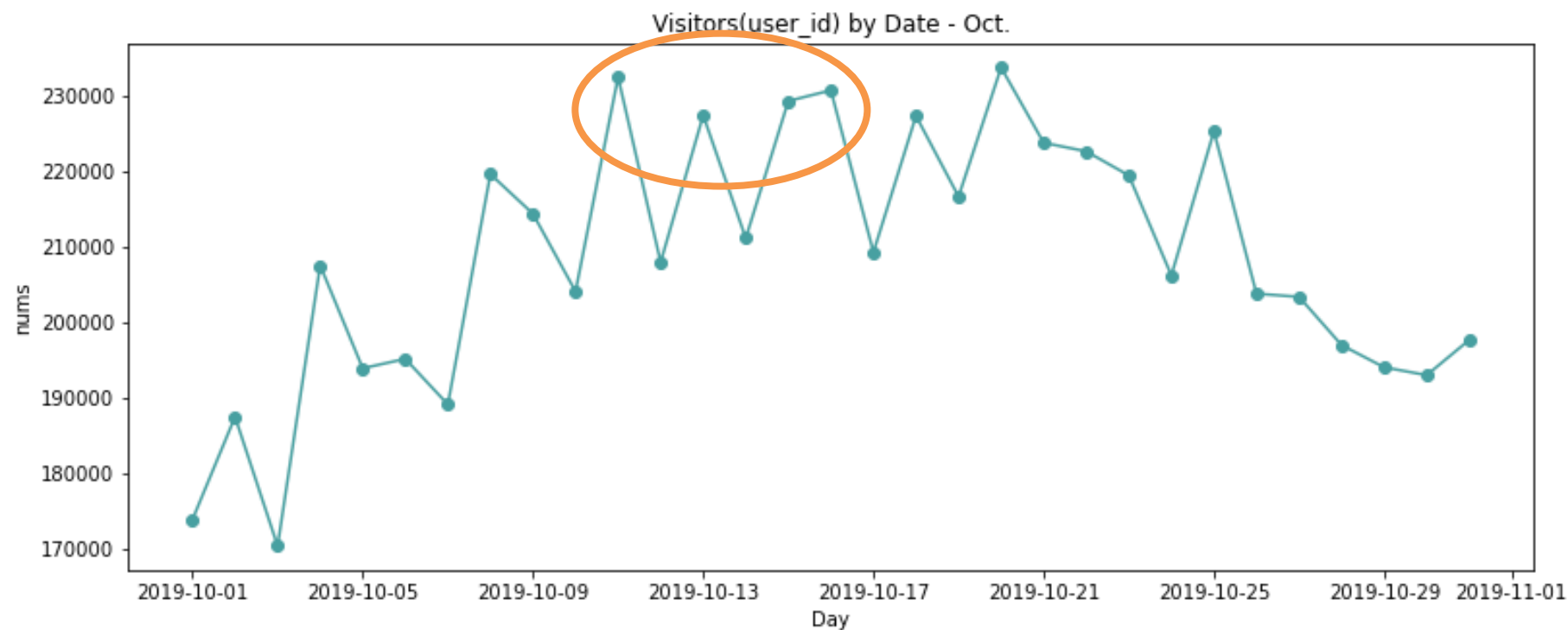
👊 EDA

👉 10~11월 일별 방문자 수

✅ 10월: 11~20일 동안의 방문자 수가 많음

✅ 11월: 15~17일 동안의 방문자 수가 가장 많음

▶ Kaspi Zhuma(현지판 블랙프라이데이 세일, Zhuma = Friday): 15~17일



EDA 및 데이터 분석

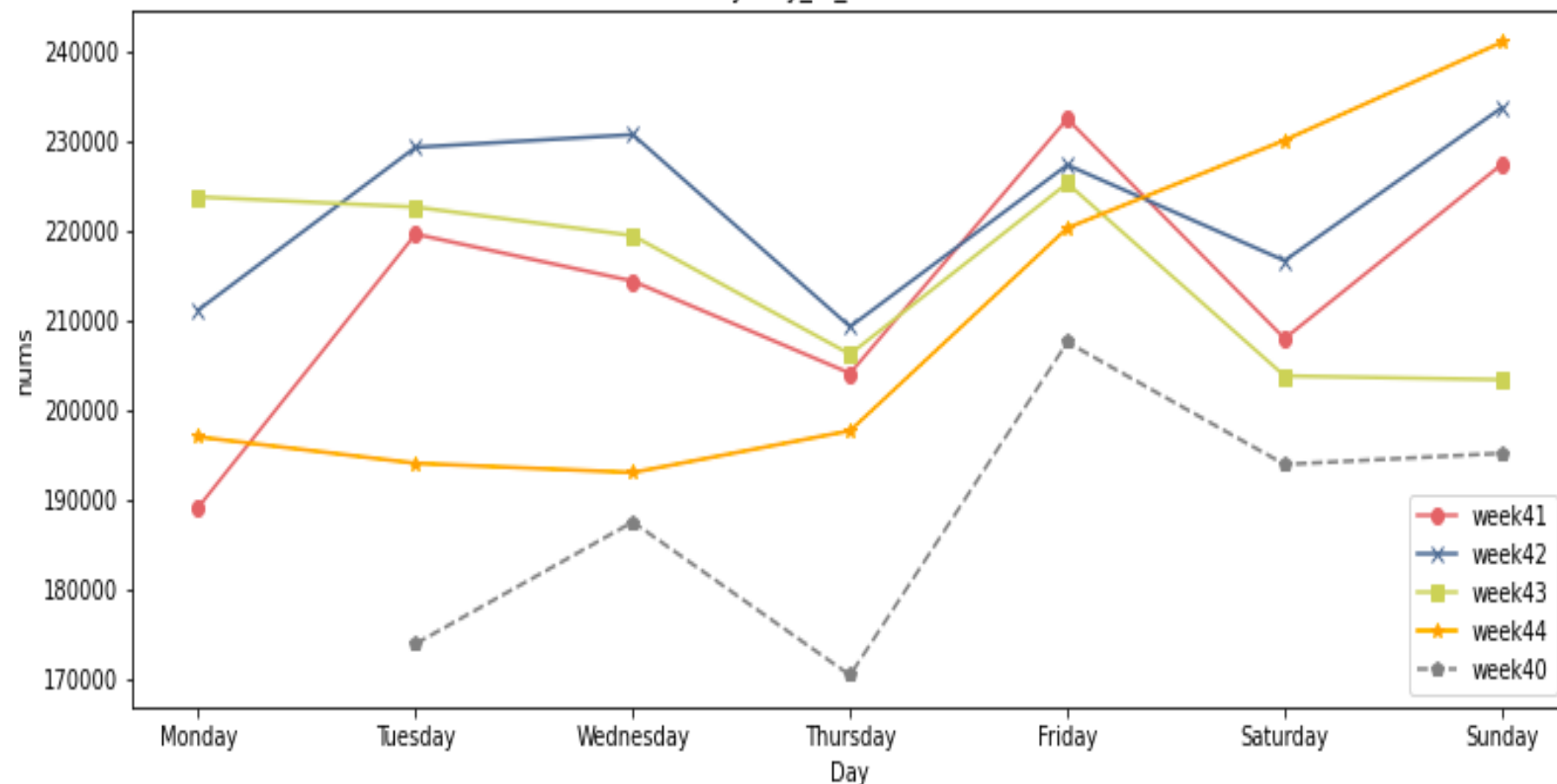
👊 EDA

👉 10월(40~44주차)의 요일별 방문자 및 구매 수

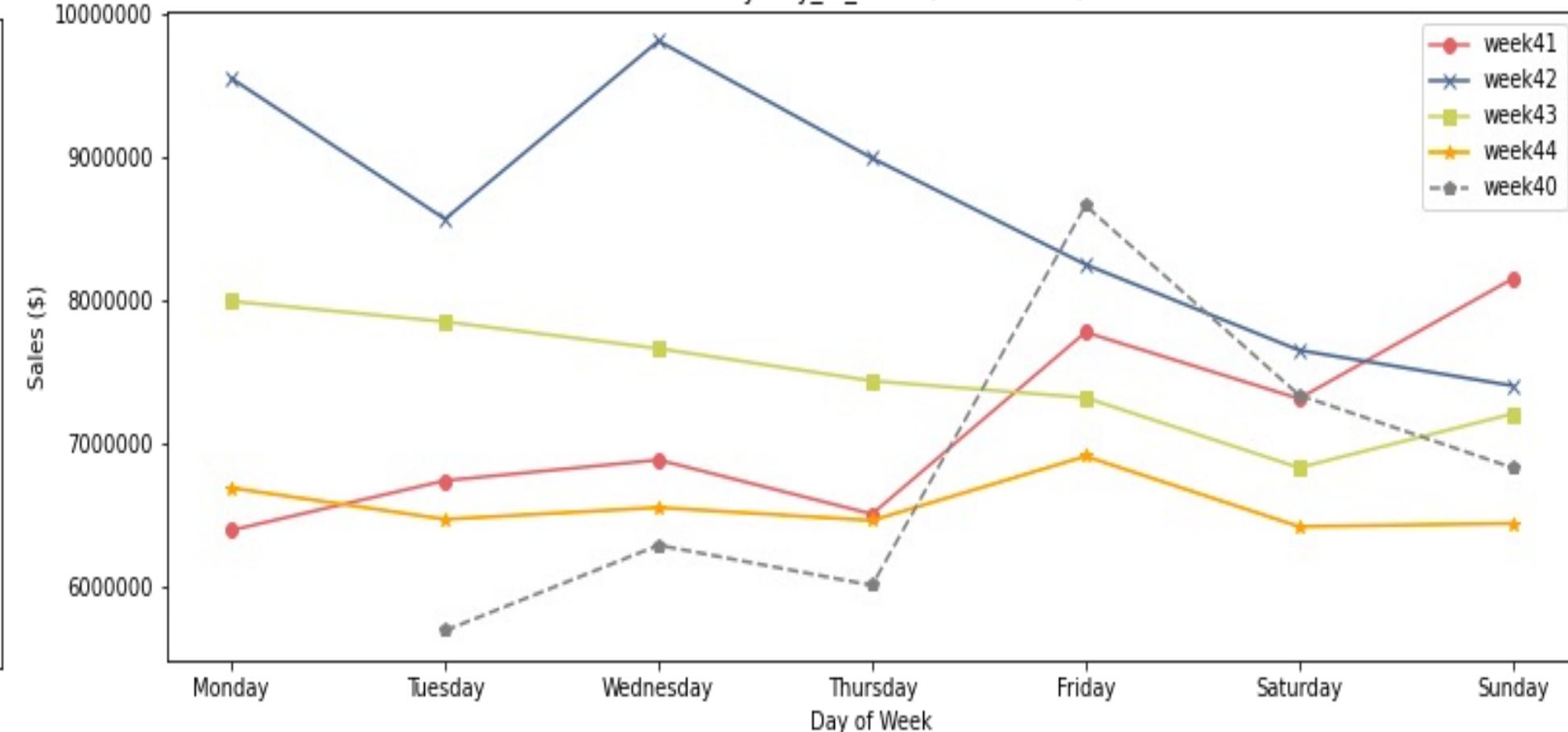
✅ 방문자 수: 대체로 금/토/일요일에 방문자 수가 많고, 41~43주차에서 방문자 수 증가

✅ 구매 수: 대체로 금/토/일요일에 구매 수가 증가하며, 대체로 42주차의 구매 수가 가장 많음

Visitors by Day_of_Week (week40~44)



Sales by Day_of_Week (week40~44)



EDA 및 데이터 분석

👊 EDA

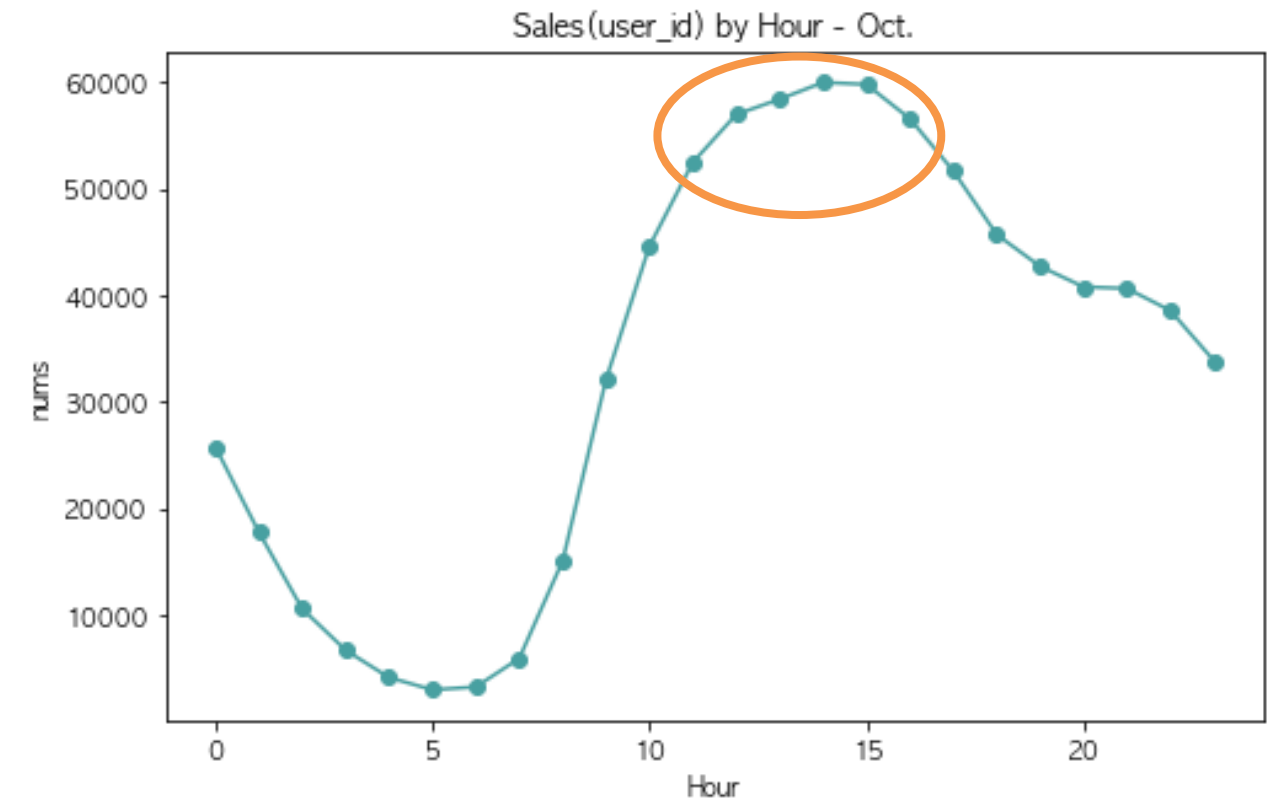
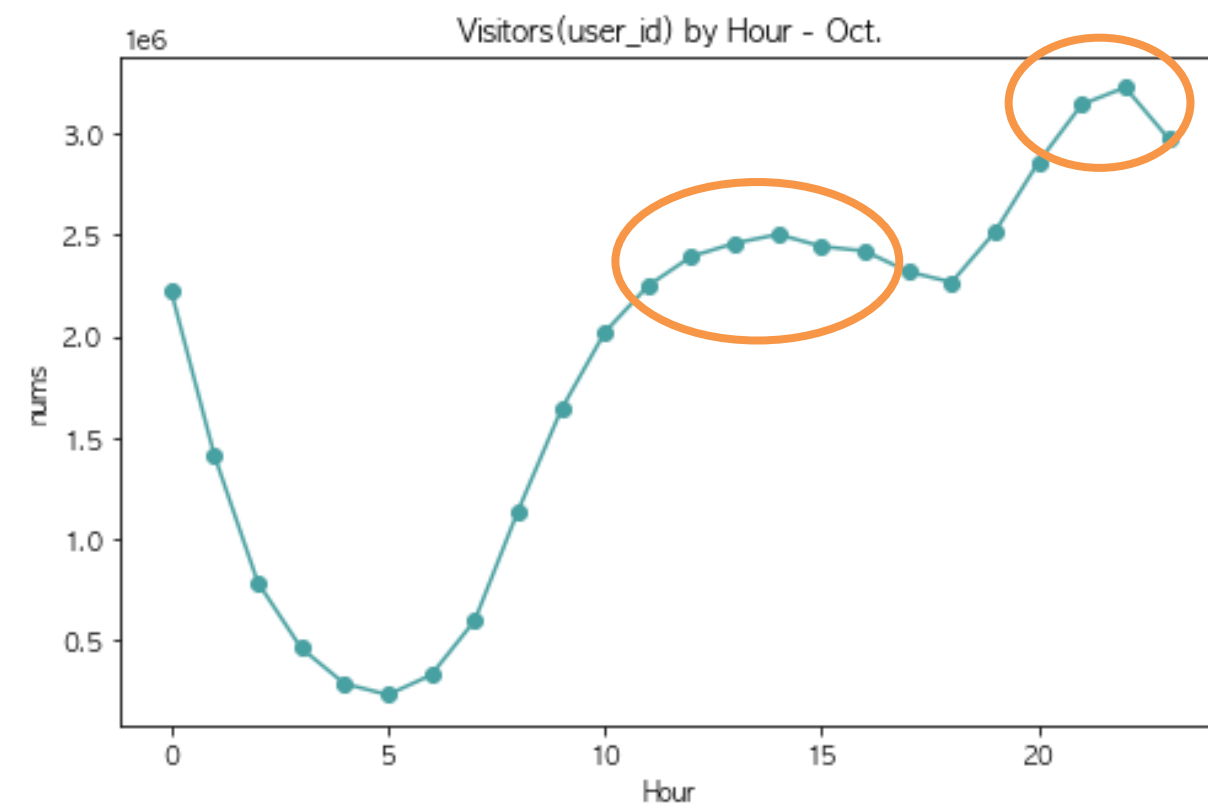
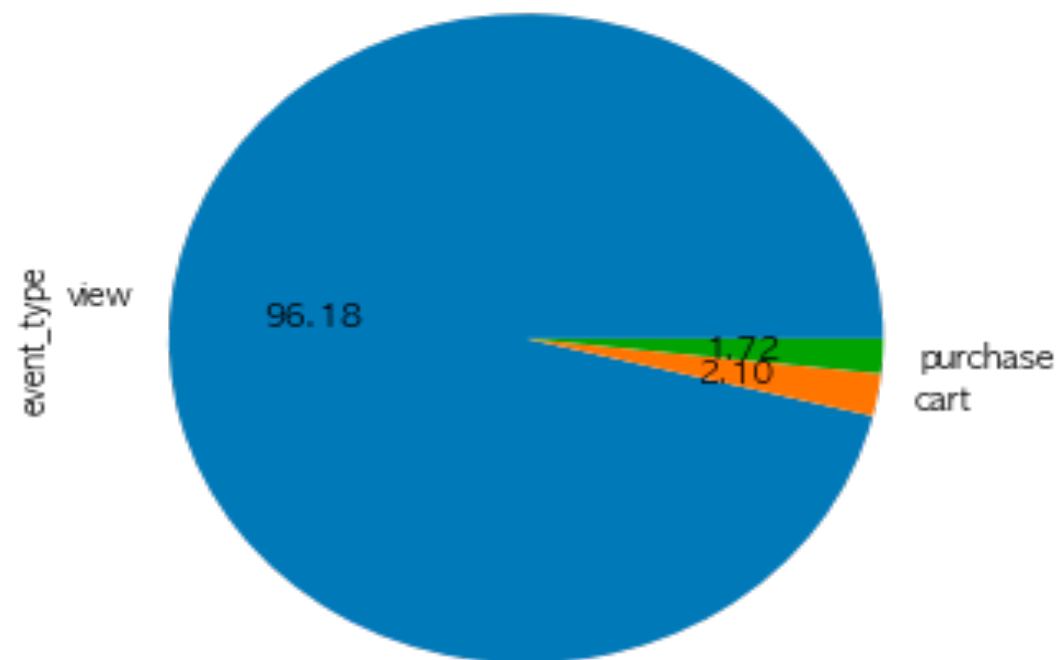
👉 **event_type** 비중: view가 가장 많고, cart, purchase 순으로 분포 (좌측 그래프)

👉 시간별 방문자/구매 수 분포도

✅ 11~14시, 19~23시에 가장 많은 방문자가 방문 (중앙 그래프)

✅ 13~16시: 구매가 가장 활발 (우측 그래프)

유저 행동별 pie chart

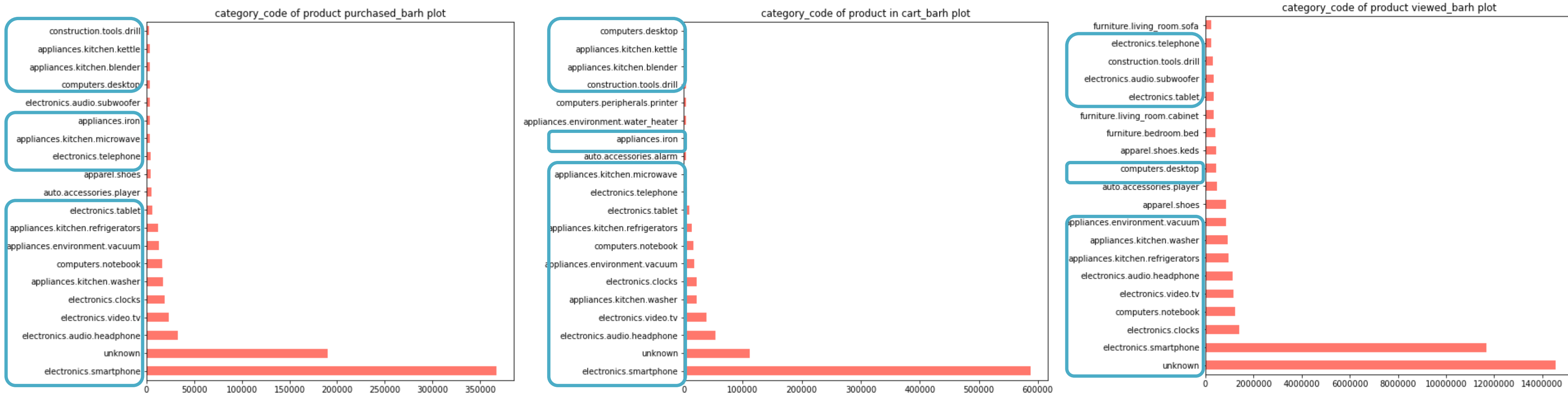


EDA 및 데이터 분석

👊 EDA

👉 event_type에 따른 카테고리 top 20

✅ 세 가지 카테고리 top 20에서 전자/가전제품이 가장 많이 포진



EDA 및 데이터 분석

👊 EDA

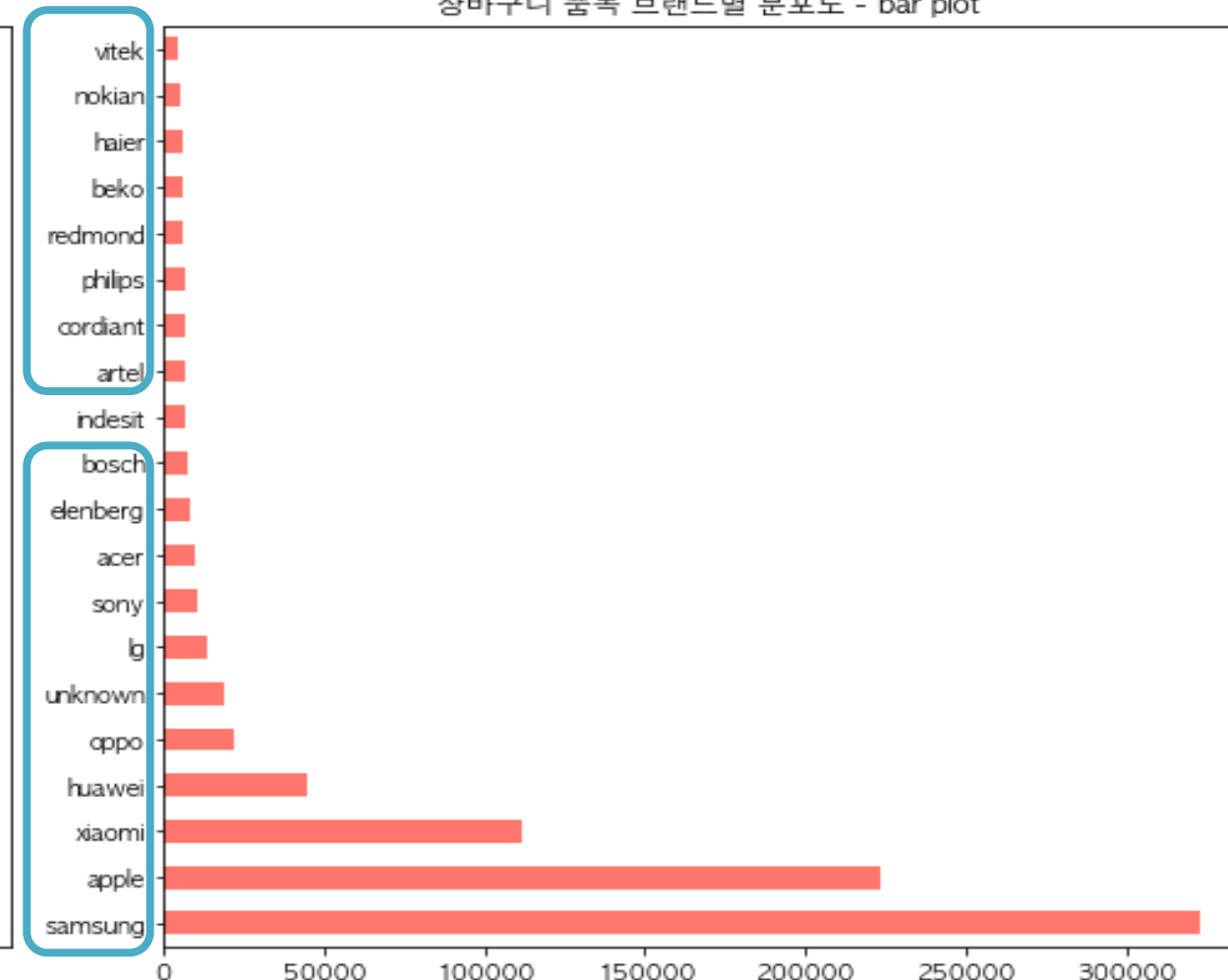
👉 event_type에 따른 브랜드 top 20

✅ 카테고리과 마찬가지로 = 세 가지 brand top 20에서 대부분 전자/가전제품 브랜드

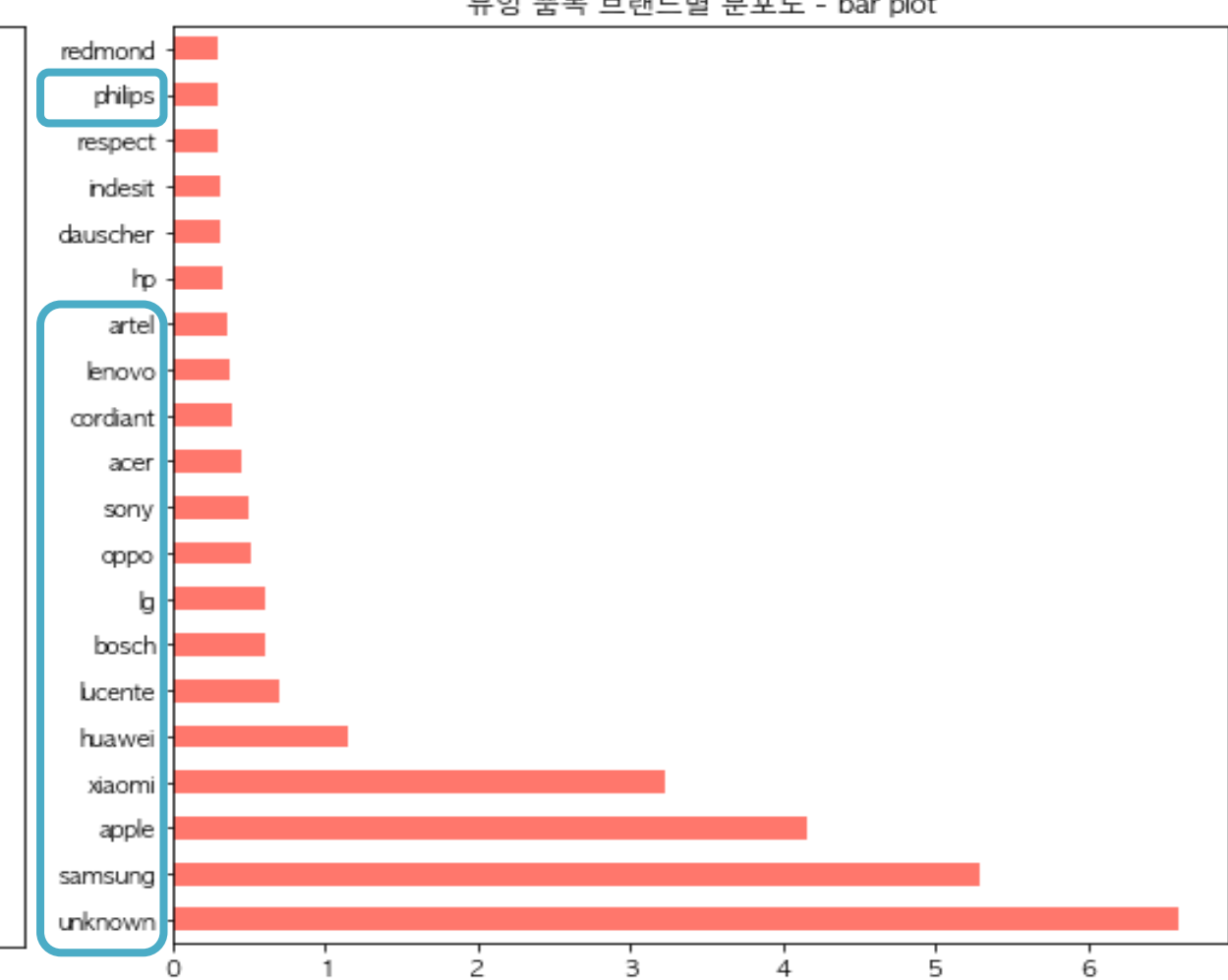
구매 품목 브랜드별 분포도 - bar plot



장바구니 품목 브랜드별 분포도 - bar plot



뷰잉 품목 브랜드별 분포도 - bar plot



EDA 및 데이터 분석

👊 데이터 분석

👉 구매 전환율 top 20 상품:

✅ 구매 전환율이 가장 높은 상위 20개 상품의 카테고리
브랜드

▶ 20개 상품 중 약 18개의 상품이 전자/가전 제품

⇔ 전자/가전제품의 구매가 활발히 발생해 구매 전환율이 아
주 높은 것을 알 수 있음

category_code	brand
apparel.shoes	marcomen
appliances.environment.vacuum	samsung
appliances.environment.air_heater	unknown
electronics.smartphone	apple
electronics.smartphone	xiaomi
appliances.personal.scales	polaris
electronics.smartphone	apple
electronics.smartphone	xiaomi
electronics.smartphone	huawei
computers.desktop	pulser
electronics.smartphone	apple
electronics.smartphone	samsung
electronics.video.tv	artel
computers.peripherals.printer	epson
computers.notebook	hp
appliances.kitchen.oven	dauscher
apparel.shoes.keds	strobbs
electronics.smartphone	samsung
computers.peripherals.mouse	msi
computers.notebook	asus

EDA 및 데이터 분석

👊 데이터 분석

👉 상품을 보기만 하고 구매 이력이 없는 고객:

✅ 해당 고객이 관심을 갖는 카테고리 top 20에는 마찬가지로 **전자/가전제품**이 주를 이룸

+ 그 외에는 **신발, 가구** 등이 포함

✅ 브랜드 top20에도 **전자/가전제품**의 브랜드가 다수 포진

electronics.smartphone	6447101	samsung	2925094
electronics.clocks	837634	apple	2286470
computers.notebook	728769	xiaomi	1746091
electronics.video.tv	649765	huawei	644130
electronics.audio.headphone	588577	lucente	444628
appliances.kitchen.refrigerators	583866	bosch	358987
apparel.shoes	559709	lg	344303
appliances.kitchen.washer	521075	oppo	289170
appliances.environment.vacuum	483926	sony	282464
auto.accessories.player	314503	acer	266790
apparel.shoes.keds	288906	cordiant	223609
computers.desktop	285158	lenovo	216266
furniture.bedroom.bed	275424	artel	206512
furniture.living_room.cabinet	226800	respect	195869
electronics.audio.subwoofer	195443	hp	195259
electronics.tablet	192986	dauscher	185115
furniture.living_room.sofa	183841	casio	181060
construction.tools.drill	164436	indesit	174397
kids.carriage	145662	stels	167189
electronics.telephone	142586	philips	164819

EDA 및 데이터 분석

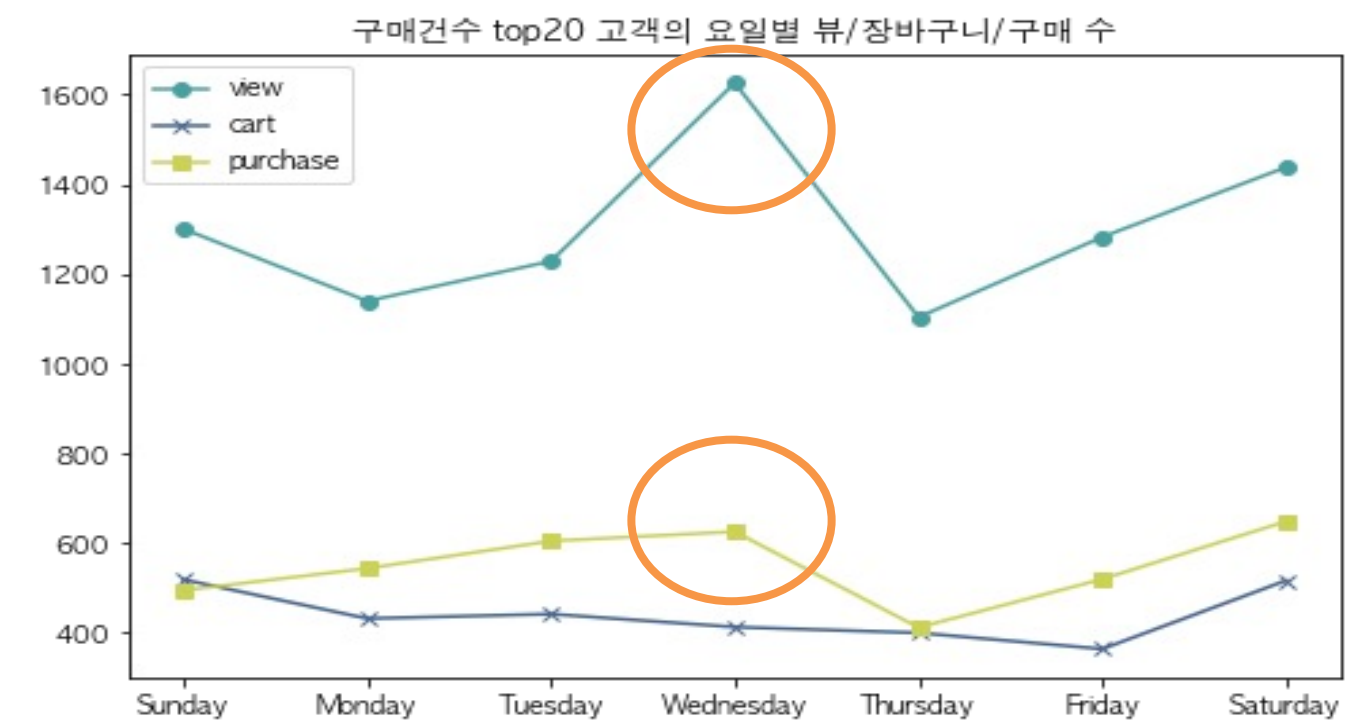
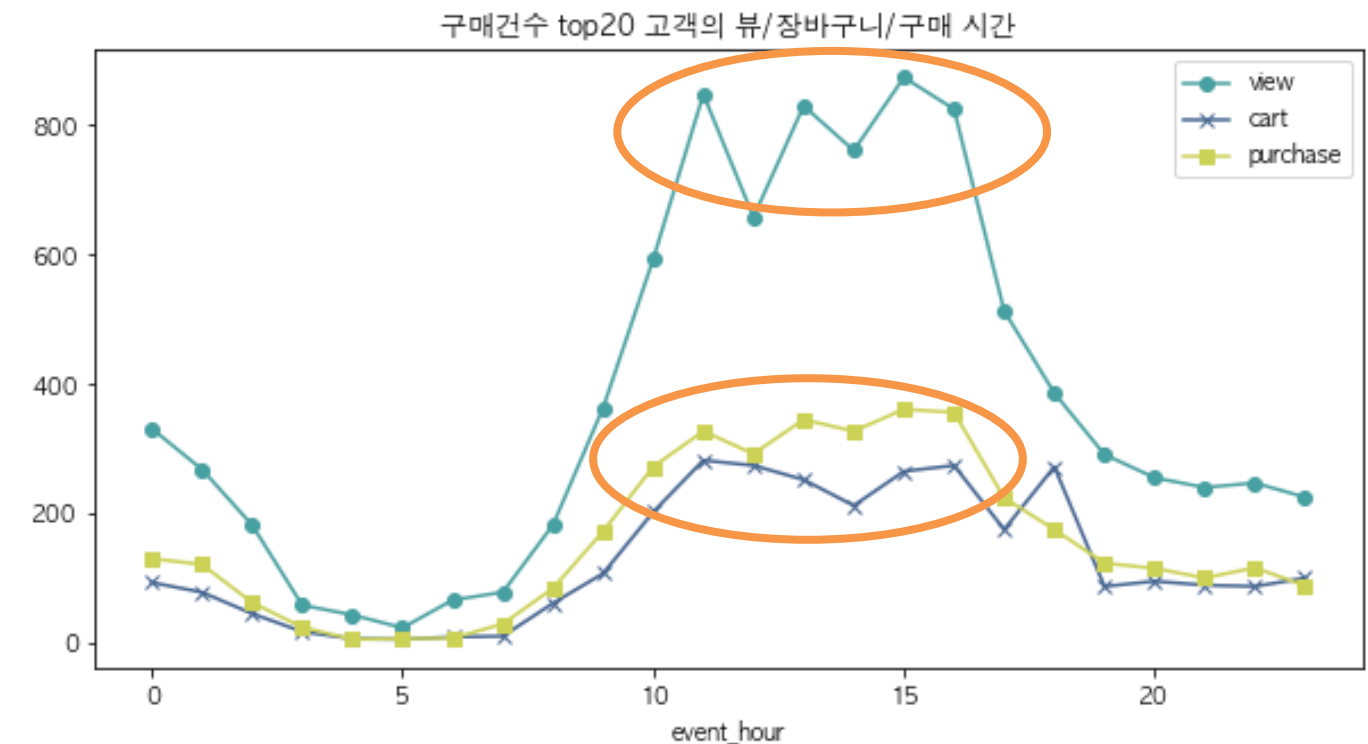
👊 데이터 분석

👉 구매 건수 top 20의 충성 고객:

✅ 시간별: 방문과 구매가 가장 활발한 시간대가 거의 일치

✅ 요일별: 방문이 가장 활발한 요일과 구매가 가장 많이 발생하는 요일이 같으며, 요일별 방문 및 구매의 변화 추이가 유사

▶ “수요일”에 가장 많은 방문과 구매가 발생



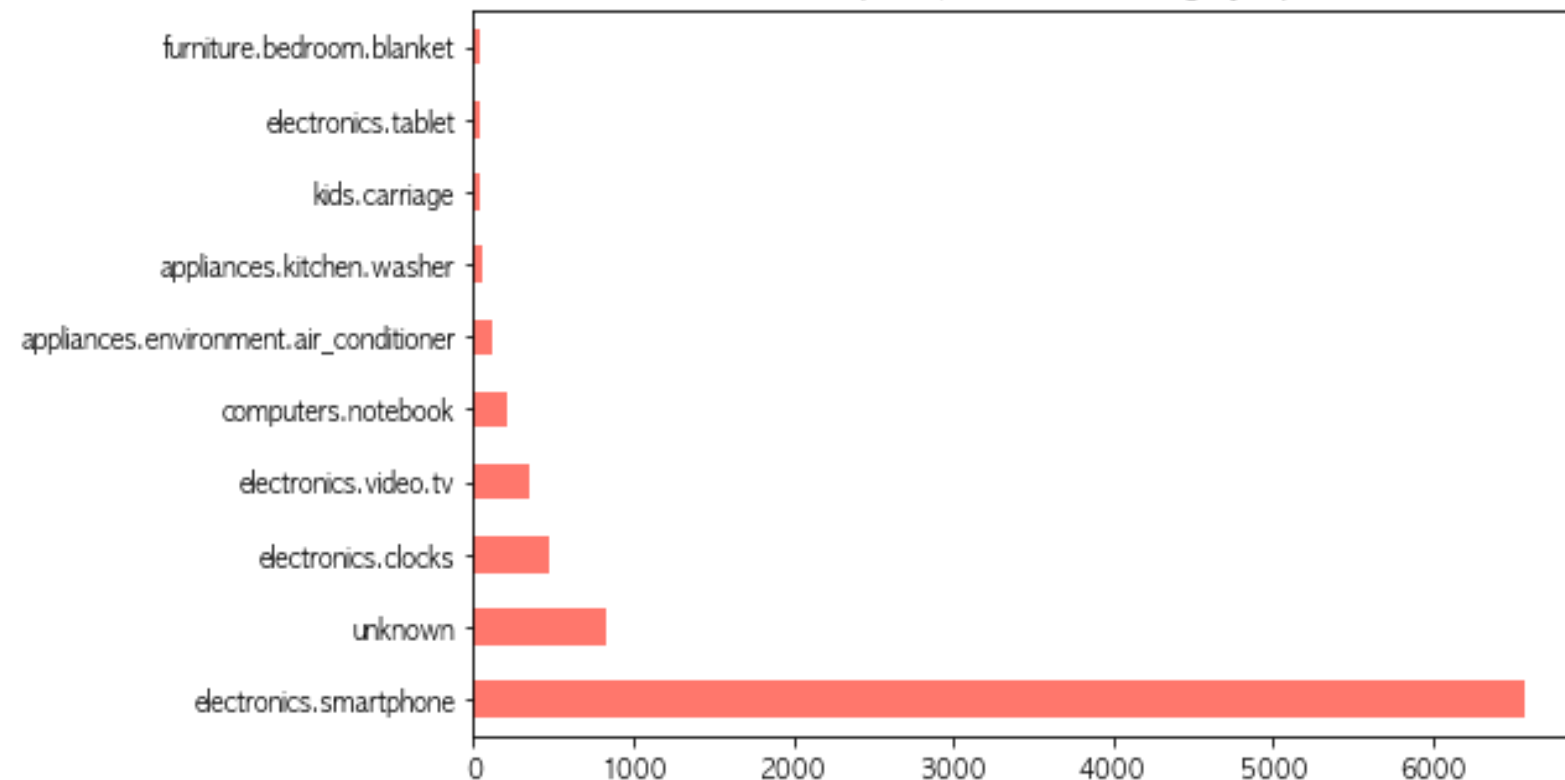
EDA 및 데이터 분석

👊 데이터 분석

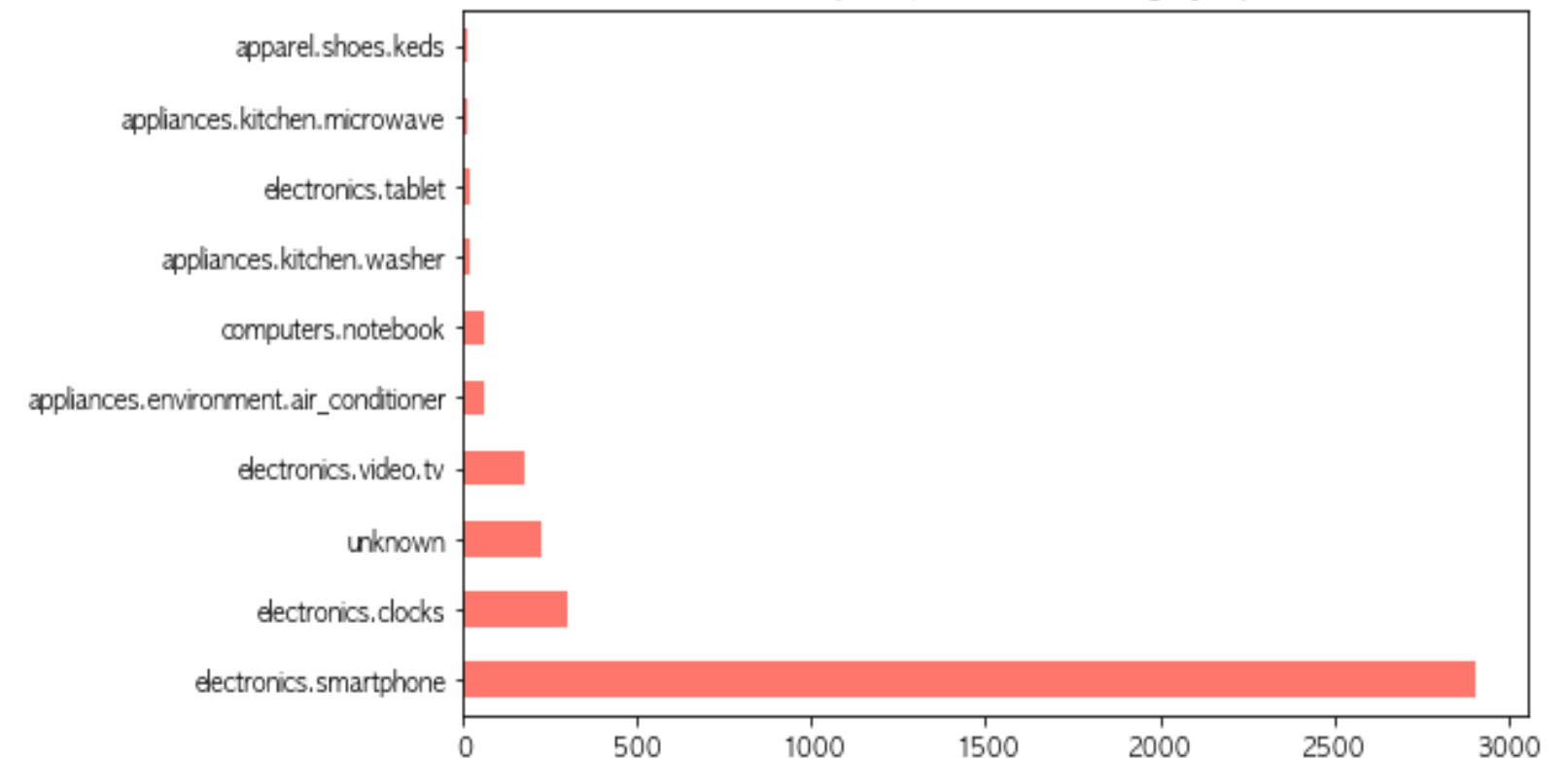
👉 구매 건수 top 20의 충성 고객:

✅ 구매 상위 20의 충성 고객이 가장 많이 보거나 구매한 상품 카테고리: 동일하게 전자/가전제품의 비중이 가장 큼

구매건수 top20의 뷰잉한 상품 category top 10



구매건수 top20의 구매한 상품 category top 10



EDA 및 데이터 분석

👊 데이터 분석

👉 퍼널 분석:

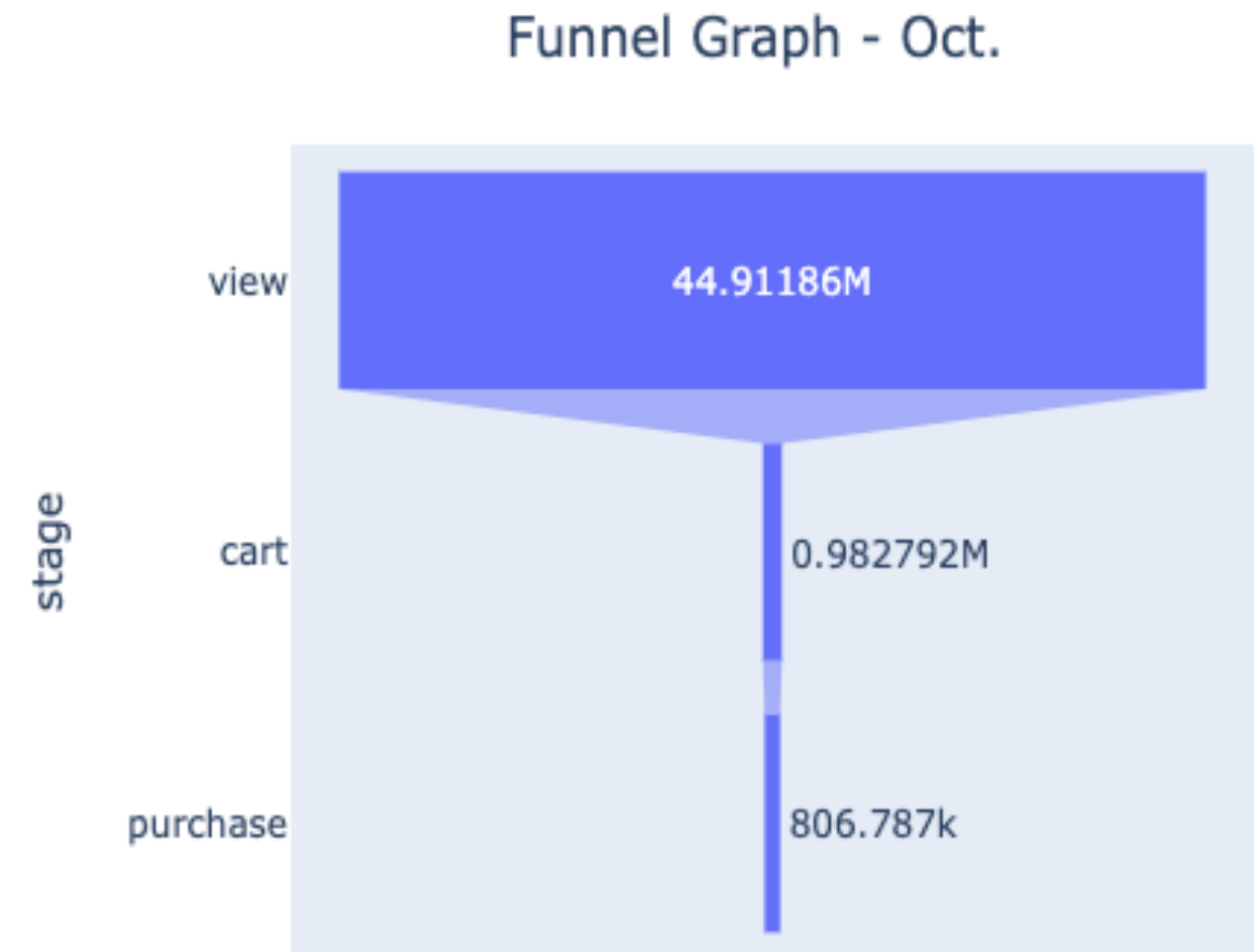
▶ 10월 한 달 동안 방문에서 장바구니로 이어지는 전환율 매우 ↓↓

↔ 장바구니에서 구매로의 전환율은 비교적 ↑↑

▶ view → cart 전환율 : 2.19 %

cart → purchase 전환율 : 82.09 %

최종 전환율 : 1.8 %



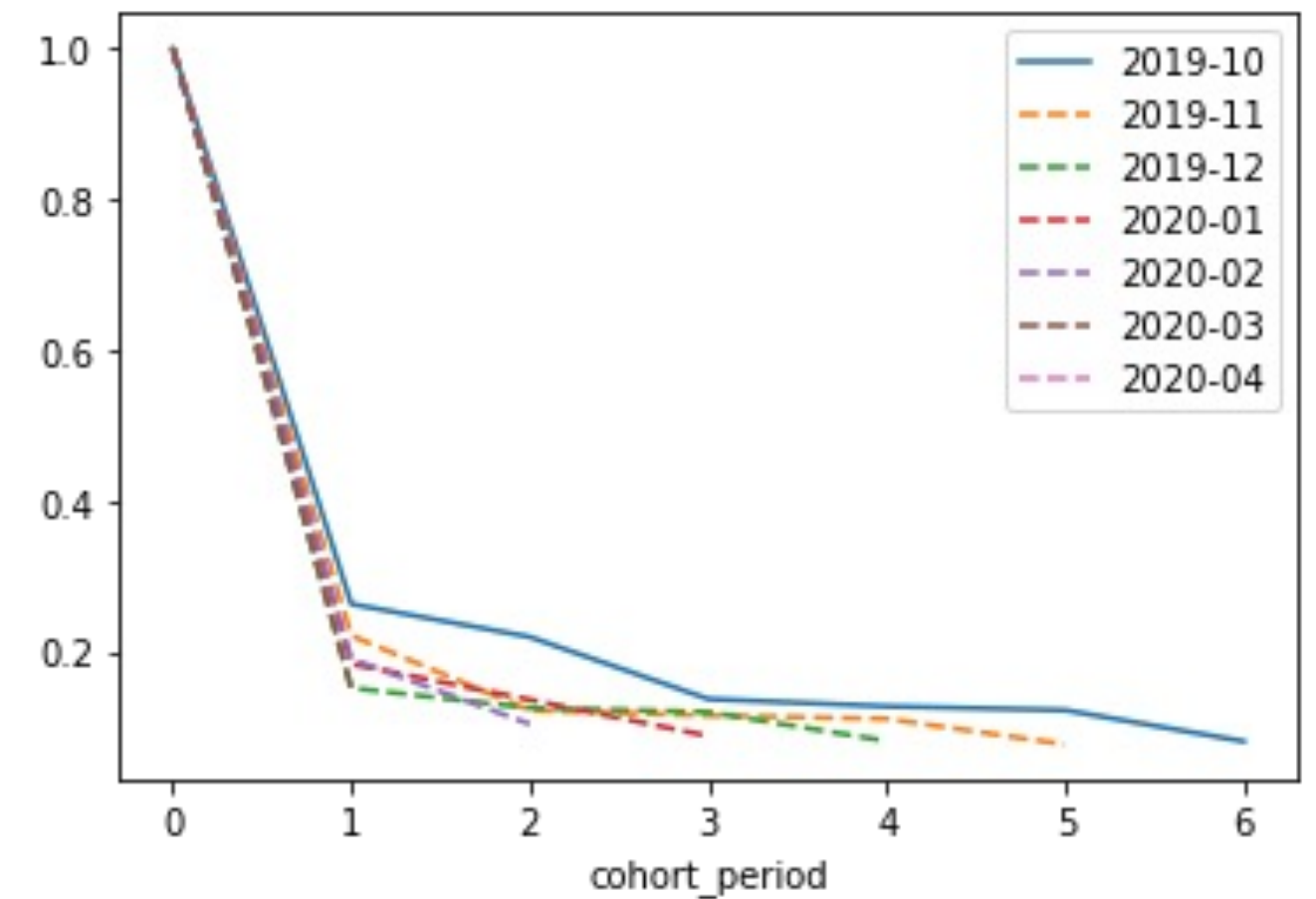
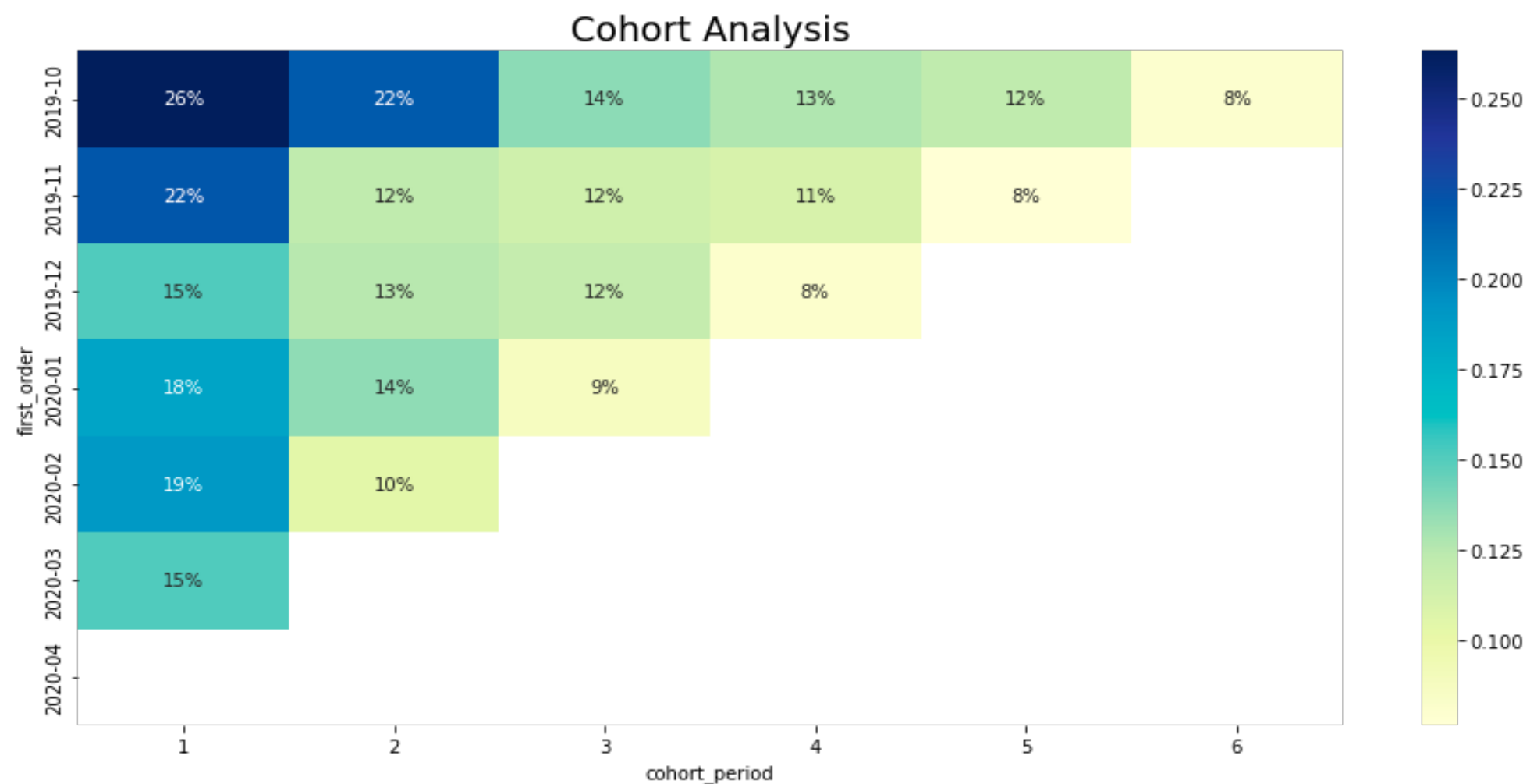
EDA 및 데이터 분석

👊 데이터 분석

👉 코호트 분석: 2019.10~2020.04까지의 재구매율

▶ 지속적으로 감소하는 경향을 보임

+ (우측 그래프를 통해) 재구매율이 반등되는 지점이 없음을 파악 ⇒ 재구매 주기 확인 어려움

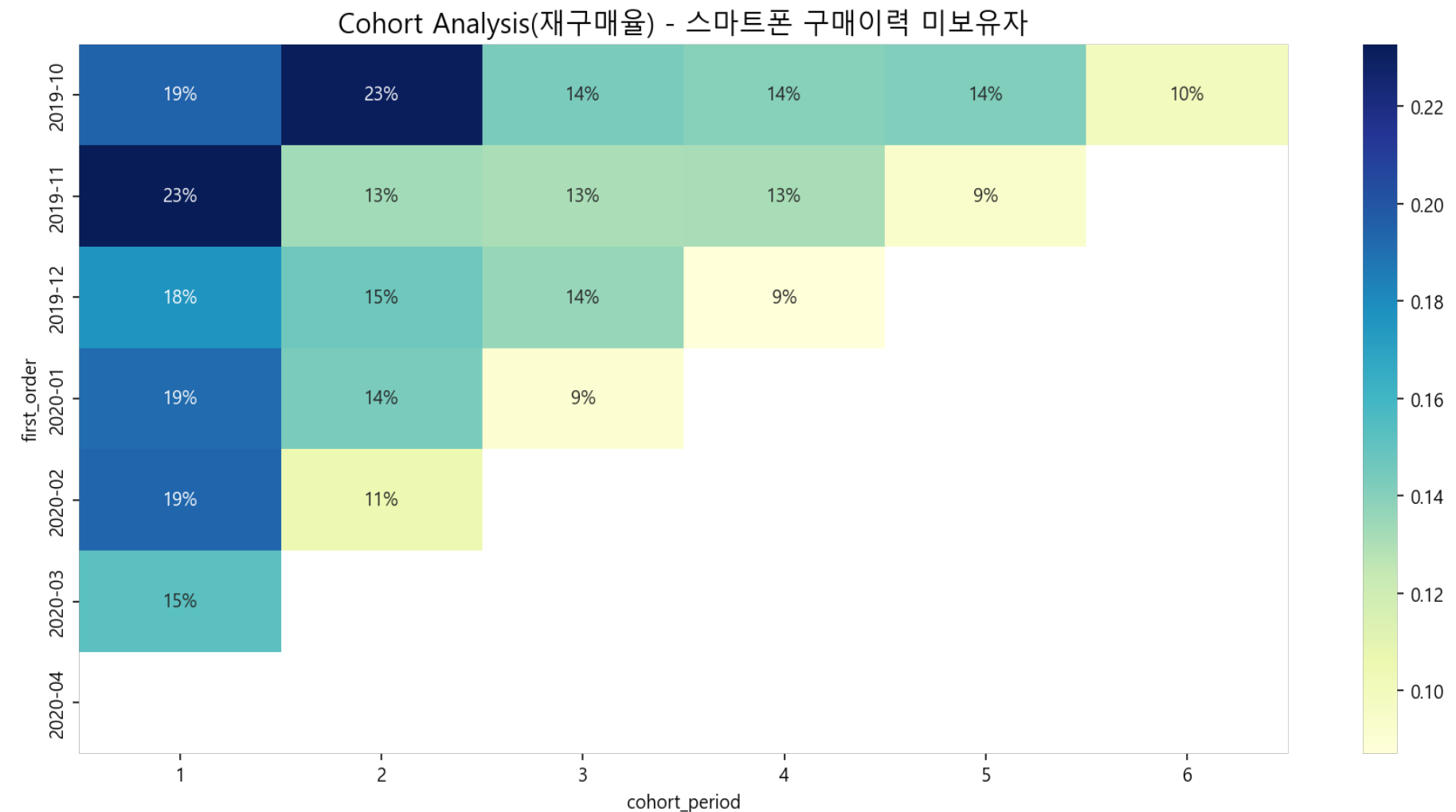
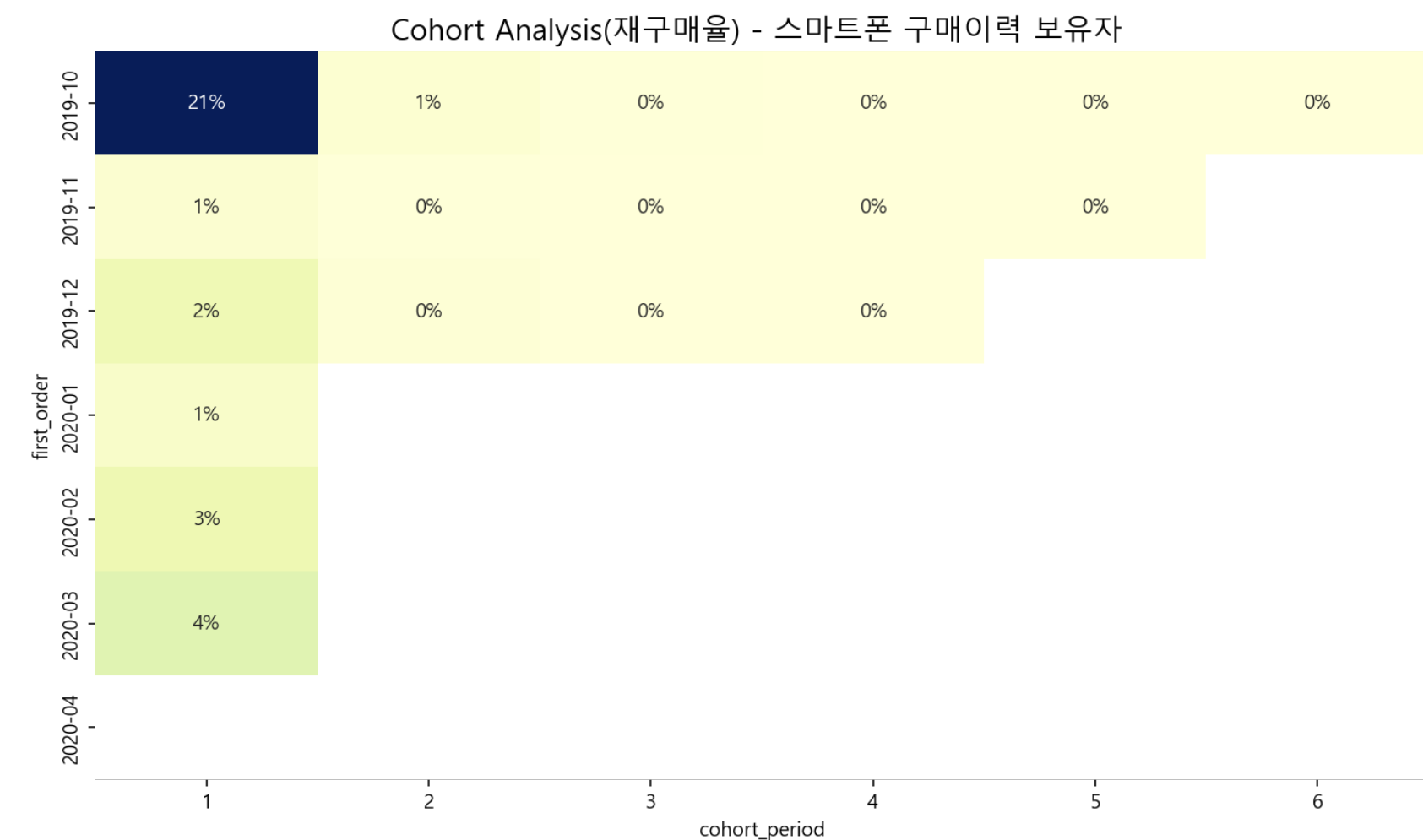


EDA 및 데이터 분석

👊 데이터 분석

👉 코호트 분석: 스마트폰 구매자 vs 그외 상품 구매자 재구매율 비교

- ▶ 스마트폰 구매자(좌측 그래프): 2019.10~2020.04 동안 재구매율이 “0%”에 수렴
- ▶ 그외 상품 구매자(우측 그래프): 재구매율 감소하고 있지만 상대적으로 재구매가 더 활발



EDA 및 데이터 분석

👊 데이터 분석

👉 기존 고객 vs 신규 고객 간 구매액 분석

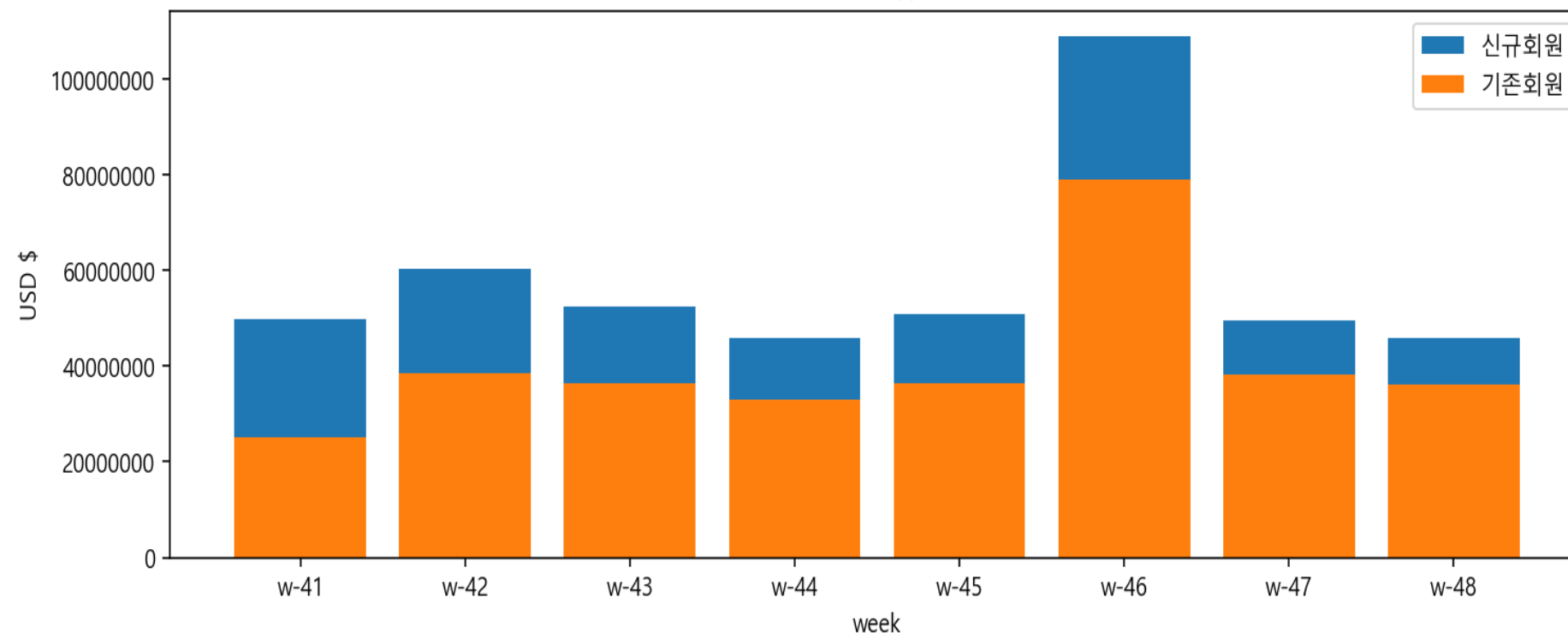
** 신규고객: 해당 주차에 처음 방문하여 구매한 고객

▶ 대체로 **기존 고객 > 신규 고객**에 대한 구매액 (좌측 그래프)

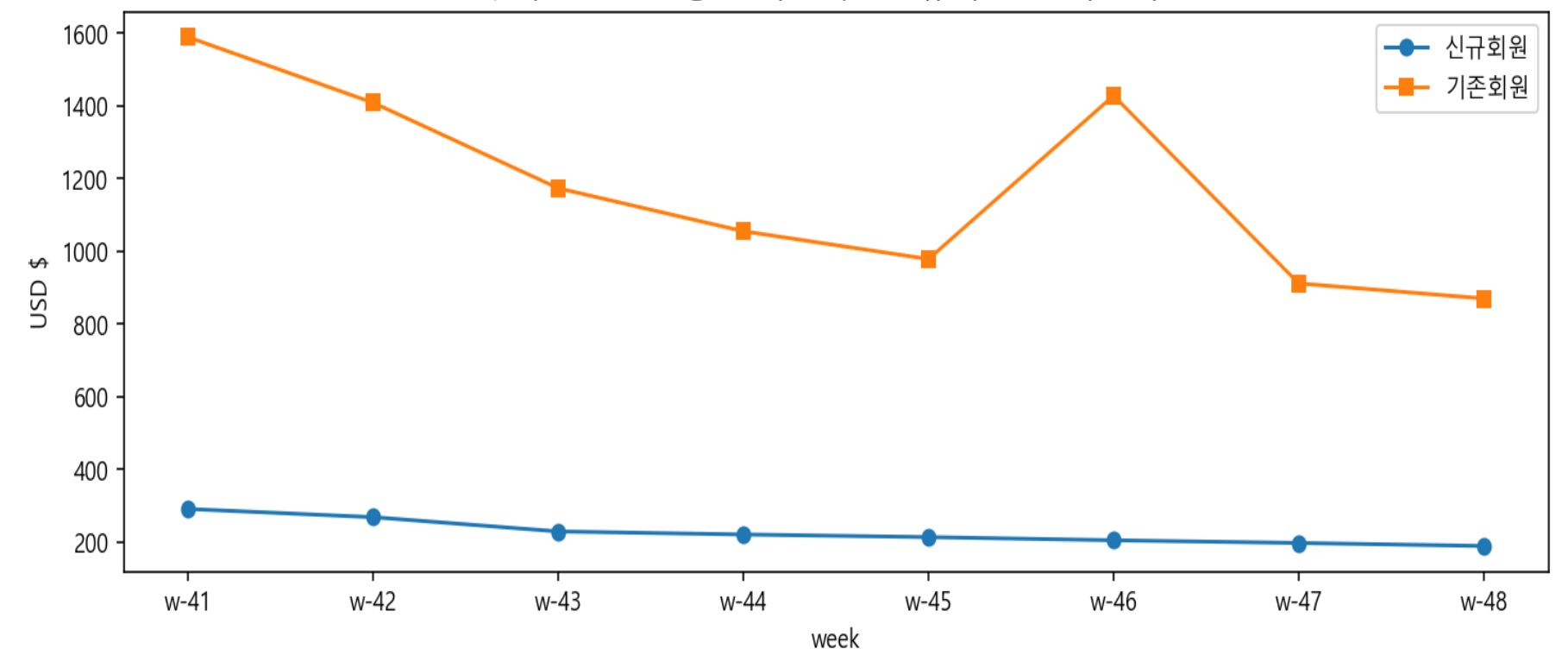
+ **46주차**의 경우 **3배 이상** 차이 남

▶ 1인당 구매액에 대해 기존 고객이 신규 고객에 비해 최소 3배 이상임을 확인 (우측 그래프)

주차별 총매출액 - 신규회원 vs 기존회원



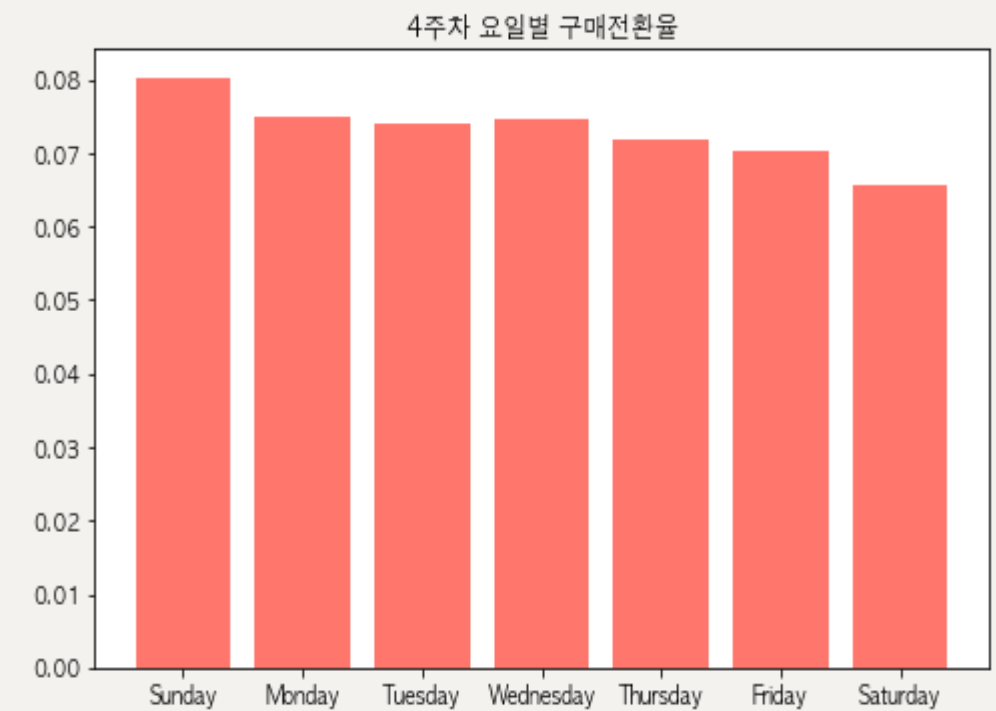
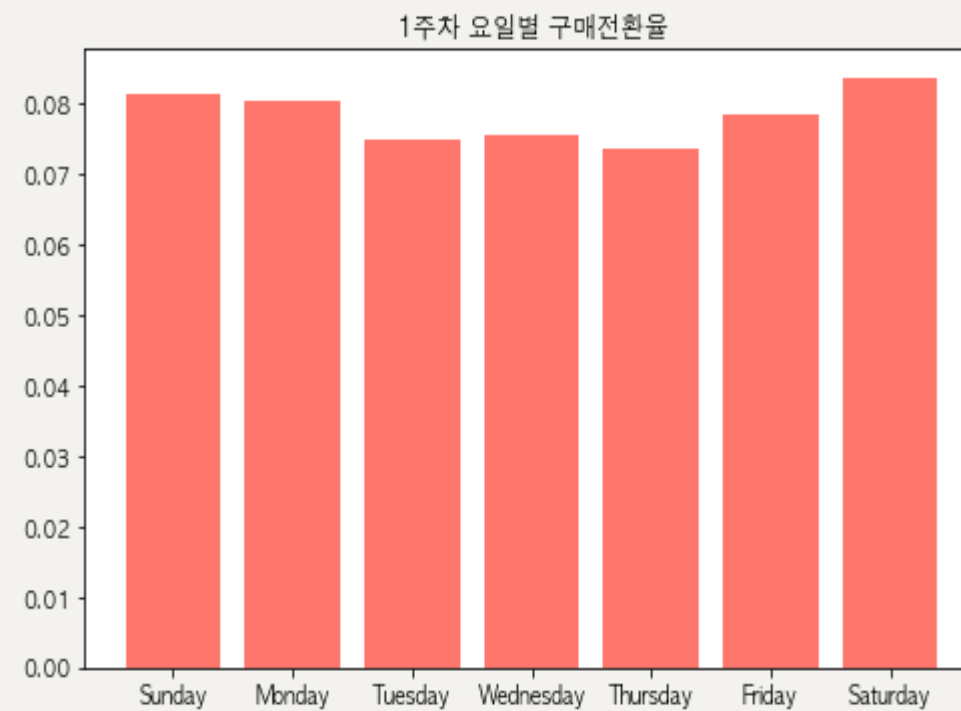
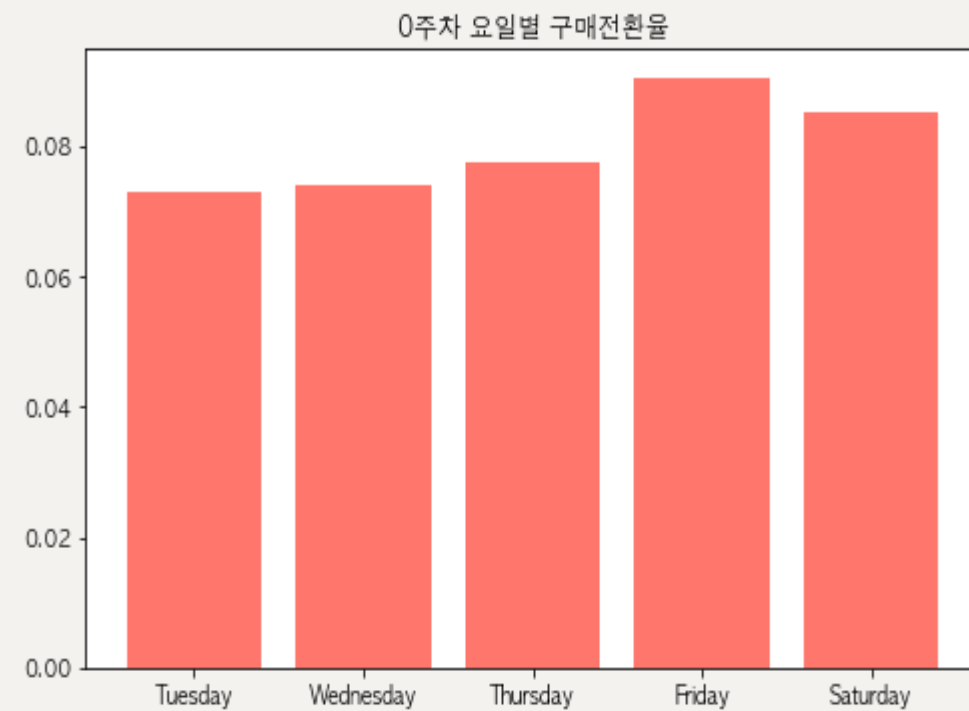
주차별 1인당 평균 매출액 - 신규회원 vs 기존회원



가설 검증

1 주말 구매전환율은 **주중** 구매 전환율보다 높을 것이다.

- ▶ 10월 0주차(=9월 마지막주), 1주차: 주말인 **토/일요일의 구매 전환율이 높은 편**
- ▶ 10월 4주차: 주말 중 일요일의 전환율만 높음



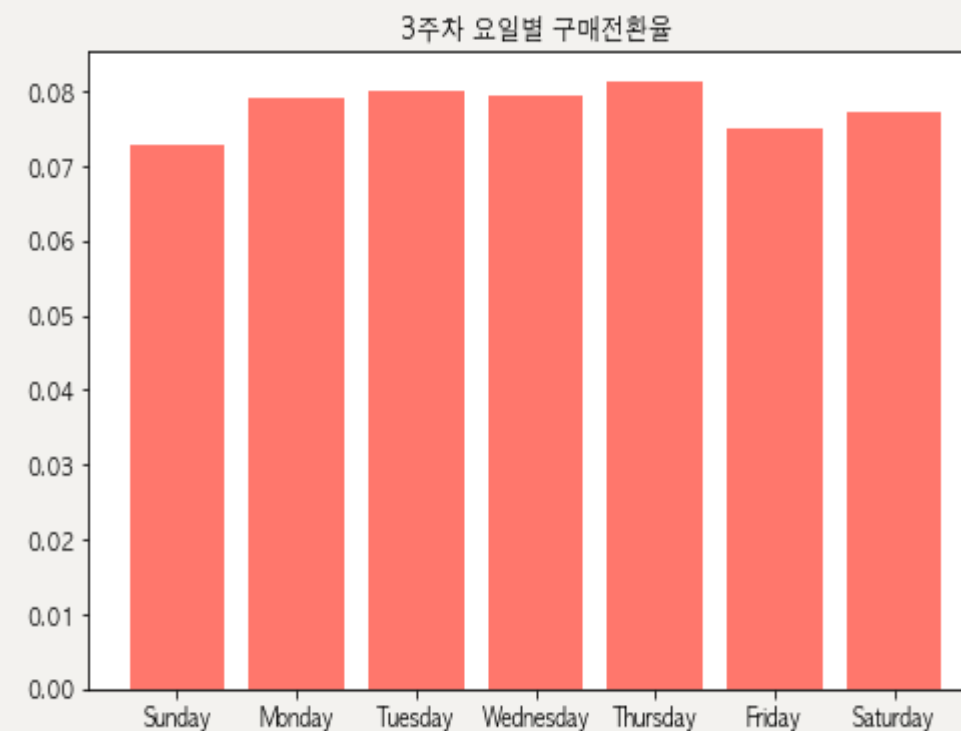
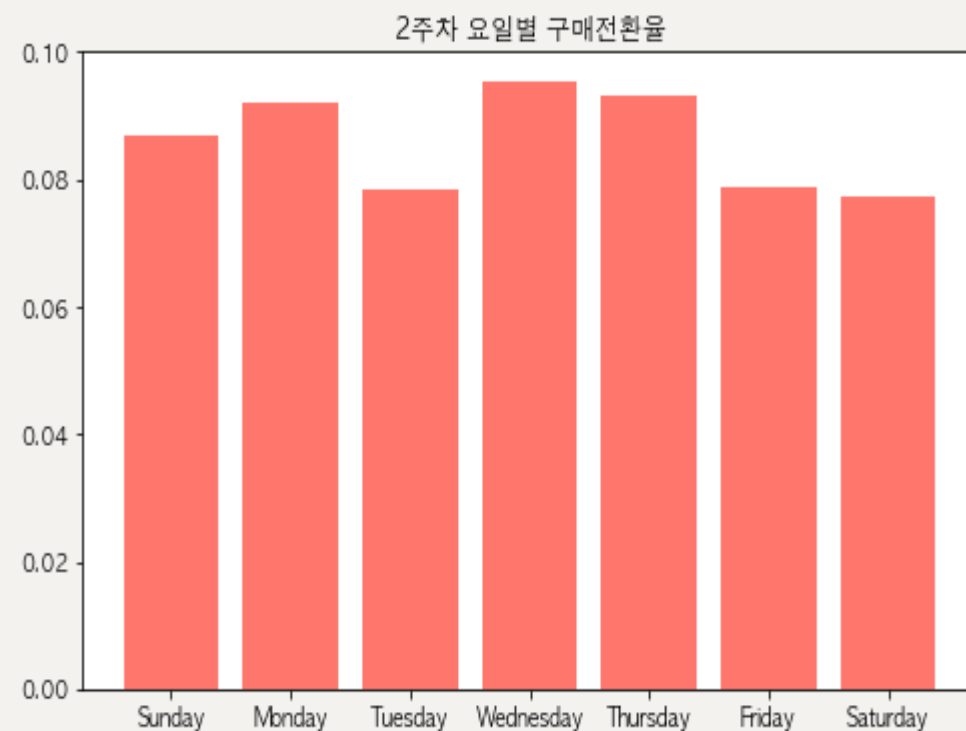
가설 검정

1 주말 구매전환율은 주중 구매 전환율보다 높을 것이다.

▶ 10월 2, 3주차: 토/일요일의 구매 전환율 낮음

▶ 주말/주중 평균 구매 전환율이 같은지 t-test 검정: p-value=0.5425116558539653으로 유의수준 5%에서 두 평균 전환율이 같다고 할 수 있음

👉 “가설 1 기각할 수 있음”



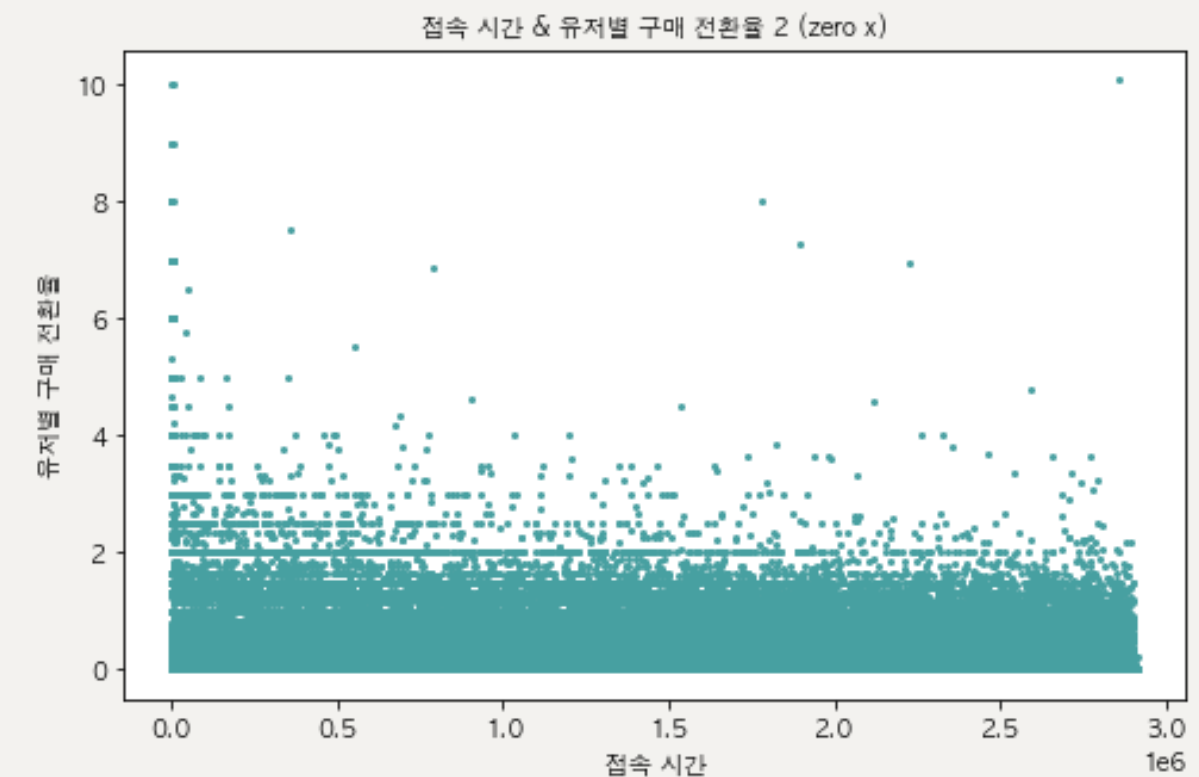
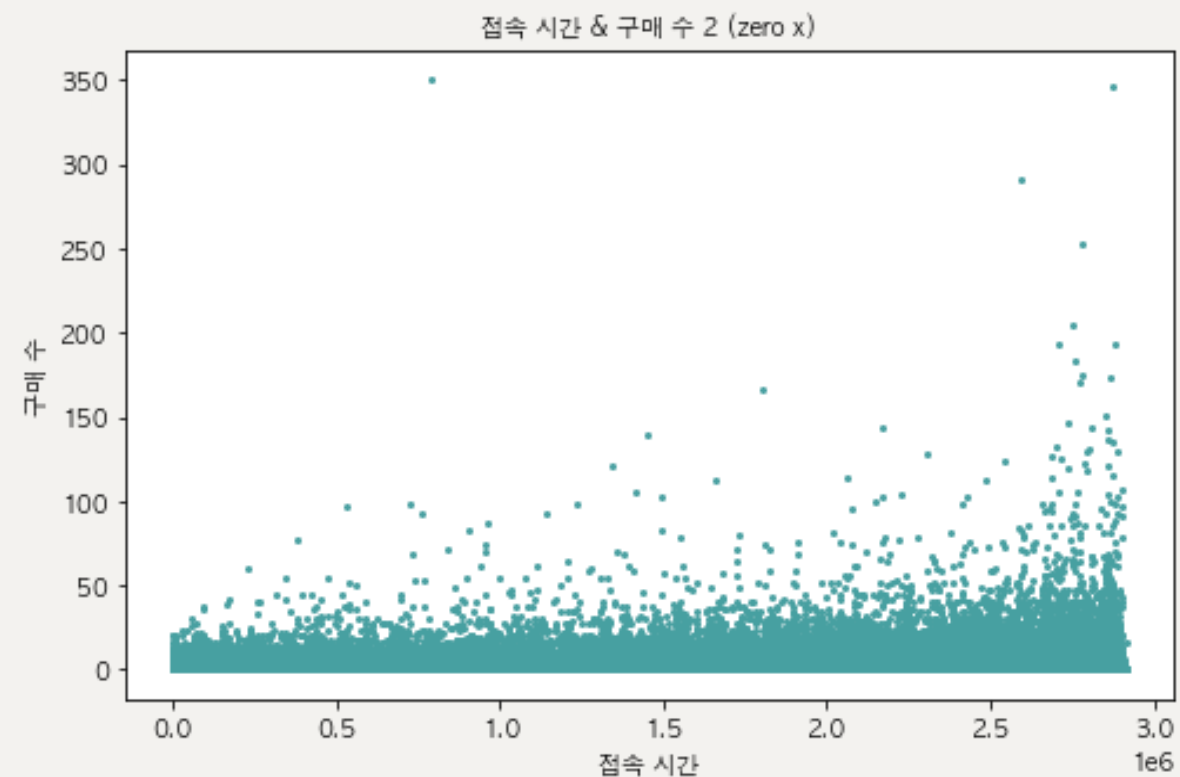
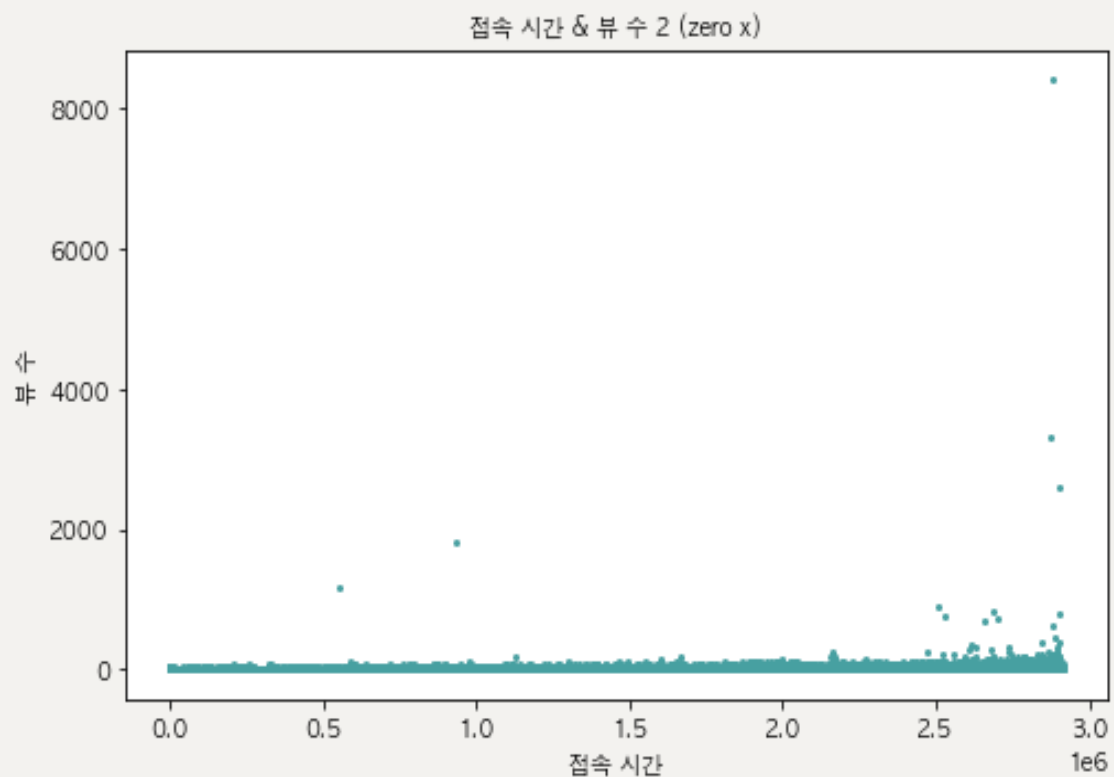
가설 검증

2 사이트에 오래 머물수록 구매 전환율이 더 높을 것이다.

**접속 시간: user_id를 기준으로 계산

- ▶ 방문/구매 건수=0인 데이터 제거하여 진행함
- ▶ 접속시간이 길어질수록 구매 수 증가하나, 뷰의 증가율 비교적 작음 \Leftrightarrow 구매 전환율 감소

👉 “가설 2 기각할 수 있음”

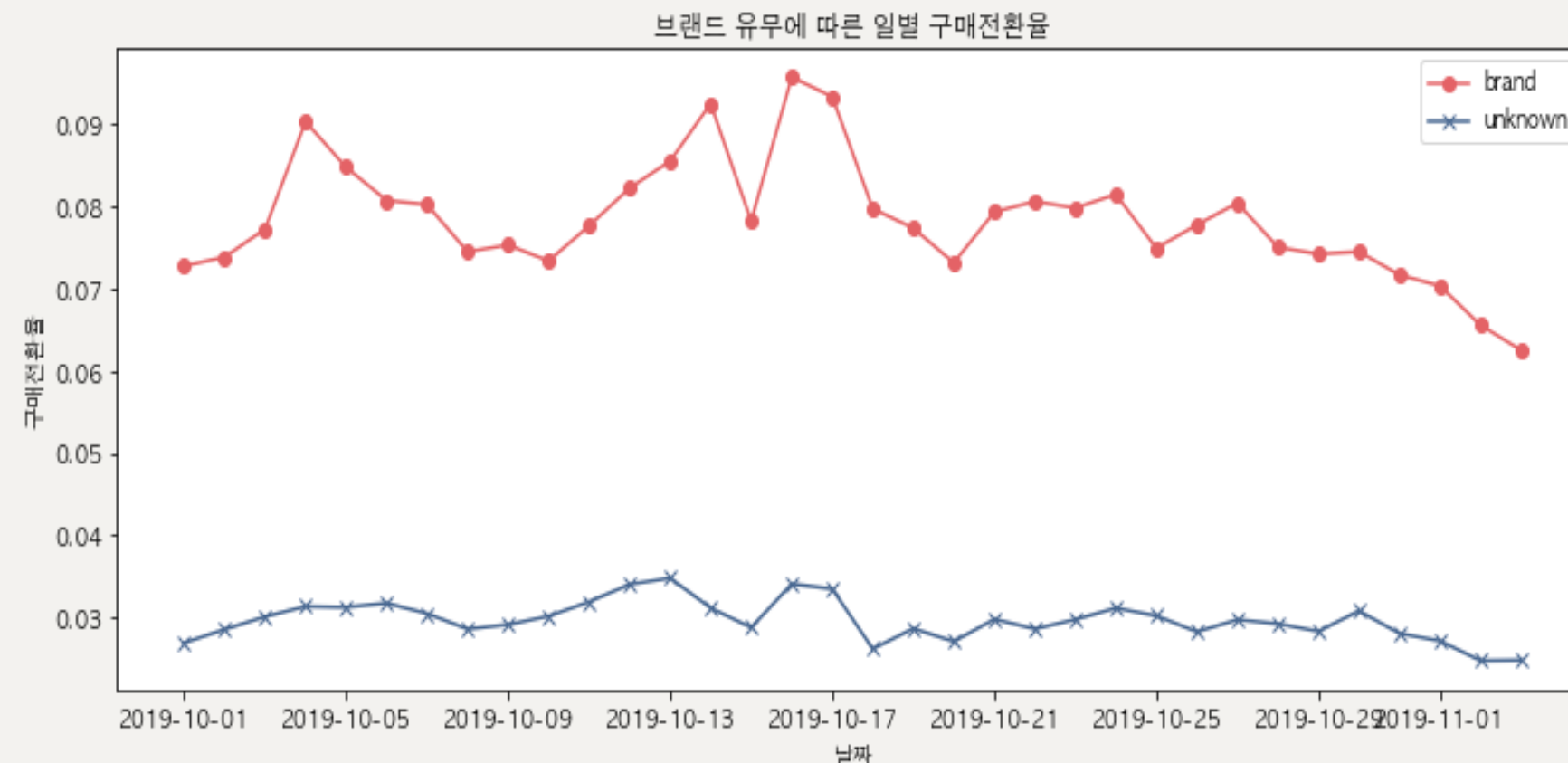


가설 검정

3 **brand**가 있는 상품일 경우의 평균 구매 전환율은 **brand**가 없는 상품보다 더 높을 것이다.

- ▶ 브랜드가 있는 상품 구매 전환율이 약 3% 정도 ↑
- ▶ 브랜드의 유무에 따른 평균 구매 전환율이 같은지 **t-test 검정**: 유의수준 5%에서 p-value=5.634619828049219e-4로 0에 수렴하는 값 → **두 평균 구매 전환율 같지 x**

👉 “가설 3 기각할 수 없음”

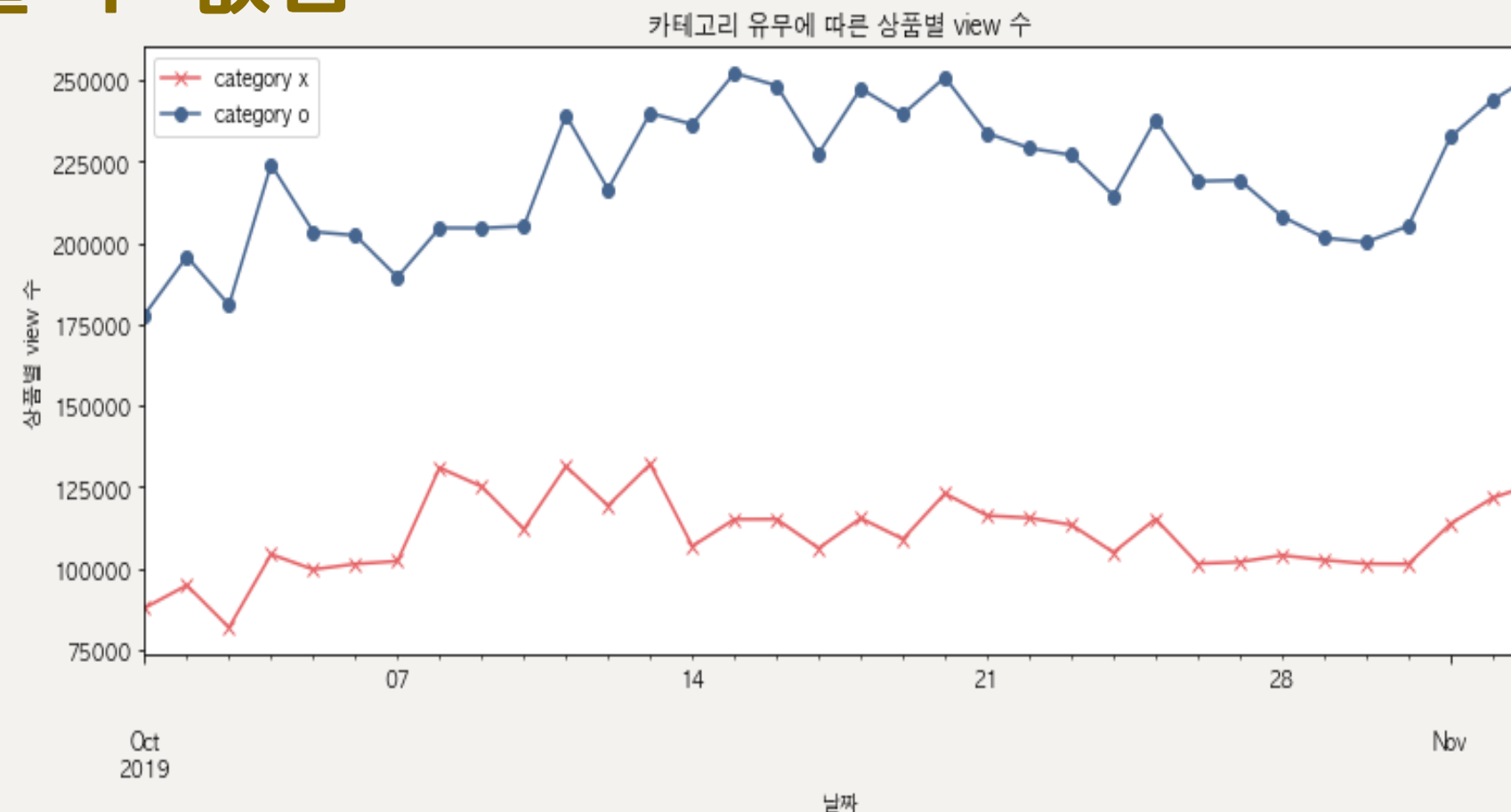


가설 검증

4 categorized되어 있지 않은 상품의 View 수는 categorized되어 있는 상품보다 더 적을 것이다.

- ▶ 카테고리화된 상품의 View 수가 약 5만건 정도 ↑
- ▶ 카테고리의 유무에 따른 상품별 평균 View 수 같은지 t-test 검증: 유의수준 5%에서 p-value=1.5138770184568762e-37로 0에 수렴하는 값 → 두 평균 View 수 같지 x

👉 “가설 4 기각할 수 없음”



Action Plan

1 신규고객 확보를 위한 Action Plan

👉 카테고리 등록 활성화

- 가설 검정을 통해 카테고리에 따라 상품 view 수에 차이가 존재함을 확인
- **카테고리화**가 되어있을 때, 유저가 필요한 상품을 **효율적으로 검색**할 수 있음
- E-Commerce가 낯선 고객에게 E-Commerce만의 편리성을 느낄 수 있도록 **검색 편리성 제공**

👉 옴니채널 서비스 활용 (* 옴니채널 서비스: 소비자가 온/오프라인에서 상품을 검색/경험/구매할 수 있는 서비스)

- 카자흐스탄: **전자상거래의 인식이 비교적 낮으며**, 관련 **인프라도 부족 + 도심 지역과 청년층**을 중심으로 E-Commerce 이용률이 ↑↑
+ 소비자의 1) 인터넷 결제와 2) 상품 배송의 불완전하고 안정적이지 못한 시스템, 그리고 3) 전자 쇼핑에 대한 **신뢰수준 ↓↓**
 - 소비자가 **오프라인으로 상품을 보고** 편리하게 **온라인으로 구매 가능**하도록 하여 **신규고객 유입 유도**
 - ▶ 오프라인 매장의 기존 고객을 유지하면서 온라인 시장의 새로운 고객 유치 가능

Action Plan

1 신규고객 확보를 위한 Action Plan

👉 신규 고객을 위한 프로모션 및 할인행사 진행

- 신규고객: 구매 이력이 없는 고객
- 방문 수는 많지만 구매는 활발하지 않은 18~23시를 공략해 첫 구매 시 쿠폰 및 할인을 제공하고, 구매이력이 없는 고객들이 관심 가지는 상품에 대한 프로모션 행사 진행
- 카자흐스탄 전자상거래 트렌드: 모바일 쇼핑이 급속도로 확산
 - 모바일로 상품을 주문하는 카자흐스탄 인구 비율 65%로 집계됨
 - ▶ 첫 in-app(=모바일 앱 내에서) 구매 시 할인 쿠폰 제공해 모바일 앱 신규 고객 확보

Action Plan

2 기존 고객의 구매 장려를 위한 Action Plan

- 기존 고객의 총 구매액 및 1인당 구매액 > 신규 고객 총 구매액 및 1인당 구매액
 - 기존 고객의 구매율 증가 = 마케팅 비용을 비교적 적게 들여 더 많은 구매를 유도할 수 있음
- 👉 구매한 상품과 연관된 상품을 추천하는 서비스 제공
 - 예) 핸드폰 구매 시 핸드폰 케이스, 필름, 이어폰 등 추천
 - 단순히 핸드폰을 구매하는 것이 아니라 관련 상품을 추천하여 또 다른 상품 구매를 유도
- 👉 기존 고객을 타겟팅한 프로모션 진행
 - 기존 고객의 방문과 구매가 가장 활발한 12~17시에 등급별 할인쿠폰 및 마일리지 등 제공

Action Plan

2 기존 고객의 구매 장려를 위한 Action Plan

👉 저관여제품을 중심으로 상품 다양성 확보

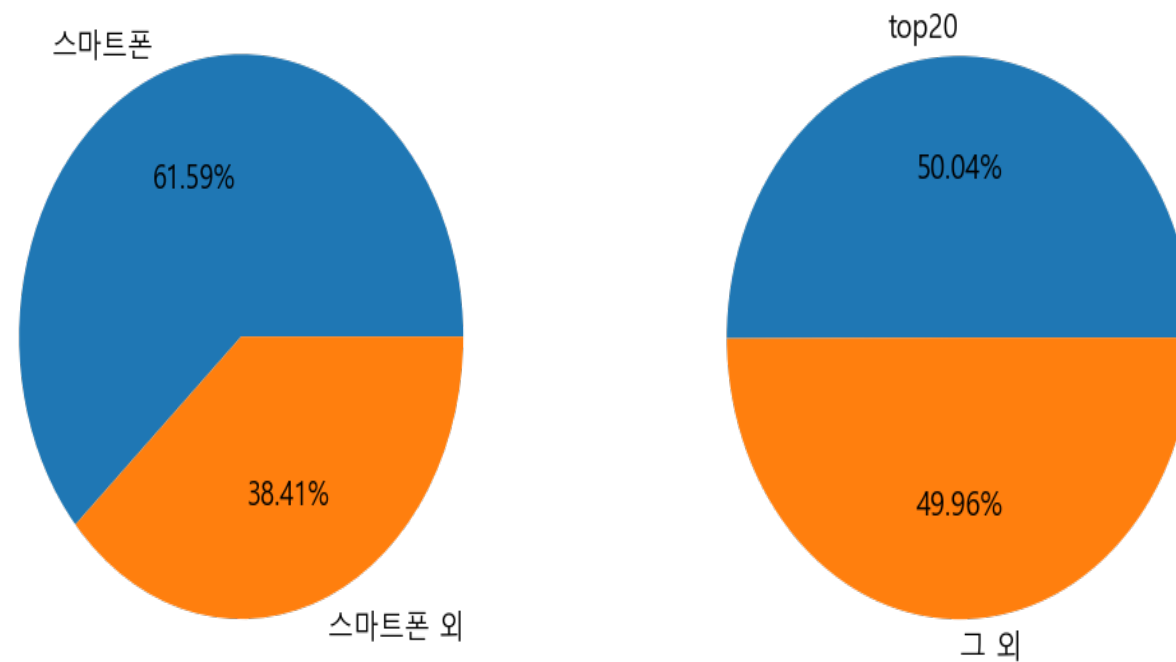
- 주어진 데이터를 통해 해당 E-Commerce에 **저관여제품이 많지 않음**을 파악함
 - 저관여제품: 제품의 중요도가 낮고 가격이 저렴하며 상표 간의 차이가 별로 없을 뿐만 아니라 잘못 구매해도 상관없다고 생각하여 **소비자가 별다른 고민 없이 구매**하는 제품
 - **광고의 반복적 노출**: 저관여제품 소비자의 **태도 및 행동 변화에 효과 O**
 - 제품별 재구매 주기
 - 생수: 약 한 달
 - 기저귀: 약 2주
 - 식품/음료: 5~14일 이내
 - 건강보조식품, 뷰티, 패션잡화: 1~3개월 이내
 - 생활용품: 3개월 이상
 - 저관여제품 중 **필수재 & 소비 주기가 짧은 제품**(예) 생필품, 식품, 개인위생 등등)을 위주로 확보
- ▶ 기존 고객에게 **필수재/생필품 또한 온라인/모바일로 쉽게 구매할 수 있다고 어필하는 반복적인 광고 노출**을 통해 구매 전환을 유도할 것

추천시스템

✓ 모델 및 평가지표 선정 이유

- 👉 Baseline model : event가 많은 상위 20개의 product에 대해 일괄 추천
 - ▶ 2019년 10월 판매된 아이템 중 스마트폰이 전체의 61.59%를 차지
 - ▶ 인기 상위 20개의 스마트폰이 전체 스마트폰 판매의 절반을 차지

판매 아이템 수 비교 - 스마트폰 vs 비스마트폰 스마트폰 판매 수 비교 - top20 vs 그 외 (19년 10월)



인기 상위 20개 product 정보

	product_id	category_code	brand		product_id	category_code	brand
0	1004739	electronics.smartphone	xiaomi	10	1004833	electronics.smartphone	samsung
1	1004767	electronics.smartphone	samsung	11	1002524	electronics.smartphone	apple
2	1002544	electronics.smartphone	apple	12	1004957	electronics.smartphone	xiaomi
3	1004873	electronics.smartphone	samsung	13	1004741	electronics.smartphone	xiaomi
4	1004870	electronics.smartphone	samsung	14	1004777	electronics.smartphone	xiaomi
5	1004856	electronics.smartphone	samsung	15	1005159	electronics.smartphone	xiaomi
6	1004836	electronics.smartphone	samsung	16	1005160	electronics.smartphone	xiaomi
7	1004785	electronics.smartphone	huawei	17	1005031	electronics.smartphone	xiaomi
8	1005115	electronics.smartphone	apple	18	1005161	electronics.smartphone	xiaomi
9	1004249	electronics.smartphone	apple	19	1005239	electronics.smartphone	xiaomi

- 👉 Advanced model : Content-Based model (TF-IDF, word2vec)
 - ▶ 유저의 이전 행동(view, cart, purchase)과 유사한 상품들을 추천
 - ▶ 다른 유저의 평점 데이터가 없어, CF 모델은 적용하기 어렵다고 판단하여 CB 채택
 - ▶ 신규 유저에게도 적용이 가능한 모델이라 판단하여 채택 (=콜드스타트 문제해결)
- 👉 Evaluation metrics : MAP@K, NDCG@K
 - ▶ 랭킹 기반(인기, 이벤트 수) 시스템에 사용되는 MAP@K, NDCG@K를 평가지표로 사용

추천시스템

✓ 모델별 성능

👉 Baseline model

▶ MAP@20: 0.03470936329312459 / NDCG@20: 0.08250075125861263

👉 Content-based : TF-IDF

- 각 상품의 메타정보(category_code, brand 등)를 활용하여 유사도 계산

- cosine 유사도를 사용해 유사도가 높은 순으로 유저별 20개의 상품 추천

▶ MAP@20: 0.017148624592693736

▶ NDCG@20: 0.04917596785249336

👉 Content-based : Word2Vec

- doc2vec를 사용해 각 상품의 product_id와 category_code 벡터화

- 상품별 유사도기준 유저가 본 상품과 연관성이 높은 20개의 상품 추천

▶ MAP@20: 0.001588630171801049

▶ NDCG@20: 0.005573220462657566

📌 doc2vec.Doc2Vec 파라미터: 레퍼런스에 나온 그대로를 활용하여 모델 구현

dm=0, dbow_words=1, window=10, vector_size=100, alpha=0.025, seed=1234, min_count=5,
min_alpha=0.025, workers=4, hs = 1, negative = 10

✓ 분석

👉 결론

- 1) **카테고리** 등록을 활성화하고 2) 아직 E-Commerce가 낯선 중장년 고객층을 위해 **옴니채널** 서비스를 적극 활용하며, 3) 고객이 가장 많이 방문하는 **18~23시에 in-app 구매 시 적용되는 쿠폰 등** 제공하여 **신규 고객 확보**할 것
- 또한 1) 유저의 구매 상품에 대한 **관련 상품을 추천**하는 서비스를 제공하고 2) 기존 고객이 가장 많이 방문하는 **12~17시에 등급별 할인 및 프로모션**을 진행함과 동시에 3) **저관여제품 중심으로 다양한 상품을 확보**한 후 해당 제품에 대한 **반복적 광고 노출로 기존 고객의 구매 장려**를 유도할 것

👉 회고

- 첫 E-Commerce 데이터 분석 프로젝트라 E-Commerce에 대한 도메인 및 인사이트를 습득함으로써 많은 것을 배울 수 있는 기회였음
- session 정보를 활용해 보다 다양한 분석기법 및 심도깊은 분석까지는 하지 못해 아쉬움

결론 및 회고

✅ 추천

👉 결론:

- 베이스라인 모델이 성능이 가장 우수한 결과를 보임

Baseline model > CB model(TF-IDF) > CB model(Word2Vec)

👉 회고

- 시간적 여유가 부족해 추천알고리즘 학습과 다양한 추천 모델을 적용해보지 못한 점이 아쉬움
 - 사용한 개선모델이 최선의 개선모델인가?에 대한 검토 필요
 - 제한된 정보(제품명, 평점 등)로 인하여 적절한 모델을 적용하지 못함
 - 유사도 방법 변경, 메타 데이터 추가 등을 통한 개선 필요
 - 다양한 데이터셋을 활용한 추천시스템 학습 필요

결론 및 회고

참고 문헌

- https://dream.kotra.or.kr/kotranews/cms/news/actionKotraBoardDetail.do?SITE_NO=3&MENU_ID=180&CONTENTS_NO=1&bbsSn=243&pNttSn=183671
- <https://workingwithpython.com/pythoncohortanalysis/>
- https://dream.kotra.or.kr/kotranews/cms/news/actionKotraBoardDetail.do?SITE_NO=3&MENU_ID=180&CONTENTS_NO=1&bbsSn=243&pNttSn=183671
- <http://www.openads.co.kr/content/contentDetail?contsId=4386>, <https://corp.tmon.co.kr/media/detail/15277>
- <https://4ir.kisti.re.kr/ick/cmmn/viewPost/20180207000009>, <https://corp.tmon.co.kr/media/detail/15277>, <http://www.openads.co.kr/content/contentDetail?contsId=4386>

감사합니다!