

Code States AIB 17

---

# 패션이커머스 문제진단 프로젝트

team DA

임나래 유하민 이현준 이재은

---

# 목차

- 
- |    |              |
|----|--------------|
| 01 | 프로젝트 개요 및 구성 |
|----|--------------|
- 
- |    |         |
|----|---------|
| 02 | 데이터셋 안내 |
|----|---------|
- 
- |    |                 |
|----|-----------------|
| 03 | 문제진단 및 액션플랜 (1) |
|----|-----------------|
- 
- |    |                 |
|----|-----------------|
| 04 | 문제진단 및 액션플랜 (2) |
|----|-----------------|
- 
- |    |          |
|----|----------|
| 05 | 이탈 모델 안내 |
|----|----------|
- 
- |    |         |
|----|---------|
| 06 | 결론 및 제언 |
|----|---------|
-

---

# 01

## 프로젝트 개요 및 구성

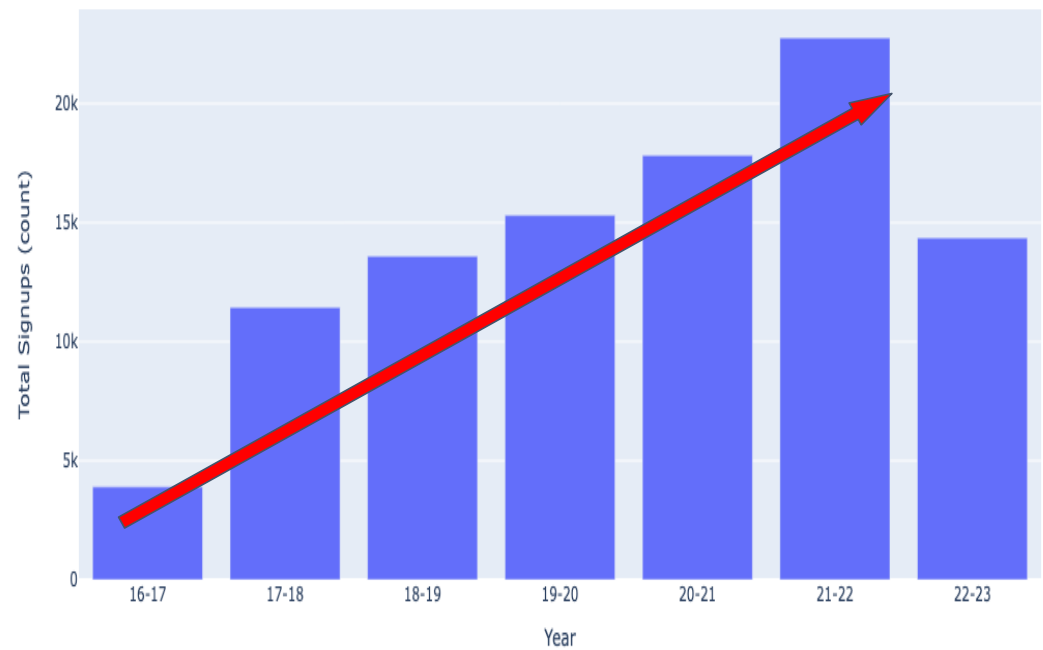
## 분석 배경 (1)

인도네시아 패션 이커머스회사  
패션캠퍼스

2016년 설립 이래  
신규 고객 수 매년 꾸준히 증가  
(2022.07 집계)

➡ 기존 고객 유지 및  
신규고객 유입을 통한  
매출 상승을 위한 프로모션 시행

Total Sign-ups by Year



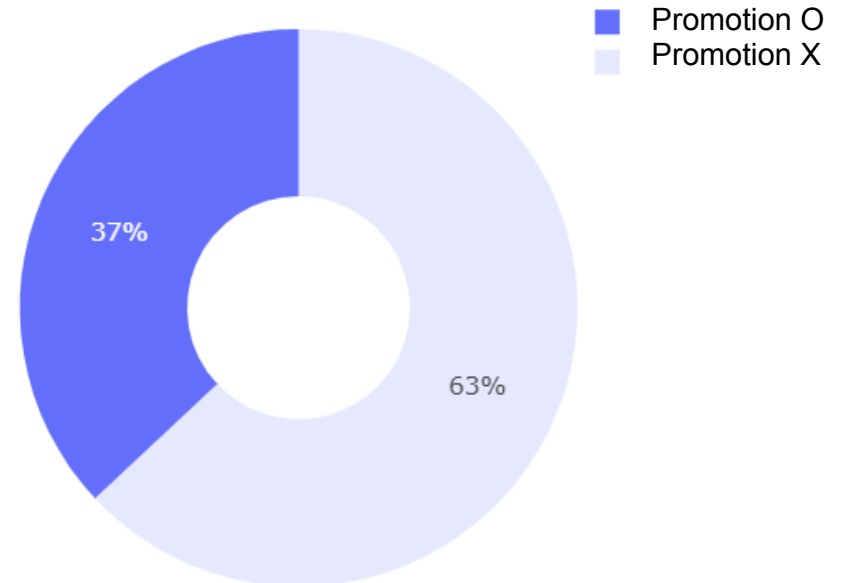
## 분석 배경 (2)

### 프로모션 진행 결과

마지막 구매에서 프로모션 사용 후 이탈한 고객  
의 비율  
: 전체 이탈 고객 중 37%

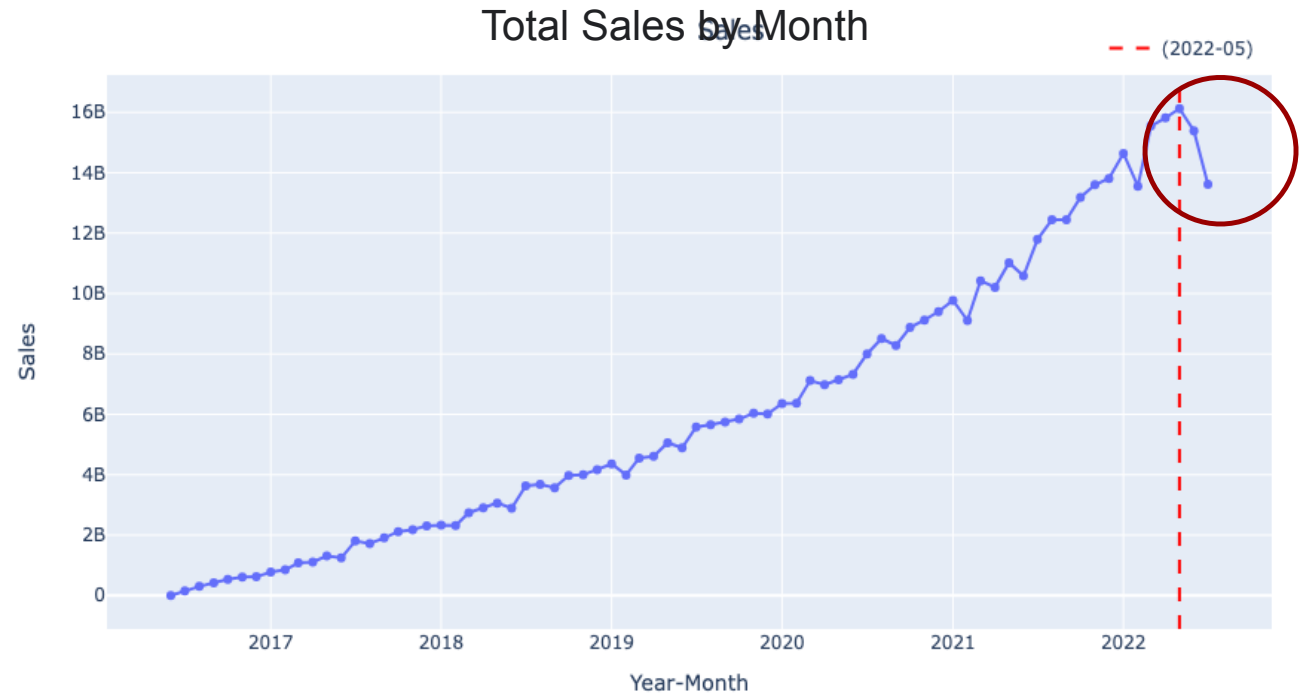
➡ 많은 사용자가 프로모션 사용 후,  
플랫폼으로 돌아오지 않아  
사용자 이탈률 상승

Promotion from last transaction  
Used churn consumer



## 분석 배경 (3)

매출 분석 결과  
꾸준한 매출 상승 중,  
**최근 2개월 매출  
15.5% 급격한 하락**  
(현재시점 2022-08 기준)



## 문제 도출 및 분석 목표

---

### 문제 도출

최근 2개월간의 급격한 매출 하락

### 목표 1

급격한 매출 하락세 회복

➡ 고객 세분화를 통해 각 세그먼트에 맞는 타겟 액션 플랜 수립

### 목표 2

고객 이탈을 방지하기 위한 고객 이탈 예측 모델 제작

➡ 이탈 가능성이 높은 고객들에 대한 적시적인 대응 방안 마련

## 추가 목표

### 배경 분석

최근 글로벌 중고 옷 시장이 새 옷 시장에 비해 21배 성장  
소비자가 중고 쇼핑 경험에 만족 시, 재구매 발생율 상승 (논문 소비자 분석 근거)  
인도네시아 패션 이커머스 마켓 'Tinkerlust'  
: 2018년 중고 의류 타겟을 통해 매달 Revenue 15%~20% 성장

Source: Determinants of Secondhand Clothes Repurchase Intention:  
Indonesian Consumer's Perspective (2020)

### 목표

고객 유치, 유지 및 성장을 위해 고객의 중고 제품에 대한 관심도 분석

➡ 신사업으로 중고 사업 진행 방향 수립



## 담당 업무

구성원	역할	담당 업무
임나래	팀장	<ul style="list-style-type: none"><li>▶ RFM분석(RFM점수, 연령에 따른 고객 세분화)</li><li>▶ 데이터 분석(상품에 따른 매출 분석, 인구통계학적 분석)</li></ul>
이현준	팀원	<ul style="list-style-type: none"><li>▶ 외부자료 수집</li><li>▶ 데이터 분석(중고, 구매, 방문 주기 및 전환율 분석)</li></ul>
이재은	팀원	<ul style="list-style-type: none"><li>▶ 외부자료 수집</li><li>▶ 데이터 분석 (프로모션, 매출)</li></ul>
유하민	팀원	<ul style="list-style-type: none"><li>▶ 세션 분할 진행과 세션 관련 고객의 행동 분석</li><li>▶ 학습 데이터 생성 및 모델 제작</li></ul>

## 수행 절차

프로젝트 수행 기간 : 22.06.07 ~ 22.06.30

6.7 – 6.10

6.11 – 6.15

6.15 – 6.25

6.26 – 6.30

### 사전 기획

데이터파악 및 문제정의  
상품/비즈니스/유저 측면  
데이터 EDA

외부데이터 수집  
인도네시아 시장조사  
및  
메가트렌드 파악

### 데이터 탐색

데이터 정제 및 정규화  
데이터 내 컬럼 재정의

### 데이터 분석

세션 데이터를 통한 퍼널분석  
고객세분화

Customer Segment 별  
상품/비즈니스/유저 측면  
데이터 분석

이탈 함수 제작

### 이탈모형 제작

주요 지표 선정  
액션아이템 설계

이탈모형 모델링

모형 성능 최적화  
및 결과 시각화

---

02

# 데이터셋 안내

# 데이터셋 정보

출처

캐글, 'Fashion Campus'

데이터셋

전처리

click\_stream.csv

- Null 값 → 0 변환
- 라벨인코딩  
['Promo code']  
: 사용 1, 미사용 0
- ['Payment Status']  
: Success 1, Failed 0
- 중고 검색어 추출  
['search\_keywords']  
=='Second Hand'

product.csv

- ['brand'] 생성  
['productDisplayNa  
me'].split().str[0]
- 데이터타입 변경  
['gender']  
['masterCategory']  
['subCategory']  
['baseColour']  
['season']  
['usage']  
⇒ Category타입
- ['year'] 제거

customer.csv

- 데이터타입 변경  
['gender'],  
['home\_location']  
⇒ Category타입
- ['age\_now'] 생성  
2022년(현재)  
- birthdate의 year

transaction.csv

- 라벨인코딩  
['Promo code']  
: 사용 1, 미사용 0
- ['Payment Status']  
: Success 1, Failed 0
- 'Shipment' 특성 제거

# 데이터셋 정보

출처

캐글, 'Fashion Campus'

데이터셋

click\_stream.csv

product.csv

customer.csv

transaction.csv

사용  
컬럼

'session\_id', 'event\_name',  
'event\_time',  
'traffic\_source',  
'product\_id', 'quantity',  
'item\_price',  
'payment\_status',  
'search\_keywords',  
'promo\_code',  
'promo\_amount'

'id', 'gender',  
'masterCategory',  
'subCategory',  
'articleType',  
'baseColour', 'season',  
'usage', 'brand',  
'productDisplayName'

'customer\_id',  
'gender', 'birthdate',  
'first\_join\_date',  
'home\_location',  
'age\_now',  
'age\_segment',  
'RFM\_Segment',  
'join\_year',  
'join\_month'

'created\_at', 'customer\_id',  
'booking\_id', 'session\_id',  
'payment\_method',  
'payment\_status',  
'promo\_amount',  
'promo\_code', 'total\_amount',  
'shipment\_fee',  
'product\_id', 'quantity',  
'item\_price', 'with\_promo'

최종사이즈

(12833602, 11)

(44424, 8)

(99999, 7)

(1254578, 14)

## RFM세분화

RFM 분석 : 고객의 구매 기록을 기반으로 고객의 가치를 평가하는 기법

R

Recency, 최근성

고객이 마지막으로 구매한 시점 : 값이 낮을수록 최근 구매를 의미

F

Frequency, 구매빈도

고객의 특정 기간 동안의 구매 빈도 : 값이 높을수록 잦은 구매를 의미

M

Monetary, 지출금액

고객의 특정 기간 동안의 지출 금액 : 값이 높을수록 높은 매출 기여도를 의미

## RFM 등급에 따른 고객 세분화

---

3

R 최근성

상위, 중위, 하위 각각 약 33.3 % 기준으로 ,  
Recency값이 낮은 (최근성이 높은) 분위 순으로 3,2,1 등급 부여

2

F 구매빈도

M 지출금액

상위, 하위 각각 50 % 기준으로 ,  
값이 높은 (빈도와 금액이 높은) 분위 순으로 2,1 등급 부여

## RFM 등급에 따른 고객 세분화



### R 최근성

상위, 중위, 하위 각각 약 33.3 % 기준으로 ,  
Recency값이 낮은 (최근성이 높은) 분위 순으로 3,2,1 등급 부여

'R(최근성)' 만 3개의 등급으로 세분화 : 현재 상황에서는 'R 지표'가 중요하다고 판단

- 최근에 구매하지 않고 이탈하는 고객 수 상승 ➡ 최근 2개월 매출 급감
- 'R(최근성)'이 높은 고객들은 재구매 가능성이 높고 이탈 가능성이 낮다고 판단 가능



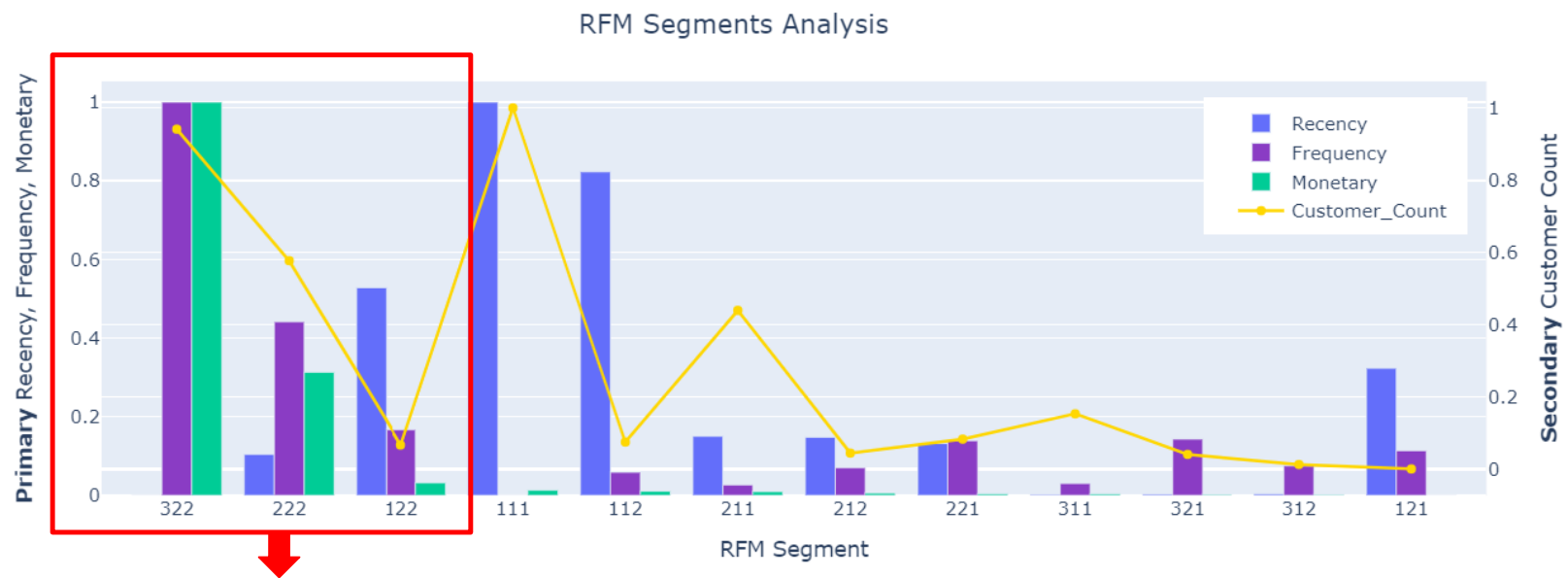
## RFM 등급에 따른 고객 세분화

RFM 등급에 따른 고객 세분화를 통해  
전체 고객 중 구매 이력이 있는 고객을  
총 12개의 그룹으로 세분화

➡ 타겟 고객 그룹 선정  
322, 222, 122 그룹

	Recency	Frequency	Monetary	Customer_Count
RFM_Segment				
322	25.06	64.63	1051400934164	13739
222	104.64	29.78	329386070692	8451
122	426.99	12.64	33761330954	1039
111	786.08	2.21	14109390664	14596
112	651.30	5.88	11742206358	1157
211	139.49	3.88	10291703222	6454
212	137.52	6.61	6453703713	713
221	125.64	10.89	4838510640	1274
311	26.78	4.11	3887520590	2292
321	27.32	11.14	2517934499	661
312	27.90	6.90	2175956958	253
121	271.25	9.31	278768273	75

# RFM세분화 : 타겟 Segment 선정

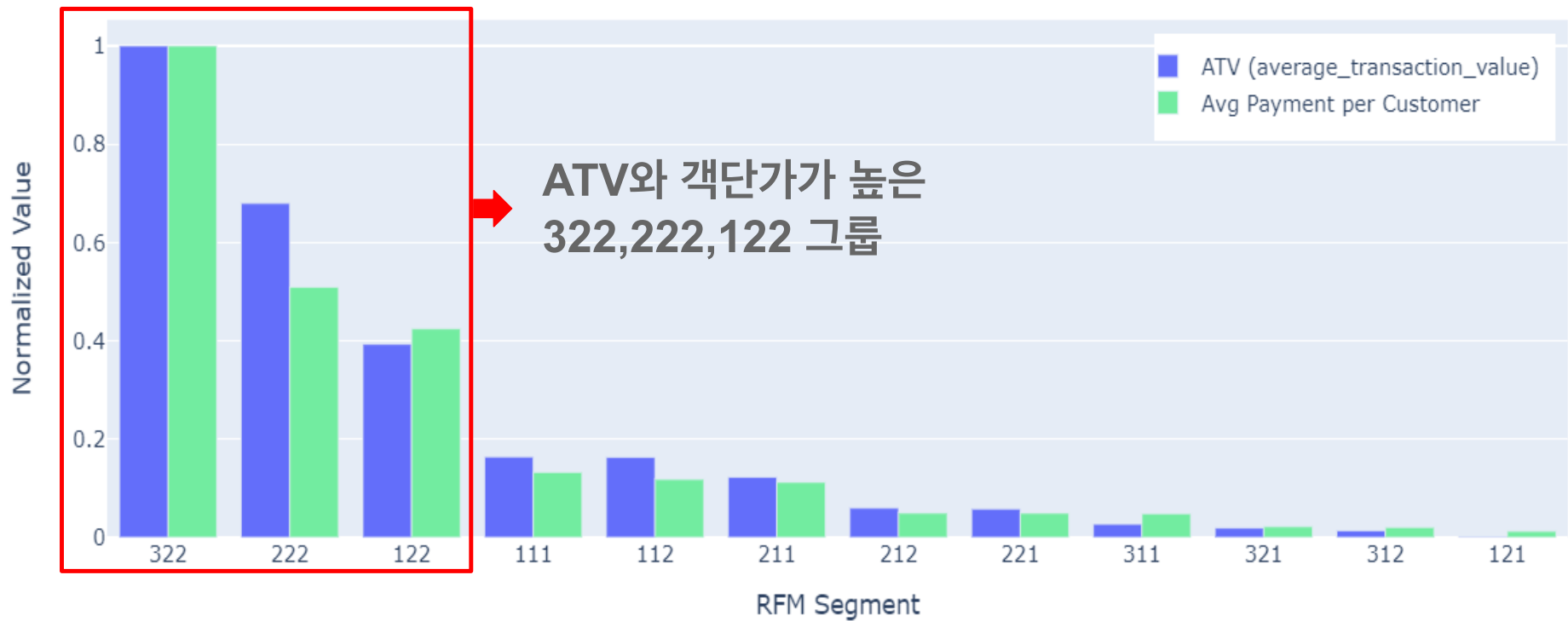


322 : 최근성, 빈도, 금액 모두 높은 충성도 높고 가치 있는 고객 ➡ 이탈 방지 및 충성도 유지 중요성

222 : 중간 등급의 최근성, 높은 빈도와 금액을 가진 성장 가능성이 있는 고객 ➡ 타겟 마케팅을 통해 322화

122 : 상대적으로 낮은 최근성 대비 높은 빈도와 금액을 가진 고객 ➡ 다시 활성고객이 된다면 높은 지출 가능성

## RFM세분화 : 거래 건 당 평균 금액과 객단가



---

# 03

## 문제진단 및 액션플랜 (1)

## 문제진단 (1) 322 그룹의 이탈

---

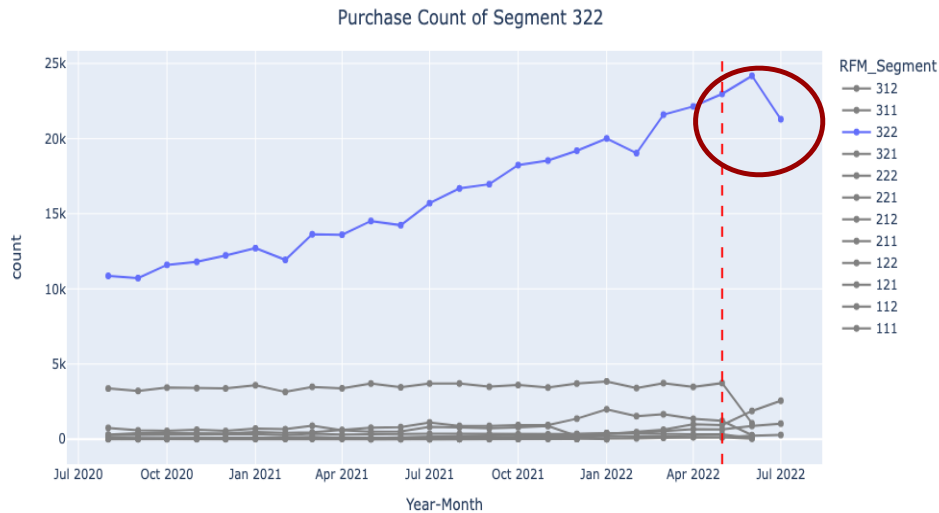
| 매출에 미치는 영향력이 큰 그룹의 이탈

➡ 322 그룹의 이탈 방지 대책 시급

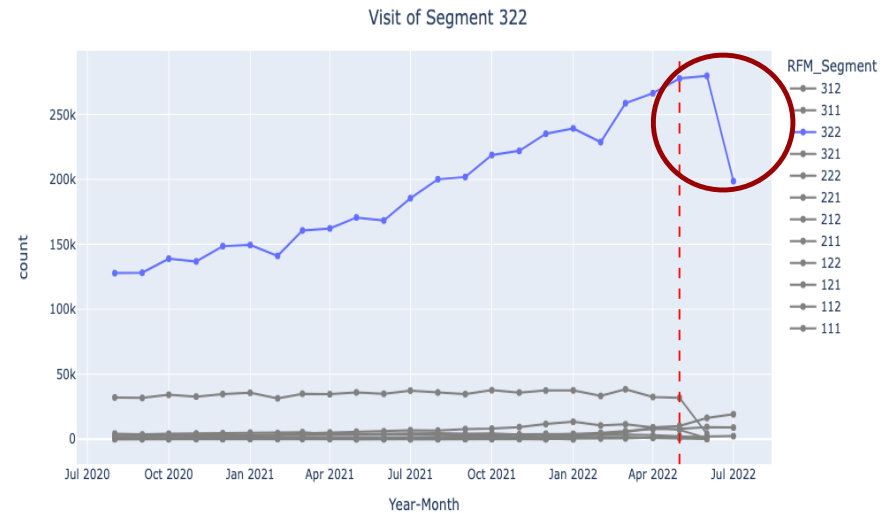
- 기존 고객 유지가 신규 고객 유치보다 비용 측면에서 효율적

# 문제진단 (1) 322 그룹의 이탈

## 1. 문제 기간 322 그룹의 구매 및 방문횟수 감소



구매횟수 12%

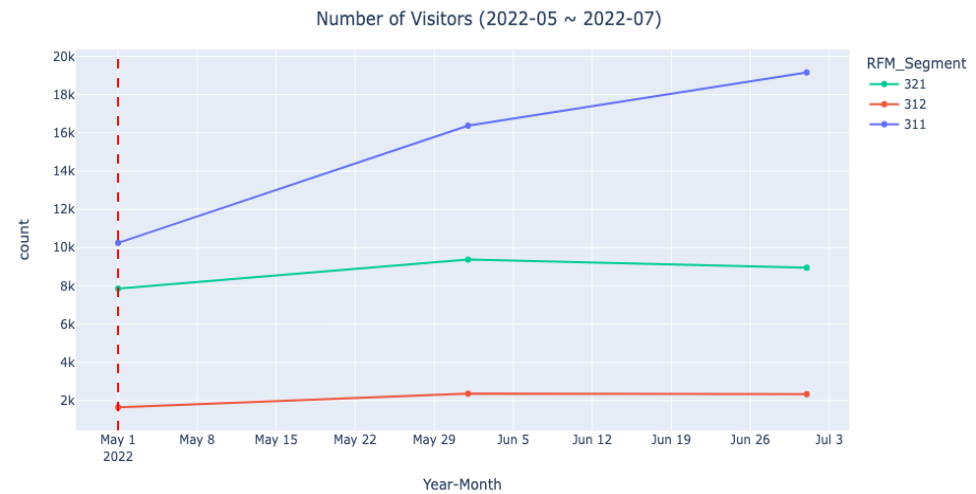
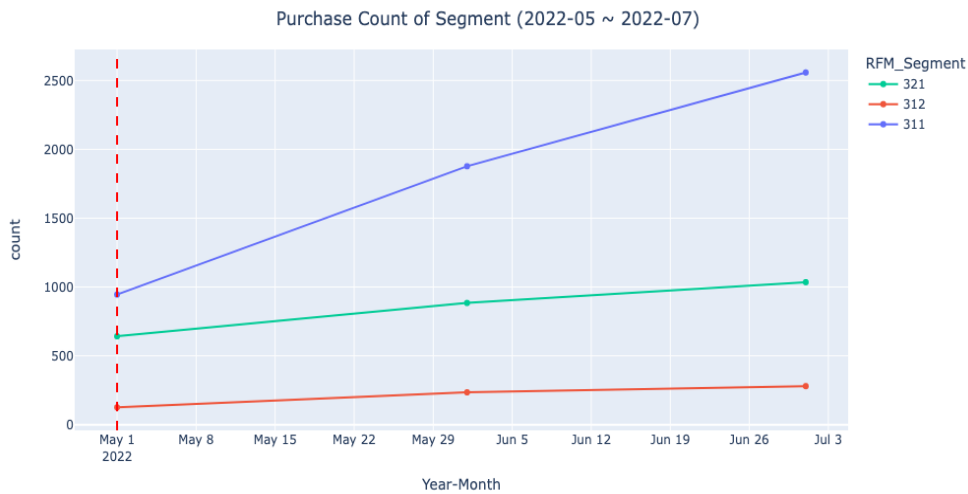


방문횟수 29%



# 문제진단 (1) 322 그룹의 이탈

## 1. 문제 기간 322 그룹의 구매 및 방문횟수 감소

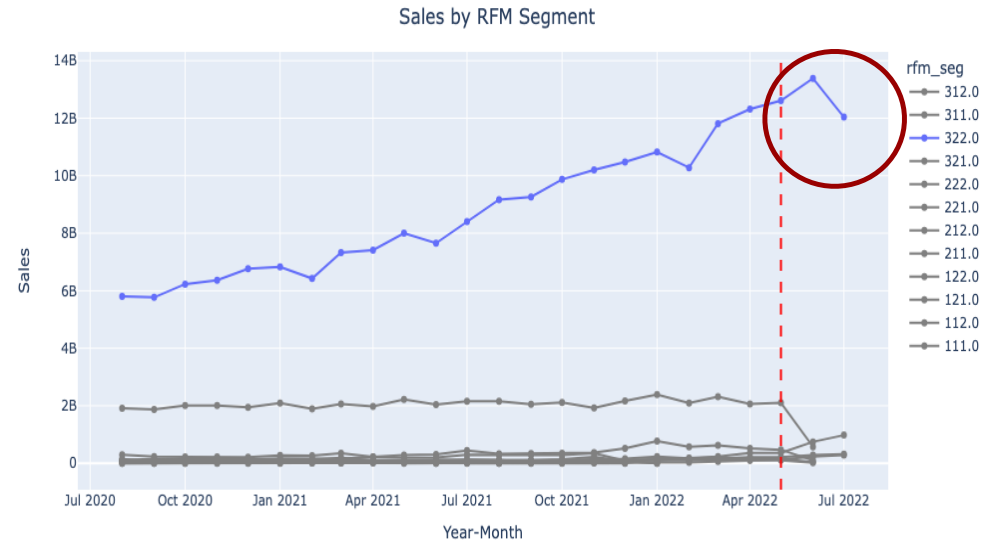


Recency가 3인 그룹 중 322 외 다른 그룹들의 급격한 감소 X

## 문제진단 (1) 322 그룹의 이탈

### 2. 문제 기간 매출 감소

- 구매 및 방문 감소 영향
- 최근 전체 매출 감소의 주 원인



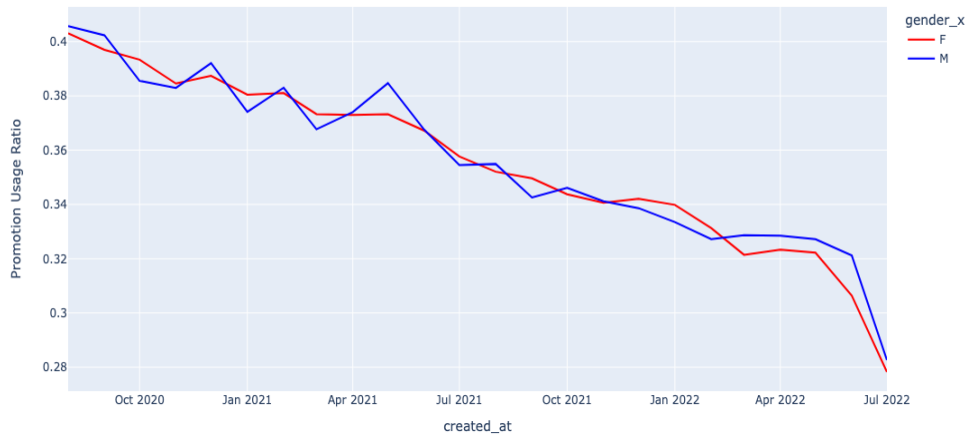


# 액션플랜 (1) 프로모션 개선

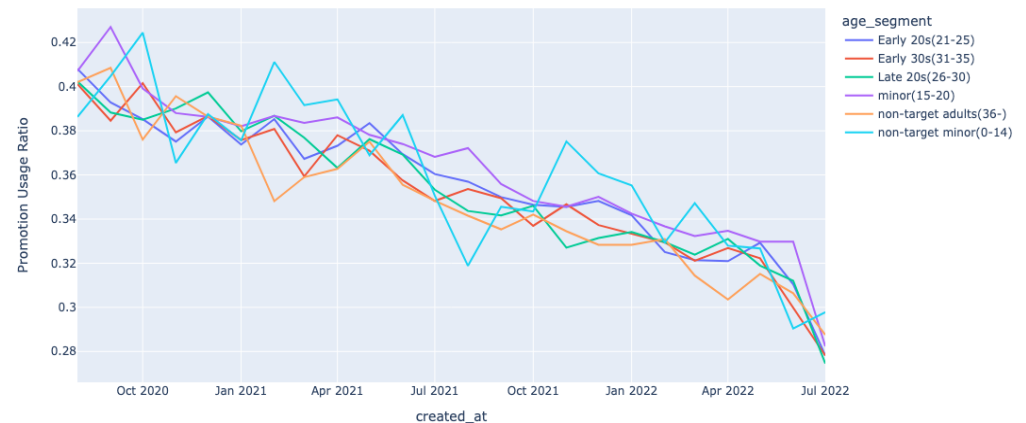
## 프로모션 문제점 1

- 코드를 활용한 프로모션을 진행하고 있으나 타겟 설정 부재

Promotion Usage Ratio by Gender

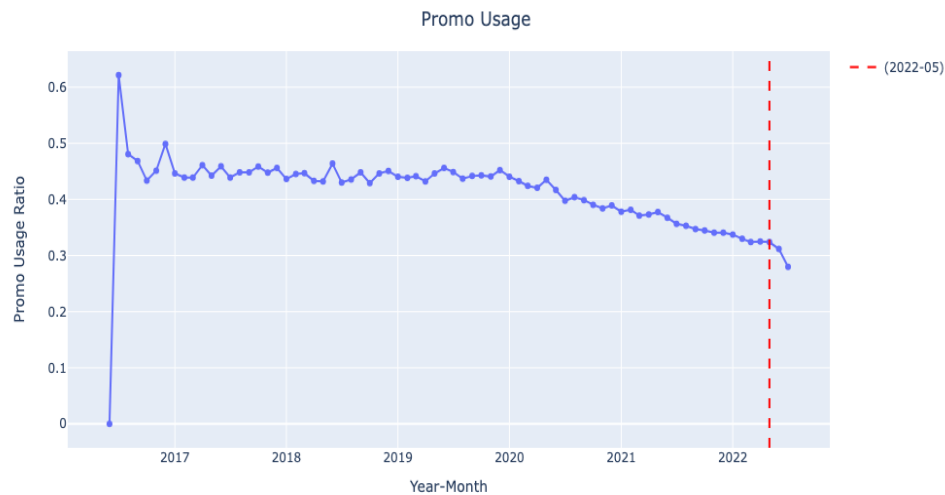


Promotion Usage Ratio by Age Segment

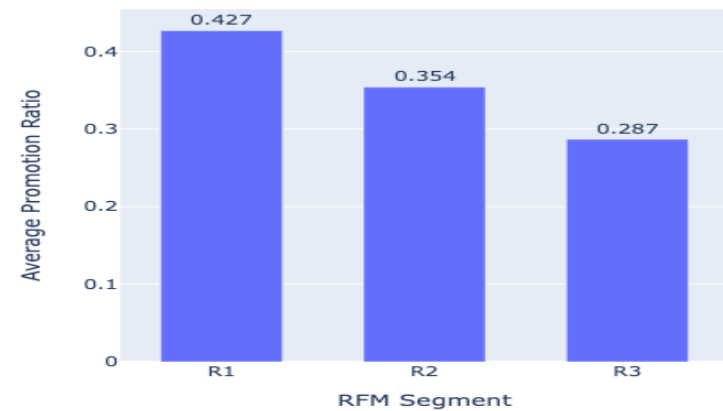


# 액션플랜 (1) 프로모션 개선

## 프로모션 문제점 1



### Average Promotion Usage by RFM Segment

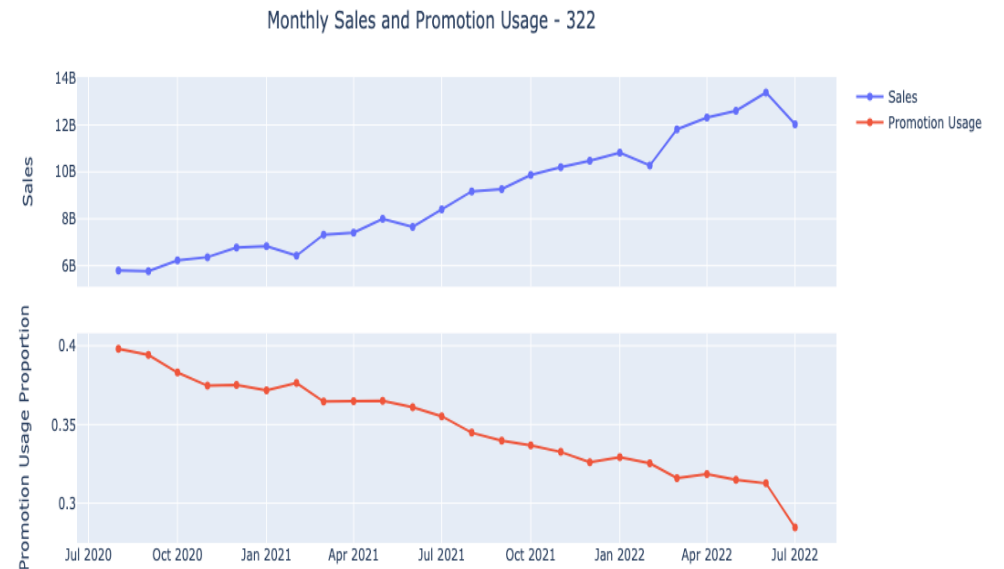


➡ 최근 프로모션에 대한 만족도 감소

# 액션플랜 (1) 프로모션 개선

## 프로모션 문제점 2

- 프로모션의 매출 영향력 미미
- 비효율적인 프로모션 비용 사용



## 액션플랜 (2) 프로모션 개선

---

목표

322  $\xrightarrow{\text{X}}$  122 & 222

액션

등급제 운영 등을 통한 로열티 부여

선호 제품 추천 및 할인 혜택 제공 등 재방문/재구매 유도

근거

특색 없는 프로모션

매출에 미치는 영향 미미

기대효과

고객 방문 증가

---

# 04

## 문제진단 및 액션플랜 (2)

## 장바구니 포기 정의

---

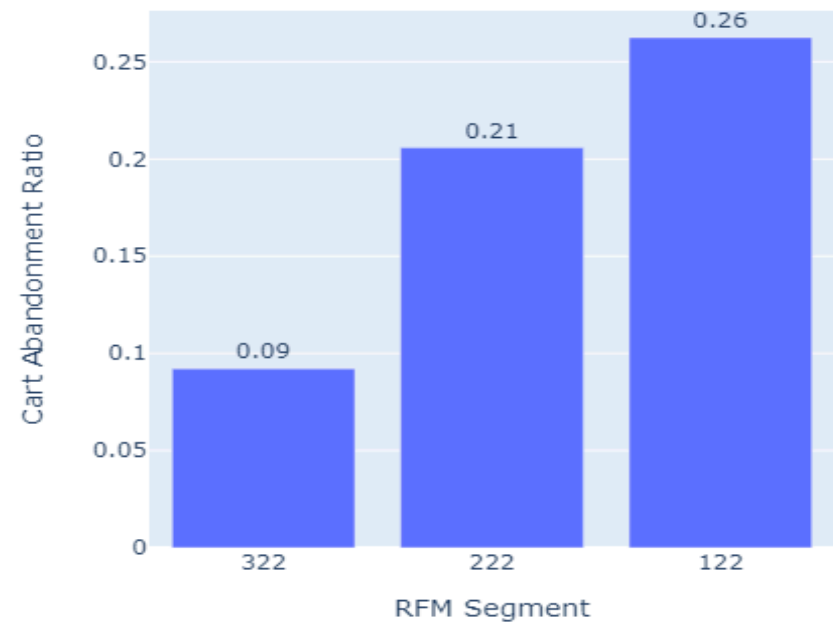
- 고객이 물건을 장바구니에 추가했지만 거래를 완료하지 않은 경우
- `last_event == add_to_cart`인 경우

## 문제 진단 (2) 장바구니 포기

타겟 고객별 장바구니 포기 비율

➔ 122 > 222 > 322

Cart Abandonment for the Session

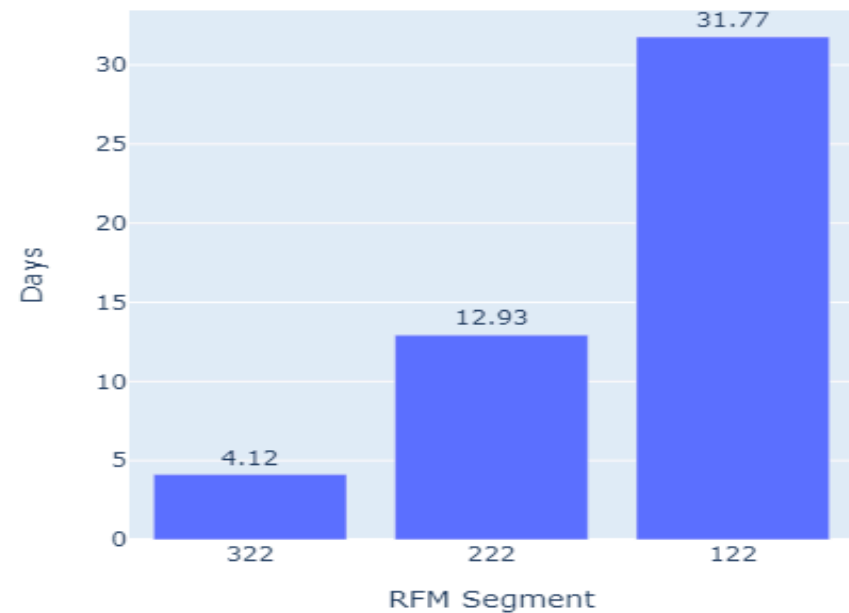


## 문제진단 (2) 장바구니 포기

평균 장바구니 구매 전환 기간

➔ 122 > 222 > 322

Average Cart-to-Booking Conversion Time(Days)





## 액션플랜 (2) 장바구니 포기

목표

122 & 222 → 322

액션

장바구니 리마인드 알림

근거

휴대폰 push 알람은 이메일보다

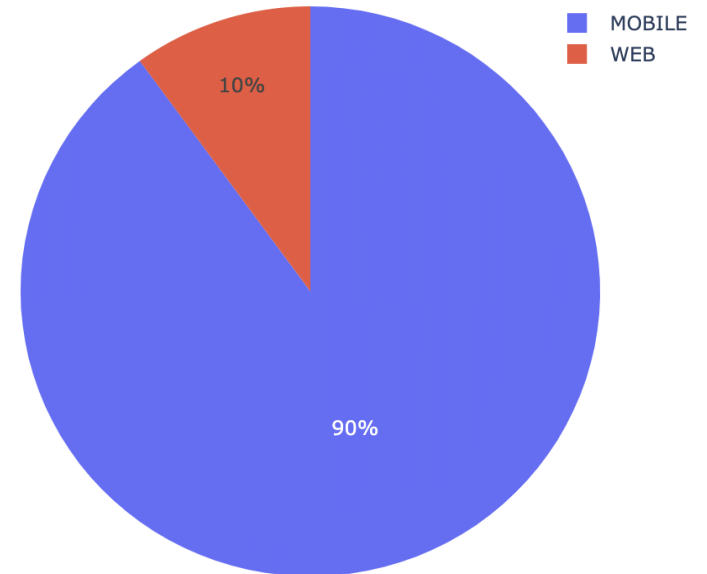
30% 높은 오픈율을 보임

(source: shopify 기사)

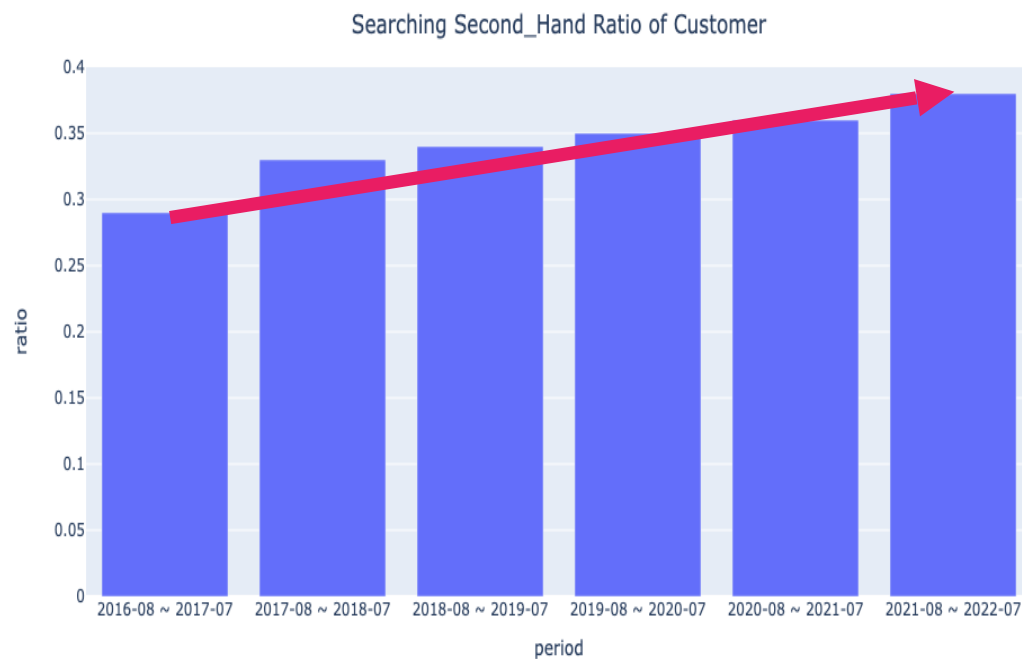
기대효과

고객 방문 증가

Traffic Source Distribution



## 문제 진단 (2) 중고 검색 비율



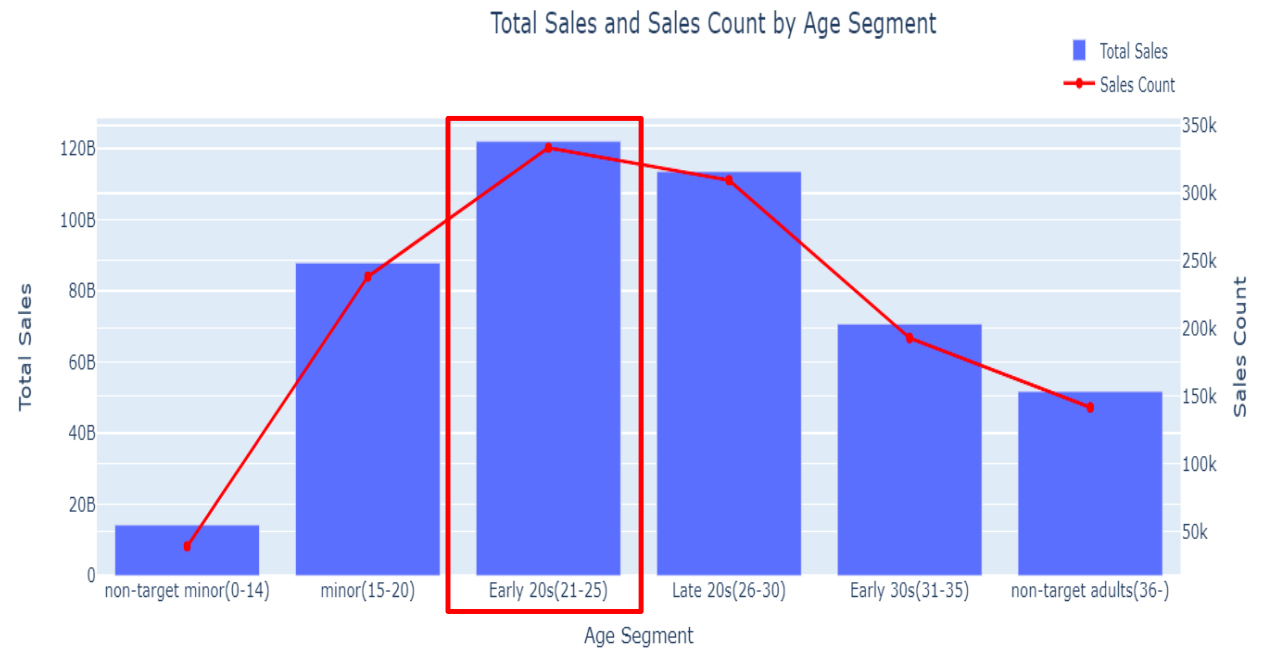
매년 중고에 대한  
검색 비율 상승

# 연령대별 중고 거래 및 매출

## Sampling Research

나이	중고 거래 경험 비율 (%)
19~25세	82.6
12~18세	7.5
26~35세	4.7
36~50세	1.2

Source: Determinants of Secondhand Clothes Repurchase Intention: Indonesian Consumer's Perspective(2018)



## 중고 사업 액션 아이템

---

목표

연령 타겟층 10대 ~ 20대의 구매율 및 쇼핑 만족도 상승

액션

중고 사업 시작

근거

중고에 대한 타겟 연령층의 관심 상승,  
(경쟁사) Tinkerlust 매출 상승

기대효과

매출 상승

---

05

## 고객 이탈모델

# 이탈의 정의

- 25일 동안 구매가 이루어지지 않은 경우를 이탈로 정의
- click\_stream\_new 데이터에서 'BOOKING' 이벤트가  
**25일** 동안 발생하지 않은 고객
- 최근 2년 (2020/08 ~ 2022/07) 데이터 사용

RFM_Segment	Recency	Frequency	Monetary	Customer_Count
322	25.07	64.56	77036622.61	13746
222	104.68	29.76	39198156.48	8463
122	426.93	12.64	32562655.21	1040
112	651.66	5.89	10240577.98	1146
212	137.57	6.63	9177086.09	705
312	27.95	6.90	8718052.29	249
321	27.23	11.12	3892859.70	652
221	125.58	10.86	3882560.66	1258
121	271.25	9.31	3802710.31	75
311	26.78	4.11	1739848.60	2297
211	139.50	3.88	1635310.08	6464
111	785.82	2.21	989615.81	14604

## (1) 이탈 판별 함수(타겟 라벨링)

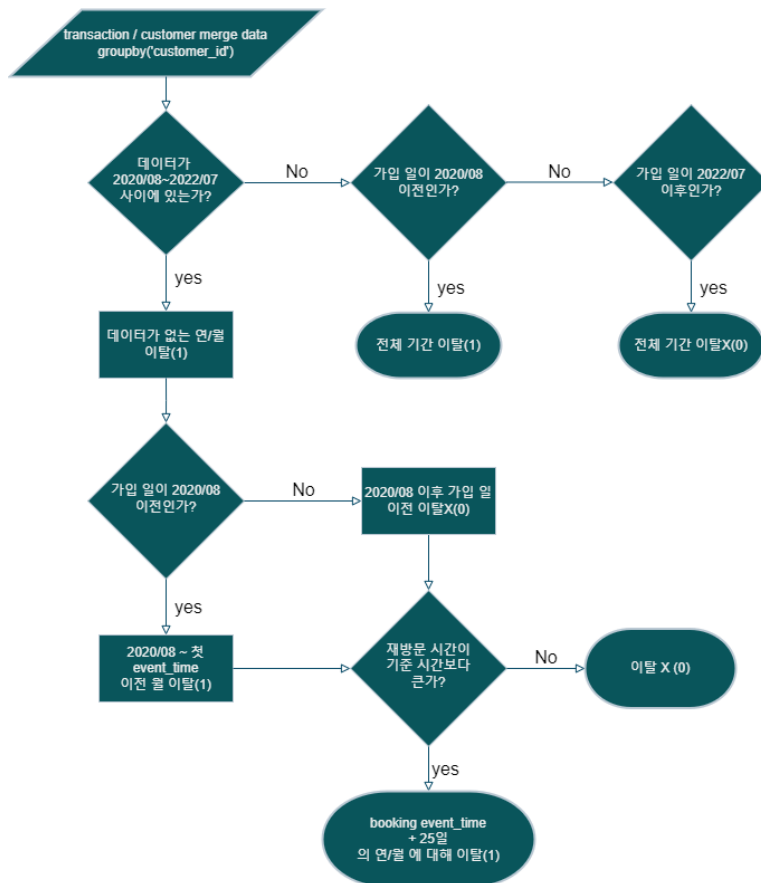
---

- 모델 학습을 위한 타겟 생성 함수
  - 마지막 구매로부터 25일 이내에 재구매가 이루어지지 않는 시점을 이탈로 판별
  - 활성 (0) / 이탈 (1)
  - 이진 분류 모델
- 

[코드 Github 링크](#) 클릭

---

## (2) 이탈 판별 함수 순서도



### 함수 적용 결과

	2020-08	2020-09	2020-10	2020-11	2020-12	2021-01	2021-02	2021-03	2021-04	2021-05	...
65539.0	1	1	1	1	1	1	1	1	1	1	...
98311.0	1	1	1	1	1	1	1	1	1	1	...
98312.0	1	1	1	1	1	1	1	1	1	1	...
98323.0	1	1	1	1	1	1	1	1	1	1	...
32789.0	1	1	1	1	1	1	1	1	1	1	...
...	...	...	...	...	...	...	...	...	...	...	...
99989.0	0	0	0	0	0	0	0	0	1	1	...
99991.0	0	0	0	0	0	0	0	1	1	1	...
99992.0	1	1	1	1	1	1	1	0	1	0	...
99995.0	0	0	0	0	0	0	0	0	0	0	...
99998.0	1	1	1	1	1	1	1	1	1	1	...



### (3) 이탈 모델 소개 및 목적

---

- 각 고객의 행동 정보를 통해 이탈의 가능성 예측 및 대처
  - 고객 이탈에 영향을 끼치는 주요한 지표를 확인하고 해당 지표를 개선
  - 1종 오류: 이탈하지 않은 고객에게 이탈 방지를 위한 액션을 제공
  - 2종 오류: 이탈한 고객에게 적절한 때에 이탈 방지를 위한 액션 취하지 못함
  - 평가 지표로 Recall(재현율)을 확인하고 Precision(정밀도)와 정확도, F1 score 등을 고려
-

## (4) 학습 데이터 소개

---

- 카테고리 : 지역, 나이, RFM
  - 수치형 : 홈페이지 방문 수, 평균 재방문 주기, 평균 구매 횟수, 프로모션 사용 비율, ATV
- 

상위가격 10% 구매 비율, 하위가격 10% 구매 비율, 평균 이벤트 지속시간

---

장바구니 포기 비율, 장바구니 - 구매 전환 기간 평균

---

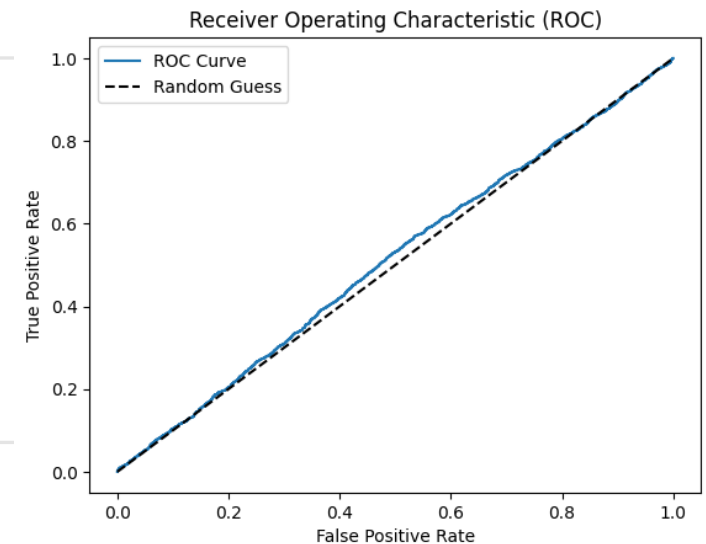
## (5) 기준모델(전체 기간)

- Logistic Regression 모델
  - 하이퍼 파라미터 (타겟 불균형 해소 `class_weight`)

- 성능

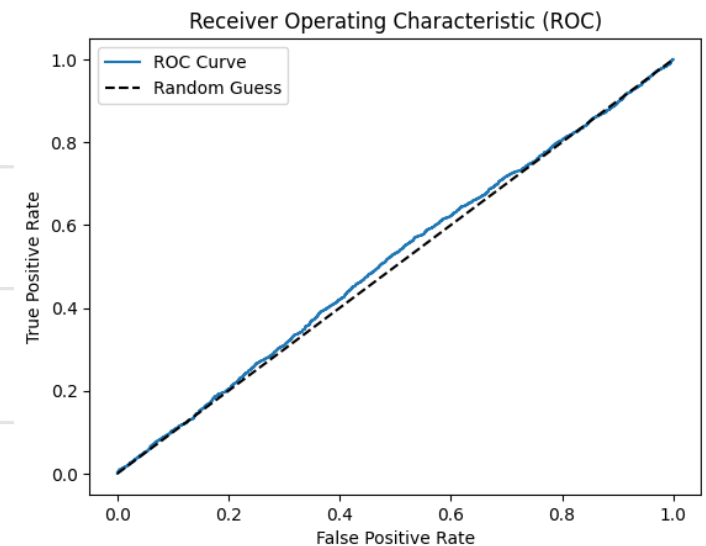
	precision	recall	f1-score	support
0	0.00	0.00	0.00	1770
1	0.83	1.00	0.90	8370
accuracy			0.83	10140
macro avg	0.41	0.50	0.45	10140
weighted avg	0.68	0.83	0.75	10140

- ROC-AUC score : 0.51



## (6) 기준 모델 해석

- 직선 ROC Curve: 모델이 랜덤으로 예측하고 있음
- Target 1 Recall : 1 / Target 0 Recall : 0
- 모든 데이터를 1로 예측하는 모델
- ROC-AUC score : 0.51



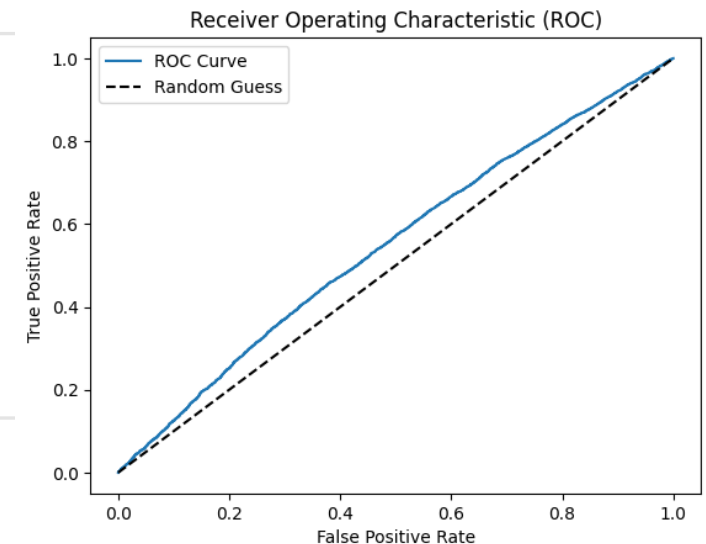
## (7) 최종모델 (2022/05 ~ 2022/07)

- Logistic Regression 모델
  - 하이퍼 파라미터 (타겟 불균형 해소 class\_weight)

- 성능

	precision	recall	f1-score	support
0	0.32	0.26	0.29	5327
1	0.74	0.79	0.76	14203
accuracy			0.64	19530
macro avg	0.53	0.53	0.53	19530
weighted avg	0.62	0.64	0.63	19530

- ROC-AUC score : 0.55



---

# 06

## 결론 및 제언

## 문제진단 결론

---

- 문제상황 : 꾸준한 매출 성장 이후 최근 2개월의 매출 하락
  - 문제진단 :
    1. 뚜렷한 목적이 없이 제공되는 프로모션
    2. 높은 장바구니 포기율
    3. 높아지는 중고 물품 수요에 대한 늦은 대응
-

## 액션플랜 결론

---

- 액션플랜 :
    1. 활성화 고객을 대상으로 하는 등급 프로모션 진행
    2. 선호 제품 추천 및 할인 혜택 제공
    3. 장바구니 알림 서비스 도입
    4. 중고 물품 거래 서비스 도입
-



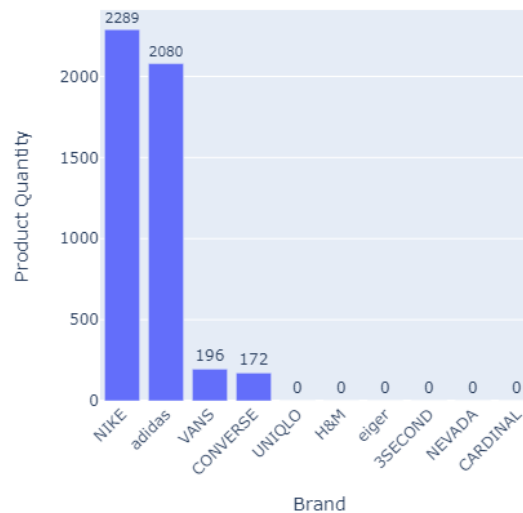
## 문제 진단 및 액션플랜 제언

### ● 고객의 선호도 고려를 통한 매출 개선



[그림 3] 인도네시아 선호 패션 브랜드  
(출처: 포퓰릭스(Populix), 2021.12)

### top 10 브랜드 물품 등록 수



총 44424 개의 물품 등록

인도네시아 선호 패션 브랜드

브랜드 상위 10개 물품 수

4737개

## 문제 진단 및 액션플랜 제언

---

- 한계점: 데이터에서 주어지지 않은 정보로 인한 해석의 어려움

---

ex)      Product.csv : year 특성, transaction.csv : 프로모션에 대한 설명

---

홈페이지 UX/UI 와 event에 대한 정보 부족

---

같은 물건 가격에 대해 5~6배 정도 차이나는 가격에 대한 설명 부족

---

> 액션플랜의 기대효과 대한 분석에 어려움

---

## 문제 진단 및 액션플랜 제언

---

- 한계점: RFM 세부 구분

---

현업에서 각 특성 별로 20% 간격으로 5개로 구분하는 것이 일반적

---

시간적인 한계로 분석하는 고객 그룹을 최소화 하기 위해 3/2/2 개로 구분

---

> 각 특성별 그룹의 특징이 뚜렷하게 구별되지 않아 분석에 어려움

## 이탈 판별 모델 한계점

---

- 모델 학습 데이터 구축

---

그룹별(322/222/122) 전체 기간 데이터에서 **평균**이 차이가 나는 지표 선택

---

이탈 기간과 활성 기간의 특징이 평균내는 과정에서 **상쇄**

---

> 이탈에 영향을 미치는 주요한 특성 분석에 대한 신뢰도 낮음

---

- 문제 정의를 위한 분석 과정이 길어져 모델 실험 **시간 부족**

---

## 이탈 판별 모델 개선

---

- 그룹별 **월간** 데이터에서 차이가 나는 지표를 선정, **월별 이탈 여부**와 결합한 학습 데이터 생성
  - 모델의 **신뢰도를 확보**한 후 해석을 통해 지표에 대한 유의미한 개선방향을 제시
-

패션이커머스 문제진단 프로젝트

---

**경청해주셔서 감사합니다 !**

team DA

임나래 유하민 이현준 이재은

---

## 프로젝트 내 용어 설명

---

- 현재(시점) : 2022년 08월 01일
  - RFM 분석 : 고객의 구매 기록을 최신성, 구매빈도, 지출금액으로 나누어 고객의 가치를 평가하는 지표
  - 322그룹 : 최근성, 구매빈도, 지출금액 모두 상위권으로, 충성도 높은 가치있는 그룹
  - 222그룹 : 최근성은 중간등급이지만, 구매빈도, 지출금액이 상위권으로, 성장 가능성 높은 그룹
  - 122그룹 : 상대적으로 최근성이 낮지만, 구매빈도와 지출금액은 높은 그룹
  - ATV (Average Transaction Value) : 각 거래 당 발생하는 평균 매출 금액 (여러개의 상품 동시 구매 경우도 고려)
  - 객단가 : 고객 당 평균 구매 금액 (한 번의 거래에서 고객의 평균 지출 금액)
  - 구매 이탈 판별 기준 : 25일 이내 재구매 하지 않으면 구매 이탈
-