

PORTFOLIO

비즈니스를 읽는 데이터 분석가 이성우

이성우

<https://github.com/seongulalala>

010-5491-7781

이성우

데이터 분석가



Birth. 1998.07.10
E-mail. iceberg2018@naver.com
Tel. 010-5491-7781
Web. github.com/seongulalala/

나를 한 마디로
표현한다면?

경영학적 사고와 데이터 분석 능력을 모두 갖춘 하이브리드형 분석가입니다

학력사항

2017.02	영일고등학교 졸업
2025.02	상명대학교 경영학부 졸업
2025.02	상명대학교 빅데이터애널리 틱스 졸업

자격사항

데이터분석준전문가(ADsP)	2024.03
빅데이터 분석기사	2024.07
SQL개발자(SQLD)	2024.12
TOEIC 820점	2025.04
컴퓨터활용능력 1급	2023.11

보유스킬

Tableau	★★★★★
Power BI	★★★
Looker Studio	★★★
Python	★★★★
MYSQL	★★★★
EXCEL	★★★★★

협업 툴

GIT	★★★★
Notion	★★★★
Slack	★★★★★

CONTENTS

여행플랫폼 '트리플' 국내 여행지 추천 서비스 도입 제안 03

트리플의 AI추천 서비스 개선방안 제시
이를 위한 프로토타입 추천 서비스 개발

클래스플래닛 사용자 구독 여정 개선 프로젝트 04

가상의 데이터로 퍼널분석과 ab테스트,
코호트 분석을 통한 전략 제시

Blinkit 수익성 중심 마케팅 전략 대시보드 05

Blinkit 데이터를 기반으로 수익성 중심 전략 대시
보드 생성

Sakila Never Die - Sakila 구독형 OTT 서비스 전환 아이디어 06

비즈니스 모델 변경의 제안과 이를 위한 고객 세분화

Olist의 재구매율 향상을 위한 아이디어 제안 07

Olist의 병목구간 식별 및 요인 분석과 전략 제시

여행플랫폼 '트리플' 국내 여행지 추천 서비스 도입 제안

위기 속에서도 포기하지 않고 프로젝트를 완수하며, 프로젝트 전 과정을 주도하였습니다

사용 기술 분석 기법: 군집분석(K-MEANS), 회귀분석(LINEAR REGRESSION, RANDOM FOREST ,ENSEMBLE) 데이터 처리: PYTHON ERR생성 : DBDIAGRAM.IO 프로토타입: STREAMLIT	담당 역할 데이터 분석 GITHURB주소 HTTPS://GITHUB.COM/ALEXKWJ/PROJECT_TRIPLE 기여도 40% / 5명
---	--

프로젝트 배경
여행 산업에서는 단순 인기 여행지 추천을 넘어, 사용자의 취향과 데이터를 반영한 AI 기반 맞춤형 추천 서비스 수요가 증가하고 있습니다. 이를 반영하여 AIHURB에서 여행 데이터를 수집 분석하고 트리플의 AI추천 서비스를 개선한 UX*UI와 프로토타입을 구현하고자 하였습니다.

프로젝트 기간: 2025.08.25~2025.09.01

직면한 어려움
팀 리더의 갑작스러운 이탈로 전략 혼란 발생
데이터셋 내 결측치·잡음 등 불완전한 데이터 환경
제한된 기간 내 데이터 분석부터 모델·프로토타입 구현까지 완수해야 하는 압박

여행플랫폼 '트리플' 국내 여행지 추천 서비스 도입 제안

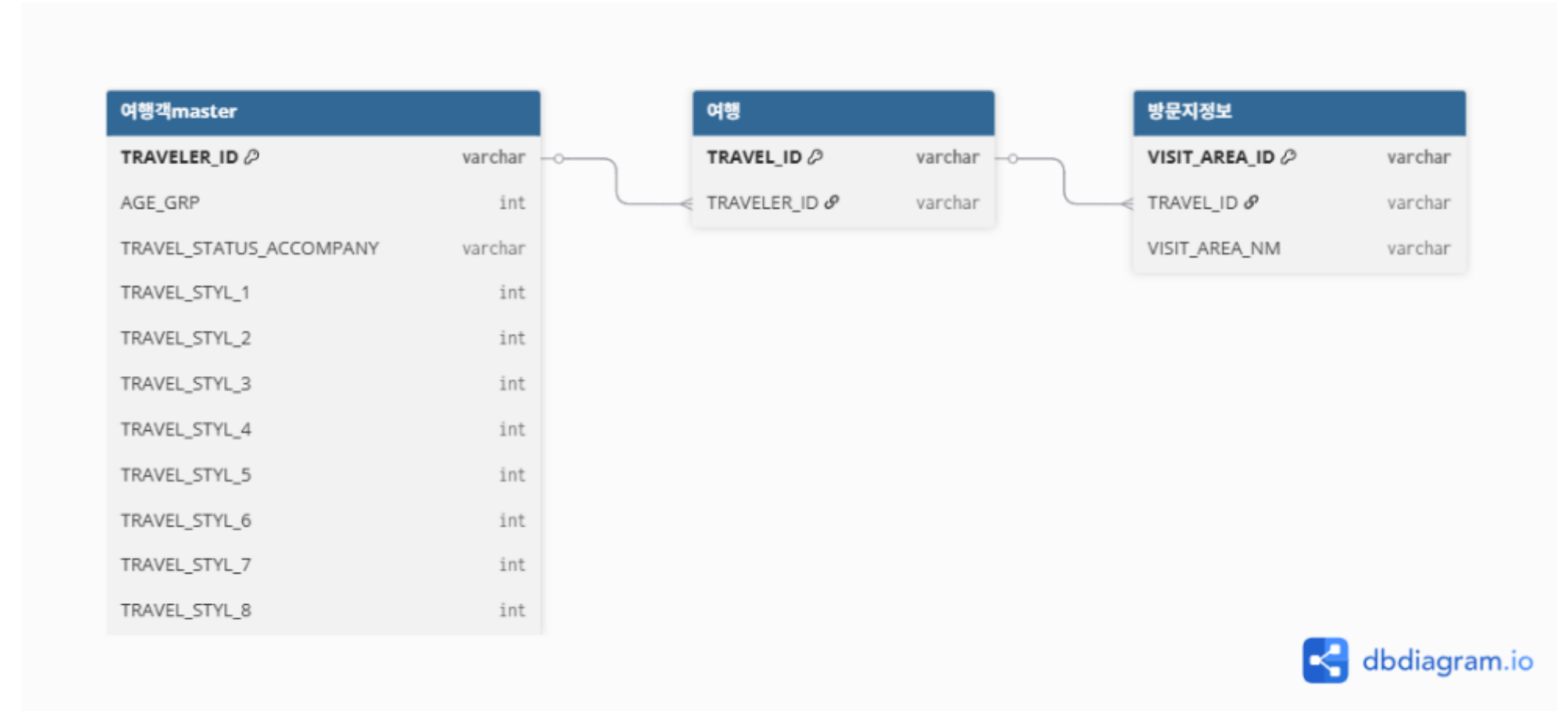
프로젝트 세부 내용

프로젝트 초반, **팀 리더의 갑작스러운 이탈과 불완전한 데이터셋**으로 프로젝트 방향이 크게 흔들렸고, 데이터는 결측치와 잡음이 많아 신뢰성이 떨어졌습니다.
여행 방문지가 “친구 집, 집”과 같이 의미 없는 값으로 기록되거나 지역명이 행안부 코드(예: 서울=11)로만 제공되어 활용하기 어려웠습니다.
게다가 제한된 기간 안에 분석부터 모델 구축, 프로토타입 개발까지 완성해야 했기 때문에, 압박감이 있었습니다.

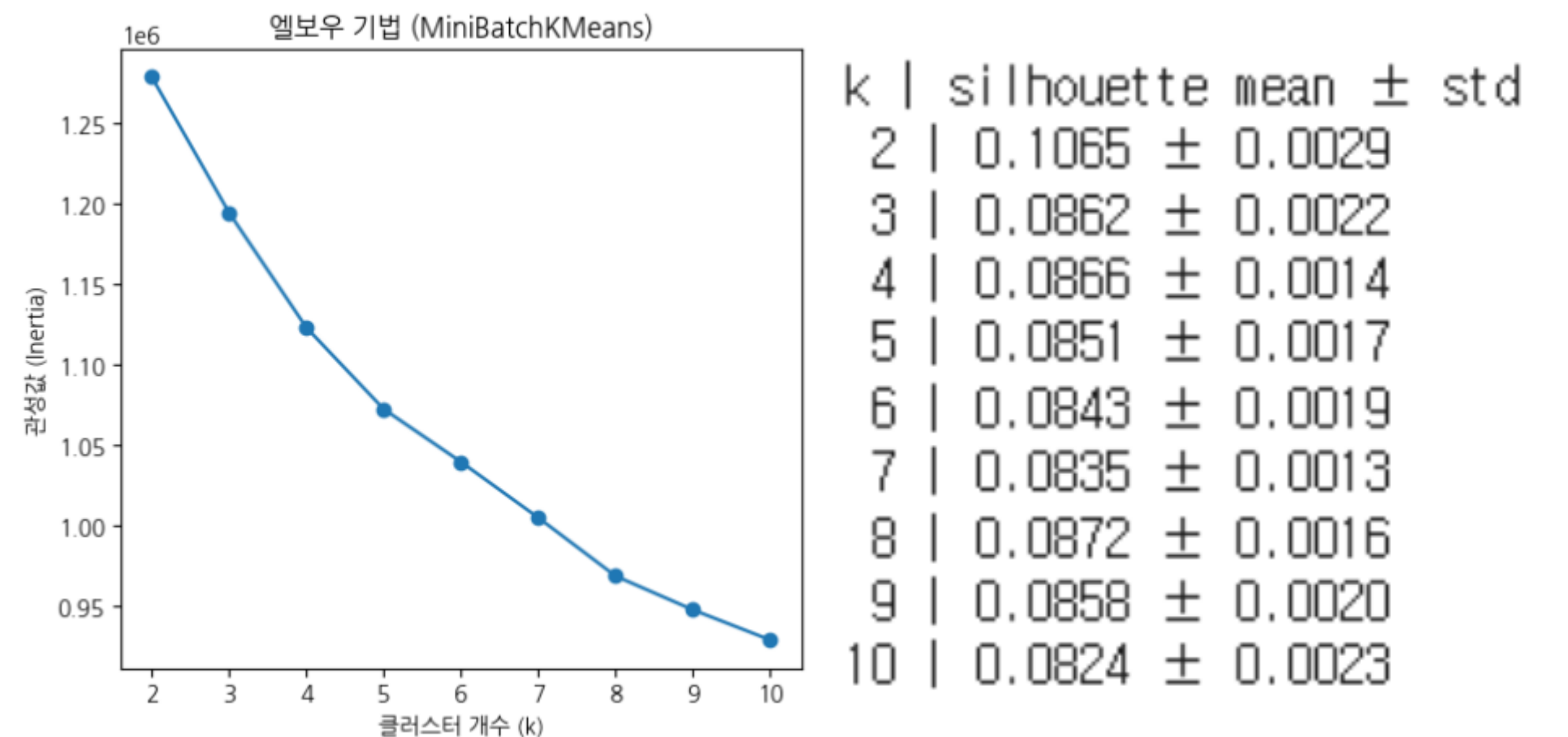
저는 이러한 상황을 극복하기 위해 데이터 구조를 재정비하는 데 집중했습니다.
ERD를 설계한 뒤, AIHUB 데이터를 수집·병합하여 테이블 간 관계를 명확히 정의했습니다.
불필요한 데이터를 제거하고 코드값을 사람이 이해할 수 있는 지역명으로 매핑했으며,
비즈니스 인사이트와 컴퓨팅 사고를 함께 고려하여 군집의 개수와 피처를 선정하였고
이후 **파이프라인을 구축**하고 **스텐다드 스케일링**, **원핫 인코딩**을 통해 모델 학습에 적합한 정제 데이터셋을 구축했습니다.

정제된 데이터를 기반으로 **K-Means 군집화** 모델을 적용해 고객을 유형별로 세분화했고, **각 군집의 여행 패턴을 분석**하여 맞춤형 추천이 가능한 기반을 마련했습니다.
마지막으로, 분석 결과를 활용해 **Streamlit 기반 프로토타입을 제작**하여, 사용자가 입력한 정보를 바탕으로 즉시 맞춤형 여행지를 추천할 수 있는 기능을 구현했습니다.

ERD



컴퓨팅 사고와 비즈니스 인사이트를 모두 고려하여 k = 4로 설정



클래스플래닛 사용자 구독 여정 개선 프로젝트

시계열 분석, 퍼널 분석, 감정분석, A/B·코호트 등 다양한 분석 기법을 활용한 경험을 보유하고 있습니다.

사용 기술

- 분석 기법: 퍼널분석 , AB테스트 , 코호트분석
- 데이터 처리: PANDAS, SCIKIT-LEARN
- 시각화: MATPLOTLIB, SEABORN

담당 역할

- 데이터 분석

GITHURB주소

- [HTTPS://GITHUB.COM/ALEXKWJ/PROJECT_LOG_ANALYSIS](https://github.com/ALEXKWJ/PROJECT_LOG_ANALYSIS)

기여도

- 25%/5명

프로젝트 목표

가상의 기업 로그 데이터를 활용한 자율 프로젝트로, 저희 팀은 퍼널 분석, A/B 테스트, 페르소나 설정, 사용자 여정 분석 등을 통해 더 나은 서비스 전략을 제시하고자 했습니다.

프로젝트 기간 : 2025.08.08 ~ 2025.08.14

직면한 어려움

프로젝트 초기에 사용자 데이터를 고려하지 않고, 기획팀 의도에만 맞춰 페르소나(PERSONA) 설정과 퍼널 구간 정의를 진행 → 초기 분석 방향의 오류 발생
기획 의도와 실제 데이터가 맞지 않아 분석 효율성이 떨어지고 시간이 지연됨
목표 수정 과정에서 분석 시간이 단축되는 제약이 발생

여행플랫폼 '트리플' 국내 여행지 추천 서비스 도입 제안

프로젝트 세부 내용

빠른 분석으로 불가능을 가능으로

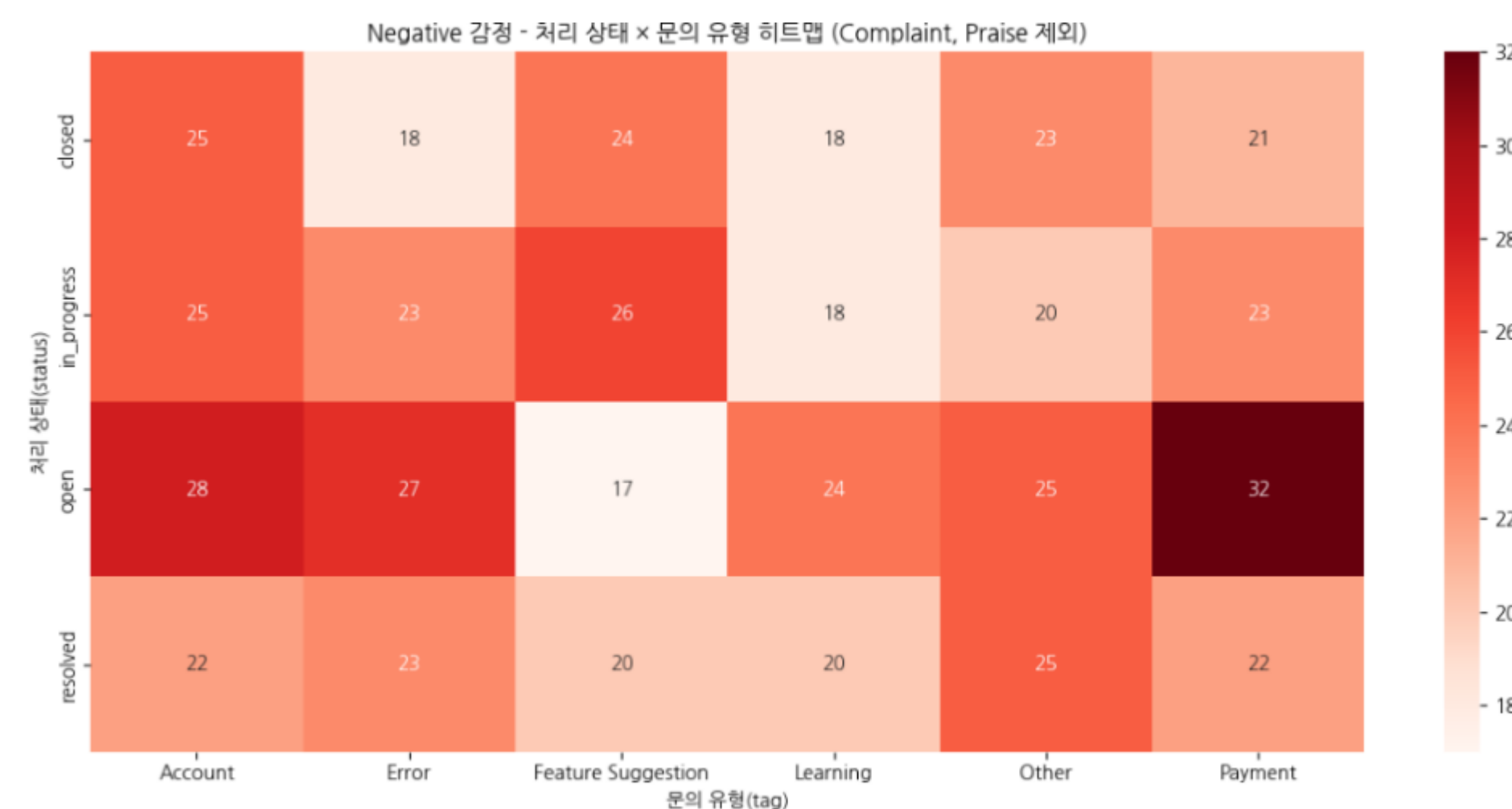
초기 데이터를 충분히 고려하지 않고 설정한 목표로 인해 분석 방향이 잘못 되었고, 이를 수정하느라 **시간이 촉박해지는 어려움**이 있었습니다.

재정의된 목적(유료 전환율 개선)에 따라 **3차례의 A/B 테스트를 진행**했으나, 데이터의 한계(데이터의 수 부족)로 유의미한 차이를 확인하기는 어려웠습니다.

이에 **고객 관점에 다시 집중**하여, 부정적인 문의 데이터를 유형별·처리 상태별로 **히트맵 시각화**하여 전략 도출에 활용했습니다.

또한 **코호트 분석**을 통해 구독 전환율이 급격히 감소하는 구간을 식별하고, 이를 바탕으로 새로운 **개선 전략을 제안**했습니다.

부정적 감정을 지닌 문의 내용 횟수 히트맵
(x=문의 유형, y = 처리 상태, 기간 = 2025.06월)



여행플랫폼 '트리플' 국내 여행지 추천 서비스 도입 제안

프로젝트 세부 내용

문제 해결의 열쇠는 결국 데이터입니다

성과

총 3번의 A/B 테스트는 모두 유의미한 차이를 보이지 않았음
그러나 코호트 분석을 통해 프로모션 푸시 알림의 최적 발송 시점을 도출
최종적으로, 서비스 전략 도출이라는 성과를 달성

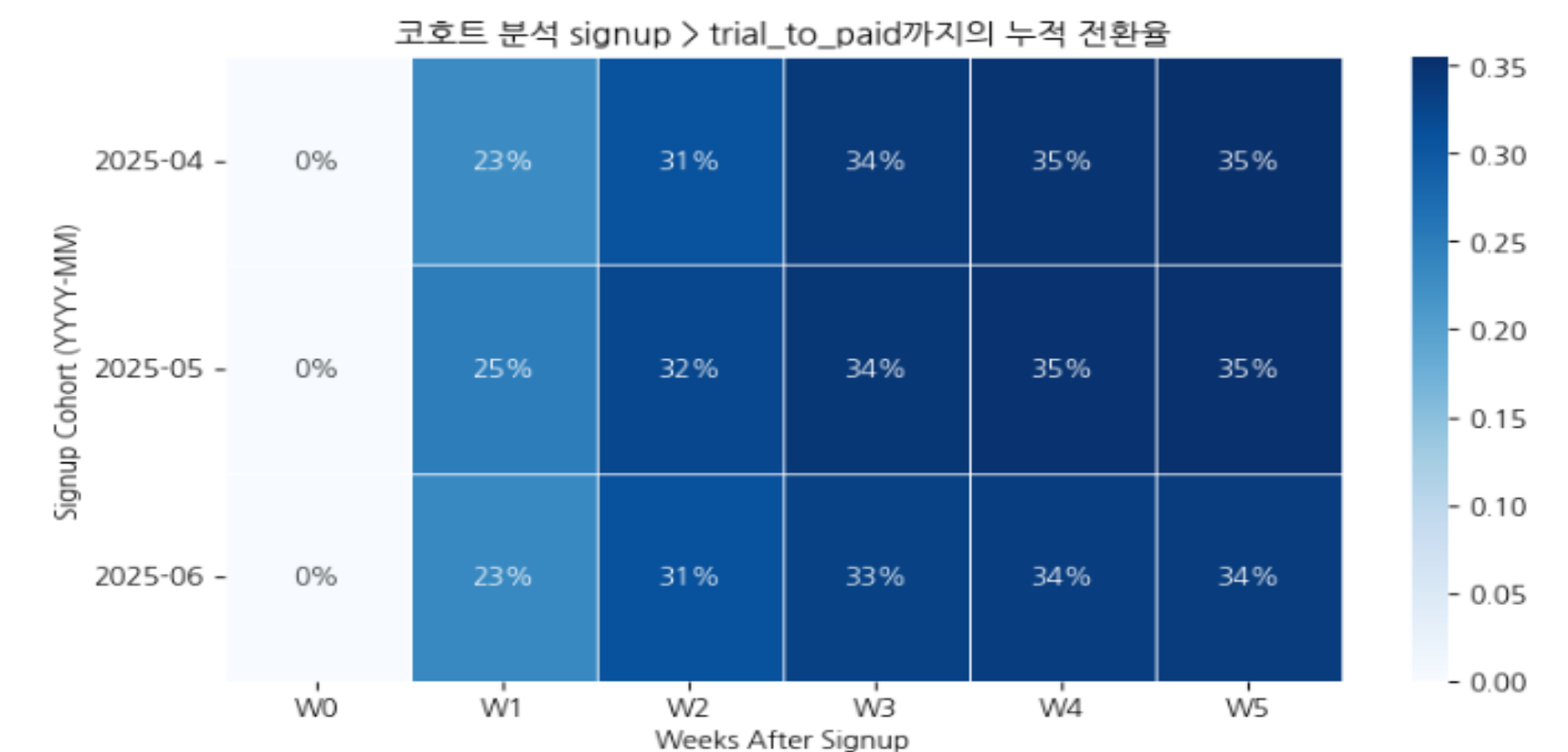
발휘한 역량

데이터 탐색 역량: 제한된 시간 안에 핵심 분석을 빠르게 수행
문제 해결력: A/B 테스트 기각 후, 코호트 분석으로 전환하여 대안을 제시
분석 도구 활용 역량: 퍼널·감정 분석, 시계열 그래프 등 다양한 기법 활용

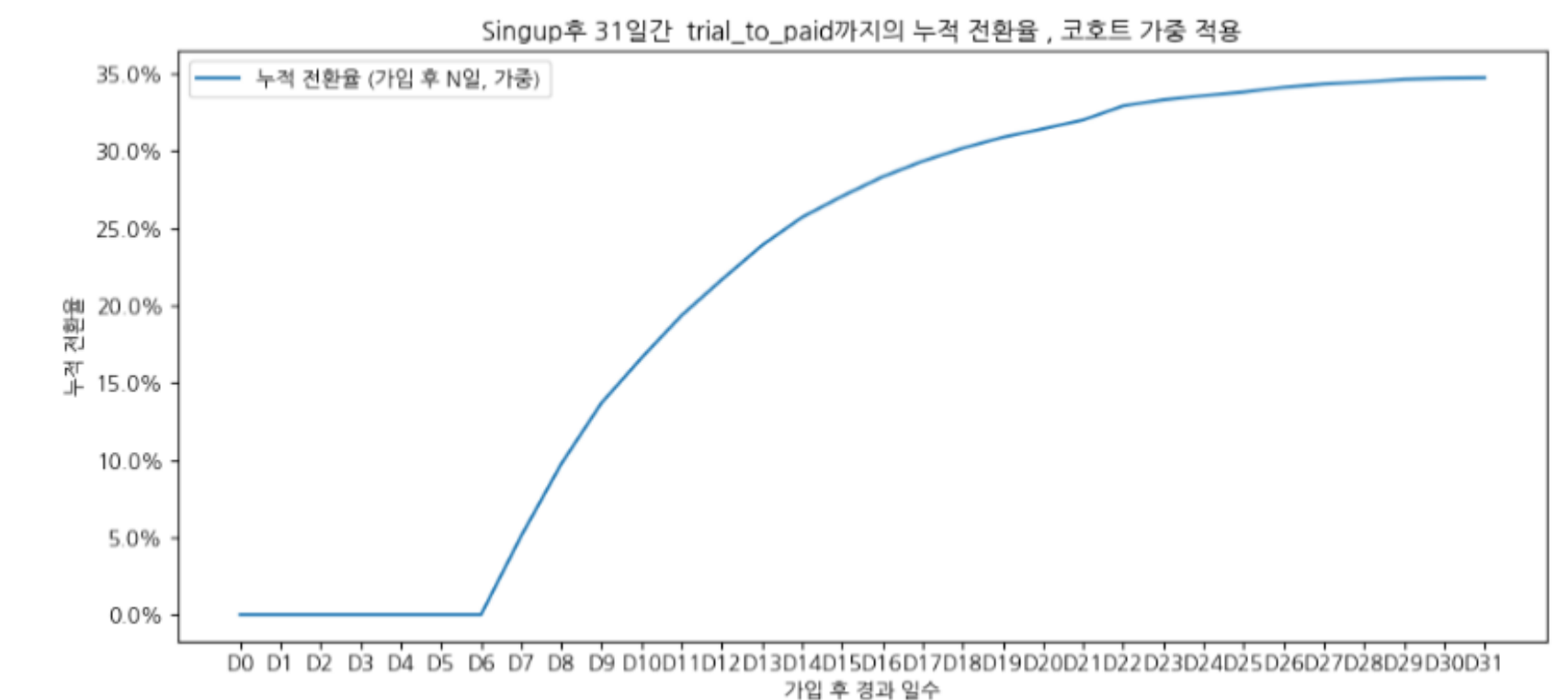
배운 점

기획 방향도 중요하지만, 데이터를 충분히 고려하는 것이 분석의 출발점임을 깨달았습니다.
데이터의 수와 질이 분석 결과의 신뢰성과 활용 가치를 결정한다는 점을 직접 확인했습니다.
여러 방법을 유연하게 적용해 대안을 찾는 태도가 필요하다는 점을 배웠습니다.

유료 전환율에 대한 코호트 분석(x=주,y=월)



코호트 분석 시계열 시각화(x=일,y=전환율)



BLINKIT 수익성 중심 마케팅 전략 대시보드

MYSQL 기반 데이터 가공과 시각화 도구 활용 역량으로 경영진의 의사결정을 지원하겠습니다.

사용 기술

- 분석 기법 : CTR, CVR ,시계열 추이 분석
- 데이터 처리: MYSQL , PYTHON
- 시각화: LOOKER STUDIO
- ERR생성 : DBDIAGRAM.IO

담당 역할

- 데이터 분석

GITHURB주소

- [HTTPS://GITHUB.COM/YHJ2898/PJT3_BLINKIT](https://github.com/YHJ2898/PJT3_BLINKIT)

기여도

- 25% / 5명

프로젝트 배경

인도의 쿼커머스 기업 BLINKIT 데이터를 활용해, 수익성 중심의 마케팅 전략 대시보드를 구축하기로 하였습니다
대시보드는 총 3페이지(성과·고객·상품)로 구성되었으며, 매출액·총수익·마진율을 핵심 지표로 삼아,
이를 기준으로 성과를 측정하고 비교할 수 있는 기준을 마련했습니다.

프로젝트 기간 : 2025.07.25 ~ 2025.07.30

직면한 어려움

대시보드 제작을 팀원별로 분업한 결과, 데이터 스키마 불일치 문제 발생
3페이지의 결과물을 통합하는 과정에서 테이블 구조와 집계 단위가 달라 병합이 불가능한 상황 발생

BLINKIT 수익성 중심 마케팅 전략 대시보드

프로젝트 세부 내용

데이터 표준화를 통한 문제 해결

저는 MySQL로 원천 데이터를 추출하고, Pandas를 활용해 전처리 및 집계를 진행했습니다. 그러나 대시보드 **통합 과정에서 스키마 불일치 문제가 발생되었고**, 이는 최종 산출물의 신뢰성을 크게 저해할 수 있는 위험 요소였습니다.

이를 해결하기 위해 즉시 팀원들과 협의하여 데이터 표준 정의(테이블 **구조·컬럼명·집계 단위**)를 수립했습니다.

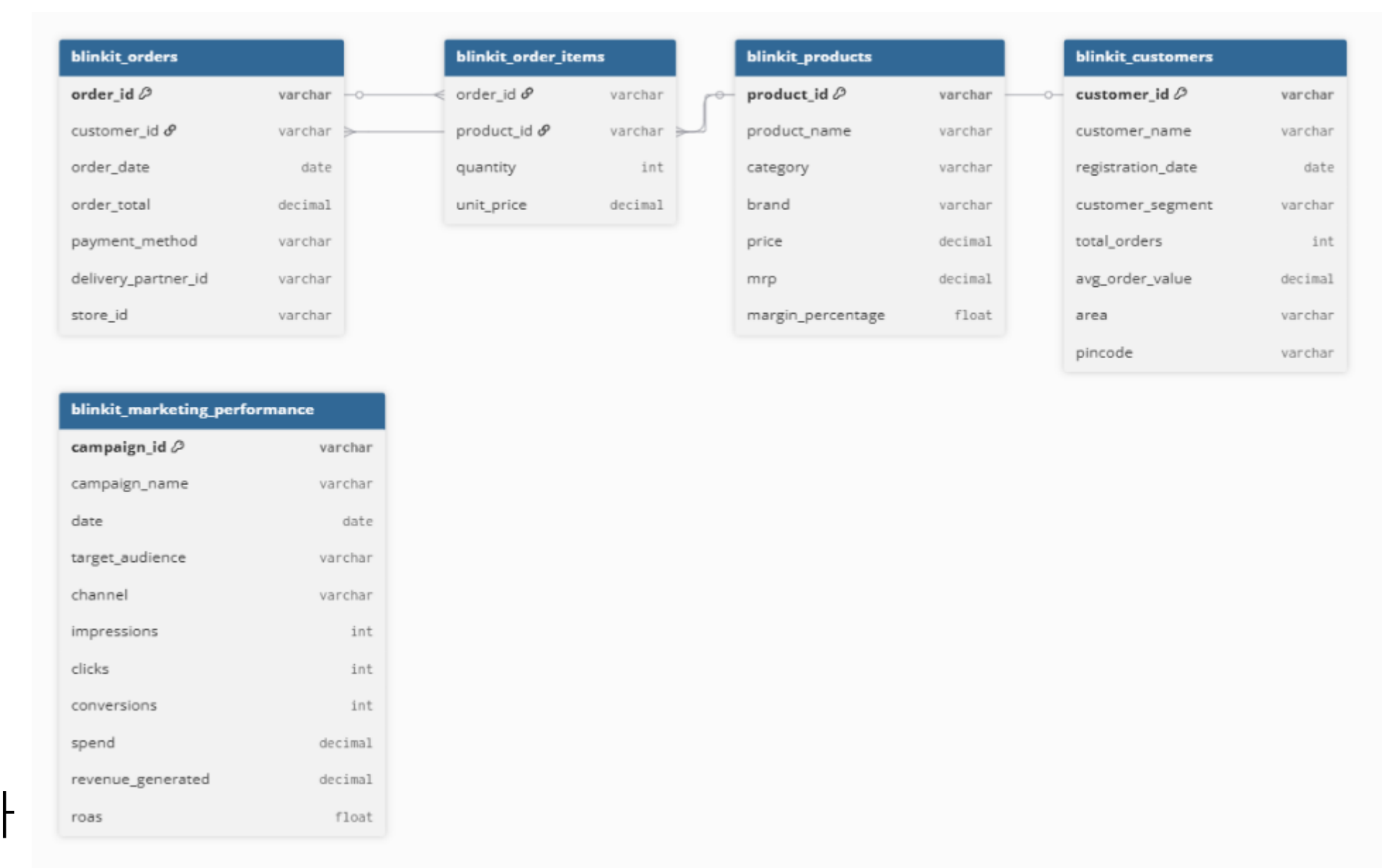
제가 직접 병합 규칙을 설계하고 통일 작업을 주도하여, **세부 대시보드 간의** 스키마 차이를 조정했습니다.

제가 **담당한 상품** 대시보드에서는 카테고리·브랜드별로 총매출·총수익·마진율 KPI를 집계하고 시각화했습니다.

대시보드는 Z-기법 레이아웃을 적용하여, **사용자가 좌측 상단에서** 우측 하단으로 자연스럽게 흐름을 따라가며 매출·수익·마진율을 비교할 수 있도록 구성했습니다.

이를 통해 경영진이 카테고리/브랜드별 수익성과 마진 기여도를 직관적으로 비교할 수 있도록 했습니다.

스키마 통합을 위해 생성한 ERD



BLINKIT 수익성 중심 마케팅 전략 대시보드

프로젝트 세부 내용

핵심 지표를 기준으로, 대상을 고려한 시각화를 완성

성과

KPI 체계(총매출·총수익·마진율) 확립

경영진 의사결정용 3페이지 대시보드를 완성하여, 수익성 중심 예산 재배분 근거를 제공

발휘한 역량

SQL 활용 역량: **MySQL** 기반 데이터 추출 및 집계

데이터 처리 역량: Pandas를 활용한 데이터 전처리 및 스키마 통일

시각화 역량: KPI를 중심으로 **Looker Studio**를 사용해 대쉬보드 제작

협업·조율 역량: 팀원들과 데이터 표준을 정의하고, 산출물 병합을 주도

배운 점

데이터 시각화는 단순 그래프 제작이 아니라, 의사결정자가 무엇을 보고 싶어 하는지를 고려한 **KPI 설계가 핵심**임을 배웠습니다

협업 프로젝트에서는 **데이터 표준화와 스키마 정의가 성공적인 결과물을 위한 필수 조건**임을 실감했습니다.

내가 담당한 대쉬보드 구성



SAKILA NEVER DIE - SAKILA 구독형 OTT 서비스 전환 아이디어

고객 세분화와 지출 분석을 통해 비즈니스 인사이트와 전략 제시 역량을 강화하였습니다.

사용 기술

- 분석 기법: RFM분석, 군집분석(K-MEANS), ERD
- 데이터 처리: PANDAS, SCIKIT-LEARN
- 시각화: MATPLOTLIB, SEABORN

담당 역할

- 데이터 분석

GITHURB주소

- [HTTPS://GITHUB.COM/ZOEYKIM77/SAKILA_2T](https://github.com/ZOEYKIM77/SAKILA_2T)

기여도

- 30% / 5명

프로젝트 배경

SAKILA 데이터베이스를 기반으로, 지속 가능한 발전을 위해 SAKILA의 비즈니스 모델을 OTT 구독 모델로 전환하고자 하였습니다. 목표는 데이터를 분석하여 구독제 가격 체계와 포지셔닝 전략을 수립하는 것이었습니다.

프로젝트 기간 : 2025.07.17~2025.07.22

직면한 어려움

장르·배우 인기도가 데이터 전반에서 균일하게 분포 → 고객군 간 차이가 드러나지 않는 문제
제한된 고객 수(600명)로 세분화가 어렵고, 차별화 요인을 찾기 위한 추가 지표 설계 필요

SAKILA NEVER DIE - SAKILA 구독형 OTT 서비스 전환 아이디어

프로젝트 세부 내용

세분화를 통해 타겟팅할 고객 식별

저는 데이터 분석 담당으로, 비즈니스 모델 전환 시 발생할 고객 이탈 가능성을 고려해, 이탈 위험군을 명확히 타겟팅할 수 있도록 고객 세분화 작업을 진행하였고 요금제 근거 도출을 맡았습니다.

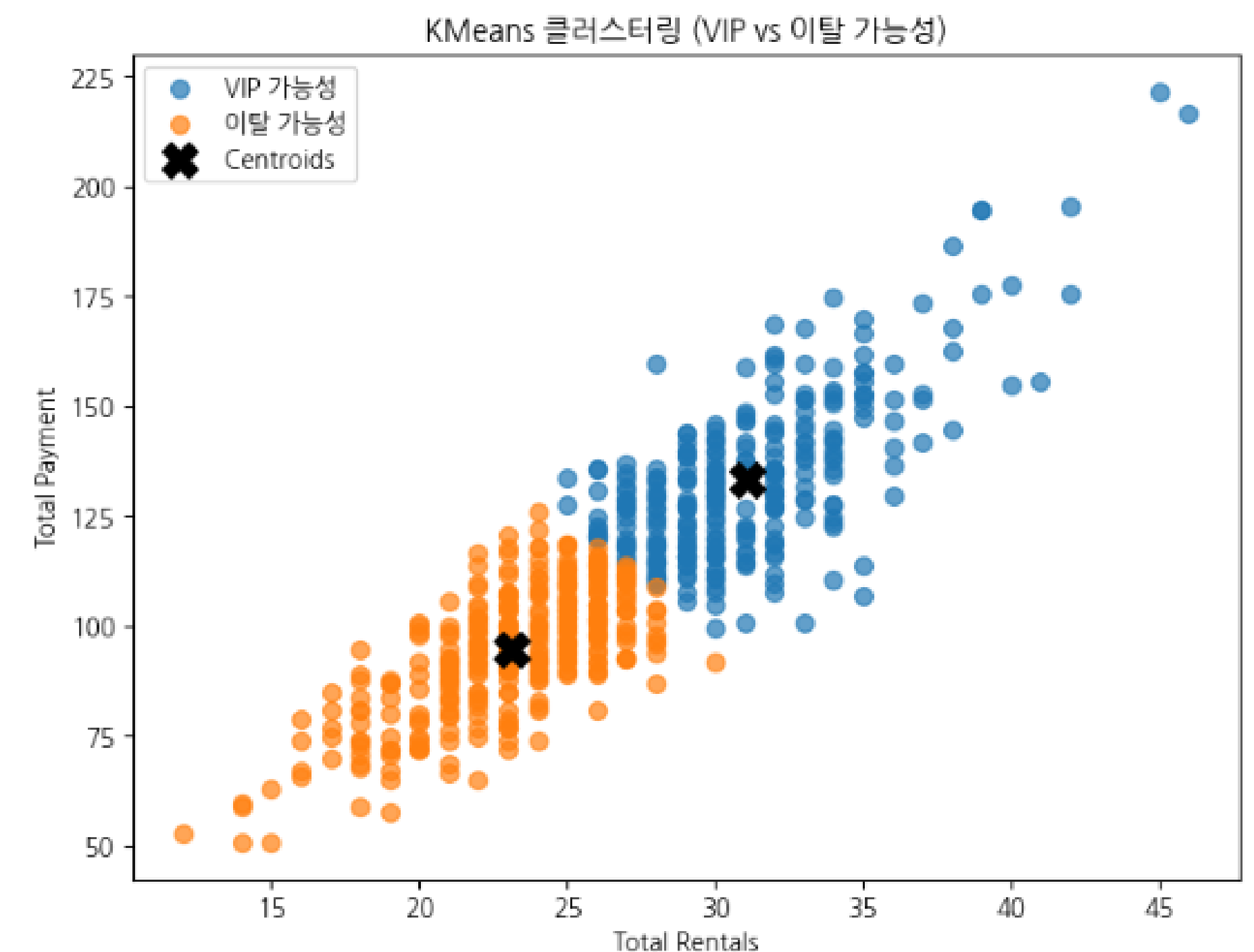
Sakila DB의 데이터를 Mysql을 통해 고객·대여·결제 정보를 추출하고 Pandas를 통해 고객 세분화를 진행했습니다

RFM 분석을 적용했으나, 데이터 기간이 1년 미만으로 짧아 R의 값이 유의미하지 않아 한계가 있었기 때문에 R(최근 대여 일수)를 제외하고 F(총 대여 횟수)와 M(총 지불액)으로 K-Means 군집화를 진행해 고객 세분화를 다시 시도하였습니다.

실루엣 기법과 비즈니스 인사이트를 함께 고려하여 최적의 군집 수를 k=2로 설정하였고

고객별 월평균 지출액을 산출하여, 구독제 가격 설정의 근거 마련하였고 군집별 소비 패턴과 충성도를 분석해, 맞춤 전략의 기반 확보했습니다

군집분석 시각화(k=2, x= 총 대여횟수 , y=총 지불액)



OLIST의 재구매율 향상을 위한 아이디어 제안

방대한 데이터 속에서도 분석 목적과 방향성을 일관되게 유지하는 태도를 지켜가겠습니다.

사용 기술

- 분석 기법: 재구매율 지표 정의 및 산출, 고객 세분화, 시계열 추이 분석
- 데이터 처리: PANDAS, SCIKIT-LEARN
- 시각화: MATPLOTLIB, SEABORN
- 프로젝트 기간 : 2025.07.02~2025.07.07

담당 역할

- 데이터 분석

GITHURB주소

- [HTTPS://GITHUB.COM/ALEXKWJ/PROJECT_LOG_ANALYSIS](https://github.com/AlexKWJ/project_log_analysis)

기여도

- 50% / 5명

프로젝트 배경

브라질 이커머스 플랫폼 OLIST의 공개 데이터를 분석하여, 낮은 재구매율(약 3%) 문제를 해결하기 위한 전략을 도출하고자 했습니다.

내 액션

저는 먼저 OLIST의 고객, 주문, 결제, 상품 데이터셋을 PANDAS로 불러와 병합하고, 고객 단위로 데이터를 집계했습니다.

퍼널분석 결과 재구매율이 낮다는 병목현상을 파악하였고 이를 해결하기 위해 CUSTOMER_UNIQUE_ID 기준으로

재구매 고객과 미재구매 고객을 분리하여,

배송·결제·상품/가격·구매 시점 네 가지 요인이 재구매를 설명할 수 있는지 가설을 세웠습니다.

DELIVERED 상태의 주문만 필터링한 뒤, 배송 지연 여부·결제 방식·카테고리별 성향·첫 구매 후 30일 내 재구매 여부를 중심으로 교차 분석을 수행했습니다.

결과 도출

배송 지연은 재구매와 뚜렷한 상관관계를 보이지 않음

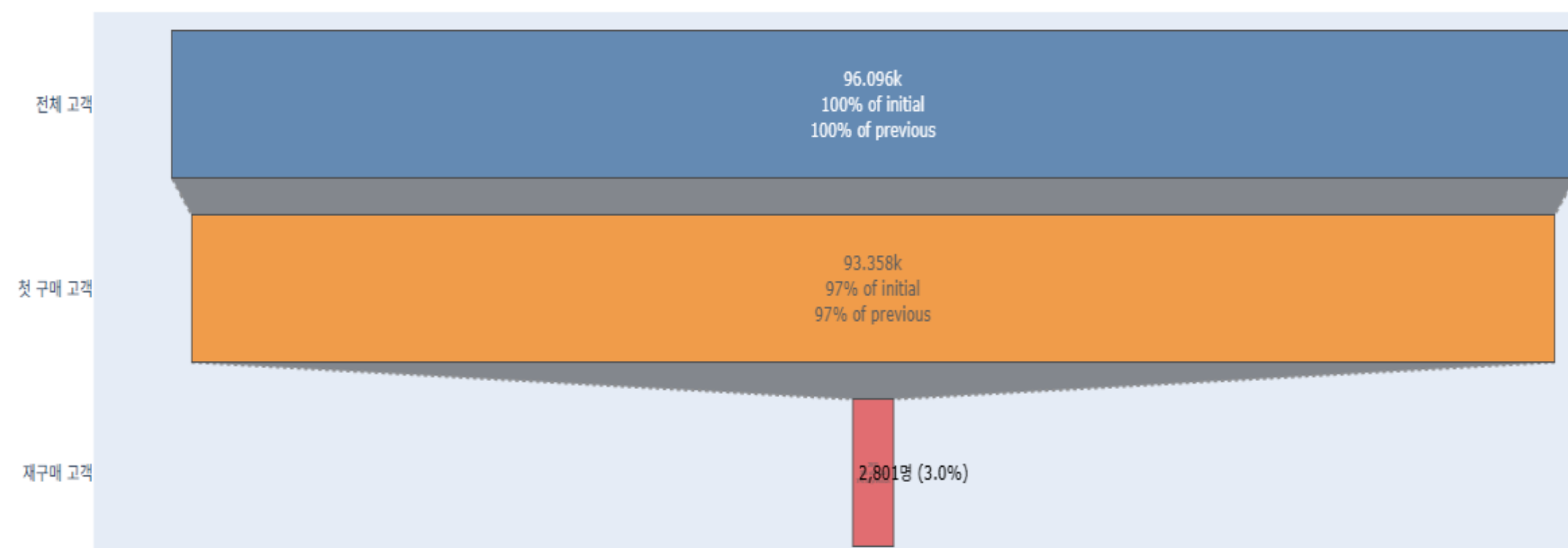
저가 카테고리에서 재구매율이 상대적으로 높음

재구매는 첫 구매 후 2-4주에 집중되는 패턴을 확인

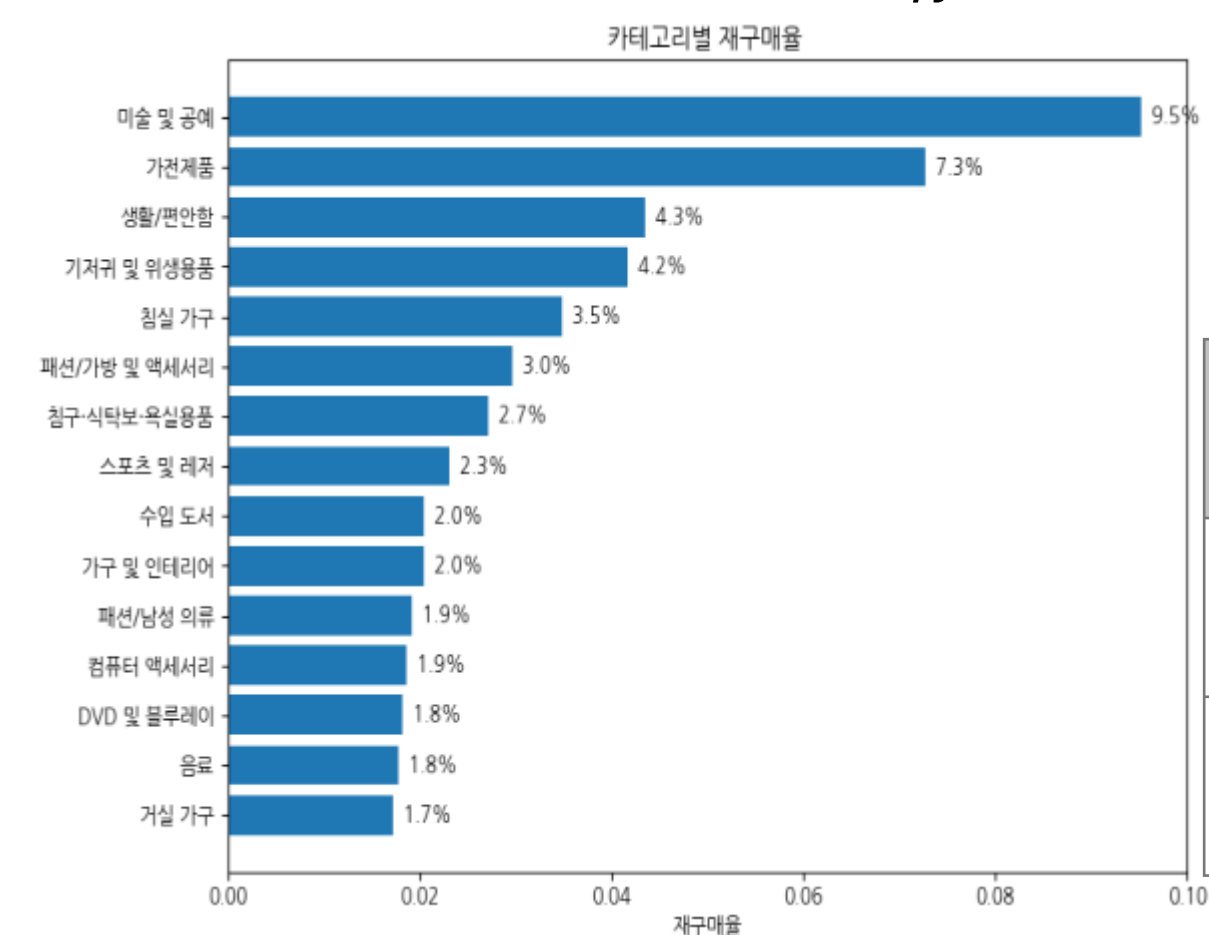
OLIST의 재구매율 향상을 위한 아이디어 제안

프로젝트 세부 내용

퍼널 분석 시각화(전체 고객 > 첫 구매 > 재구매(3%))



카테고리별 재구매율(x=재구매율, y=카테고리 내림차순)



카테고리별 평균 가격	카테고리 가 격	전체 가격
가전제품	103BRL(헤 알)	120BRL(헤 알)
미술 및 공예	75BRL(헤알)	120BRL(헤 알)

성과

- 재구매 집중 구간과 카테고리 특성을 근거로, 초기 2-4주 집중 마케팅 전략과 저가 카테고리 타겟팅 방안을 제안
- 바우처·구독형 모델 등 실질적인 재구매 촉진 전략을 도출

발휘한 역량

데이터 처리 역량:

Pandas를 활용해 고객·주문·결제 데이터를 병합, 재구매율 산출을 위한 전처리 수행

분석 설계 역량:

배송, 결제 방식, 상품 카테고리, 구매 시점 등 **다양한 요인을 비교하여 가설을 검증**

패턴 인식 역량:

재구매가 특정 시점(2-4주)에 집중되는 패턴을 발견하고 전략으로 연결

문제 해결 역량:

단순 데이터 분석에 그치지 않고, 실제 비즈니스에서 적용 가능한 전략으로 해석

배운 점

- 첫 프로젝트에서 분석에만 치중해 기획 관점을 놓친 경험을 통해, **분석은 반드시 비즈니스 인사이트와 연결되어야 의미가 있다는 점**을 배웠습니다.
- 방대한 데이터 속에서도 목적과 방향을 잃지 않고 분석을 설계하는 태도가 중요함을 실감했습니다.