

MIDDLE PROJECT

공유 오피스 결제 전환율 향상을 위한 데이터 분석 및 전략 제안



시계열 예측과 앙상블 모델을 활용한 최적 전략 도출

1팀 박준혁 이은지 장은희 조재윤

01

분석 배경

공유 오피스 동향
분석 주제

02

EDA

데이터 셋
1차 전처리
가설 정의
현황 분석

03

예측 및 세부 분석

시계열 예측 분석
앙상블 모델을 활용한 예측 분석
주요 변수 세부 분석

04

제안

신규 요금제 기획
무료체험 타당성 검토

05

제언 및 시사점

시사점
한계점
추가 분석 제안

분석 배경

01

공유 오피스 동향

분석 주제

시장 성장과 위축

코로나19 팬데믹 이후 공유 오피스 시장은 급성장했지만, 스타트업 투자 위축, 위워크 파산 신청 등의 영향으로 출점이 지속적으로 감소하고 있다.

현재 서울 내 주요 공유 오피스 기업으로는 패스트파이브, 스파크플러스, 위워크가 있으며,
이중 패스트파이브는 매출에서 우위를 차지하고 있음에도 불구하고, 2018년 이후 6년째 적자가 지속되고 있다.

비즈한국 (2024-03-08)

**“위워크 파산 4개월,
토종 공유 오피스는 어떻게 살아남았나”**

글로벌 최대 규모 공유오피스 기업 위워크가 코로나19 여파를 이기지 못하고
지난해 파산을 선언하면서 국내 공유오피스 기업들의 생존에도 관심이 쏠렸다.
파산보호 조치가 복미에 있는 위워크 사무실에만 적용된다고는 해도
국내 시장 역시 영향을 피하기가 어려울 것이라ں 시각이었다.
‘운영할수록 손해’인 공유 오피스 사업 모델의 한계가 명확해졌다는 평가도 이어졌다.

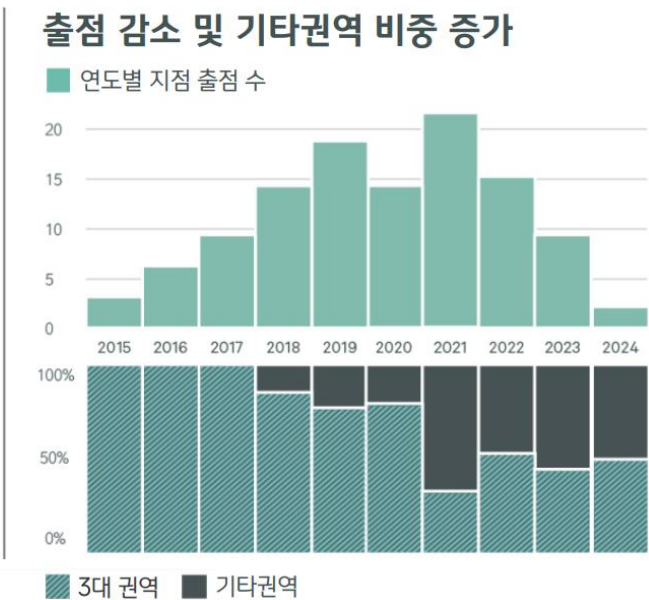
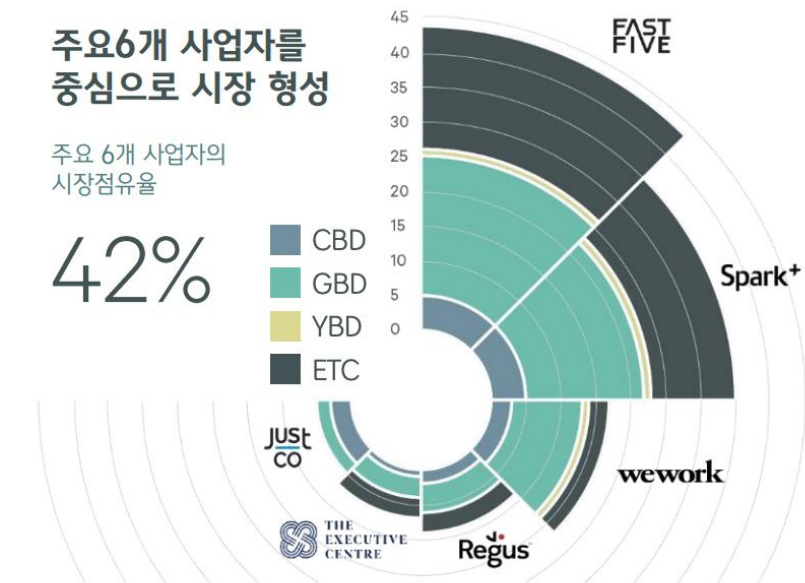
디지털타임스 (2024-09-18)

**“건물주 우위 시장에서 공유오피스 매력 급감...
올해 출점 2곳뿐”**

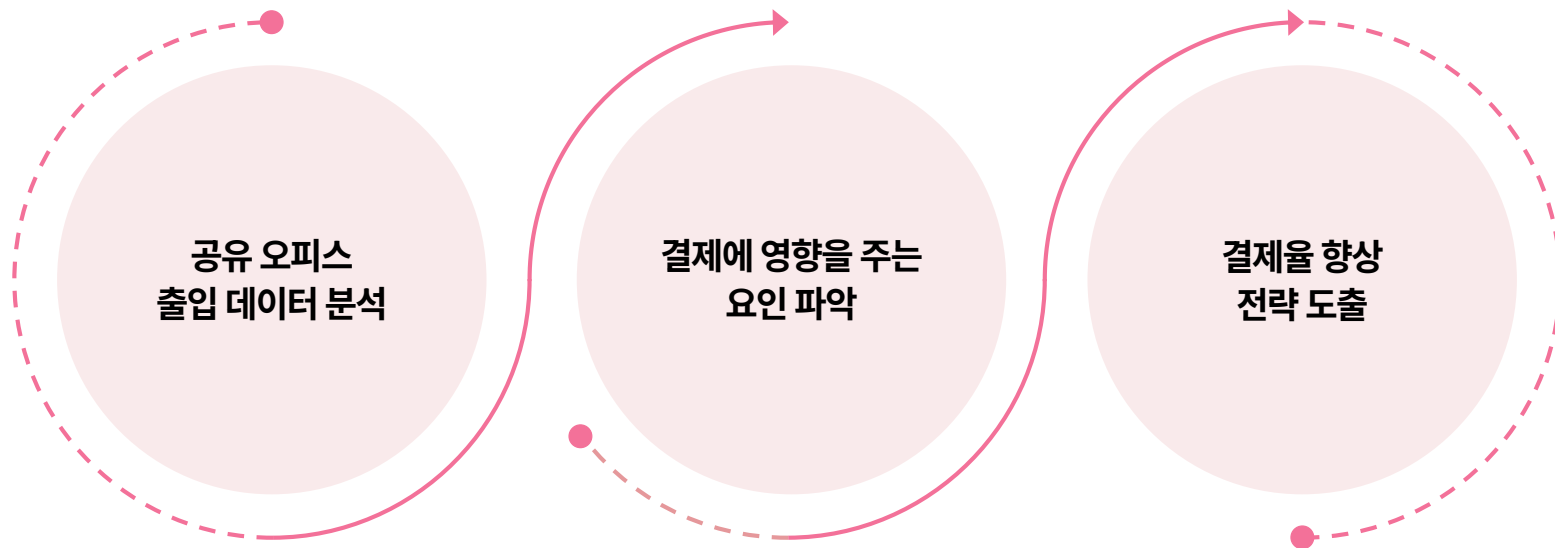
코로나19 팬데믹 당시재택근무 일상화로 빠르게 늘었던서울 시내 공유 오피스들이
최근 출점 수가 지속적으로 감소하는 등 부진한 흐름을 보이고 있다.
고금리 시기 스타트업 시장의 투자 위축으로 인한 수요 감소,
1세대 공유 오피스 미국 위워크 파산,
임대인 위주 시장으로의 개편 등이 그 원인으로 꼽힌다.

변화하는 수요와 업계 대응

2025년 2월 기준, 1~3개월 단기 계약 선호도가 3% 증가하자 패스트파이브는 일주일 단위 초단기 임대 서비스를 도입했다.
이는 경제·사회적 불확실성 속에서 기업들이 유연한 비즈니스 운영을 선호한 결과로, 공유 오피스 업계도 이에 맞춰 혁신을 이어갈 것으로 예상된다.



01 분석 주제



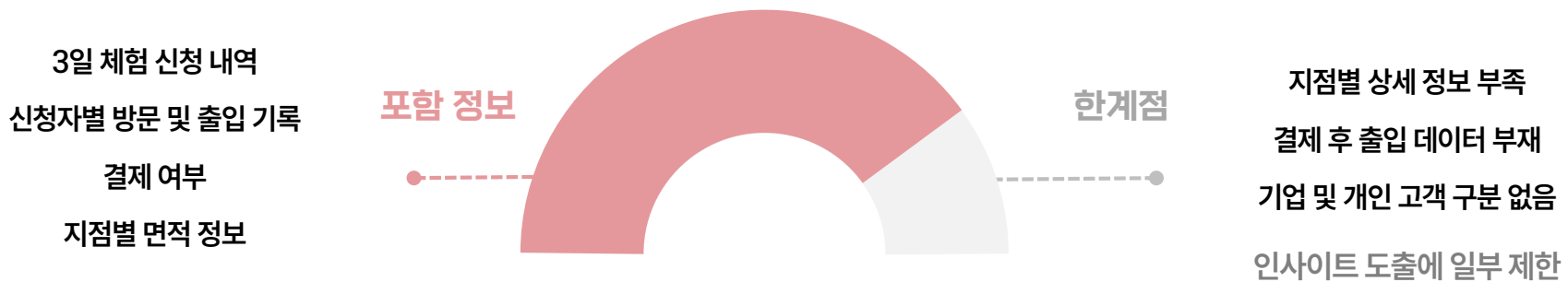
EDA

02

데이터 셋
1차 전처리
가설 정의
현황 분석

분석 데이터 개요

2021.05.01 ~ 2023.12.31, 2년 8개월간의 공유 오피스 **출입 데이터** 활용

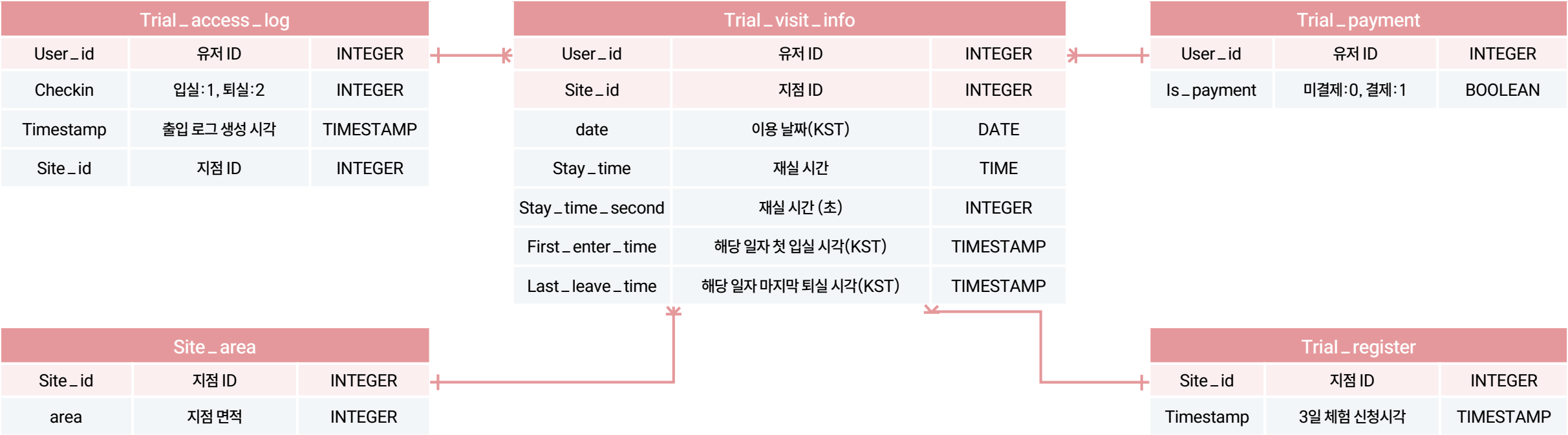


데이터 구성

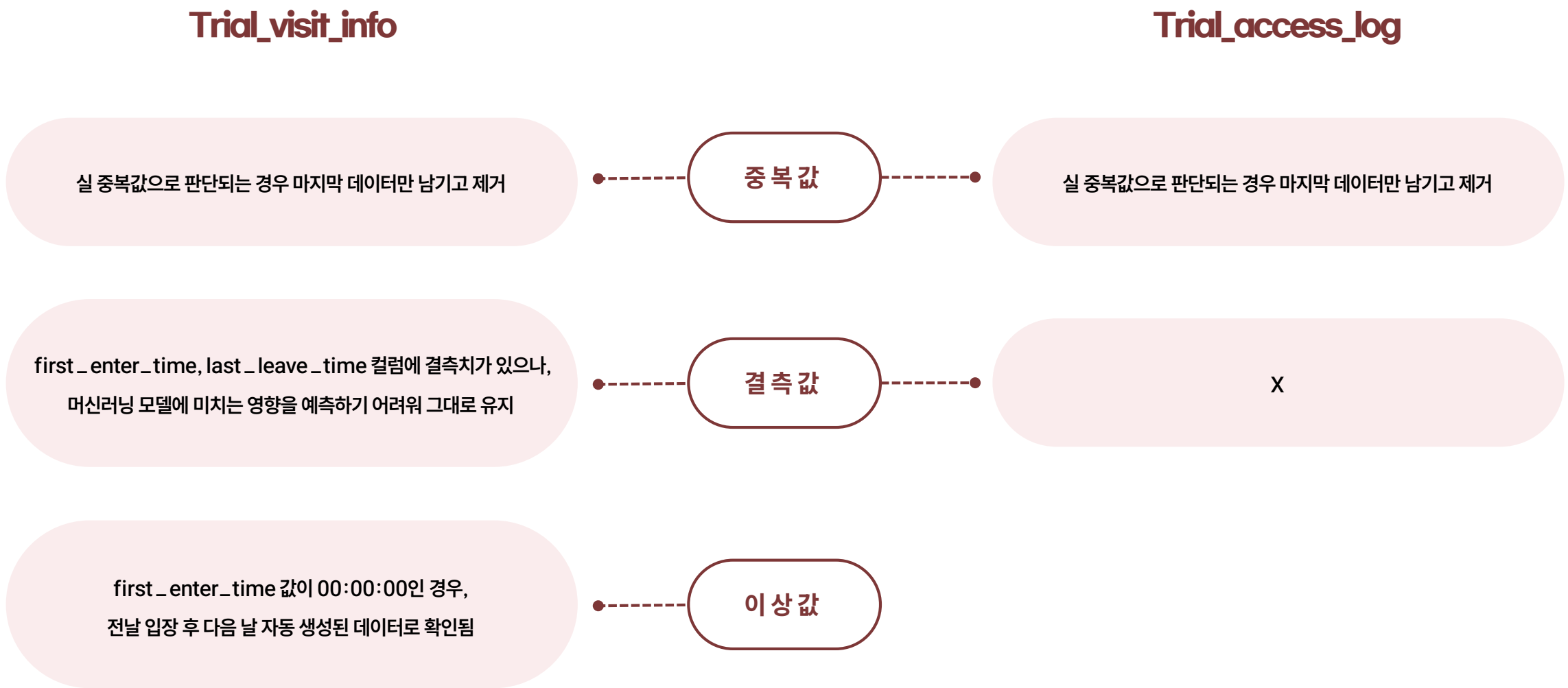
4개의 이벤트 데이터와 1개의 지점 정보 데이터프레임으로 구성

	테이블 명	설명	주요 컬럼
1	trial_register	3일 무료 체험 신청	유저 ID, 3일 무료 체험 신청 일시
2	trial_visit_info	3일 무료 체험 신청자 일자별 방문 기록	유저 ID,지점 ID, 날짜, 최초 입실 시각, 최종 퇴실 시각, 체류 시간
3	trial_access_log	3일 무료 체험 신청자 출입 기록	유저 ID, 지점 ID, 타임스탬프, 출입 방향
4	trial_payment	3일 무료 체험 신청자 결제 여부	유저 ID, 결제 여부
5	site_area	지점별 면적	지점 ID, 지점 면적

ERD



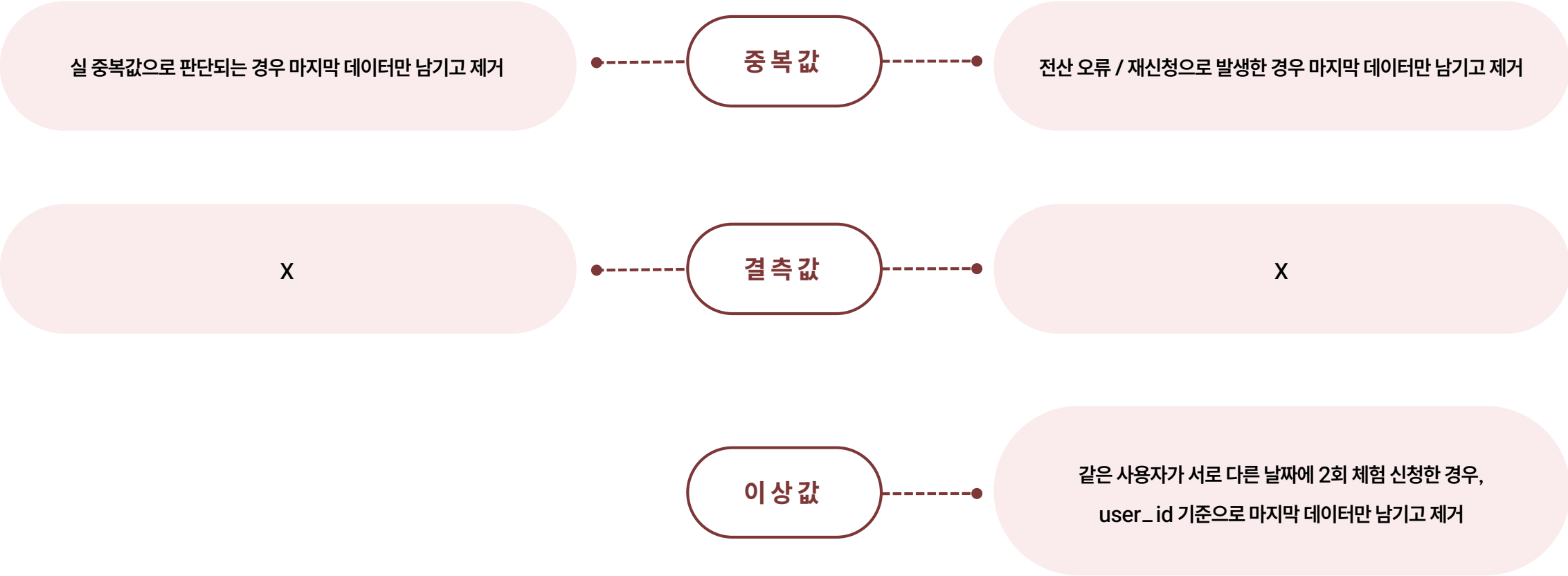
02 1차 전처리



02 1차 전처리

Trial_payment

Trial_register



테이블 병합

병합 방식

user_uuid를 기준으로 trial_visit_info
trial_register, trial_payment를 LEFT JOIN 방식으로 병합

이상치 처리

무료체험 기간(3일)을 초과한 총 방문 일수 4일 이상 데이터 삭제

컬럼 추가

trial_access_log 데이터를 date와 user_uuid로 그룹화하여
방문 횟수(count 컬럼) 추가

결측치 처리

count 컬럼에 결측치가 존재하지만, 추후 머신러닝 전처리 과정에서 처리 방안 고려

02 가설 정의

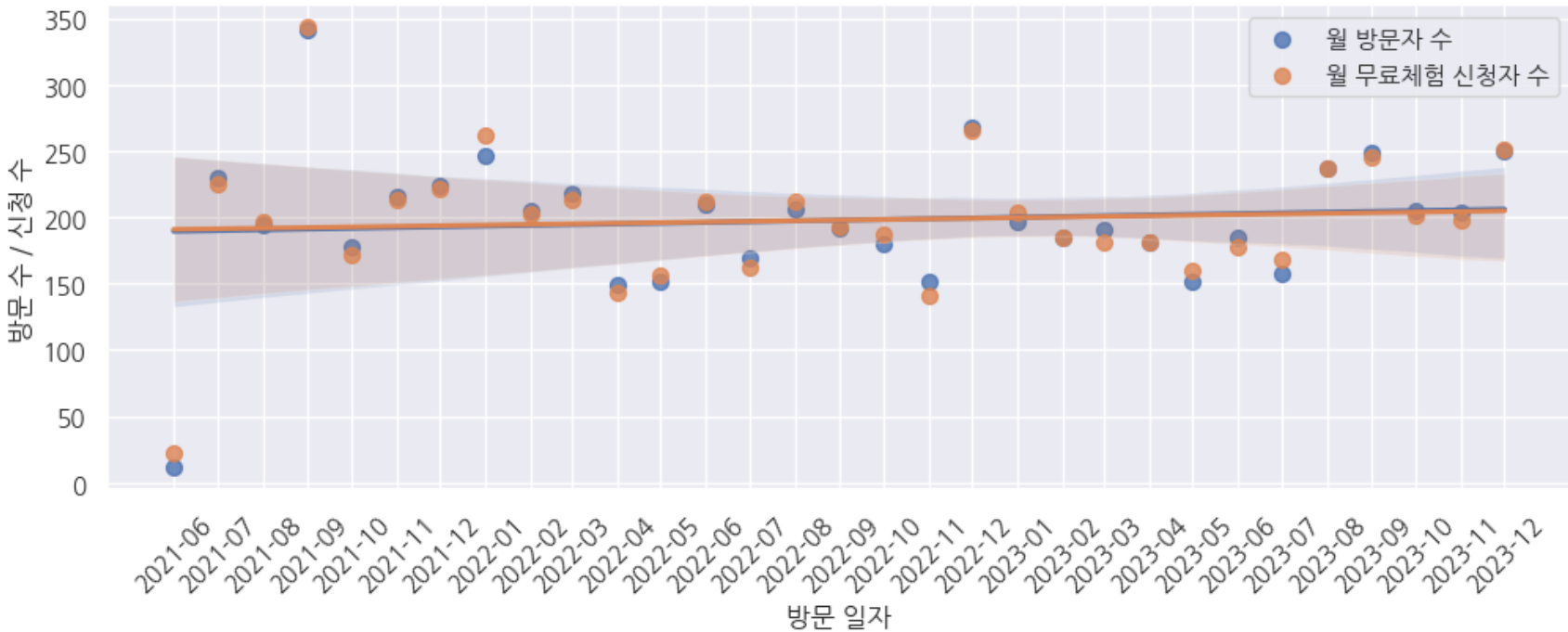
체류 시간에 따라 결제율에 차이가 있을 것이다.

방문 일 수에 따라 결제율에 차이가 있을 것이다.

방문 지점의 면적에 따라 결제율에 차이가 있을 것이다.

서비스 이용자 현황 - 월별 방문자 및 신청자

2021.06 ~ 2023.12 기준, 방문자 및 신청자 수는 150~260명 내외로 유지



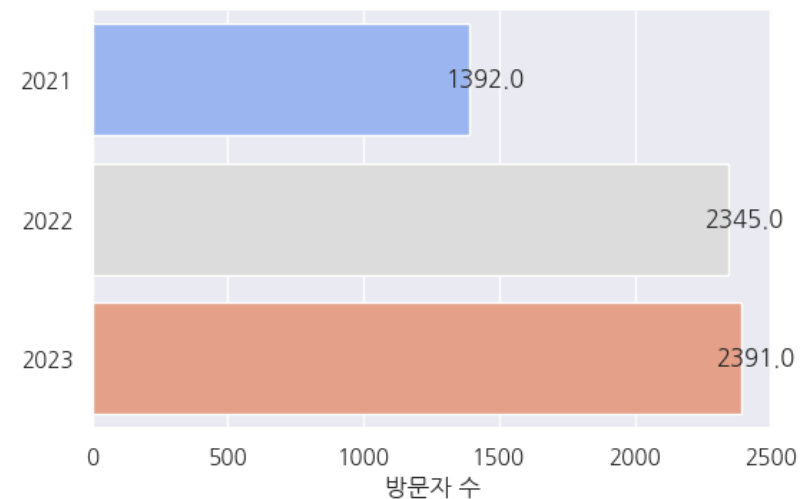
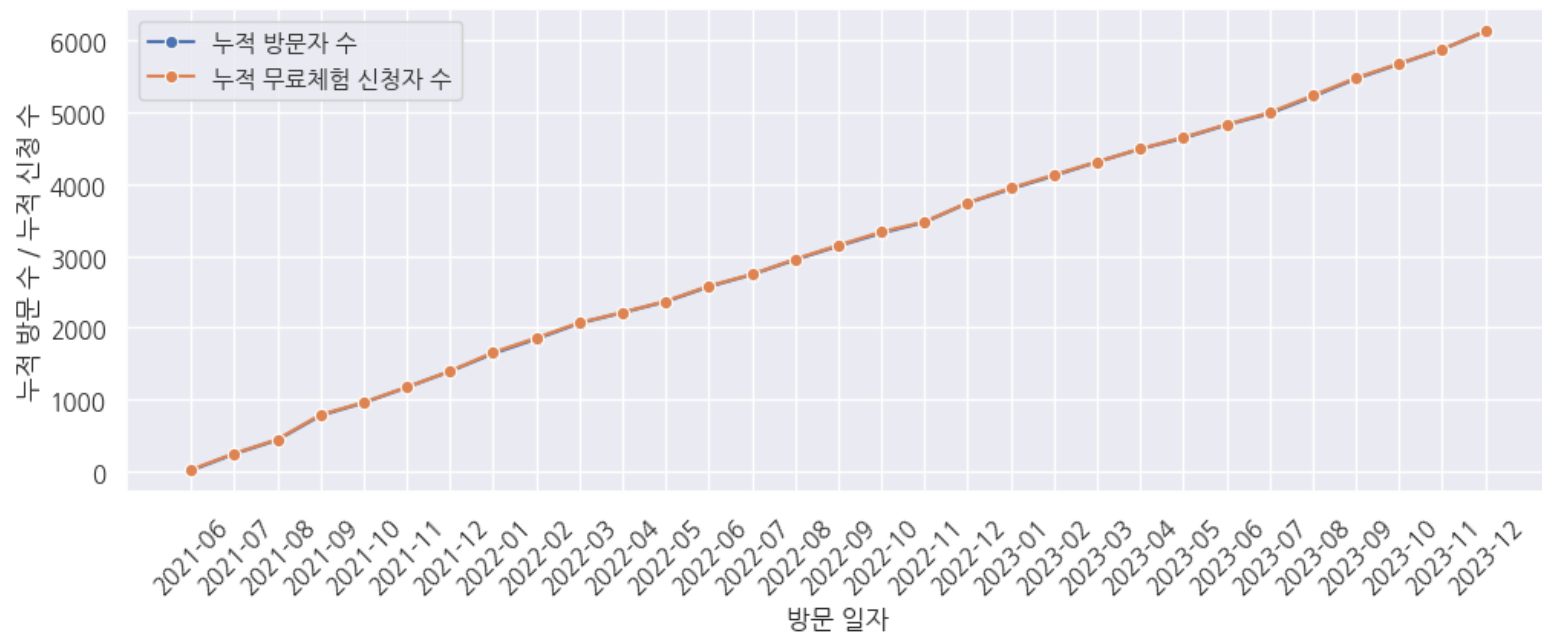
* User별 첫 방문일 기준

서비스 이용자 현황 - 연도별 방문자 수 증가율

2023.12 기준, 누적 방문자 및 신청자 수는 약 6,100명

2021 → 2022년 방문자 수 68.4% 증가

2022 → 2023년 방문자 수 1.9% 증가 (전년 대비 증가율 둔화)



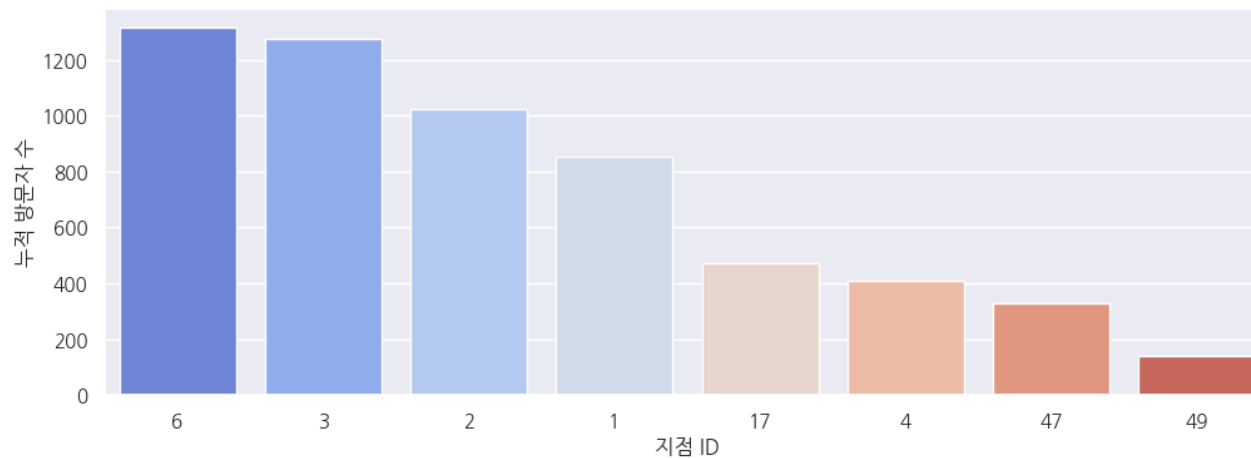
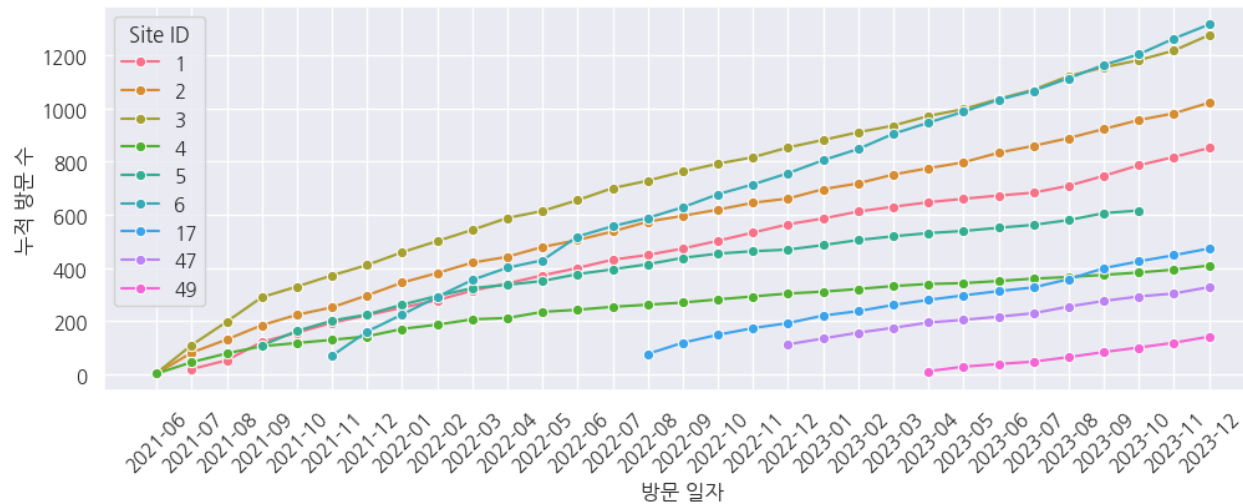
지점별 현황 - 지점 오픈 순서

지점 ID	면적	오픈 일자
1	50	2021.07.02
2	100	2021.06.30
3	150	2021.06.30
4	100	2021.06.30
5	150	2021.09.03
6	150	2021.11.15
17	50	2022.12.09
47	50	2022.12.09
49	50	2023.04.22

1, 2, 3, 4번 지점이 동일 시점 (2021.06~07)에 오픈
이후 5, 6, 17, 47, 49번 지점 순으로 개점
2022년 이후 면적 **50평** 지점만 추가 오픈

지점별 현황 - 누적 방문자 수

누적 방문자가 가장 많은 지점은 6번 (이후 3, 2, 1 순)



면적별 평균 방문 유저 수

150평 지점(1,104명) > 100평 지점(809명) > 50평 지점(462명)
150평 지점 방문 유저 수가 50평 지점보다 약 2.3배 많음

면적별 보유 지점 수

면적 (평)	보유 지점 수 (개)
50	4
100	3
150	2

면적별 평균 방문 유저 수

면적 (평)	평균 방문 유저 수 (명)
50	462
100	809
150	1,104

면적과 결제율의 관계

귀무가설 (H_0)

지점의 면적에 따른 구독 전환율 차이가 없다.

대립가설 (H_1)

지점의 면적에 따른 구독 전환율 차이가 있다.

카이제곱 검정 결과,

카이제곱 통계량 : 6.9730

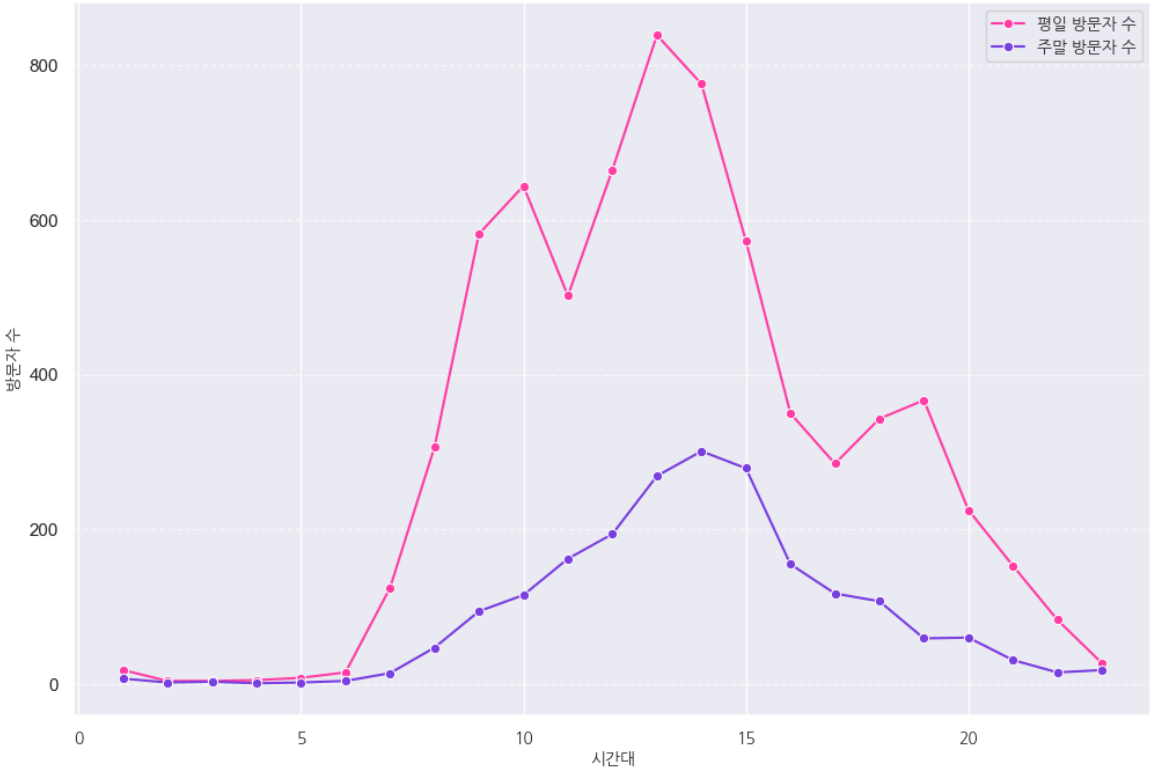
P-value : 0.0306

면적이 넓은 지점에서 결제 전환율이 높게 관측

방문 패턴 - 시간대별 방문자 수

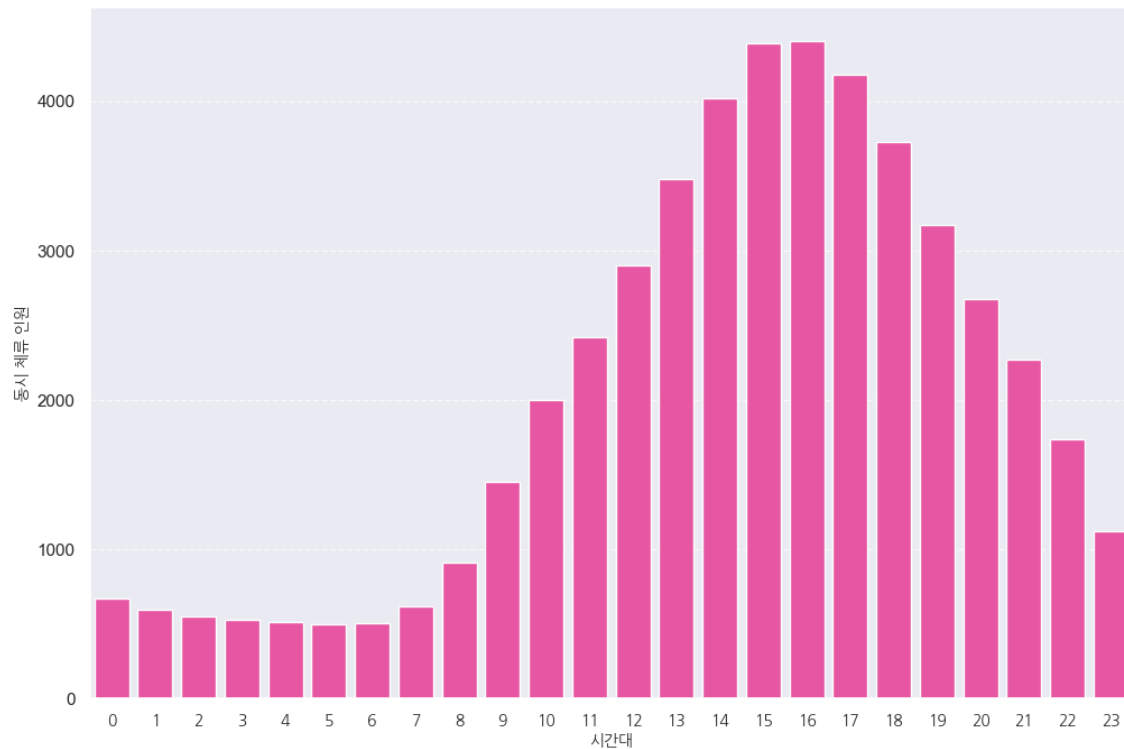
평일 방문자 수 > 주말 방문자 수

10시~15시 방문자가 가장 많으며, 이후 점차 감소



방문 패턴 - 시간대별 혼잡도

15~16시가 동시 체류 인원이 가장 많으며, 이후 점차 감소

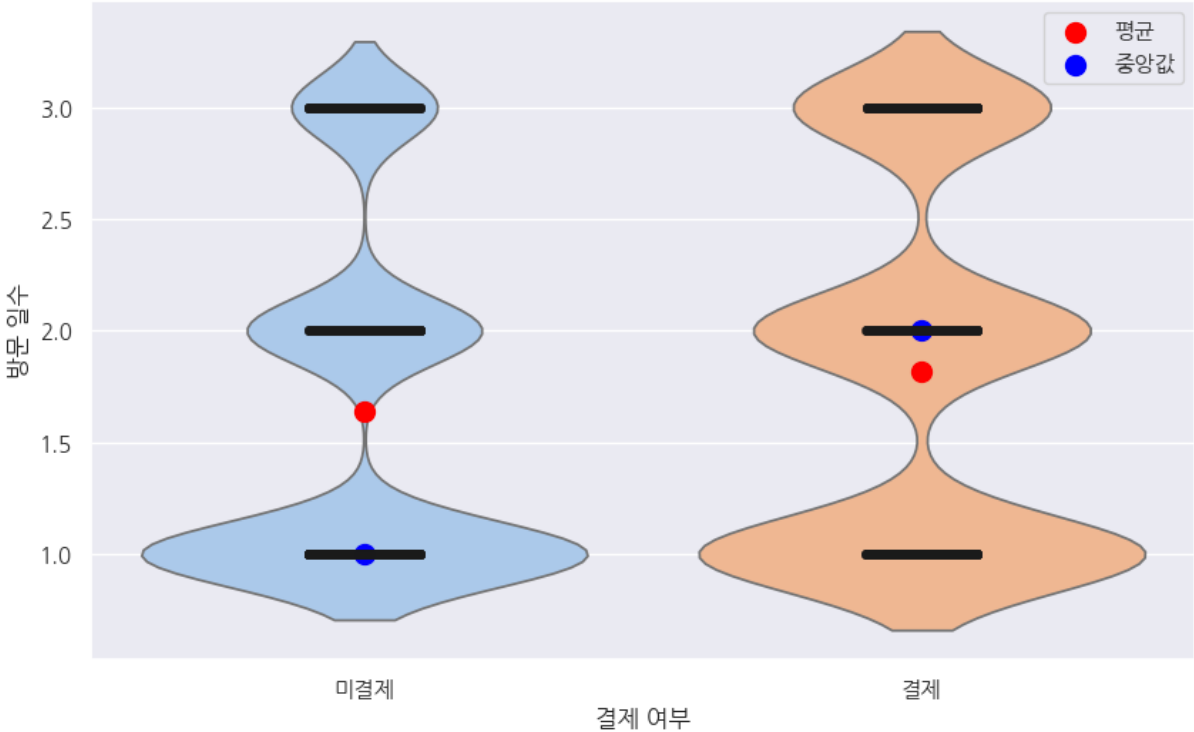


* 혼잡도: 동시간대 체류 인원

방문 패턴 - 결제 여부별 방문일 수

결제자 평균 방문일 (1.81일) > 미결제자 평균 방문일 (1.63일)

결제 고객의 평균 방문 일 수가 더 많은 것으로 확인



결제 여부	중양값	평균
결제	2.0	1.8186
미결제	1.0	1.6375

예측 및 세부 분석

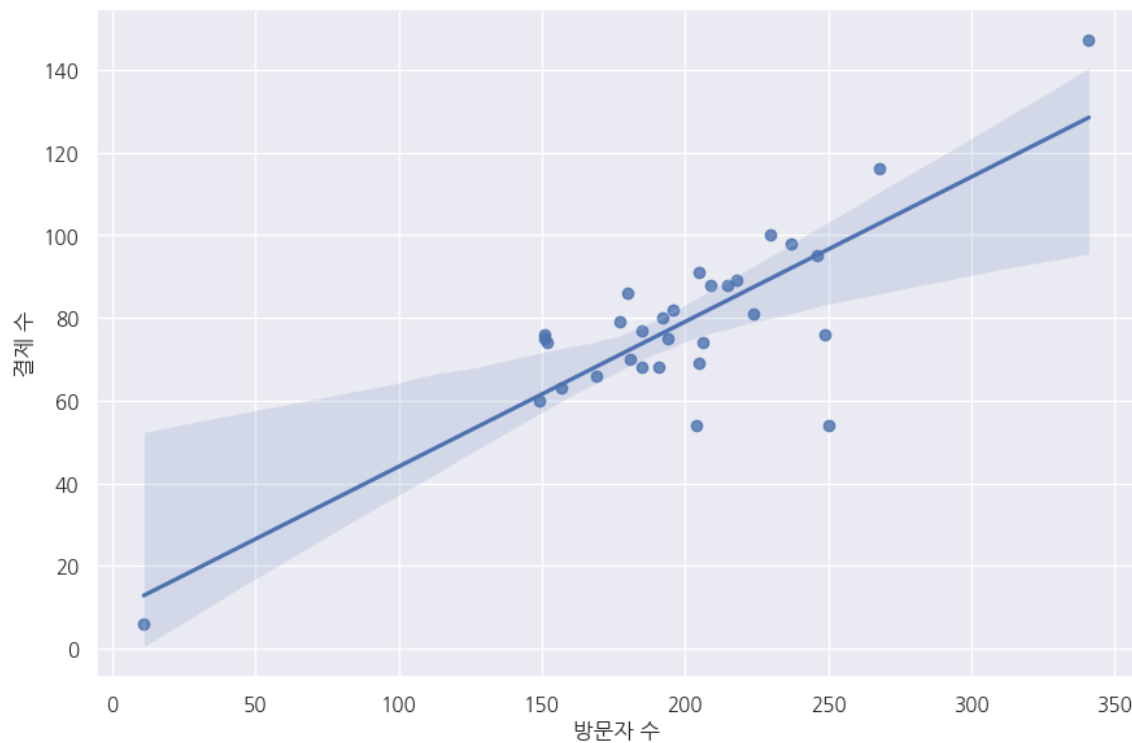
03

시계열 예측 분석

양상블 모델을 활용한 예측 분석

주요 변수 세부 분석

방문자 수와 결제자 수의 관계



방문자 수와 결제 수 간 뚜렷한 **선형 관계** 확인

(상관계수: 0.8237)

방문자 수 증가 → 결제 수 증가

시계열 예측 모델을 활용한 방문자 수 예측을 통해
향후 180일간 방문자 수 변화 예측

사용 모델

Prophet

* Facebook 개발 시계열 예측 모델

장점

계절성과 추세 모델링 가능

이상치에 강건

결측 데이터 처리 가능

공유 오피스 방문 데이터의 요일별, 월별 계절성 및 이벤트 (휴일, 프로모션 등) 변동성 반영

모델 평가 지표 선정

항목	설명	
주요 평가 지표	SMAPE	MAPE의 대칭 버전으로, 예측값과 실제값 간 차이를 공평하게 반영하여 상대적 오차를 균형 있게 평가
보조 평가 지표	MAPE	예측값이 실제값과 평균적으로 몇 퍼센트(%) 차이가 나는지 측정하는 지표

예측 기간 설정

1, 2, 3, 4, 5, 6 지점: 중장기 계획 수립을 위해 180일 설정

17, 45, 47 지점: 데이터 보유 기간이 상대적으로 짧아 90일 설정

모델 평가 계획

cross_validation 함수를 사용해서 교차 검증

initial	horizon	period
80%	20%	Horizon의 절반 기간

1차 전처리 및 하이퍼파라미터 설정

전처리

first_enter_time, last_leave_time 결측치 삭제
지점별 방문 일자별 방문자 수 집계
date는 ds, 결과 변수인 방문자 수는 y로 컬럼 수정

하이퍼파라미터 설정

changepoint_prior_scale : 기본값 0.05 설정
주간 계절성 추가 (period: 7일 지정)
월간 계절성 추가 (period: 30.5일 지정)

1차 성능 비교

지표	지점 1	지점 2	지점 3	지점 4	지점 5	지점 6	지점 17	지점 47	지점 49
SMAPE	0.590	0.490	0.601	0.429	0.451	0.551	0.464	1.153	0.461
MAPE	0.464	0.532	0.730	0.454	0.523	0.602	0.659	0.858	0.498

2차 하이퍼파라미터 튜닝

하이퍼파라미터 설정

changepoint_prior_scale : 0.1로 조정
weekly_seasonality : True 설정
주간 계절성, 월간 계절성 : 1차와 상동

2차 성능 비교

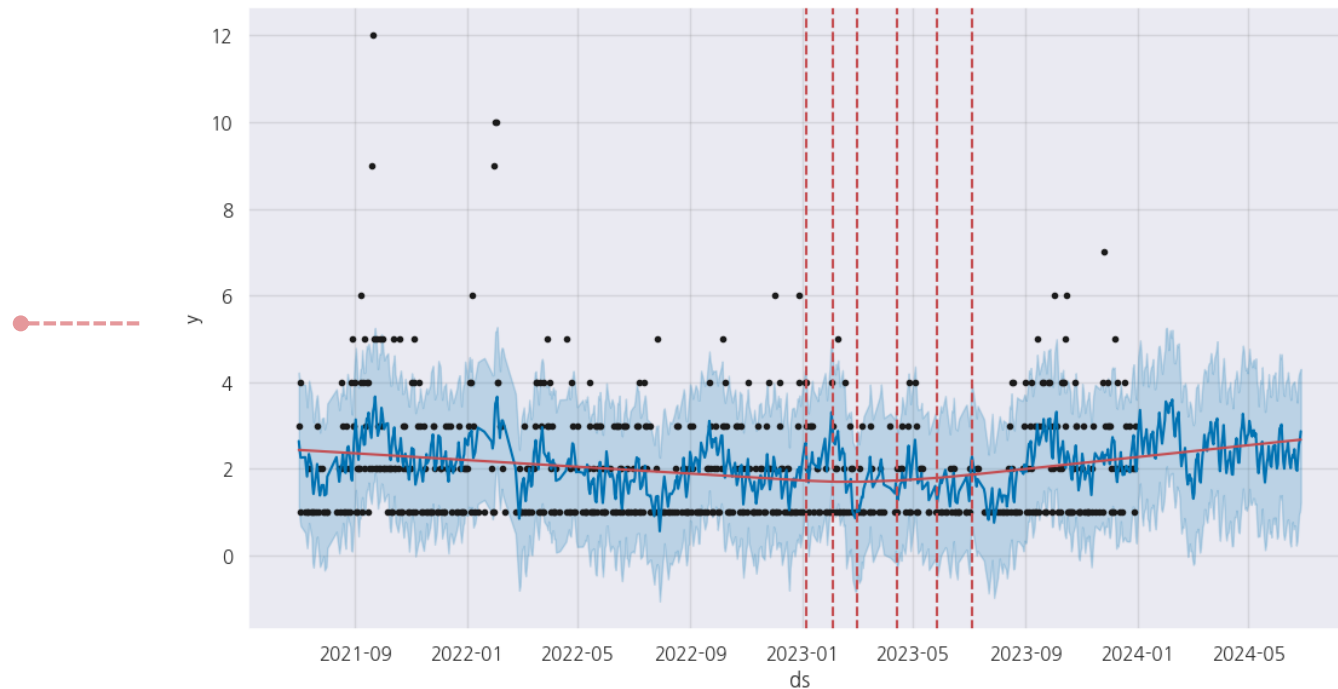
지표	지점 1	지점 2	지점 3	지점 4	지점 5	지점 6	지점 17	지점 47	지점 49
SMAPE	0.587	0.477	0.575	0.433	0.469	0.566	0.494	0.936	0.464
MAPE	0.463	0.555	0.743	0.513	0.611	0.581	0.787	0.733	0.506

모델 해석 및 결론

붉은 선: 트렌드 라인 → 시간이 지남에 따라 방문자 수의 변화 패턴을 나타냄

검정/파란 선: 실제 데이터/예측 데이터

빨간 점선: 추세 변화점 → 특정 시점에서 트렌드가 변화하는 지점을 의미



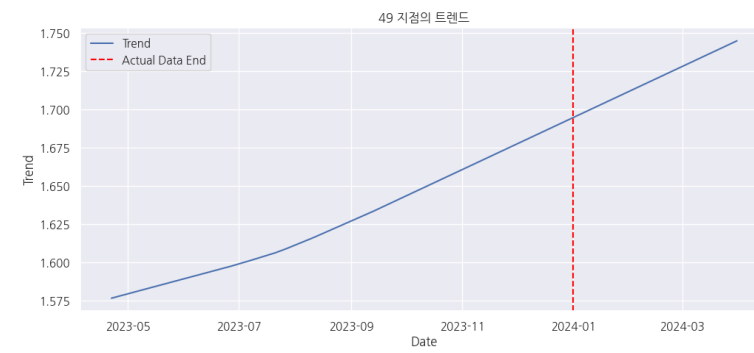
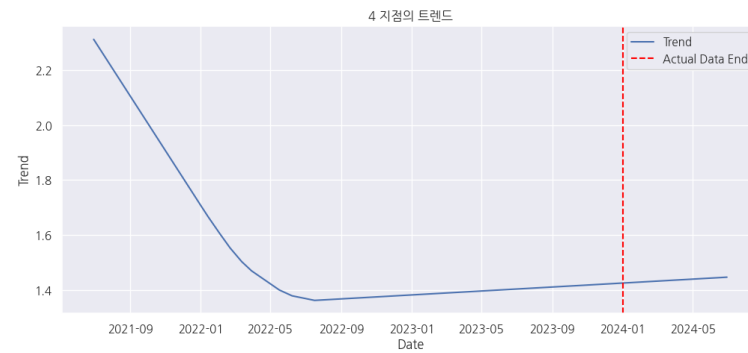
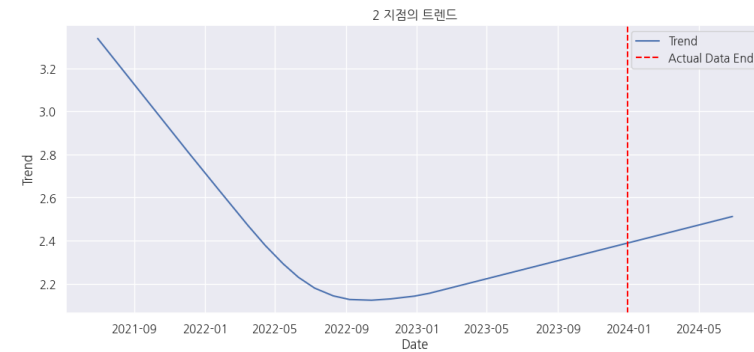
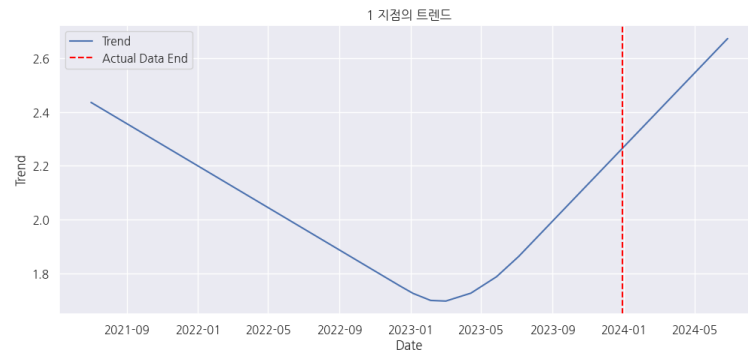
03 시계열 예측 분석

트렌드 (Trend)

지점 1, 2: 방문자 감소 후 회복세 → 향후 180일간 **증가** 예상

지점 4: 오픈 후 방문자 급감 → 2명 이내 유지

지점 49: 오픈 후 꾸준한 상승세 → 향후 90일간 방문자 **증가** 지속 예상

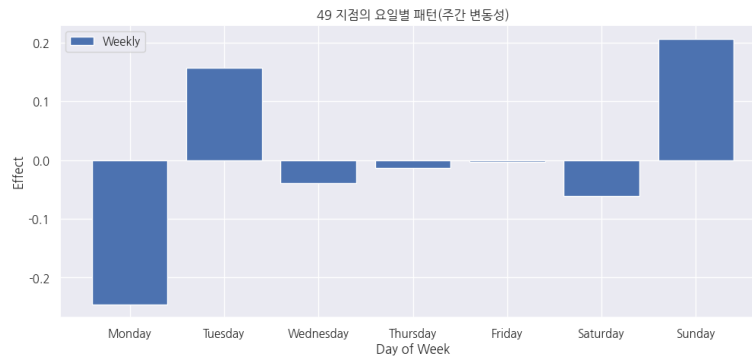
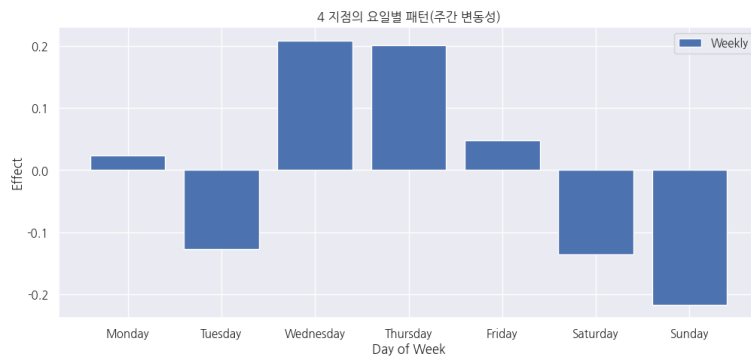
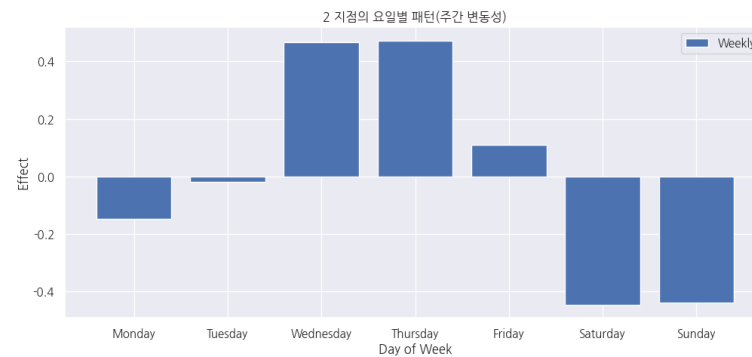
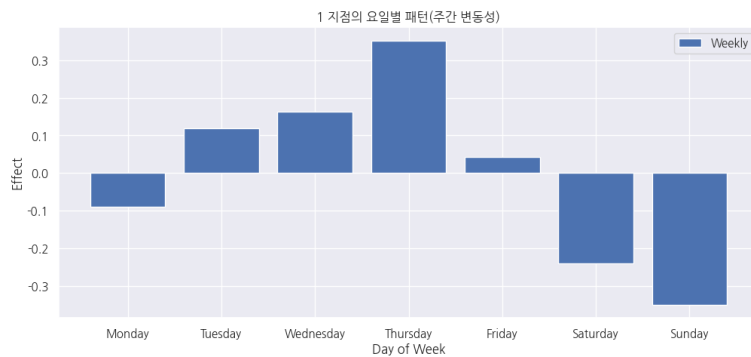


* 평가 지표가 상대적으로 낮은 지점은 해석에서 제외함

요일별 패턴 / 주간 변동성

지점 1, 2, 4: 수·목요일 방문자 집중, 주말 방문을 낮음 → 기업 고객·고정 방문자 비중 높음으로 예상

지점 49: 화·일요일 방문자 많음 → 프리랜서·일반 고객 비중 높을 것으로 추정



Prophet 모델을 통한 시계열 예측 분석 결론

지점 1, 2, 49

방문자 증가 추세 → 무료체험 방문자 증가를 지속하기 위해 상품 개선 및 전략적 관리 필요

지점 4

방문자 정체 상태 → 방문자 증가 저해 요인을 분석하고 구체적인 개선 방안 마련 필요

요일별 방문 패턴을 기반으로 지점별 주요 고객층을 분석하고,
맞춤형 접근 전략 필요

사용 모델



항목	설명
목적	체험 고객의 구매 전환 예측 및 핵심 영향 변수 도출
주요 지표	AUC-PR (불균형 데이터에 대한 정밀도-재현율 균형 평가)
보조 지표	Accuracy
전처리 과정	사이클링 인코딩, 원핫인코딩, 오버 샘플링 적용, 2단계 파생 변수 생성

TRIAL_DATE **FIRST_ENTER_TIME**
TOTAL_VISIT_COUNT **SITE_ID** **USER_UUID**
STAY_TIME **STAY_TIME_SECOND** **COUNT**
DATE **IS_PAYMENT** **LAST_LEAVE_TIME**

TRIAL_DATE **FIRST_ENTER_TIME**
TOTAL_VISIT_COUNT **SITE_ID** **USER_UUID**
STAY_TIME **STAY_TIME_SECOND** **COUNT**
DATE **IS_PAYMENT** **LAST_LEAVE_TIME**

03 앙상블 모델을 활용한 예측 분석

1차 전처리

결측치 처리

first_enter_time, last_leave_time → 각 520개 결측치 삭제

count 변수 → 2436개 결측치 발생 (데이터 손실 방지를 위해 **중앙값**으로 대체)

피처 엔지니어링

site_id → **원-핫 인코딩** 적용

파생 변수 추가 (trial_diff, weektype, total_visit_count, year, month, hour)

변수 제거

date, first_enter_time, trial_date, last_leave_time

2차 전처리

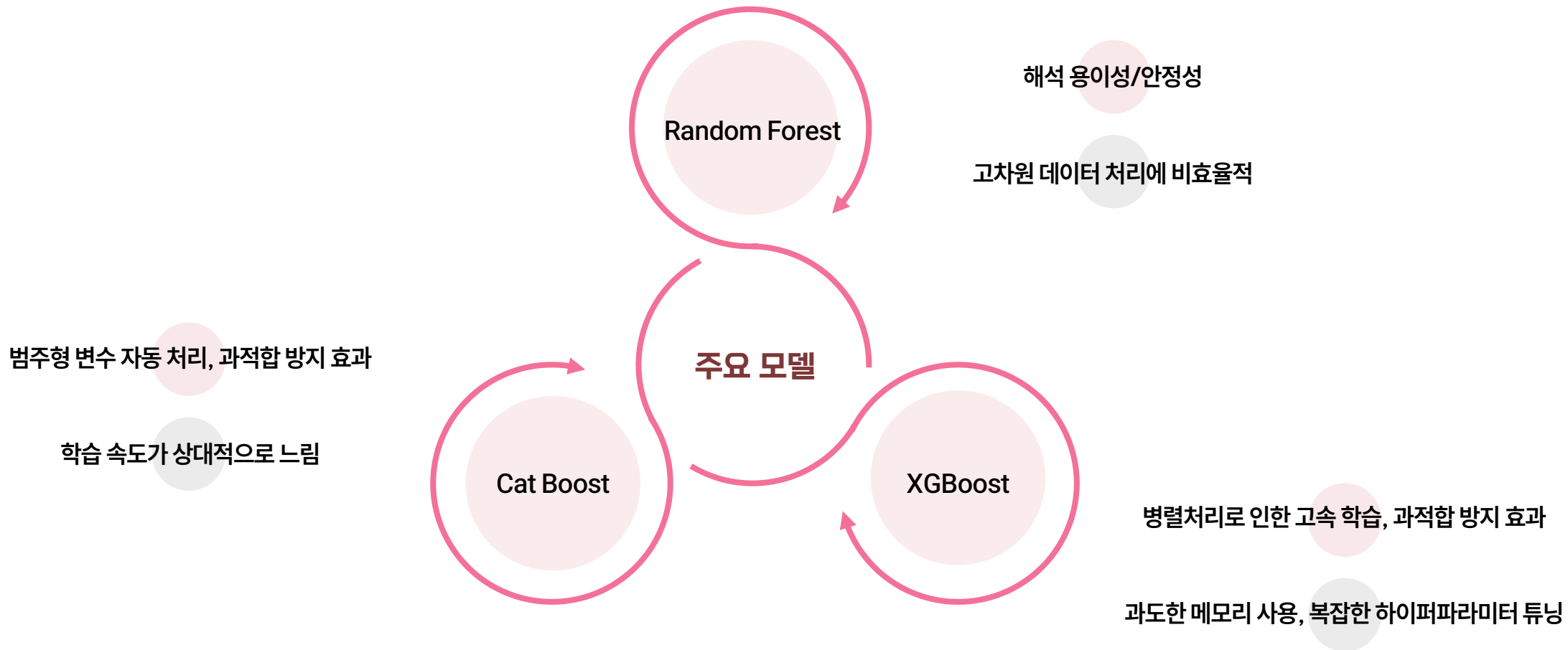
변수 추가

hour_sin/hour_cos → **사이클링 인코딩** 적용 (시간 데이터 주기성 반영)

unique_site_count → 사용자가 방문한 공유 오피스 지점 수

area_pyeong → 공유 오피스 면적 정보 추가

03 앙상블 모델을 활용한 예측 분석



03 앙상블 모델을 활용한 예측 분석

모델 성능 비교

	RandomForest	XGBoost	LightGBM	CatBoost	RUSBoost	HistGradientBoost
1차 AUC-PR	0.54	0.46	0.54	0.47	0.48	0.47
1차 Accuracy	0.62	0.62	0.63	0.63	0.56	0.59
2차 AUC-PR	0.55	0.54	0.54	0.61	0.53	0.59
2차 Accuracy	0.62	0.62	0.62	0.65	0.59	0.64

모델 순위

	AUC-PR	Accuracy	순위
CatBoost	0.61	0.65	1
RandomForest	0.55	0.62	2
XGBoost	0.54	0.62	3

03 앙상블 모델을 활용한 예측 분석

CatBoost	RandomForest
day_sin	day_sin
month_sin	stay_time_second
stay_time_second	first_enter_hour_sin
first_enter_hour_sin	month_sin
last_leave_hour_sin	last_leave_hour_sin
total_visit_count	weekday_sin
weekday_sin	count
count	total_visit_count
trial_first_diff	date_year
date_year	trail_first_diff

속성중요도 Top 10

두 모델의 교집합 속성들을 뽑아 세부 분석에 사용

date_month, date_day,
stay_time_second, weekday,
first_enter_hour

03 주요 변수 세부 분석

분석 목적

ML 모델에서 속성 중요도가 높은 **5가지** 변수를 중심으로 세부 분석하여 인사이트 도출 및 전략 제시

date _ month

weekday

date _ day

first _ enter _ hour

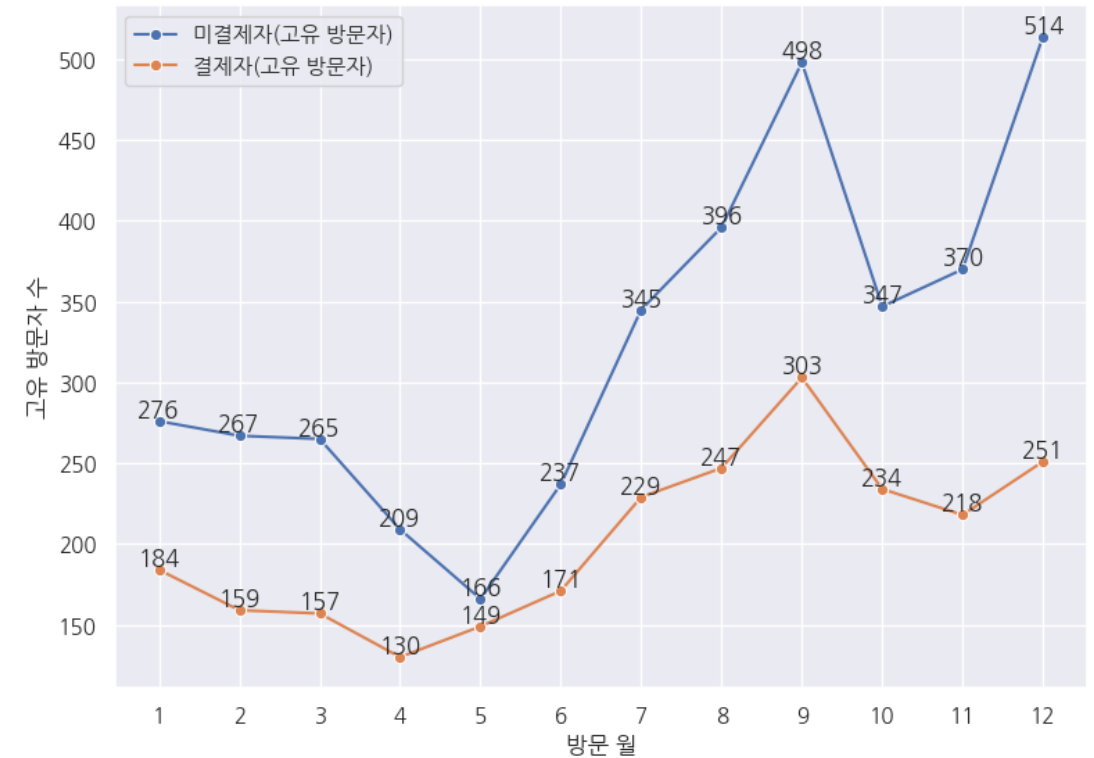
stay _ time _ second

03 주요 변수 세부 분석

방문 월에 따른 결제 여부 분석 (date _ month)

하반기에 결제 전환율 증가 경향 확인

정부·투자 기관의 예산 편성 및 집행 일정 → 하반기에 자금 확보 후 결제 증가
기업의 운영 안정화 → 상반기보다는 하반기에 오피스 이전 및 서비스 결제 선호



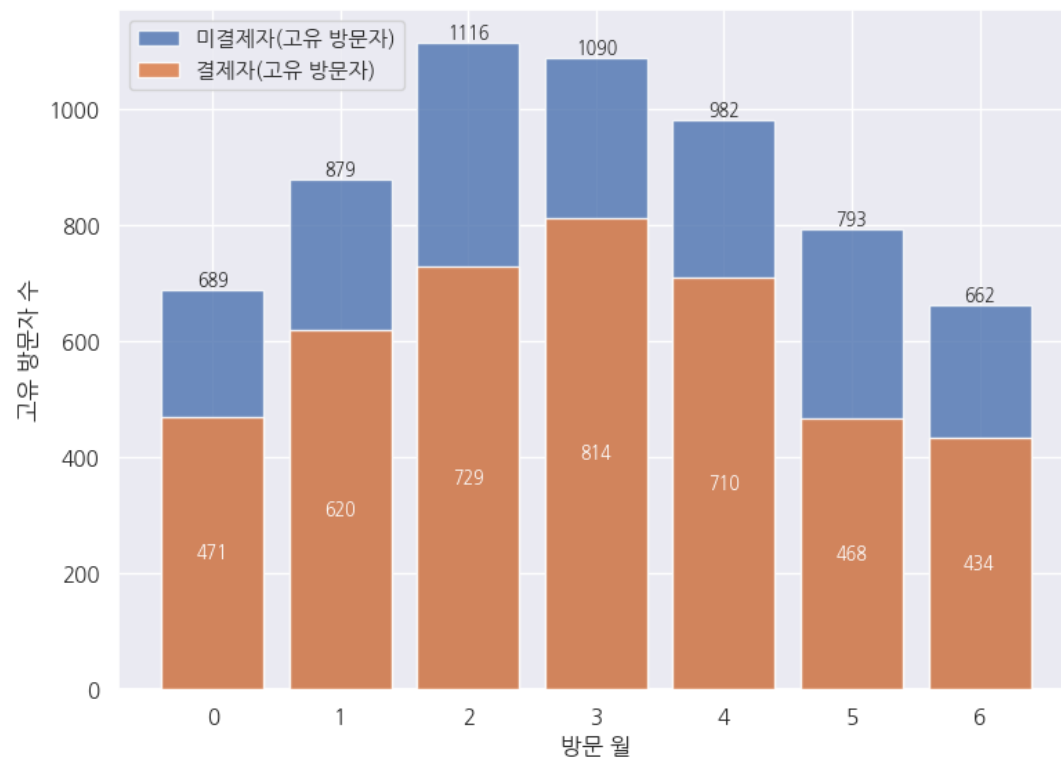
* 고유 방문자 기준

03 주요 변수 세부 분석

방문 요일에 따른 결제 여부 분석 (weekday)

목, 금요일의 결제 전환율 70%로 가장 높음
주말 방문자 수 증가, 그러나 결제 전환율은 감소

월요일 방문자는 주로 회사원으로 구성된다고 가정
화~금요일 방문자 증가는 프리랜서 및 일반 고객,
또는 **체험 목적** 이용자 비중이 높은 것으로 추정



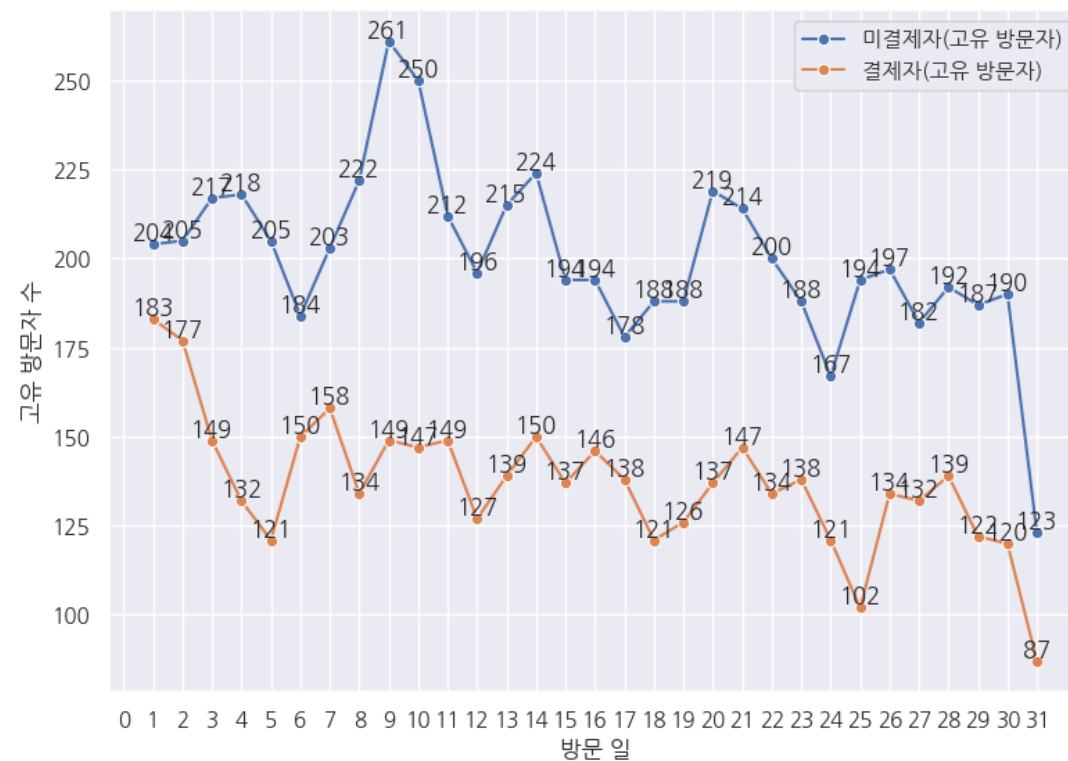
* 고유 방문자 기준

03 주요 변수 세부 분석

방문 일자별 결제 여부 분석 (date_day)

일자별 패턴으로는 요인 판단 어려움

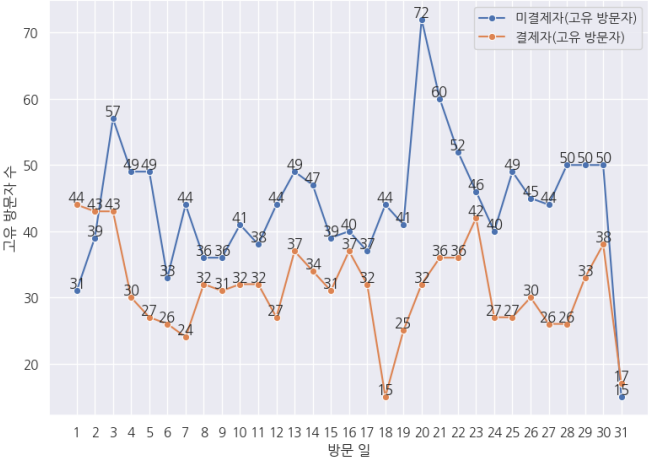
1차적으로 연도별로 나누어 패턴 분석
이후 2차로 월별/지점별로 구분하여 요인 탐색



* 고유 방문자 기준

03 주요 변수 세부 분석

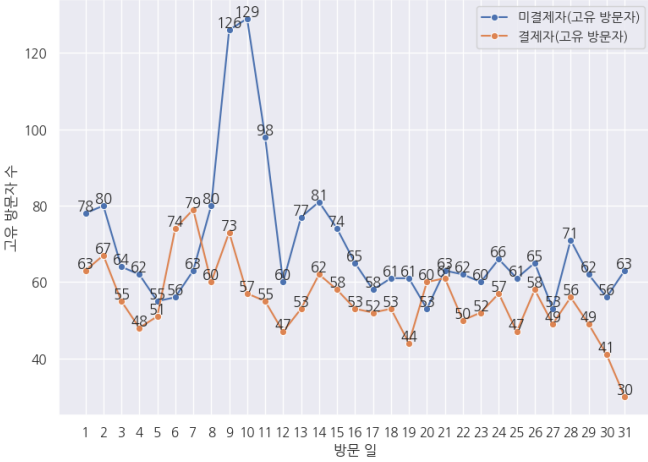
2021년



20일 방문자 급증

코로나19로 인한 방역 대책

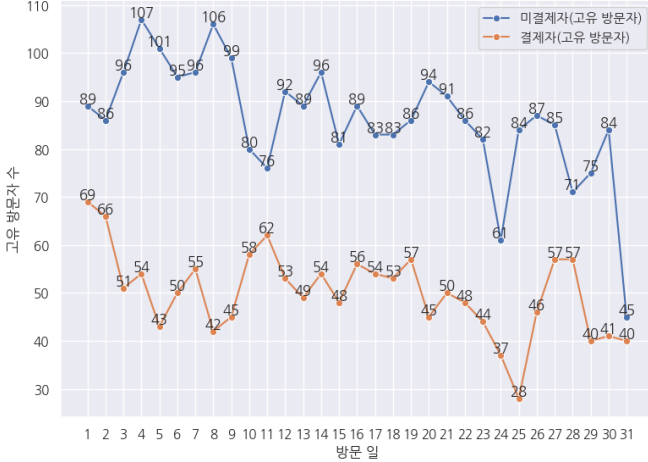
2022년



9~11일 방문자 급증

새로운 지점 오픈 영향

2023년



24~25일 결제/미결제 급감

전체적으로 24~25일 방문자 저조

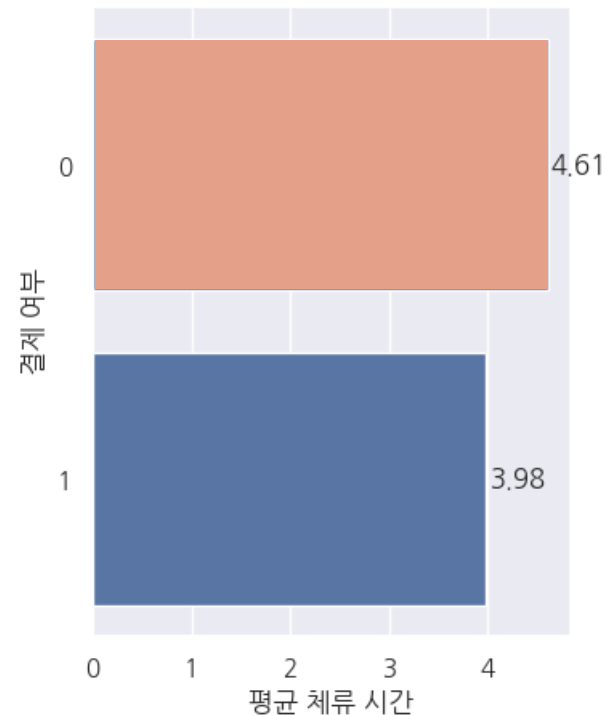
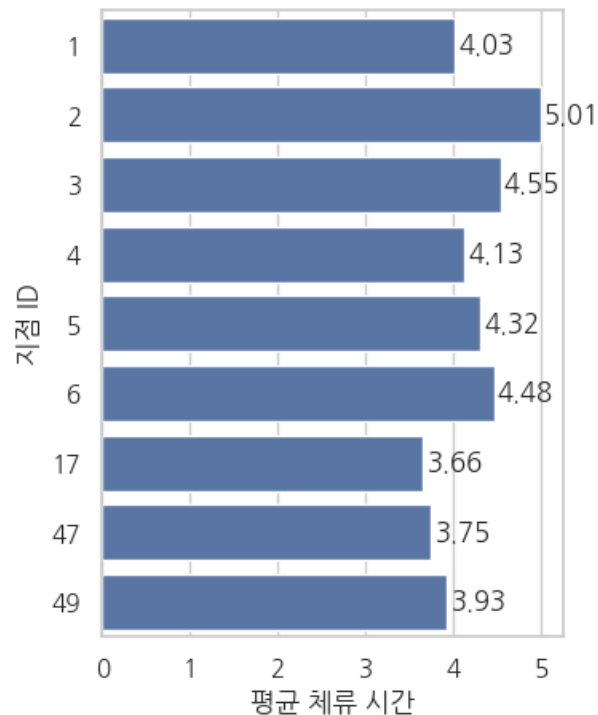
* 고유 방문자 기준

03 주요 변수 세부 분석

체류 시간 분석 (stay_time_second)

면적이 큰 지점일수록 체류 시간이 길어지는 경향
미결제자 체류 시간 (4.61시간) > 결제자 체류 시간 (3.98시간)

기업 고객이 많이 이용하는 지점일수록 체류 시간이 길어질 가능성
미결제자는 체험 기간 동안 충분히 사용 후 결제 없이 이탈
결제자는 서비스 가치를 확인한 후 즉시 결정하여 비교적 짧은 체류



03 주요 변수 세부 분석

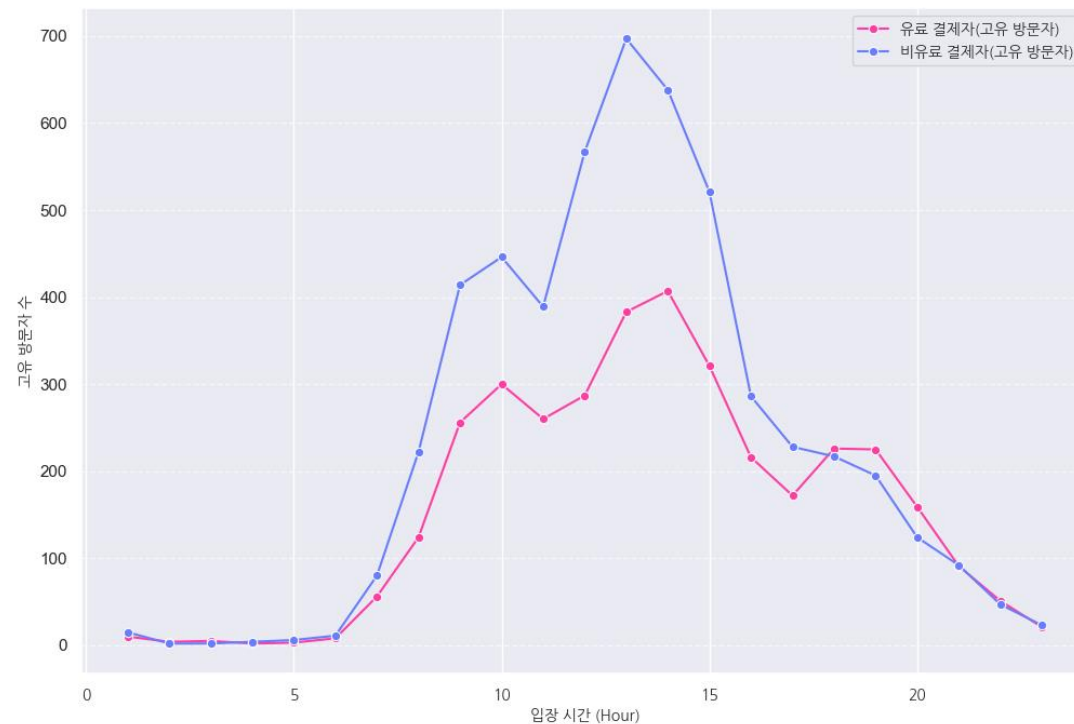
입장 시간별 결제 여부 분석 (first_enter_hour)

출근 시간대: 방문자 수 증가 (미결제자 다소 많음)

점심 시간대: 방문자 증가, 그러나 미결제자 증가

야간 시간대(19시 이후): 방문자 급감, 결제 전환율 상대적으로 높음

오후 1시 이전에는 기업 고객 비중이 높으며,
이후부터는 프리랜서 및 일반 고객의 비중이 증가할 가능성

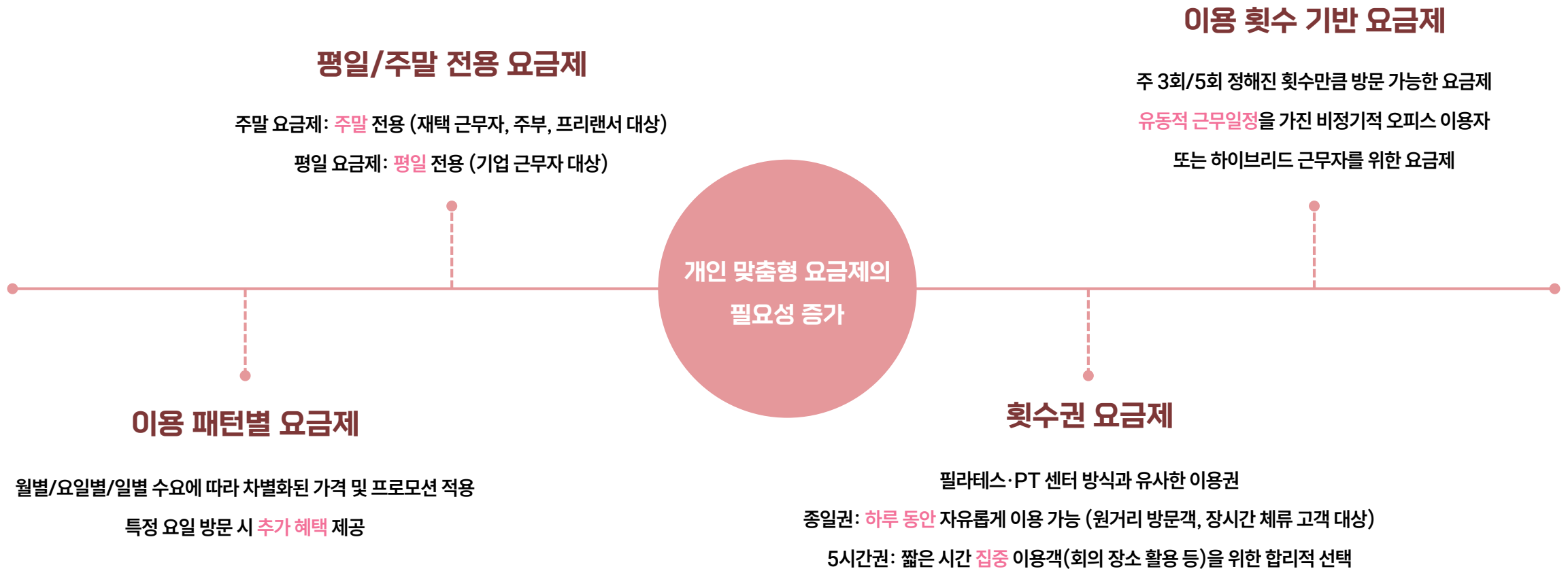


* 고유 방문자 기준 / 전날 입장하여 다음 날 퇴장한 경우, 데이터가 0으로 기록되어 분석에서 제외함

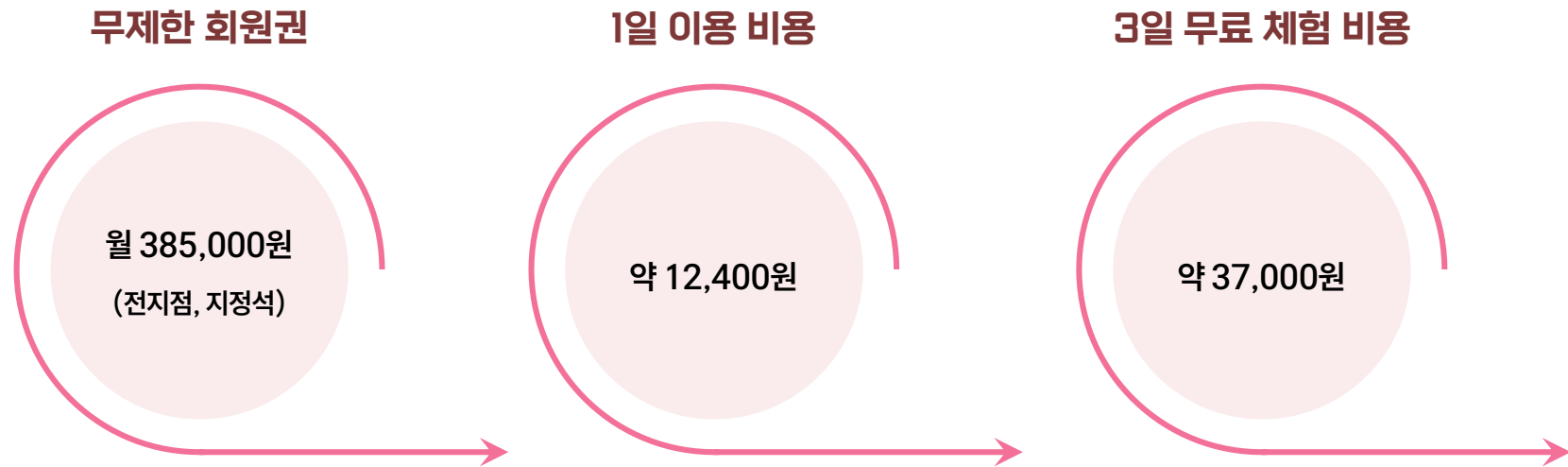
제안 04

신규 요금제 기획
무료 체험 타당성 검토

04 신규 요금제 기획



무료 체험 및 월 회원권 비용 분석



연도별 서비스 비용 및 수익 분석

연도	미결제 인원(명)	결제 인원(명)	전체 인원(명)	결제 비율(%)	3days 비용(원)	수익 (1개월 요금 기준)	투자 비율(%)
2021	1,001	703	1,704	41.3	63,388,800	270,655,000	23.4
2022	1,376	963	2,339	41.2	87,010,800	370,755,000	23.5
2023	1,571	821	2,392	34.3	88,982,400	316,085,000	28.2

방문일 수 비교

결제 여부	방문일 수 (평균)	방문일 수 (중앙값)
결제	1.818571	2.0
미결제	1.637490	1.0

04 무료 체험 타당성 검토

결제 전환율이 약 40%로 높은 수준이며,
최소 1개월 요금 기준으로
투자 비용(25%) 대비 수익성이 있음

무료 체험의 타당성

방문자들의 평균 방문일이
결제 인원 1.8일, 미결제 인원 1.6일로,
2일을 넘지 않음

체험일 수 조정 제안

투자 비용 절감 및 운영 효율성 증대를 위해
무료체험을 3일에서 **2일**로 단축

제언 및 시사점

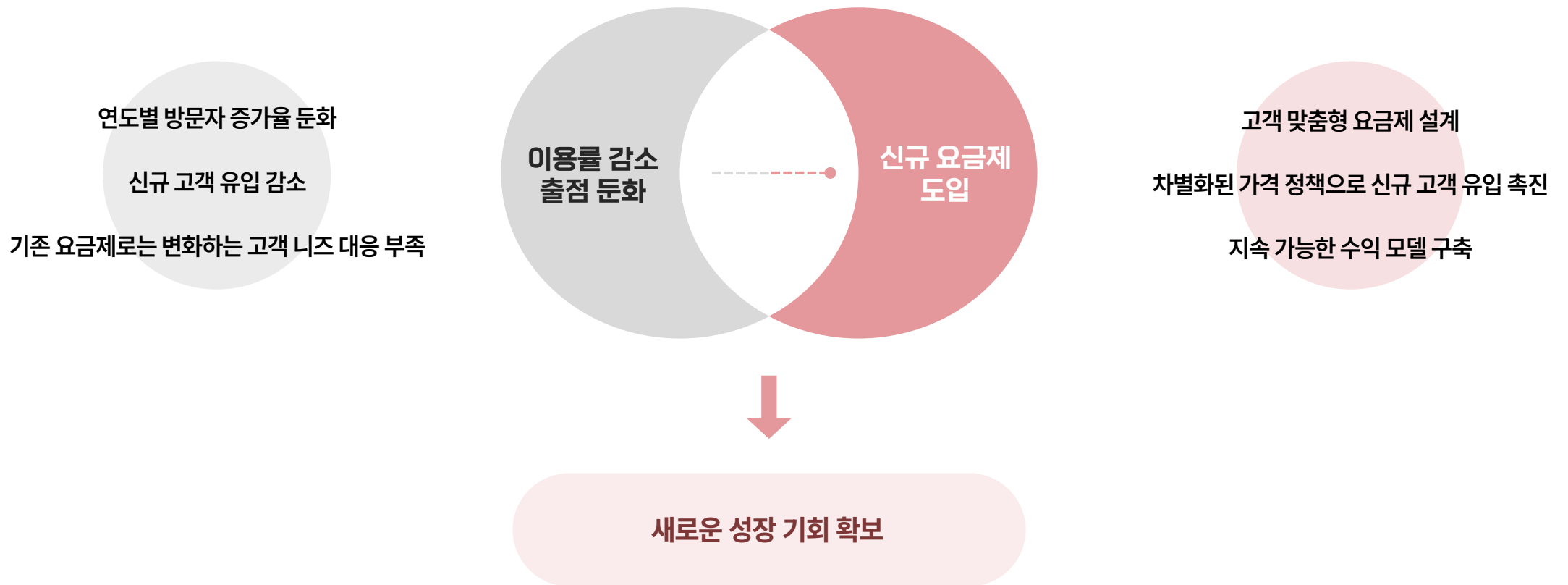
05

시사점

한계점

추가 분석 제언

05 시사점



05 한계점

방문자 흐름과 결제 여부 분석 가능



더 깊이 있는 인사이트 도출을 위해 추가 데이터 수집이 필요

3일 체험 데이터

방문 일자 정보만 포함 / 사용자 활동 세부 정보 부족

지점별 정보 부족

위치, 수용 인원 데이터 없음 → 교통 날씨 등 외부 데이터 활용 어려움

데이터 형식

대부분 Datetime 데이터로 구성되어 다양한 속성 분석에 한계

결제 데이터 한계

결제 여부만 포함 (실 결제 금액, 결제 수단, 환불 정보 없음) → 세부 분석 한계

추가 데이터 확보를 통해
방문자 분석 정밀화 & 운영 최적화 가능

05 추가 분석 제안

회의실, 핫 데스크, 지정석 등 공간별 사용 빈도 분석
좌석 수 & 이용 요금 최적화를 위한 데이터 기반 의사결정

혼잡도 분석

출퇴근
패턴 분석

지하철/버스 정류장 근처 지점의 방문율 분석

출퇴근 시간대 이용 패턴과 연계하여 맞춤형 멤버십 상품 최적화



1팀 박준혁 이은지 장은희 조재윤

Q & A

MIDDLE PROJECT

감사합니다
