

“이 점수로 는 못 팔죠”

예측 모델이 말 안 듣는 상황에서 찾아낸 전략 4가지

DA09기 3팀

김정섭
김한샘
고우혁
문치호
마재민

1. Executive Summary.....	4
2. Background & Problem Definition.....	5
2-1. 데이터 배경.....	5
2-2. 주제 선정 : 무료체험 정말 효과적인가?.....	6
2-3. 문제 정의.....	7
2-4. 분석 목적.....	7
3. Analysis Framework & Methodology.....	7
3-1. 분석 환경 및 데이터 파이프라인.....	7
3-2. 데이터 품질 관리.....	8
3-3. trial_visit_info 대체 테이블 생성.....	8
3-4. 모델링 구조.....	9
3-5. 평가 지표 채택 및 모델별 성능 비교.....	11
4. Key Findings.....	12
4-1. 결제 전환율: 방문 일수별 결제 전환율 차이.....	12
4-2. 결제 전환율 : 방문/미방문 그룹별 결제 전환율 차이.....	13
4-3. 결제 전환율 : 공간 면적과 전환율의 비례성 부재.....	14
4-4. 결제 전환율 : 시즌별 결제 전환율 분석.....	15
4-5. 결제 전환율 : 체류시간 별 결제 전환율 차이.....	16
4-6. 이용 시간대 분석: 출입 시간의 비대칭적 이중 피크 패턴.....	17
5. Modeling Results & interpretation.....	17
5-1. 피쳐 엔지니어링 및 마스터 테이블 구성.....	17
5-2. 사용 모델 비교 및 최종 모델 선정.....	18
5-3. 피쳐 중요도 분석.....	19
5-4. 결과 해석.....	20
6. Strategic Recommendations.....	20
6-1. 분석 결론: 3일 무료가 아닌 정교한 경험 설계로의 전환.....	20
6-2. Effort-Impact Matrix 기반 실행 우선순위 설정.....	21
6-3. 전략별 세부 설계.....	22
7. Expected Impact & Measurement Plan.....	23
7-1. Expected Impact.....	23
7-2. 성과 측정 계획.....	25
8. Appendix & Future Works.....	27
8-1. 분석 한계점.....	27
8-2. 향후 과제.....	28
8-3. 시각화.....	29

1. Executive Summary

1-1. Problem (핵심 문제)

공유 오피스 시장 점유율 확대를 위한 3일 무료 체험 전략이 기대와 달리 높은 비효율성 문제에 직면해 있습니다. 전체 신청자의 37.81%가 실제 방문하지 않는 No-Show 문제와, 방문객의 58.4%가 단 하루만 이용하고 이탈하는 초기 경험 실패가 핵심 원인으로 작용하고 있습니다. 세부 분석 결과, 미방문 그룹의 전환율(36.57%)과 방문 그룹의 전환율(38.77%)에 큰 차이가 없다는 점 이 확인되었습니다. 이는 마케팅 리소스 낭비가 심각하며, 문제의 본질이 최종 결제 과정보다 방문/재방문 단계의 실패에 있음을 시사합니다. 따라서 단순 모객이 아닌 실질적 전환을 위한 행동 유도 전략의 재설계가 시급합니다.

1-2. Analysis (분석 개요)

본 분석은 trial_visit_info 테이블의 정합성 이슈를 확인하고, trial_access_log를 기준으로 대체 테이블을 구축하여 분석의 신뢰성을 확보했습니다.

로지스틱 회귀 및 트리 기반 앙상블 모델(XGBoost, LightGBM 등)을 활용하여 유료 전환 예측을 시도했으나, 단순 행동 로그만으로는 예측 정확도에 한계가 있음을 확인하고, 분석 방향을 모델 예측에서 핵심 행동 패턴 분석으로 전환하여 전환의 결정 요인을 규명하였습니다.

1-3. Key Finding (주요 발견)

분석 결과, 고객의 전환 행동은 단순 방문 횟수나 체류 시간이 아닌, 행동의 밀도와 시점이라는 두 가지 핵심 임계점에 의해 결정됩니다.

첫 번째, Day 2 분기점입니다. 방문 일수가 늘어날수록 전환율이 상승하지만, 대다수(약 60%)는 Day 1에 이탈합니다. 따라서 2일 차 재방문 성공이 유료 전환을 위한 핵심 분기점입니다.

두 번째, 4시간 임계점입니다. 체류 시간은 1~4시간 구간에서 결제 전환율이 약 45%로 정점을 찍으며 4시간을 초과하면 오히려 효용이 감소합니다. 이는 밀도 높은 4시간 몰입 경험이 중요함을 시사합니다

세 번째, No-Show 역설입니다. 무료 체험 미방문 그룹의 전환율(36.57%)이 실제 방문 그룹(38.77%)과 큰 차이가 없으며 이는 마케팅 자원 낭비와 함께 방문이 필수가 아님을 시사합니다.

네 번째, 공간 및 시점 역설입니다. 방문객 수가 적은 비수기(봄/여름)의 결제 전환율(약 40.9%)이 성수기보다 높으며, 공간 면적과 방문객 수 역시 전환율과 비례하지 않습니다

마지막으로, 이용 시간대 집중입니다. 고객의 핵심 이용이 오후 1시~6시 사이에 집중되고 있으며 이로 인해 발생하는 오전/심야 유휴 시간대가 공간 효율성 극대화 전략의 근거가 됩니다.

1-4. Recommendation (핵심 제안)

체험형 사용자의 행동 패턴을 성공적으로 설계하고, 37.81%의 비효율을 해결하기 위해 온보딩/경험 디자인에 초점을 맞춘 4가지 전략을 제안합니다.

첫 번째로 온보딩 강화는 신청 직후 상세 가이드와 재예약 링크를 제공하여 37.81%의 No-Show 고객을 방문 단계로 유도하고, 마케팅 자원 낭비를 방지합니다.

두 번째로 Day 2 Lock-in은 첫 방문 시 4시간 내 핵심 경험을 유도하는 웰컴 크레딧을 제공하고, 퇴실 시 내일 방문 조건 인센티브를 제공하여 2일 차 재방문을 강하게 유도합니다.

세 번째로 데이터 인프라 구축은 고객 유형을 식별할 수 있는 데이터 수집 체계를 근본적으로 개선하여, 모든 전략의 타겟팅 기반을 마련해야 합니다.

마지막으로 유희 시간대 활용은 EDA 분석 기반으로 비어있는 오전·심야 시간대에 틈새 시간제 상품을 개발하여 공간 활용 효율을 극대화합니다.

2. Background & Problem Definition

2-1. 데이터 배경

현재 서울 시내 전통 오피스 평균 공실률이 3.06%로 작년 1.72% 대비 역대최고치를 기록하고 있습니다. 반면 공유오피스 시장은 다시 부흥하고 있으며 기존 스타트업과 프리랜서의 전유물이 아닌 대기업들의 서브오피스로 각광받고 있습니다.

본 데이터는 고객의 3일 무료체험 신청부터 방문, 결제까지 전 과정을 추적할 수 있어 전환율 개선을 위한 인사이트 도출에 적합합니다. 본 프로젝트의 분석 대상인 Z사는 365일 24시간 무인으로 운영하고 있는 연간 00명 방문하는 규모의 공유 오피스입니다. AI와 IoT기술을 접목하여 자사 앱과 QR코드를 통해 유저의 출입을 관리를 하고 있습니다.

- 운영 방식: 다른 공유오피스, 스터디 카페와 차별화된 자사 전용 크레딧 시스템¹을 운영 중입니다 기본적으로 커피, 차, 다과 및 프린트 서비스는 무료로 제공되지만 안마의자, 회의실, 자판기 등 일부 편의 시설을 크레딧으로 이용하는 구조입니다.
- 비용: 아래의 표는 Z사의 요금을 나타내며 요금결제시 N만큼의 크레딧으로 변환됩니다. 입장 후 일정 시간이 지나면 크레딧은 이용한 시간만큼 차감됩니다. 크레딧을 사용하여 유료음료, 안마의자, 회의실 등 유료서비스 결제에도 사용할수 있습니다.

정액 시간권 (단위: 원)	1시간	2시간	4시간	6시간	9시간	12시간	24시간	50시간
	1,500	3,000	5,000	6,000	8,000	10,000	18,000	35,000

[표 1] Z사의 요금표

본 데이터는 분석 대상자인 Z 사의 공유 오피스 서비스의 2년 8개월(2021.05.01 ~ 2023.12.31)기간의 데이터를 활용하였으며 고객의 여정을 다음과 같이 정의 했습니다.

- 무료체험 신청자: trial_register 테이블의 user_uuid 기준
- 실제 방문자: trial_access_log 테이블의 checkin한 user_uuid기준
- 최종 결제자: trial_payment 테이블의 is_payment 기준

¹ 자사의 화폐 시스템으로, 결제를 통해 자사 화폐인 '크레딧'을 충전하고, 해당 화폐를 통해 오피스 사용료 및 편의시설 이용료를 지불한다.

2-2. 주제 선정 : 무료체험 정말 효과적인가?

	방문 인원	미방문 인원
잔존 인원	39.85% (2,325명)	36.47% (1,327명)
이탈 인원	61.15% (3,660명)	63.53% (2,312명)
전체 인원	62.19% (5,985명)	37.81% (3,639명)

[표 2] 방문/미방문 인원 전환율

공유 오피스의 이용자 확보를 위한 무료 체험 프로그램이 비용 효율성이 매우 낮은 두 가지 핵심 문제에 직면해 있습니다.

1. 예약 후 미사용(No-Show) 비율의 증가는 자원 낭비와 기회비용 상실을 초래한다.

현재 공유 오피스는 혼잡 방지를 위해 무료 체험 신청을 사전에 예약받고 있지만, 예약자 중 상당수가 약속을 지키지 않고 있습니다. 전체 신청자 중 **37.81%**라는 높은 비율이 예약을 해놓고도 실제 방문 및 체험을 이행하지 않았습니다.

이러한 미사용 비율 증가는 해당 시간에 유료 고객으로 전환될 가능성이 높은 잠재 고객에게 공간을 제공할 수 있는 **기회비용을 상실**시킵니다. 또한, 예약 확인 및 관리 등에 투입된 운영 인력의 시간과 공간 리소스까지도 불필요하게 소모하여, 마케팅 활동의 효율성을 크게 떨어뜨리는 주요 원인이 되고 있습니다.

2. 방문 대비 현저히 낮은 유료 고객 전환율은 마케팅 투자의 수익성을 악화시킨다.

무료 체험 프로그램은 잠재 고객을 유료 구독으로 전환하는 것을 목표로 하지만, 현재는 체험 제공 단계에서 수익 창출 단계로의 연결이 매우 미흡합니다. 공간을 실제로 방문한 인원 중 대다수인 **3,660명(61.15%)**이 체험 후 유료 구독으로 전환되지 않고 이탈했습니다.

이처럼 높은 비전환율은 고객 유치 및 체험 제공에 들어간 모든 마케팅 투입 비용이 실질적인 매출로 회수되지 못하고 소진되었음을 의미합니다.

이는 마케팅 투자 대비 수익(ROI)을 현저히 낮추어, 프로그램이 본연의 목적인 이용자 확보와 수익 증대에 기여하지 못하고 비효율성을 야기하는 핵심 요인으로 작용하고 있습니다.

따라서 본 분석은 낮은 전환율 문제를 해결하고 마케팅 효율을 높이기 위해, **실제 구독으로 전환될 가능성이 높은 고객을 사전에 식별하는 것을 핵심 목표**로 삼습니다. 구체적으로는 고객 행동 데이터를 심층 분석하여 구독 전환에 영향을 미치는 핵심 변수를 파악하고, 이를 기반으로 전환 예측 모델을 구축하고자 합니다.

이 예측 모델을 활용하여 낭비되던 마케팅 비용을 유료 고객이 될 가능성이 높은 그룹에게 선별적으로 집중함으로써, 전체적인 전환율과 마케팅 효율성을 혁신적으로 높일 수 있을 것입니다.

2-3. 문제 정의

2021년부터 2023년 까지 무료체험을 신청하고 전환한 유저 3,652명,전환하지 않은 유저 5,972명으로 약 4:6정도의 비율이 나타납니다. 따라서 유료 구독을 결제한 사용자의 특징 중 '총 방문 일수', '체류시간', '최초 방문 지점' 과 같은 행동 패턴을 확인하여 실제로 유료 구독 결제로 이어지게 한 행동패턴을 파악합니다. 확인 된 행동패턴을 통해 유료 결제로 전환할 가능성이 높은 사용자들을 식별하여 집중 마케팅을 통해 가입 전환율을 높이는 마케팅 방안이 필요합니다.

2-4. 분석 목적

본 분석의 목적은 공유 오피스 무료 체험의 **유료 구독 전환율과 마케팅 효율성을 최대화**하는 것입니다.

이를 달성하기 위해, 기존의 비효율적인 마케팅 비용 지출을 초래했던 문제의 근원을 해결하는 데 중점을 두었습니다. 무료 체험 기간 동안 축적된 고객의 행동 로그 데이터를 활용하여 **유료 전환에 결정적인 영향을 미치는 주요 행동 변수를 탐색적 데이터 분석(EDA)를 통해 확인**하였다.

이를 기반으로 예측 모델링에 최적화된 파생 변수(Feature)를 생성하였으며, 모델의 신뢰성을 확보하기 위해 로지스틱 회귀와 같은 베이스라인 모델부터 RandomForest, XGBoost, LightGBM 등 다양한 트리 기반의 앙상블 모델을 단계적으로 구축하고 성능을 비교·검증하였다.

이 분석은 **영향력 있는 행동 변수**들을 파악하고, 이 인사이트를 기반으로 **마케팅 아이템과 전략**을 도출함으로써 공유 오피스의 **유료 결제율을 향상시키고 마케팅 ROI를 개선**하는 것을 최종 목표로 합니다.

3. Analysis Framework & Methodology

3-1. 분석 환경 및 데이터 파이프라인

본 분석에서 사용된 데이터는 공유 오피스 서비스의 2년 8개월(2021.05.01 ~ 2023.12.31) 동안의 데이터로 구조는 아래 표와 같습니다.

	테이블명	테이블 설명	주요컬럼
1	trial_register	3일체험 신청	유저id, 3일체험신청일시
2	trial_visit_info	3일체험 신청자 일자별 방문기록	유저id, 날짜, 지점id, 최초입실시각, 최종퇴실시각, 체류시간
3	trial_access_log	3일체험 신청자 출입기록	유저id, 지점id, timestamp, 출입방향
4	trial_payment	3일체험 신청자 결제 여부	유저id, 결제여부
5	site_area	지점별 면적	지점id, 지점 면적

[표 3] 데이터 테이블 설명 및 주요컬럼

데이터를 이용하여 Google Colab 환경에서 Pandas와 Scikit 등 라이브러리를 사용하여 데이터 전처리를 수행하고, EDA와 머신러닝을 진행하였습니다.

3-2. 데이터 품질 관리

분석의 신뢰도 확보를 위해 원본 데이터에 존재하는 품질 관련 이슈를 식별하여 결측치, 중복값, 컬럼의 단위 변환을 진행하였습니다.

테이블	문제 유형	해결 기준	제거 건수(감소율)
trial_register	1. 모든 컬럼 완전 중복 2. user_uuid 중복이지만 기간이 다른 행	1. 중복제거 2. 오래된 로그 삭제	35행 (약 0.36%)
trial_payment	1. 모든 컬럼 완전 중복	1. 중복제거	35행 (약 0.36%)
trial_access_log	1. 모든 컬럼 완전 중복 2. timestamp 변환 필요 3. 입실, 퇴실 비정상 로그 발생	1. 중복제거 2. KST 변환 3. 입실, 퇴실 페어링 후 제외된 행 삭제	3,342행 (약 5.25%)
trial_visit_info	1. first/last_enter_time 컬럼 결측치 555행 존재 2. trial_access_log와 데이터 불일치	1. 대체 테이블 생성	

[표 4] 데이터 전처리 내역

3-3. trial_visit_info 대체 테이블 생성

3-3-1 trial_visit_info의 데이터 불일치

무료 체험 신청자의 일자별 방문 기록을 요약한 trial_visit_info 테이블은 원본 이벤트 로그(trial_access_log)를 기반으로 생성되었음을 확인했습니다. 그러나 이 테이블을 구성하는 과정에서 **데이터 추출 또는 요약 문제**가 발생하여, 두 데이터 간에 동일 사용자에게 대한 방문 횟수, 체류 시간 등 핵심 정보가 일치하지 않는 **심각한 데이터 불일치가 발견**되었습니다. 세부적인 데이터 검토 결과, trial_visit_info 테이블에는 first_enter_time이 자정(00:00:00)에 기록되거나(747건), last_leave_time이 23:59:59에 기록된(751건) 비정상적이고 가공된 로그들이 확인되었습니다. 이러한 로그들이 원본으로 가정한 trial_access_log에는 존재하지 않았습니다. 이는 데이터 가공 과정에서 요약 오류나 손실이 발생했음을 명확히 시사합니다. 이러한 불일치 상황에서 trial_visit_info를 그대로 활용할 경우 분석 결과가 왜곡될 위험이 매우 높다고 판단하였습니다.

따라서 trial_visit_info의 사용을 전면 배제하고, trial_access_log를 원본 기준으로 삼아 방문 관련 파생 변수들을 직접 생성하여 분석에 최적화된 **대체 요약 테이블을 구축하는 전략을 채택**하였습니다.

3-3-2 trial_access_log 기반 분석용 대체 테이블 구축

앞서 trial_visit_info의 데이터 신뢰도 문제로 인해, 본 분석은 원본 이벤트 로그(trial_access_log)를 기반으로 분석용 **고신뢰도 대체 테이블을 직접 구축**하였습니다. 먼저, trial_access_log 내부에 존재하는 완전 중복 데이터를 제거하고 초기 검토를 통해 확인된 비정상적인 출입 로그를 정제하여 신뢰할 수 있는 이벤트만을 확보하는 원천 데이터 정제 작업을 수행하였습니다. 이후, trial_access_log의 이벤트 로그 구조적 특성을 반영하여 정확한 방문 행동 측정을 위해 입실(Check-in) 이벤트와 퇴실(Check-out) 이벤트를 논리적으로 대응시켜 **세션 단위로 결합하는 과정**을 거쳤습니다. 이 과정에서 유저와 시간을 기준으로 로그를 정렬한 후, 입실 없이 발생한 퇴실, 연속된 중복 입실과 퇴실, 퇴실 없이 끝나버린 입실 등 비정상적인 패턴으로 기록된 행은 모두 제거하였습니다.

또한, 데이터 탐색 과정에서 확인된 14시간 이상의 초장기 체류 데이터는 실제 체류가 아닌 시스템 오류나 태깅 누락으로 인한 잘못된 입퇴실 기록으로 판단하여 **이상치로 규정**하였습니다.

이에 따라, 분석의 정확도를 위해 **단일 방문 세션의 최대 유효 시간을 4시간으로 설정**하여 데이터를 보정하였습니다. 아울러, 날짜가 변경되는 심야 이용 건에 대해서는 입실 당일의 활동으로 간주하기 위해, 전일 입실 후 익일 새벽 5시 이전에 퇴실한 경우 **해당 세션의 체류 시간을 전일(Check-in Date)의 총 체류 시간에 합산되도록 처리**하였습니다.

이렇게 보정된 세션 정보를 기반으로 user_uuid를 기준으로 테이블 구조를 변경하고, 입실/퇴실 세션 시간 차이의 합산인 총 체류 시간(total_stay_time), 세션 개수를 통한 총 방문 횟수(total_visit_cnt)를 포함한 행동 집계 지표를 생성하였습니다. 또한, 첫 입실 시간을 기준으로 계절(season), 연도(year), 월(month), 시간대(hour), 요일(dayofweek) 등의 시계열 파생 변수를 추출하였으며, 최종적으로 결제 테이블과 조인하여 유료 결제 전환 여부(is_payment)를 종속 변수로 확보하였습니다.

이와 같이 trial_access_log를 기반으로 엄격한 정제 및 결합 과정을 거친 대체 테이블은 기존 trial_visit_info로는 불가능했던 방문 시점, 방문 빈도, 누적 체류 시간 같은 다각적이고 신뢰도 높은 행동 지표를 제공하여 분석에 사용할 수 있는 테이블을 성공적으로 마련하였습니다.

3-4. 모델링 구조

본 분석에서는 결제 전환 여부를 예측하는 분류 모델 구축을 위해 모델링 전략을 세우고, 여러 머신러닝 알고리즘의 성능을 비교하여 **사용자의 행동 패턴을 기반으로 전환 가능성을 설명**하고자 하였습니다.

3-4-1. 모델링 전략

모델은 선형 모델(Logistic Regression)부터 트리 기반의 앙상블 모델(Random Forest, AdaBoost, XGBoost, LightGBM)까지 총 5개의 분류 모델을 비교하였습니다.

이는 모델 구조 다양성을 확보하고, 모델 편향을 줄여 전환 행동을 설명하는 핵심 변수를 보다 신뢰성 있게 도출하기 위함입니다. 사용한 모델은 다음과 같습니다.

(1) Logistic Regression : 기본 모델인 Logistic Regression 모델은 해석력 중심 모델로, 성능 최적화를 위한 튜닝 없이 사용하였으며, 보다 복잡한 비선형 구조를 학습하는 앙상블 모델들과의 비교를 위한 기준점 역할을 수행합니다.

(2) Random Forest : 랜덤 포레스트 모델은 Bagging 기반의 모델로, 여러 개의 의사결정 나무(Decision Tree)를 독립적으로 학습시켜 평균화함으로써 과적합을 방지하는 효과가 있습니다. 또한 비선형 패턴과 피쳐 간 복잡한 상호작용을 학습할 수 있어, 기본 모델 대비 높은 예측 성능을 기대할 수 있기 때문에 사용하였습니다.

(3) Boosting Model : 이전 모델의 오차를 보완하면서 점진적으로 성능을 개선하는 부스팅 모델을 사용함으로써 더 높은 표현력과 예측 성능을 기대할 수 있습니다. 사용한 모델은 부스팅 모델 중 가장 기반이 되는 AdaBoost, 과적합을 억제하면서 높은 예측 성능을 보여주는 XGBoost, 대규모 데이터 및 범주형 변수 처리에 강점을 갖고 있는 LightGBM을 사용하였습니다.

3-4-2. 클래스 불균형 처리

결제 전환과 미전환 간의 비율에 차이가 존재(1:1.574)하는 불균형 데이터 구조를 고려하여, 모델 학습 시 소수 클래스(전환 고객)의 손실을 더 크게 반영하는 가중치 조정 방식(class_weight / scale_pos_weight)을 각 모델에 적용하였습니다.

(1) Logistic Regression / RandomForest / DecisionTree(AdaBoost의 base_estimator):

class_weight='balanced' 적용하여 클래스 빈도(1 / 클래스 빈도) 기반 자동 가중치 부여

(2) XGBoost / LightGBM : scale_pos_weight = (미전환 수 / 전환 수) 방식으로 양성 클래스의 중요도를 조정

이를 통해 모델이 소수 클래스(결제 전환)를 더 잘 탐지할 수 있도록 설계하였습니다.

3-4-3. 하이퍼파라미터 자동 탐색

모델 성능을 효율적으로 개선하기 위해 본 분석에서는 Optuna²를 활용한 자동화 하이퍼파라미터 탐색을 적용하였습니다.

	모델명	하이퍼파라미터
1	RandomForest	트리 개수(n_estimators), 트리 깊이(max_depth), 분기 조건(min_samples_split·leaf)
2	AdaBoost	약한 학습기(DecisionTree)의 max_depth, 학습률(learning_rate), 반복 횟수(n_estimators)
3	XGBoost	학습률(learning_rate), 추출 비율(subsample), 정규화(reg_alpha, reg_lambda)
4	LightGBM	학습률(learning_rate), 추출 비율(subsample), 최소 샘플 수(min_child_samples)

[표 5] 모델별 주요 하이퍼 파라미터

² Optuna: 탐색 공간을 설정해두면 이 중 성능을 극대화하는 조합을 반복적으로 찾아주는 Bayesian Optimization 기반 프레임워크로, 모델별로 가장 적합한 구조적 파라미터를 자동으로 탐색하는 데 강점을 지닌다.

3-4-4. 교차 검증

모델의 일반화 성능을 안정적으로 평가하기 위해 Stratified K-Fold(5-Fold)³ 방식을 사용하였습니다. 모든 모델은 동일한 교차검증 환경에서 학습 및 검증되었으며, 최종 평가는 cross_val_predict 기반 확률 예측을 활용하여 일관된 비교가 가능하도록 구성하였습니다.

3-5. 평가 지표 채택 및 모델별 성능 비교

평가지표는 무료 체험 고객 중 유료 전환 가능성이 있는 고객을 놓치지 않는 것을 목표로 하면서 모델이 일정 수준 이상의 성능을 확보하면, 해당 모델의 특성 중요도(feature importance)를 활용하여 유료 전환에 영향을 미치는 고객 행동 패턴을 규명하는 것이 궁극적인 목적이므로, 양성 클래스(유료 전환 고객)에 대한 재현율(Recall)을 주요 기준으로 설정하고, 보조지표로는 ROC-AUC와 F1 Score를 활용하였습니다.

ROC-AUC는 특정 임계값에 의존하지 않고 모델의 전반적인 분류 성능을 평가할 수 있는 지표로, 양성·음성 클래스 간 구분 능력을 종합적으로 판단하는 데 유용합니다. 이는 다양한 머신러닝 모델 간 성능 비교의 신뢰성을 높일 수 있습니다.

F1 Score는 정밀도(Precision)와 재현율의 비율을 고려하여 계산한 지표로, 모델이 단순 성능 개선을 위해 과도하게 공격적으로 임계값을 낮추어 오탐지값이 늘어나는 경우, 정밀도 역시 과도하게 낮아지게 됩니다. 이러한 상충관계를 결과로 나타내어 마케팅적 의사결정을 지원하기 위해서 보조 지표로 설정했습니다.

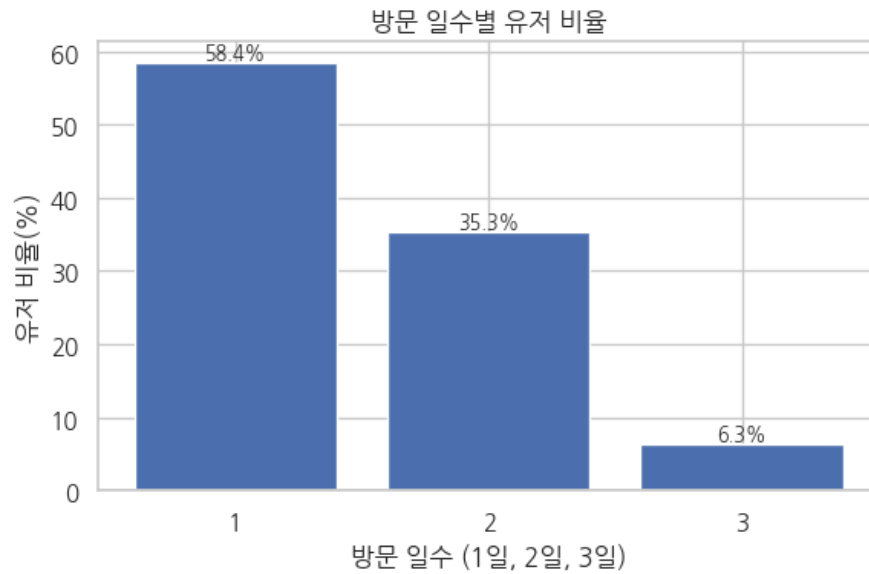
	모델명	재현율(Recall)	ROC-AUC	F1 Score
1	Logistic Regression	0.5394	0.6045	0.5004
2	RandomForest	0.5054	0.6098	0.4937
3	AdaBoost	0.7191	0.5556	0.5330
4	XGBoost	0.5471	0.6129	0.5113
5	LightGBM	0.5462	0.6082	0.5100

[표 6] 모델별 최종 성능 비교(최고점, 최저점)

³ Stratified K-Fold: Stratified K-Fold 방식은 각 fold가 원본 데이터의 클래스 비율을 동일하게 유지하도록 분할되므로, 불균형 데이터 상황에서 성능 평가의 편향을 최소화하는 데 적합하다.

4. Key Findings

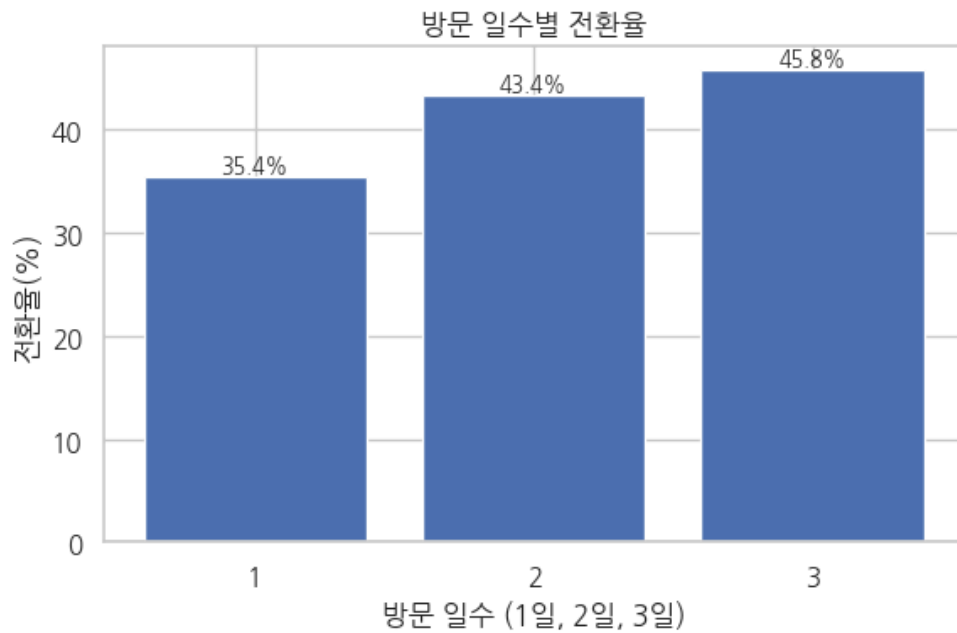
4-1. 결제 전환율: 방문 일수별 결제 전환율 차이



[그래프 1] 방문 일수별 유저 비율

탐색적 데이터 분석 결과 첫번째로 전체 방문객의 약 58%가 3일 무료 체험 기간 중 단 하루(Day 1)만 이용한 후 서비스를 이탈하는 것으로 나타났습니다. 반면, 3일 체험권을 모두 활용하는 고객은 6%에 불과했습니다.

이는 고객들이 3일권을 온전히 활용할 필요성을 느끼지 못하고 있으며, 현재의 3일 무료 체험권이라는 마케팅 문구 자체가 가치를 제공하지 못하고 비효율적으로 소모되고 있음이 확인되었습니다.



[그래프 2] 방문 일수별 전환율

두번째로 방문 일수가 1일에서 2일, 3일로 늘어날수록 유료 결제 전환율이 비례하여 상승한다는 것이 확인되었습니다. 이는 고객이 우리 공간에 머무르는 시간이 길어질수록 서비스 가치를 느끼는 아하 모먼트(Aha Moment)를 경험할 가능성이 커짐을 명확히 시사합니다.

하지만 현재 대다수의 고객(약 60%)은 이 설득의 시간을 갖기도 전에 첫날 이탈하고 있습니다. 따라서 첫 방문 이후 2일 차 재방문 성공 여부가 유료 전환을 위한 결정적인 분기점임을 확인되었습니다.

4-2. 결제 전환율 : 방문/미방문 그룹별 결제 전환율 차이

visit_group	잔존율(결제율)
미방문	36.57%
방문	38.77%

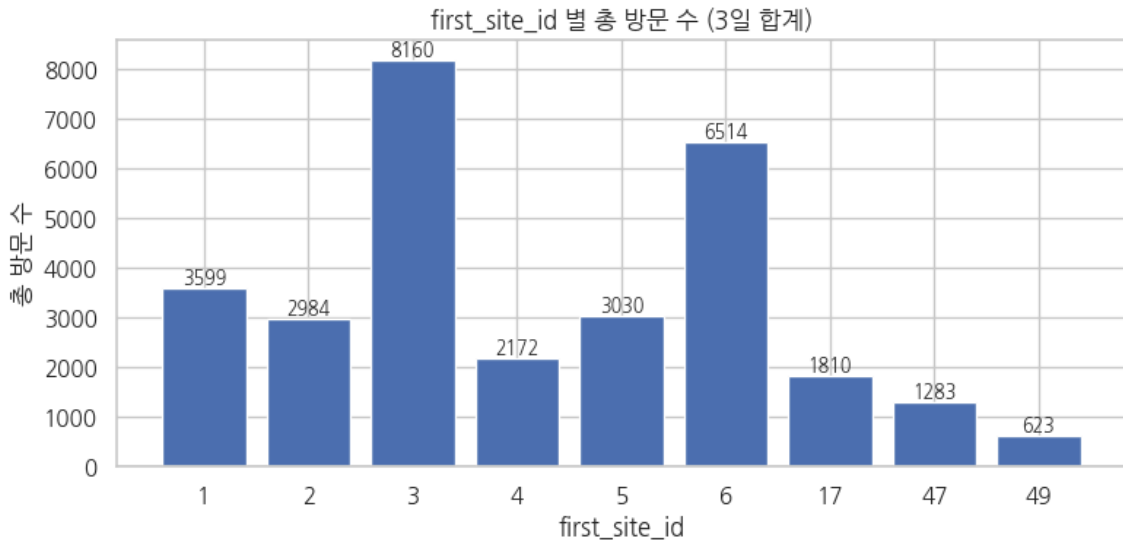
[표 6] 방문/미방문 그룹별 결제 전환율

무료 체험을 신청하고도 실제로는 단 한 번도 방문하지 않은 고객의 비율이 전체 신청자의 약 37%에 달하는 것으로 확인되었습니다.

이는 마케팅 리소스의 1/3 이상이 비효율적으로 소모되고 있음을 의미하며, 이는 마케팅 리소스의 3분의 1 이상을 비효율적으로 소모하는 동시에, 공간 운영의 효율성 및 수요 예측 가능성을 저해합니다. 또한 이 미방문 그룹의 결제 전환율(36.57%)이 실제 방문 그룹의 전환율(38.77%)과 비교했을 때 통계적으로 유의미한 차이가 없음이 확인되었습니다. 즉, 직관과 다르게 무료 체험 방문 여부 자체가 결제의 필수 조건이 아님을 의미합니다.

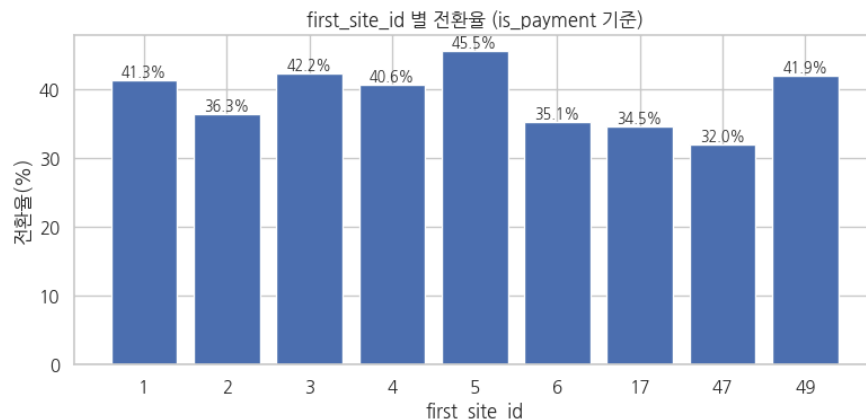
이는 고객들이 3일 체험을 통해 확신을 얻고 결제하는 것이 아라, 이미 우리 브랜드에 대한 충분한 신뢰를 가지고 있거나 온라인상의 정보만으로도 구매 결정을 내리고 있음을 분석 결과가 시사합니다.

4-3. 결제 전환율 : 공간 면적과 전환율의 비례성 부재



[그래프 3] 지점별 총 방문 수

지점별 이용률을 분석한 결과, 3호점(150평), 6호점(150평), 1호점(50평) 순으로 높았습니다. 특히 **1호점은 평수가 50평으로 작음에도 불구하고**, 대형 지점(100~150평)인 4, 5호점보다 **높은 이용률**을 보였습니다. 이는 단순히 공간이 넓다고 해서 이용자가 몰리는 것이 아니라, 지점의 위치나 내부 구조와 같은 **다른 변수가 이용률에 더 중요하게 작용**하고 있음을 시사합니다.

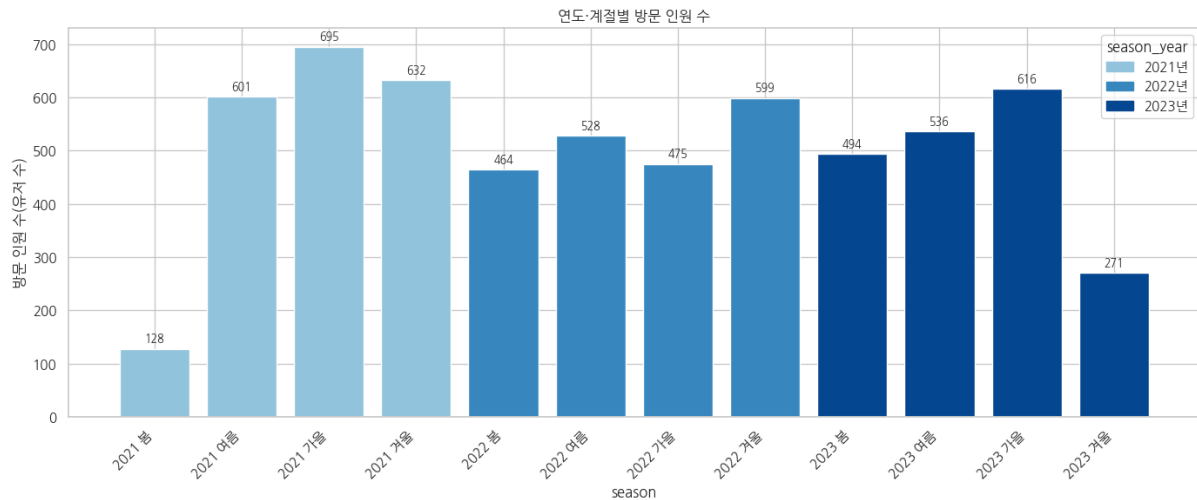


[그래프 4] 지점별 결제 전환율

지점별 성과 분석 결과, 방문객 수와 유료 결제 전환율은 비례하지 않는 것으로 나타났습니다. 대형 평수인 3호점과 6호점은 가장 많은 방문객을 유치했으나, 실제 결제로 이어지는 비율은 평균 수준에 머물러습니다. 1호점(50평, 소형) 역시 방문객은 많지만, 전환율은 5호점(150평, 대형)보다 낮았습니다.

반면, 5호점은 방문객 수는 평균 수준임에도 불구하고, 45.5%라는 압도적인 전환율 1위를 기록했습니다. 이는 단순히 사람이 많이 오는 지점이 장사가 잘되는 지점은 아니라는 것을 명확히 보여주며, 공간 확장이 무조건적인 정답이 아님을 입증합니다.

4-4. 결제 전환율 : 시즌별 결제 전환율 분석



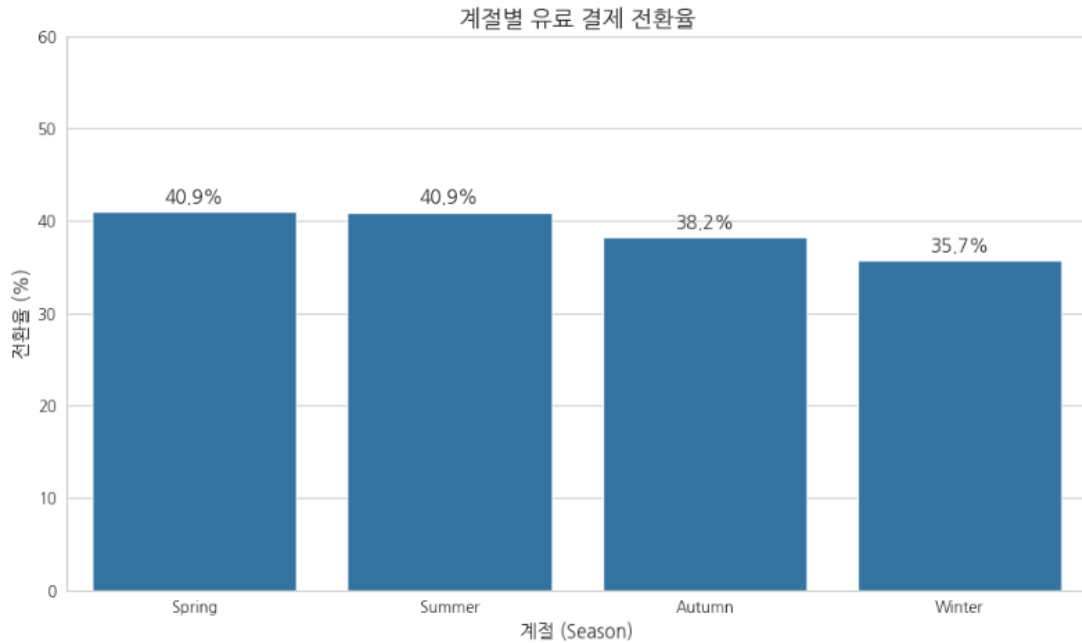
[그래프 5] 시즌별 총 방문 횟수

연도별 시계열 데이터를 상세 분석한 결과, 공유 오피스 수요는 연중 특정 시기에 따라 급변하는 뚜렷한 계절성을 보였습니다.

매년 봄 시즌에는 신규 유입 및 실제 이용률이 연중 최저 수준으로 시작되는 반면, 가을시즌에 이르러서는 이용객 수가 급증하며 연간 최대 실적을 달성하는 패턴이 3년 연속 반복해서 관측되었습니다.

이러한 패턴이 매년 일정하게 관측된다는 점은 봄 시즌의 상대적 실적 저조가 일시적인 영업 부진이 아니라, 시기에 따라 **수요가 변동하는 구조적인 계절성**임을 시사합니다.

따라서 해당 시기의 트래픽 감소를 단순한 성과 실패로 해석하기보다는, 연중 반복되는 수요 사이클로 인지하고 시기별 변동성에 선제적으로 대응할 수 있는 운영 전략 수립이 필요합니다.

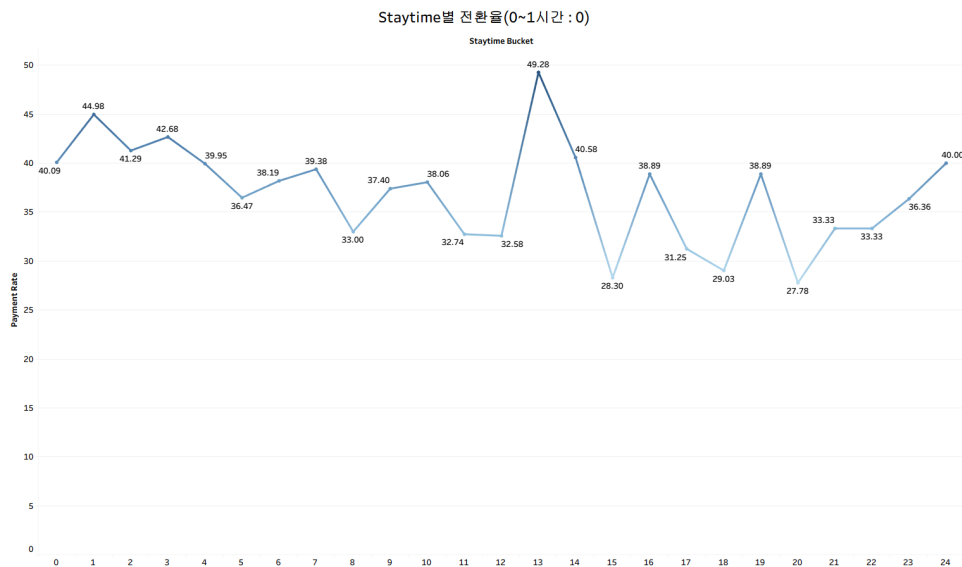


[그래프 6] 계절별 유료 결제 전환율

시계열 분석 결과, 봄과 여름은 연중 방문객 수가 가장 적은 비수기임에도 불구하고, 결제 전환율 측면에서는 오히려 성수기(가을/겨울)보다 높은 수치를 기록하는 역설적인 패턴이 확인되었습니다. 데이터상으로 방문객 유입이 급증하는 하반기(성수기)에는 전환율이 상대적으로 하락하는 반면, 방문객 수가 적은 상반기(비수기)에는 전환율이 높게 유지되는 경향이 뚜렷하게 관측됩니다.

이는 단순히 방문객 모수가 많을수록 전환 성과가 비례하여 상승하는 것이 아니며, **유입 시기에 따라 결제 전환의 효율성에 유의미한 차이가 존재함**을 데이터가 시사합니다.

4-5. 결제 전환율 : 체류시간 별 결제 전환율 차이



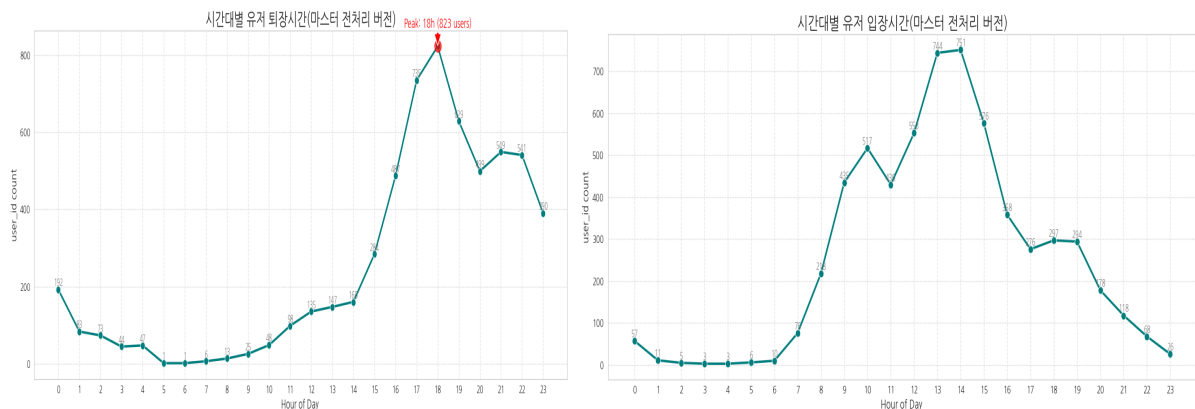
[그래프 7] 체류 시간별 결제 전환율

체류 시간이 길어질수록 결제 전환율이 비례하여 상승할 것이라는 일반적인 예상과 달리, 입실부터 퇴실까지의 체류 시간이 1시간~4시간인 구간에서 결제 전환율이 약 45%로 정점을 기록하는 패턴이

확인되었습니다. 반면, 4시간을 초과하여 장시간 체류하는 경우 전환율은 오히려 30%대로 정체되거나 하락하는 경향을 보였습니다.

이는 고객이 서비스의 효용을 충분히 체감하고 구매를 결정하기까지의 **단일 방문 유효 시간이 최대 4시간**임을 시사합니다. 즉, 4시간 이후부터는 체류 시간의 증가가 더 이상 구매 전환율 상승으로 이어지지 않으므로, 해당 시점이 구매 전환을 위한 세션 임계점(Threshold)으로 분석됩니다.

4-6. 이용 시간대 분석: 출입 시간의 비대칭적 이중 피크 패턴



[그래프 8] 출입시간대별 방문 인원

시간대별 분석 결과, 고객의 출입 행동은 입장(Check-in) 시점의 이중 피크와 퇴장(Check-out) 시점의 단일 피크라는 비대칭적인 패턴을 보였습니다.

입장 시간 분석 결과, 오전 9시~10시 사이에 첫 번째 유입 피크(439명~517명)가 나타나 주간 활동 시작 시점과 유사한 경향을 보였습니다. 그러나 하루 중 가장 높은 유입 피크는 오후 1시~2시(744명~751명) 사이에 집중되었는데, 이는 점심시간 이후 오후 집중 업무나 학습이 시작되는 패턴과 관련됨을 시사합니다.

반면, 퇴장 시간은 이와 달리 오후 6시(18시)에 일 최대인 823명으로 압도적인 단일 피크를 기록했습니다. 이는 체험 고객의 핵심 그룹이 공간 이용을 정규 업무 종료 시점(18시)에 맞춰 마무리함을 명확히 나타냅니다.

5. Modeling Results & interpretation

5-1. 피쳐 엔지니어링 및 마스터 테이블 구성

앞선 4장에서 도출된 주요 분석 인사이트를 기반으로, 결제 전환 예측에 기여도가 높을 것으로 **예상되는 핵심 변수를 선별**하였습니다.

이를 머신러닝 모델 학습에 최적화된 형태로 가공하여 최종 마스터 테이블을 구성하였습니다.

(1) 방문 빈도 변수: 무료 체험 기간동안 방문한 빈도와 관련된 지표입니다. 사용자의 방문 강도·활동성은 결제 전환 가능성을 가장 직접적으로 설명할 수 있는 행동 지표이기에 이에 따라, 일자별 방문 횟수, 참석 여부를 집계하고 사용자별 평균 방문 일 수 및 빈도를 산출하였습니다. 이는 무료 체험 기간 중 사용자가 공간을 얼마나 자주 이용했는지, 즉 서비스 적합도, 만족도, 니즈 강도를 판단하는 중요한 단서가 되며, 방문 빈도가 높을수록 유료 전환 확률이 높아질 가능성이 있을 것이라 판단하여 모델에 반영하였습니다.

(2) 체류 시간 변수: 체류 시간은 단순 방문보다 사용자의 몰입도를 나타내는 더 잘 보여주는 변수이기 때문에, 이용자가 한 번 방문했을 때 얼마나 오래 머무르는지, 그리고 이 패턴이 일정한지를 모델에게 학습시키고자 변수로 채택하였습니다.

(3) 시계열 변수: 사용자의 활동 시간대와 계절적 요인은 공간 이용 목적과 패턴을 설명하는 중요한 맥락입니다. 특히, 출근 시간대·업무 시간대·야간 시간대 등 이용 목적이 명확히 구분되는 서비스 특성을 고려하여 범주화한 변수를 사용하고, 이외에 시간 관련 변수를 주기적 특성을 가진 형태로 변환 후 사용하였습니다. 해당 변수들을 통해 사용자가 어느 시간대에 공간을 활용하고, 계절이나 월별 패턴에 수요 변동성이 존재하는지를 모델이 학습할 수 있도록 하였습니다.

(4) 방문 지점 변수: 사용자가 어떤 지점을 처음 이용했는지는 고객의 니즈와 선호도를 추정하기 위해 사용했으며, 주거지 근처 지점인지, 도심 오피스 중심지인지에 따라 이용 목적이 크게 달라질 수 있기 때문에, 변수로 사용하였습니다. 이는 특정 지점의 서비스 품질·위치적 특징이 유료 전환에 영향을 미치는지 평가하는 근거가 될 수 있을 것입니다.

5-2. 사용 모델 비교 및 최종 모델 선정

	모델명	재현율(Recall)	ROC-AUC	F1 Score
1	Logistic Regression	0.5394	0.6045	0.5004
2	RandomForest	0.5054	0.6098	0.4937
3	AdaBoost	0.7191	0.5556	0.5330
4	XGBoost	0.5471	0.6129	0.5113
5	LightGBM	0.5462	0.6082	0.5100

[표 7] 모델별 최종 성능 비교(최고점, 최저점)

앞서 3장에서 수립한 분석 프레임워크를 기반으로, 기준 모델인 로지스틱 회귀(Logistic Regression)를 포함한 총 5개의 분류 모델(Random Forest, AdaBoost, XGBoost, LightGBM)을 동일한 학습 환경 및 교차 검증 조건 하에서 비교 평가하였습니다.

평가 결과로 모든 모델의 ROC-AUC가 0.55~0.61 구간에 분포하여, 결제 전환 유무를 명확하게 판별하는 예측 성능은 다소 제한적인 수준임이 확인되었습니다.

이는 특정 모델의 알고리즘적 결함보다는 다음과 같은 데이터의 구조적 한계에서 기인한 것으로 분석됩니다.

관찰 기간의 제약 : 3일이라는 짧은 무료 체험 기간 동안 축적된 로그만으로는 고객의 구매 의사를 확정적으로 예측하기에 정보량이 제한적입니다.

정성적 핵심 변수의 부재 : 실제 결제 의사결정에 결정적인 영향을 미치는 고객의 구체적 니즈, 가격 민감도, 경쟁 서비스 비교 경험 등의 정성적 요인이 현재의 로그 데이터에는 부재합니다.

이에 따라 본 분석에서는 특정 모델을 최종 예측 솔루션으로 선정하는 것보다, 모델이 공통적으로 주목한 핵심 변수를 식별하는 것으로 분석의 주안점을 전환하였습니다.

즉, 예측의 정확도(Accuracy)보다는 전환 행동을 설명하는 설명력에 집중하여, 여러 모델이 일관되게 중요하다고 판단한 행동 지표를 도출하고 이를 마케팅 전략의 근거로 삼는 것이 타당하다고 판단하였습니다.

5-3. 피쳐 중요도 분석

순위\모델	RandomForest	AdaBoost	XGBoost	LightGBM
1	day1_staytime	avg_stay_time	visit_days	avg_stay_time
2	avg_stay_time		day2_attended	day1_staytime
3	day2_staytime		timeofday_morning_commute	day2_staytime
4	hour_cos		timeofday_evening	hour_cos
5	month_cos		hour_cos	month_cos
6	hour_sin		avg_stay_time	dayofweek_sin
7	visit_days		first_site_id_6	month_sin
8	day2_visit_cnt		day1_staytime	day1_visit_cnt
9	dayofweek_sin		day2_staytime	dayofweek_cos
10	avg_visit_cnt		first_site_id_5	day3_staytime

[표 8] 모델별 피쳐 중요도 정리표 (색깔별 변수 분류: 체류 시간 변수, 시계열 변수, 방문 빈도 변수)

앞선 5-2절에서 수립한 분석 방향에 따라, 4가지 트리 기반 앙상블 모델(Random Forest, AdaBoost, XGBoost, LightGBM)을 대상으로 각각 상위 중요 변수를 추출하여 교차 비교를 수행하였습니다.

분석 결과, 각 모델의 알고리즘적 특성에 따라 변수별 상대적 가중치에는 다소 차이가 있었으나, 체류 시간, 방문 빈도, 특정 시간대 활용과 관련된 주요 행동 피쳐들이 모든 모델에서 공통적으로 상위권에 포진하는 뚜렷한 경향성이 확인되었습니다.

이는 결제 전환 여부가 특정 단일 요인이 아닌, 고객의 몰입도와 활동성, 그리고 이용 맥락이라는 **복합적인 행동 패턴에 의해 결정됨**을 데이터가 일관되게 지지하고 있음을 시사합니다.

5-4. 결과 해석

분석 결과, 전반적인 ROC-AUC가 0.58 내외에 머무르며 결제 전환 여부를 높은 정확도로 판별하기에는 데이터의 정보량이 제한적이라는 **데이터의 구조적 한계가 확인**되었습니다. 특히 무료 체험 서비스의 특성상, 실제 전환 의사결정에 결정적인 영향을 미치는 고객의 내재적 니즈, 가격 민감도, 타 서비스 비교 경험 등 정성적 외부 변수가 로그 데이터에 부재했던 점이 **예측 성능 제약의 주된 원인**으로 판단됩니다.

그러나 이러한 정량적 예측의 한계에도 불구하고, 4가지 앙상블 모델(Random Forest, AdaBoost, XGBoost, LightGBM)이 도출한 피처 중요도(Feature Importance)는 매우 일관된 패턴을 보였습니다. 분석 결과, 고객의 전환 의사결정은 크게 몰입도, 활동성, 이용 맥락이라는 세 가지 핵심 차원에 의해 설명됨이 확인되었습니다.

구체적으로는 고객이 서비스에 얼마나 깊이 관여했는지를 보여주는 평균 체류시간(avg_stay_time) 및 일자별 체류 시간(day1_staytime, day2_staytime)이 가장 중요한 변수로 지목되었습니다. 그 뒤를 이어 적극적인 이용 의지를 대변하는 일 평균 방문 횟수(avg_visit_cnt)와 총 방문 일수(visit_days)가 주요한 영향력을 보였으며, 고객의 라이프스타일과 주기적 특성을 반영하는 주 이용 시간대(timeofday) 및 시계열 순환 변수(hour_sin, month_cos)가 상위권에 포진하며 유료 전환에 유의미한 영향을 미치는 것으로 나타났습니다.

이러한 공통 변수들의 발견은 사용자의 핵심 행동 패턴이 전환 의사결정과 밀접하게 연동되어 있음을 통계적으로 입증합니다. 즉, 누가 결제할 것인가를 완벽히 예측하지는 못했으나, **어떤 행동 패턴을 보인 고객이 결제하는가에 대한 인사이트를 도출**할 수 있었다.

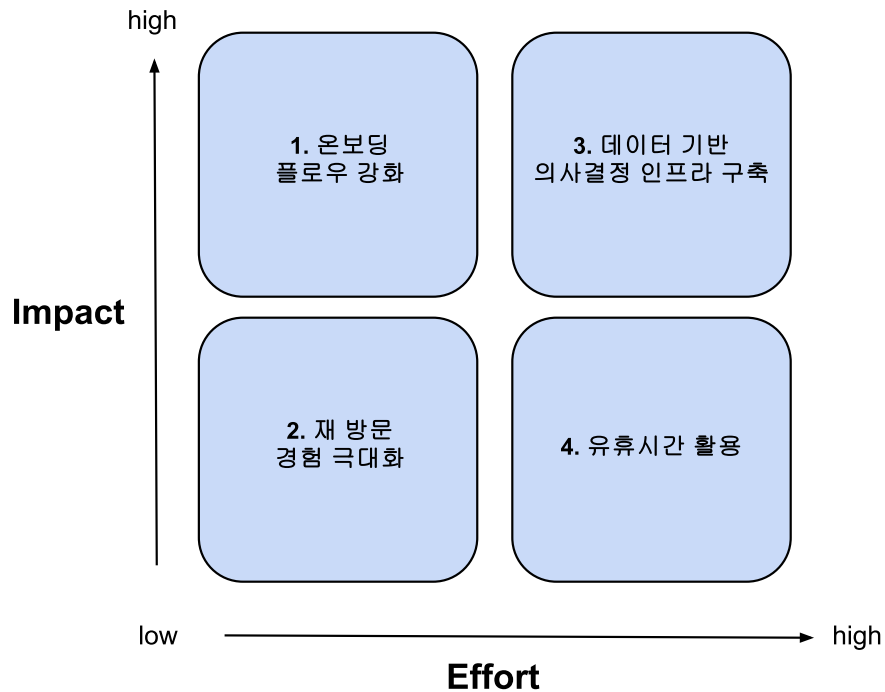
6. Strategic Recommendations

6-1. 분석 결론: 3일 무료가 아닌 정교한 경험 설계로의 전환

EDA와 머신러닝 모델링 분석을 종합한 결과, 현재의 3일 무료 체험 정책은 실제 고객의 이용 패턴과 맞지 않는 비효율적인 구조임을 확인하였습니다. 전체 방문객의 58.4%가 체험 첫날(Day 1)만 이용하고 이탈하며, 무료 체험을 신청하고도 방문하지 않는 **No-Show 비율**이 37.81%에 달해 막대한 마케팅비용이 낭비되고 있습니다. 또한, **체류 시간 4시간** 이후부터는 결제 전환율이 정체되거나 하락하는 패턴과, 방문하지 않은 그룹의 전환율(36.57%)이 방문 그룹(38.77%)과 큰 차이가 없다는 점은 단순한 공간 제공만으로는 구매를 설득할 수 없음을 시사합니다.

따라서 향후 전략의 핵심 목표는 막연하게 3일이라는 시간을 제공하는 현행 마케팅에서 벗어나, 데이터가 지목한 병목 구간을 해결하는 것입니다. 즉, No-Show 고객을 첫 방문으로 이끌고, 첫 방문 고객을 이튿날 다시 오게 만드는 **정교한 1+1일 경험 설계로의 전환**을 제안한다.

6-2. Effort-Impact Matrix 기반 실행 우선순위 설정



매트릭스 설계 기준은 방문률, 결제 전환 등 핵심 지표 개선에 직접적인 영향을 주는지에 대한 Impact와 설계, 개발, 운영 등 리소스나 시간, 비용에 대한 Effort를 통해 전략들을 상대적으로 비교하였습니다. 그리고 해당 내용을 통해 전략별 실행 순서를 선정하였습니다.

6-2-1. 선정이유

(1) No-Show 고객을 첫 방문으로 (온보딩 플로우 강화)

선정 이유: 무료 체험 신청자의 37.81%에 달하는 미방문(No-Show)은 마케팅 리소스 누수가 가장 심각한 구간입니다. 데이터 트래킹과 같은 대규모 개발 없이도, 안내 메시지와 재예약 링크 제공만으로 즉각적인 방문율 개선(Impact High, Effort Low)이 가능하기에 최우선 과제로 선정했습니다.

(2) 재방문 경험 극대화 (1일 차 이탈을 Day 2 재방문으로)

선정 이유: 온보딩으로 유입된 고객을 실제 결제로 연결하는 핵심 단계입니다. 분석 결과 2일 차 재방문이 유료 전환율 급상승의 결정적 분기점임이 확인되었습니다. 방문객의 58.4%가 1일 차에 이탈하는 문제를 해결하는 것이 매출 직결 효과(Impact High)가 가장 크므로 2순위로 배치했습니다.

(3) 초개인화 마련 (데이터 기반 의사결정 인프라 구축)

선정 이유: 앞선 1, 2순위 전략의 효율을 극대화하고 장기적인 성장을 위해서는 누가 우리의 핵심 고객인가를 식별할 수 있어야 합니다. 이는 전사적인 시스템 개편이 필요한 중장기 과제(Effort High)이므로, 단기 성과 전략(1, 2순위)과 병행하거나 후속 단계에서 타겟팅 정교화를 위해 실행하는 것이 효율적입니다.

(4) 유티 시간 활용 (공간 효율 극대화)

선정 이유: 서비스 이용이 오후 1시~6시에 편중되어 발생하는 공간의 비효율을 해결하는 전략입니다. 다만, 이는 주 고객층 파악(3순위)이 선행되어야 타겟팅이 정확해지며, 핵심 퍼널개선 이후의 추가 매출 확보 차원이므로 4순위로 배치했습니다.

6-3. 전략별 세부 설계

6-3-1. No-Show 고객을 첫 방문으로 : 온보딩 플로우 강화

EDA 분석 결과, 전체 신청자의 37.81%에 달하는 미방문 그룹은 서비스를 전혀 경험하지 않았음에도 불구하고 36.57%라는 높은 전환율을 기록했습니다. 이는 방문 그룹의 전환율(38.77%)과 비교해도 통계적으로 큰 차이가 없어, 이들이 이미 우리 브랜드에 높은 신뢰와 니즈를 가진 잠재 핵심 타겟임을 의미합니다. 따라서 이들을 실제 방문단계로 진입시키는 것만으로도, 추가적인 마케팅 비용 없이 가장 확실하게 성과를 개선할 수 있습니다.

세부 실행 방안으로 신청 직후 발생하는 **정보의 불확실성을 제거**합니다. 단순한 예약 확정 문자 대신, 지점 위치와 주차/교통편, 비치된 물품(모니터, 케이블 등) 정보가 담긴 상세 가이드를 발송하여 방문에 대한 심리적 장벽과 막연한 불안감을 해소합니다.

두 번째로 예약일 하루 전 리마인드 메시지 발송 시 **일정 변경링크를 함께 제공**하여 유연성을 부여합니다. 이는 고객이 개인 사정으로 방문하지 못할 때 체험 자체를 포기하는 대신, 날짜를 조정하여 기회를 유지하도록 유도함으로써 No-Show를 방지합니다.

마지막으로, 체험 기간 종료 시까지 방문 기록이 없는 고객에게는 자동 종료 알림이 아닌 **체험 기회 1회 연장 또는 재예약 링크를 발송**하여 다시 한번 방문 동기를 자극하는 재참여 트리거(Trigger)를 작동시킵니다.

6-3-2. 재 방문 경험 극대화: 1일 차 이탈을 Day 2 재방문으로

데이터에 따르면 방문 일수가 1일일 때의 전환율(35.4%) 대비, 2일 차 방문 시 전환율은 43.4%로 급격히 상승합니다. 또한 머신러닝 모델의 피쳐 중요도 분석에서도 visit_days(방문 일수)와 day2_attended(2일 차 출석 여부)가 전환을 예측하는 핵심 변수로 지목되었습니다. 즉, 2일 차 재방문이야말로 유료 전환을 결정짓는 가장 중요한 분기점입니다. 그러나 현재 58.4%의 고객은 1일 차 이용 후 이탈하고 있어, 이들을 2일 차 방문으로 유도하는 장치가 시급합니다.

세부 실행 방안으로 4시간이라는 골든타임 내에 서비스 가치를 체감하게 하고, 즉시 재방문 동기를 부여하는 것입니다. 먼저 첫 방문 고객에게 커피 외에도 안마의자, 회의실 등 유료 시설을 즉시 이용할 수 있는 **웰컴 크레딧(Welcome Credit)**을 제공하여, **4시간 임계점** 내에 밀도 높은 '아하 모먼트(Aha Moment)'를 경험하도록 지원합니다.

두 번째로 1일 차 퇴실 시점에는 Day 2 락인 인센티브를 제공합니다. 이는 **다음날 재방문 시에만 사용 가능한** 추가 크레딧이나 제휴 쿠폰형태로 발송되며, 사용자에게 내일 다시 와야 할 명확하고 즉각적인 이유를 제시하여 1일 차 이탈을 방지하는 강력한 유인책이 될 것입니다.

6-3-3. 초개인화 데이터 마련 : 데이터 기반 의사결정 인프라 구축

데이터 분석 과정에서 지점별 성과가 평수와 비례하지 않고, 비수기인 봄철에 오히려 전환율이 높게 나타나는 등 흥미로운 패턴들이 발견되었습니다. 그러나 현재 데이터에는 고객의 이용 목적 유형정보가 부재하여, 이러한 현상이 발생하는 구체적인 원인을 규명하거나 타겟팅 전략을 세우는 데 한계가 있었습니다. 장기적인 성장을 위해서는 감이 아닌 데이터로 의사결정 할 수 있는 인프라가 필수적입니다.

세부 실행 방안으로 **고객 프로필 데이터의 확보**입니다. 무료 체험 신청 폼을 개선하여, 고객에게 부담을 주지 않는 선에서 이용 목적, 직업군, 예상 이용 시간대 등 **핵심 식별 정보를 필수적으로 수집**할 수 있도록 체계를 개편합니다.

이렇게 확보된 데이터를 바탕으로 가설 검증 및 틈새 상품 개발을 추진합니다. key-finding에서 확인된 계절적 수요(봄 시즌)나 4시간 이용 패턴을 결합하여 시즌 한정 멤버십 같은 신규 상품을 기획·검증합니다.

6-3-4. 유휴시간 활용 : 공간 효율 극대화

데이터 분석 결과로 시간대 분석 결과는 서비스 이용의 핵심 구간이 오후 1시부터 6시까지로 편중되어 있음을 명확히 보여줍니다.

이는 24시간 운영되는 공간이 오전 시간(00시~08시)이나 저녁 피크 이후(20시~23시)에 구조적인 유휴상태에 놓여 있음을 의미합니다. **공간 자원의 가용 시간을 활용**하여 단위 시간당 매출 효율을 높이는 전략이 필수적입니다.

세부 실행방안으로 공간 자원의 비효율을 해소하기 위해 **틈새 수요를 공략하는 시간제 상품**을 개발합니다. 유입이 적었던 시간대(오전 08:00~12:00 또는 심야 22:00 이후)를 타겟으로 모닝 집중 패스나 심야 스터디 패스와 같이 시간대와 목적을 명확히 한 상품을 기획합니다.

이와 함께, 24시간 운영되는 지점에 한하여 심야 시간 이용 고객을 위한 무인 체크인/체크아웃 시스템 및 보안 강화 등 안정적인 이용 환경을 구축하여 서비스의 가동률을 높여야 합니다.

7. Expected Impact & Measurement Plan

7-1. Expected Impact

앞서 언급된 제안들이 실행되었을 때의 기대효과를 검증하고 산출하기 위해서는 데이터 트래킹 설계 개선이 반드시 선행되어야 합니다. 현재 데이터 구조만으로는 온보딩 개선의 영향, 첫 방문 시 어떤 혜택이 실제로 사용되었는지, 그리고 우리의 주 고객이 누구인지까지를 정확히 파악하기 어렵습니다. 따라서 아래에 제시된 기대 효과와 검증 계획은 데이터 트래킹 설계가 일정 수준 이상 개선되었다는 가정을 바탕으로 작성하였습니다.

기준 수치

- 무료체험 신청자 : 9,624명
- 방문율 : 62.2%(5,985명)
- 방문 잔존율 : 38.8%(2,325명)
- 미방문 잔존율 : 36.5%(1,327명)

- 결제 금액 : 24시간/18,000원 1회 결제 (현재 데이터에서 결제 금액, 형태를 추적할 수 없기 때문에 보수적으로 판단)

7-1-1. 온보딩 플로우 강화

온보딩 설계 개선의 직접적인 기대효과는 무료체험 신청 고객의 방문 전환율을 끌어올리는 것입니다. 방문률이 개선되면, 이후 첫 방문 혜택 제공, 아하 모먼트 등 전환 장치가 작동할 수 있는 모수 자체가 커지기 때문에 결제 전환율과 전체 매출에도 간접적으로 긍정적인 영향을 미칩니다. 특히 체험 기간 종료까지 한 번도 방문하지 않고 이탈한 노쇼 고객의 비율이 높은 현재 상황에서는 온보딩을 개선하여 “한 번도 공간을 경험해보지 못 하고 이탈하는 고객”에게 z사가 제공하는 공간과 서비스의 강점을 보여줄 수 있는 기회를 확대할 수 있습니다.

현재 데이터 기준으로 2021년 5월부터 2023년 12월까지 무료체험을 신청한 고객 9,624명의 방문률은 약 62.19%입니다. 온보딩 개선을 통해 이를 80%까지 약 17.81%p 상승을 이루었을 때 기대되는 추가 결제 전환 효과는 다음과 같이 계산할 수 있습니다. (추가 방문 고객이 기존 방문 고객과 동일한 잔존율을 보인다는 가정을 두었습니다.)

- 방문율 개선 효과 : 62.2% -> 80.0% (+17.8%p)
- 추가 방문자 수 : 9,624명 * (80.0% - 62.2%) = 약 1,700명
- 추가 잔존 고객 수 : 9,624명 * (80.0% - 62.2%) * 38.8%(방문 고객 잔존 수) = 약 660명
- 매출 효과 : 660명 * 18,000원으로 11,880,000원 수준의 추가 매출이 발생합니다.

실제로는 첫 방문 경험 개선과 결합될 경우, 온보딩으로 확보한 추가 방문 고객에게 더 높은 전환율을 적용할 수 있으므로, 장기적인 매출 효과는 이보다 더 커질 수 있습니다.

7-1-2. 재 방문 경험 극대화

첫 방문 시 혜택 제공 전략은 온보딩 설계를 통해 확보한 방문 고객을 실제 유료 결제와 재방문으로 연결하는 것에 초점을 둡니다. 첫 방문 고객에게 크레딧 형태의 혜택(이용 시간, 회의실, 안마의자, 커피머신 등)을 제공함으로써 고객이 실제 업무, 공부, 회의를 통해 공간을 충분히 체험하며 아하 모먼트를 겪고, 재방문 혹은 결제의 리텐션 유도라는 기대효과를 가지고 있습니다.

현재 데이터 기준으로 무료체험 신청 후 방문한 인원은 약 62.2%이지만, 온보딩 효과가 적용되었다는 가정 하에 80%가 방문을 했다는 기준을 적용했습니다. 이를 통해 9624명의 80%인 약 7700명이 방문을 한 후 현재 방문 인원의 결제율인 38.8%를 첫 방문 시 혜택 제공을 통해 48%까지 약 9.2%p 상승을 이루었을 때 기대되는 추가 결제 전환 효과를 다음과 같이 계산할 수 있습니다.

- 방문 인원의 결제율 개선 효과 : 38.8% -> 48% (+9.2%p)
- 추가 결제 고객 수 : 7700명 * (48% - 38.8%) = 약 708명
- 매출 효과 : 708명 * 18,000원으로 12,744,000원의 추가 매출이 발생합니다.

7-1-3. 데이터 기반 의사결정 인프라 구축

데이터 트래킹 설계 개선의 기대효과는 단기 매출 증대보다는 z사의 의사결정과 실험 과정을 데이터 기반으로 전환하는 데에 있습니다. 무료체험 신청 -> 방문 -> 결제까지의 사용자 여정이 일관된 키와 이벤트로 기록되면, 각 단계에서 어떤 고객이 얼마나 이탈하는지, 어떤 접점에서 개선 여지가 있는지를 정량적으로 파악할 수 있습니다.

이를 통해 지금까지는 “노쇼 고객이 많다”, “잔존 고객의 공통점을 찾기 어렵다” 이러한 문제들 때문에 확인할 수 없던 주 고객층과 이용 특징, 구체적인 전환율에 대해 명확한 진단을 하고 어떤 세그먼트가 서비스와 가장 잘 맞는지, 어떤 마케팅과 전략에 자원을 우선 할당해야 하는지에 대한 기준을 세울 수 있습니다. 이렇게 모든 실행 전략의 타겟팅과 A/B 테스트 설계를 정교하게 만들어 중장기적으로 효율적인 성장 기반을 만드는 점이 데이터 트래킹 설계 개선의 핵심 기대효과입니다.

7-1-4. 유희시간 활용

피크 타임 외 시간대 활용 전략은 24시간 운영이라는 z사의 특성을 살려, 상대적으로 공간이 비어 있는 오전, 심야 시간대에 추가 수요를 유치하는 것을 목표로 합니다. 데이터의 특성 상 주 고객층이 아직 완전히 규정되지 않은 상황에서, 이 전략은 상대적으로 시간대 제약에서 자유로운 취업 준비생, 프리랜서, 스터디 회원 등을 대상으로 새로운 유입 채널을 만들 수 있습니다. 이를 통해 “비피크 시간대에 더욱 가성비 높은 복합 공간”이라는 포지셔닝을 구축할 수 있다는 기대효과가 있습니다. 하지만 피크 타임 외 시간대 활용 전략은 무료 체험 퍼널에 직접적으로 포함되어 있지 않고, 현재 데이터로 주 고객층 규정이 불가능하기 때문에 데이터를 통해 기대효과를 계산하기보다는 가정을 명시한 시나리오로 기대 효과를 추정하여 다음과 같이 계산하였습니다.

비피크 시간대 고객 유치 효과(예시 가정)

- 일 평균 추가 유료 이용자 15명 확보
- 한 달 30일 운영, 1인당 하루 평균 결제 금액 5,000원(4시간)
- 추가 이용자 수 : 15명 * 30일 = 450명/월
- 비피크 상품을 통해 유입된 고객 중 무료체험 후 48%(첫 방문 혜택)가 실제 결제로 이어진다고 가정하면 $450 * 48\% = \text{약 } 216\text{명}$
- 매출 효과(월 기준) : $216\text{명} * 5,000\text{원} = 1,080,000\text{원/월}$ 수준의 추가 매출 발생

위 수치는 비피크 전용 상품을 통해 일 평균 15명의 추가 결제 고객 후보를 확보하고, 그 중 48%만 실제 결제까지 이어진다는 보수적인 가정에 기반한 예시입니다. 실제로는 비피크 전용 할인 요금과 스터디, 강연, 커뮤니티 등을 함께 운영하여 반복 방문과 장시간 이용을 유도할 수 있기 때문에 장기적인 매출 효과는 이보다 더 커질 수 있습니다.

7-2. 성과 측정 계획

Effort-Impact Matrix 기반 마케팅 전략에서 제안한 4가지 전략 옵션의 실행 성과를 정량적으로 검증하기 위해, 각 전략의 목표에 부합하는 핵심 성과 지표(KPI)를 다음과 같이 설정합니다.

7-2-1. 데이터 트래킹 설계 개선 성과 측정

측정 목적: 누가 우리의 고객인지, 어떤 목적과 경로로 서비스를 이용하는지 식별할 수 있을 만큼 데이터 인프라가 구축되었는지 확인합니다.

주요 지표

- 고객 유형 식별률: 전체 신청자 중 개인/기업, 직무, 이용 목적 등의 프로파일링 데이터가 확보된 유저의 비율을 측정합니다. (기존: 0% → 목표: 90% 이상)
- 로그 정합성: trial_visit_info와 access_log 간의 데이터 불일치 건수를 모니터링하여 데이터 신뢰도를 검증합니다

7-2-2. 온보딩 설계 성과 측정

측정 목적: 신청 후 미방문하는 약 37.81%의 이탈 고객을 방문 단계(Activation)로 유도했는지 확인합니다. 목표 방문률(80%)를 달성했는지, 재예약 링크가 노쇼를 줄이는지 검증합니다

주요 지표:

- 방문율: (실제 방문자 수 / 무료 체험 신청자 수) * 100. 가장 직관적으로 온보딩 메시지와 리마인드 효과를 검증하는 지표입니다.
- 재예약 링크 전환율: (재예약 완료 고객 수 / 재예약 링크 발송 대상자 수) * 100. 미방문 고객에게 발송된 재예약 링크를 통해 실제 방문 일정을 다시 잡은 비율을 측정하여, 재예약 링크의 효과를 정량적으로 검증하는 지표입니다.

7-2-3. 첫 방문 시 혜택 제공 성과 측정

측정 목적: 방문한 고객이 크레딧 혜택(이용 시간, 회의실, 안마의자, 커피머신 등)을 통해 서비스를 충분히 경험하고, 아하 모먼트에서 유료 결제, 재방문으로 이어지는지 확인합니다. 특히 2일 차 재방문이 전환의 분기점임이 확인되었으므로 2일 차 재방문과 혜택 사용 여부를 중점적으로 측정합니다.

주요 지표:

- 방문객 대비 결제 전환율: (유료 결제 고객 수 / 실제 방문 고객 수) * 100. 전체 무료체험 신청자가 아닌 방문 고객을 모수로 하여 혜택의 순수 효과를 검증합니다.
- 2일 차 재방문율: (첫 방문 후 2일 이내 재방문 고객 수 / 첫 방문 고객 수). 첫 방문 후 익일에 다시 방문한 비율. 혜택이 재방문 유도에 효과적이었는지 검증합니다.
- 쿠폰/크레딧 사용률: (혜택 사용 고객 수 / 혜택 지급 고객 수). 지급된 혜택을 실제로 사용한 비율을 통해 혜택의 매력도를 평가합니다.

7-2-4. 피크 타임 외 시간대 활용 성과 측정

측정 목적: 이용률이 낮은 오전 및 심야 시간대의 트래픽과 매출이 실제로 증가했는지 확인합니다. 방문 수 외에 좌석 가동률, 시간당 매출이 개선되고 있는지, 그리고 기존 피크 타임 수요를 잠식하지

않고 추가 수요를 만들어내는지 검증하는 것이 목표입니다.

주요 지표 :

- 시간대별 평균 재실 인원 : (해당 시간대 방문 고객 수 / 운영 일수). 08:00~10:00(오전) 및 22:00~02:00(심야) 등 특정 타겟 시간대의 평균 재실 인원 변화 추이를 측정합니다.
- 시간대별 좌석 가동률 : (해당 시간대 전용 좌석 수 / 총 좌석 수) * 100. 특정 타겟 시간대의 평균 좌석 가동률 변화 추이를 측정합니다.
- 비피크 시간대 시간당 매출 : (해당 시간대 매출 합계 / 해당 시간대 총 이용 시간). 단순 인원 증가뿐 아니라 수익성 관점에서 전략의 효과를 평가합니다.
- 특수 목적 상품 판매량: (상품별 월간 판매량, 일 평균 판매량 등). 해당 시간대를 타겟으로 한 '모닝 패스', '심야권', '강연 티켓' 등의 판매 건수를 집계합니다.

8. Appendix & Future Works

8-1. 분석 한계점

8-1-1. 원본 데이터의 정합성 문제 및 설명 변수의 구조적 결핍

원본 데이터의 정합성 문제 및 설명 변수의 구조적 결핍 본 프로젝트 진행 중 가장 큰 난관은 데이터의 신뢰성 확보였습니다. 제공된 trial_visit_info (요약 테이블)와 trial_access_log (로그 테이블) 간의 데이터 불일치가 다수 발견되어, 분석의 정확성을 담보하기 위해 로그 기반의 대체 마스터 테이블을 직접 생성해야 했습니다. 더욱 치명적인 한계는 User Profile 데이터의 부재였습니다. 현재 수집된 데이터는 단순한 출입 로그와 결제 여부에 국한되어 있어, 고객의 성별, 연령대, 직업군 같은 고객의 유형이나 이용 목적, B2B/B2C 구분 등 결제 의사결정의 핵심인 인구통계학적 정보가 부족했습니다.

이로 인해 유료 전환 유저와 미전환 유저 간의 행동 패턴 차이는 발견했으나, 그 행동 이면에 있는 사용자 특성을 규명하지 못하여 심층적인 타겟팅 전략을 수립하는 데 한계가 있었습니다.

8-1-2. 미방문 고객에 대한 분석 불가

미지의 영역으로 남은 미방문 고객 분석 결과, 전체 무료 체험 신청자의 37.81%가 실제 방문을 하지 않는다는 심각한 리소스 누수 문제가 확인되었습니다. 그러나 이들은 오프라인 방문 로그조차 남기지 않았기 때문에, 데이터상에서 어떠한 흔적도 찾을 수 없는 미지의 영역으로 남았습니다. 이들이 단순 번심인지, 위치 접근성의 문제인지, 혹은 경쟁사 서비스로 이탈했는지에 대한 원인을 파악할 수 있는 데이터가 없어, 이들을 방문으로 유도할 구체적인 솔루션이 데이터가 아닌 가설에 의존할 수밖에 없었다는 한계가 있었습니다

8-1-3. 머신러닝 예측 모델의 성능 및 신뢰도 이슈

머신러닝 예측 모델의 성능 및 신뢰도 이슈 앞서 언급한 설명 변수(Feature)의 부족은 머신러닝 모델의 성능 저하로 직결되었습니다. Random Forest, XGBoost, LightGBM 등 다양한 앙상블 모델을 적용하고 하이퍼파라미터 튜닝을 시도하였으나, 최종 모델의 Recall(재현율)과 ROC-AUC 점수는 예측 모델로서 활용하기에는 다소 낮은 수준에 머물렀습니다.

이는 고객의 결제 의도가 단순히 몇 번 방문하고 몇 시간 동안 머물렀는지에 대한 행동 양(Quantity)만으로는 설명될 수 없으며, 공간에서의 경험, 서비스 만족도 등 데이터에 담기지 않은 경험의 질(Quality)이 복합적으로 작용한다는 점에서 한계가 있었습니다.

8-2. 향후 과제

8-2-1. 고객 식별을 위한 데이터 수집 체계의 근본적 개편

향후 데이터 분석이 실질적인 비즈니스 성과로 이어지기 위한 최우선 과제는 '우리의 고객이 누구인가'를 식별할 수 있는 데이터 파이프라인의 구축입니다. 무료 체험 신청 단계에서 고객 유형(개인/기업/팀), 직무, 주 이용 목적(업무/미팅/학습) 등을 필수적으로 수집하도록 신청 폼을 고도화하고, 이를 DB에 체계적으로 적재해야 합니다.

이렇게 확보된 User Profile 데이터가 기존의 trial_access_log 데이터와 결합될 때, 비로소 모델의 예측 성능을 비약적으로 향상시키고 페르소나별 정교한 타겟 마케팅이 가능해질 것입니다.

8-2-2. 정성적 조사 병행을 통한 이탈 원인 규명

데이터가 설명해주지 못하는 이유를 파악하기 위해 정성적 조사가 병행되어야 합니다. 특히 데이터가 전무한 미방문 고객과 1일 차에 이탈한 고객을 대상으로 해피콜, 이메일 설문조사 등을 진행하여 구체적인 이탈 사유를 수집해야 한다고 생각합니다. 고객이 느끼는 구체적인 불편함(위치, 시설, 가격 등)을 직접 수집함으로써, 데이터상의 수치가 의미하는 바를 정확히 해석하고 온보딩 프로세스 개선의 실질적인 근거를 마련해야 합니다.

8-2-3. 전략 검증을 위한 A/B 테스트 및 실험 환경 구축

본 보고서에서 제안한 온보딩 메시지 및 리마인드 강화, 첫 방문 시 크레딧 혜택 제공, 피크 타임 외 공간 활용 등의 전략은 가설에 기반하고 있습니다. 따라서 실제 현업 적용 시에는 반드시 A/B 테스트를 통해 그 효과를 검증해야 합니다. 고객을 실험군과 대조군으로 나누어 전략 실행에 따른 방문율과 전환율의 변화를 측정하고, 이 결과를 바탕으로 마케팅 예산을 효율적으로 배분하는 데이터 드리븐 의사결정 사이클을 정착시키는 것이 분석 이후의 핵심 과제입니다.

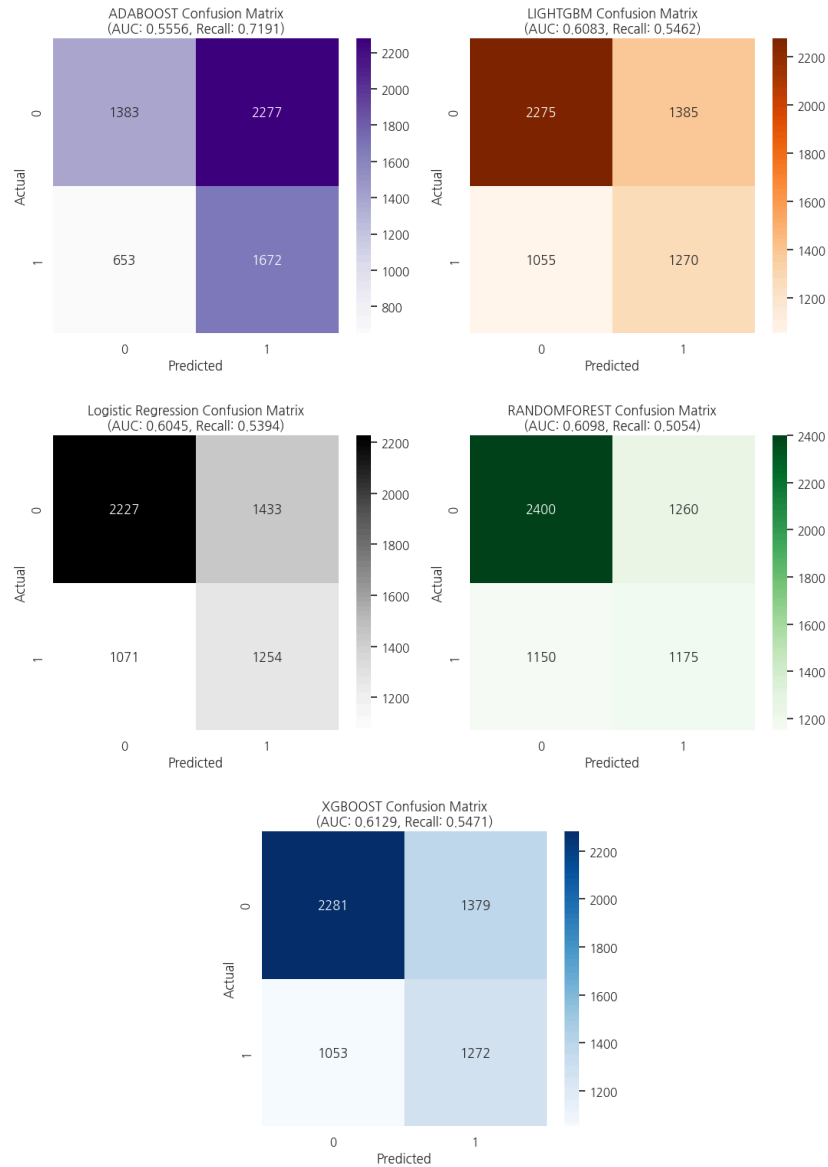
8-2-4. 이탈 예측 및 생애 가치(LTV) 모델링 확장

현재의 신규 유입 및 전환 중심 분석을 넘어, 기존 고객 관리로 분석 범위를 확장해야 한다고 생각합니다. 결제 이후 고객의 이용 패턴(방문 빈도 변화, 이용 시간 감소 등)을 모니터링하여 구독 해지 징후를 사전에 포착하는 이탈 예측 모델을 개발하고, 장기적으로 우리 서비스에 가장 많은 수익을 가져다주는 고객군을 정의하는 LTV 분석을 통해 마케팅 타겟을 더욱 정교화해야 합니다.

8-3. 시각화

8-3-1. 혼동행렬

최종적으로 결정된 모델에서 나온 최종 혼동행렬로 5개의 모델에 대한 값을 확인할 수 있습니다.



[그래프 9] 최종 모델 혼동행렬 비교 그래프