

프로젝트 개요

- 공유 오피스 3일 무료 체험 데이터를 활용하여 유저의 출입 및 방문 패턴을 분석하여 결제 전환 여부를 예측하는 모델을 구축하고 모델을 기반으로 유저 행동 특성별 타겟 마케팅 전략을 수립한다

활용 데이터

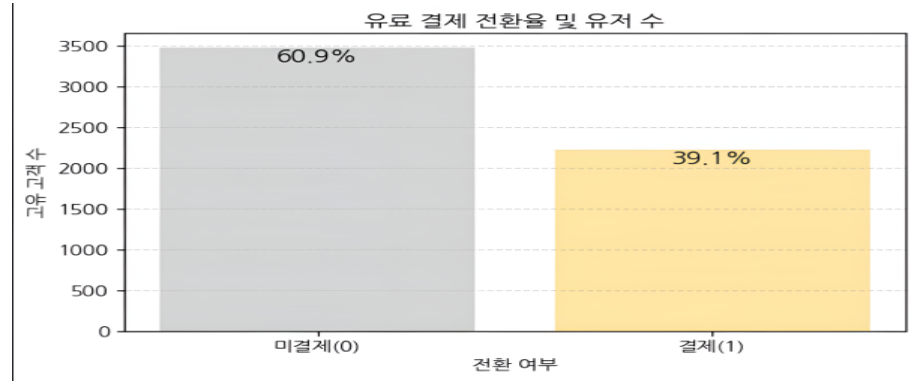
- 공유 오피스 3일 무료 체험 데이터(2021.05.01~2023.12.31)
- 데이터셋 : 3일 체험 신청자 출입 기록 로그, 3일 체험 신청, 3일 체험 신청자 일자별 방문 기록, 3일 체험 신청자 결제 여부, 지점별 면적

데이터 전처리

- 결측값 및 이상치 제거: 체류 시간 0분 혹은 비정상적으로 긴 경우 필터링
- 입실 혹은 퇴실 로그만 있는 유저 제거
- 유저별 방문 날짜별 입실(1) -> 퇴실(2) 로그 쌍을 가진 유저만 정상 방문으로 판단
- 유저별 첫 로그가 퇴실이거나 마지막 로그가 입실인 로그 시스템 오류로 판단하여 제거
- 중복되는 로그(입실->입실, 퇴실 -> 퇴실) 중 첫 행만 남기고 제거
- 방문자 6026명 중 5707명의 출입 기록 로그 데이터로 분석 진행

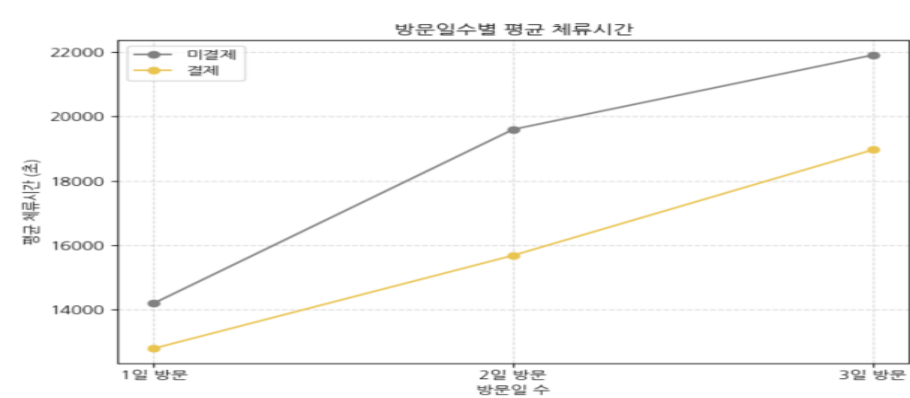
EDA

3일 무료 체험의 유료 결제 전환율



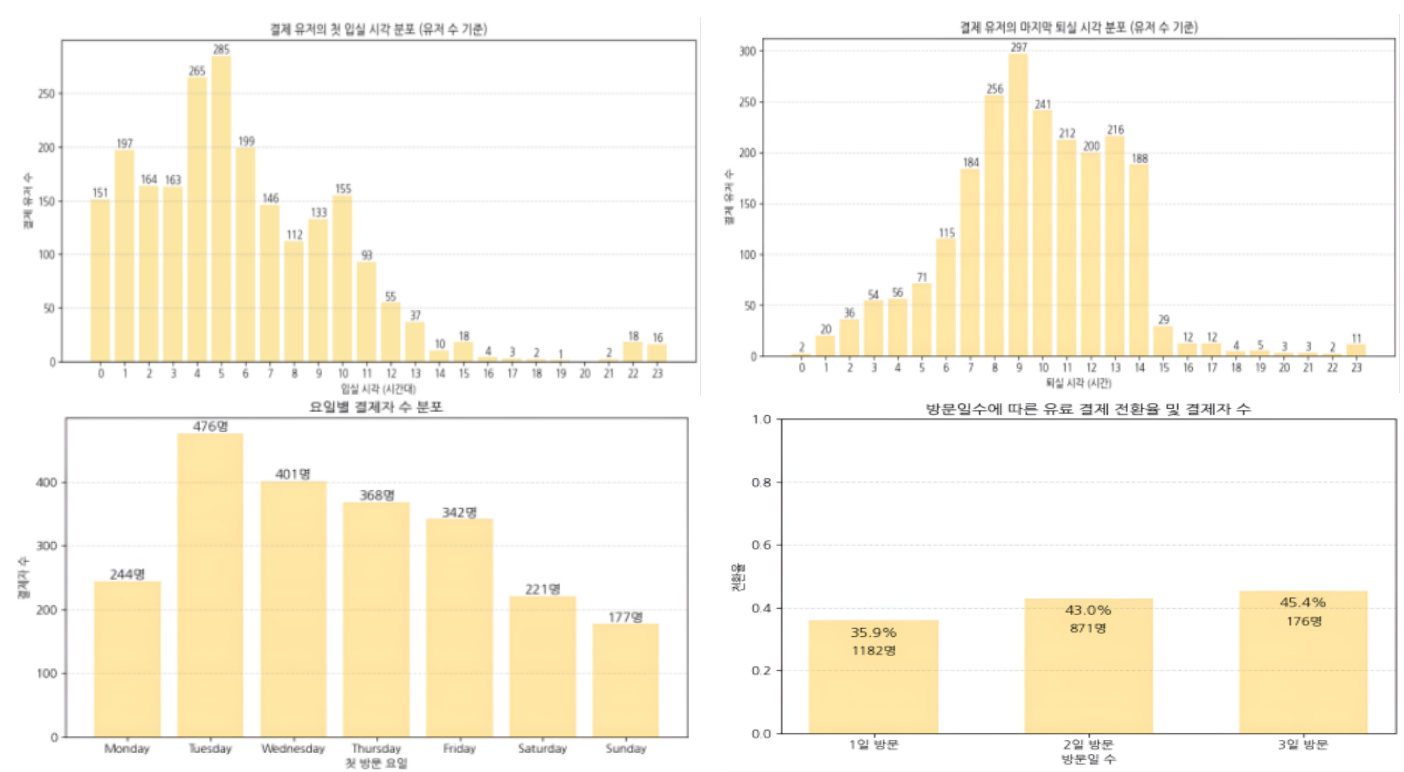
- 방문자 5707명 중 3478명(약 61%) 이 3일 무료 체험 방문 후 이탈하여 결제 전환율이 낮음
- 무료 체험 방문자는 이용 목적이 뚜렷하지만 결제 전환까지 동기가 부족했음을 시사함

결제 여부에 따른 체류 시간 비교



- 방문 일수가 증가할수록 체류 시간이 증가하지만 모든 방문 구간에서 미결제 유저의 평균 체류 시간이 더 긴 경향
- 미결제 유저가 더 오래 머물렀다는 점은 만족도는 높지만 결제 유도 설계를 보완해야함을 시사함
- 결제 유저는 짧고 효율적으로 이용하는 경향이 있어 행동 패턴 중심 개선이 중요

결제 유저의 방문 패턴



- 입실 시각은 새벽~이른 아침(0~12시)에 집중되어 있으며 새벽4~5시가 피크 타임
- 결제 유저의 퇴실 시각은 아침~ 오후 (7~14시)까지 지속되며 8-10시에 집중됨
- 주말 이용자 비중이 낮으며, 특히 화~수요일 방문 비중 높음
- 방문 일수가 늘어날수록 결제 전환율이 상승하는 경향을 보이며 결제자의 약 92%가 3일 중 1~2일 방문 후 결제
- 요일별 시간대별 맞춤 전략과 반복 방문 유도 및 2일차 결제 유도 설계가 중요

모델링

파생변수 생성

• 데이터 개수: 5,707개

방문 일수	체험 기간 중 실제 방문한 일수	첫 방문 월	체험 기간 중 처음 방문한 월
총 체류 시간	체험 기간동안 유저가 실제 머무른 시간의 총합	첫 방문 시간대	심야, 아침, 오후, 저녁으로 구간화
평균 체류 시간	총 체류 시간을 방문 일수로 나눈 값	입실 횟수	체험 기간동안의 출입 빈도
지점 면적	유저가 처음 방문한 지점의 공간 면적(단위:평)	방문 지점 수	몇 개의 지점을 방문했는지 개수
첫 방문 소요 일수	신청일로부터 실제 첫 입실까지의 일수	-	-

모델링 과정

- 랜덤포레스트는 피처 간 비선형 관계를 잘 포착하는 다수의 결정트리 기반 앙상블 모델로 초기 성능이 안정적이기 때문에 베이 스라인 모델로 선정
- 하이퍼 파라미터 튜닝에는 효율적인 탐색이 가능한 베이지안 최적화를 적용하였고, 실제 결제 유저를 놓치지 않는 것이 타겟 마케팅 전략에 중요하다고 판단하여 Precision-recall 곡선을 기준으로 recall이 가장 잘 나오는 0.4로 임계값을 조정함
- Gradient Boosting 기반의 XGBoost는 반복적으로 오차를 보완하며 학습해 복잡한 패턴을 정밀하게 포착하여 안정적인 분류 성능을 보이는 모델로, 동일 조건에서 비교한 랜덤포레스트보다 결제 유저 recall과 f1-score가 더 우수하여 최종 모델로 선정

모델링 및 성능 비교

모델	Accuracy	결제 Recall	Precision	F1-score
RandomForest(기본)	0.60	0.38	0.48	0.43
XGBoost(최종) (베이지안 최적화 +임계값 0.4)	0.58	0.56	0.47	0.51

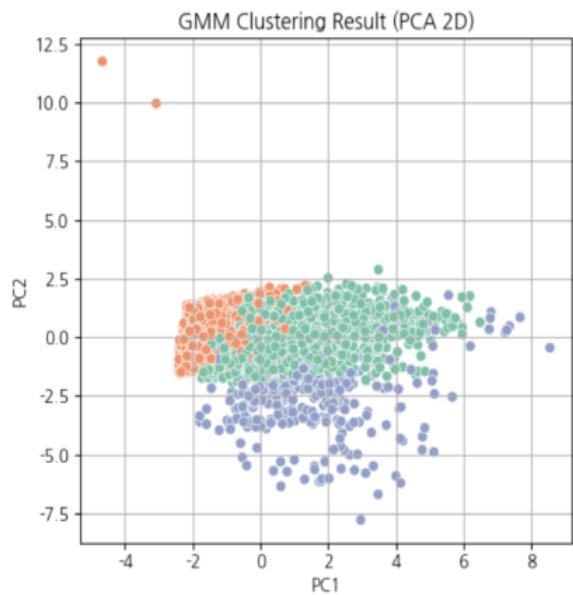
모델 해석

- 유저의 방문 및 출입 패턴을 기반으로 결제 전환 가능성 있는 유저를 예측하는데 초점을 맞춘 모델

클러스터링을 통한 유저별 타겟 마케팅 전략

PCA+GMM 클러스터링 시각화

- PCA를 통해 전체 분산의 약 90%를 유지하며 핵심 정보를 보존하면서 차원을 축소하여 군집 간 분리를 확률적으로 유연하게 수행할 수 있도록 함



	유저 수	결제 전환 가능성	특징	마케팅 전략
장기 체류 탐색형 유저	2946명	보통	체류 시간 가장 길고 입실 빈도 높음 일부 유저 지점 비교 목적의 다수 지점 탐색	-지점 비교 가이드 및 후기 정보 제공을 통한 의사결정 지원 해소
단순 체험 유저	2460명	낮음	3일 중 1일만 방문, 체류 시간 짧고 방문도 적음 짧게 체험 후 이탈	저비용 채널 활용 재방문 유도 리마인드 - 체험 기한 리마인드 알림 푸시
체험 충실형 유저	301명	높음	3일 체험 적극 활용, 체류 시간 길고 방문 빈도 높음 넓은 공간 선호하며 단일 지점 방문	3일 체험 혜택으로 2일차에 장기 요금제 및 프리미엄 서비스 등 장기 전환 및 업셀링 유도

제언

- recall 중심 모델링으로 설계 되어 전환 유저를 놓치지 않는 장점은 있으나, precision이 낮아 과잉 마케팅 대상자 포함 가능성 존재
- 추가 EDA 기반 파생 변수 생성 및 하이퍼파라미터 튜닝을 통한 예측모델 고도화 필요