PROJECT 1
구독형 강의 서비스의 첫 날 이탈 사용자 재방문 개선을 위한 데이터 분석

프로젝트 기간: 2025. 04. 07 ~ 2025. 04. 30
이름: 유윤종
yooyoon97@gmail.com

분석 배경

- 디지털 전환 가속화와 학습 주체의 자율성 확대, 시·공간 제약 해소로 온라인 교육 시장이 급성장
- 많은 오프라인 강의가 비대면 콘텐츠로 전환되며 맞춤형 강의 제공 및 학습 효과 측정 수요 증가

분석 목표

1. 구독형 강의 서비스 사용자의 전반적인 학습 여정 로그 데이터 분석
2. 사용자 세그먼트 별 행동 패턴 및 리텐션 분석
3. A/B 테스트 기반 재방문 유도 전략 제안

실험 가설

"Aha-Moment를 경험한 직후, 다음날(1일차)에 다시 방문한 사용자 세그먼트는 이탈자 세그먼트 대비 KPI가 유의미하게 높다"

사용 데이터

- **온라인 교육 콘텐츠 구독서비스의 사용자 행동 로그 데이터 테이블**
- 140,000명의 회원가입자의 활동 로그와 관련된 변수들로 구성 됨
- 주요 테이블: 레슨 시작 (enter.lesson_page), 구독 완료 (complete.subscription), 회원 가입 (complete.signup) 등

분석 과정

- **데이터 전처리**
  - 회원 가입을 한 사용자 ID를 추출하여 새로운 사용자 정보 테이블 생성
  - 사용자 행동을 모두 추출하여 사용자 정보 테이블과 병합
  - 이상 로그 · 결측 처리 및 리텐션 ·결제 파생 변수 생성

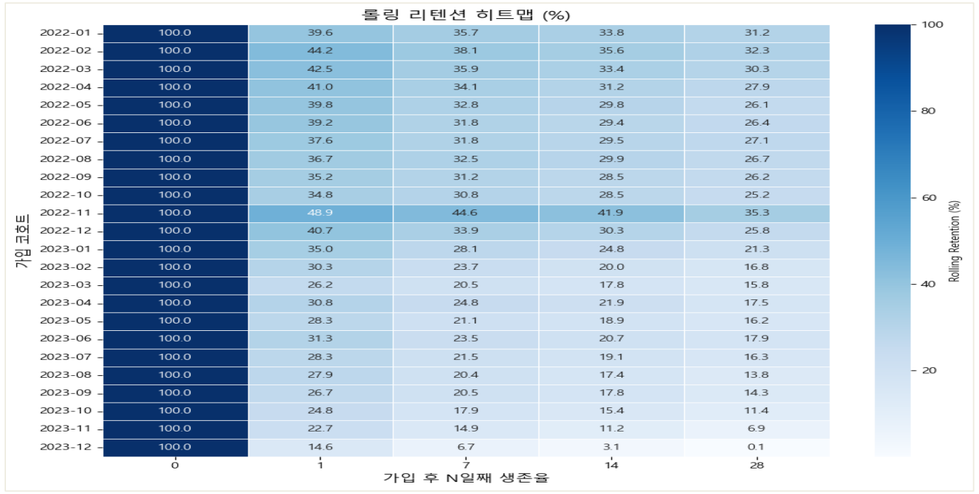
- **AARRR 프레임워크 설정**

단계	지표	설명
Acquisition	2022년 이후 가입 사용자 수	실험 대상 Pool 정의
Activation	최소 1개의 강의 수강	'Aha-Moment' 를 느낀 시점
Retention	다음 날 재방문율	Activation 다음 날 접속 비율 (%)
Revenue	구독으로 이어진 사용자 수	결제 로그가 존재하는 사용자
Referral	-	데이터 상 측정 불가

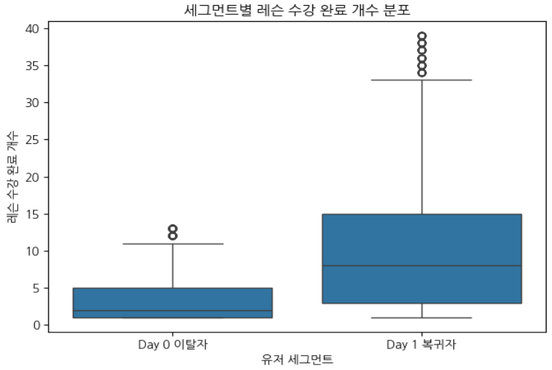
- **사용자 세그먼트 분류**

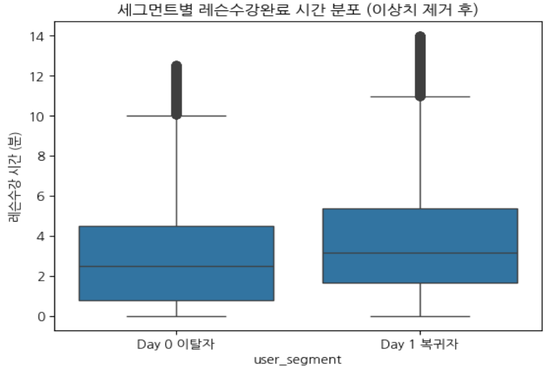
분류	설명
0일차 이탈자	'Aha-Moment' 이후 재방문 하지 않음
1일차 재방문자	'Aha-Moment' 이후 다음날 재방문함

분석 결과



- 0일차 → 1일차 평균 롤링 리텐션: 100% → 34%.
- 약 66%가 첫 날 이후 이탈하는 사용자임을 확인함
- 월 별 편차가 존재하나, 모든 코호트에서 공통적으로 **첫 날 대규모 이탈 발생**





- **1일차 재방문자** : 중앙값 약 7회, 3사분위수 약 12회 → Day 0 이탈자(중앙값 약 2회)에 비해 훨씬 많은 레슨을 완료. 평균 학습 시간이 길고 분산도 커 학습 깊이가 높음
- **0일차 이탈자** : 중앙값 약 2회, IQR 약 1~4회 → 전체 수강 횟수가 낮아 학습 몰입도가 부족. 짧은 학습 시간으로 서비스 체험이 제한적임 ($p > 0.001$)

분석 인사이트

분류	활동 일 수 (일)	결제 전환률 (%)	일 평균 세션 체류 (회)	ARPPU (원)
0일차 이탈자	1.55	8.02	0.1	76,469
1일차 재방문자	8.04	47.49	1.5	74,572

- **1일차 재방문자 :** 평균 활동 일수 8.04일, 일 평균 세션 체류 1.5회 → 적극적 서비스 이용 및 학습 몰입도 높음. 결제 전환율 47.49%로 월등히 높아, 재방문 이후 '유료 전환' 가능성이 크게 상승
- **0일차 이탈자** : 평균 활동 일수 1.55일, 일 평균 세션 체류 0.1회 → 체험 단계에서 빠르게 이탈, 서비스 이해 부족. 결제 전환율 8.02%에 불과해, 단기 체험만으로는 유료 전환 동기 부여가 미흡
- **ARPPU(사용자당 평균 매출) 비교:** 이탈자의 개별 지불액은 약간 높으나, 재방문자의 높은 전환율 덕분에 **전체 매출 기여도는 재방문자 세그먼트가 더 큼**

개선 사항 제안 및 기대 효과

전략	설명	기대 효과 및 KPI
▪ **게이미피케이션**	행동 지표를 기반으로 유저에게 시각적·심리적 보상을 제공하여 동기 강화	• 치장 아이템 획득율 • 평균 세션 수 증가율 • 결제 전환율 변화
▪ **분기·시즌 별 마케팅 및 프로모션 강화**	시즌 수요를 활용해 단기 유입을 증대, 이후 연속 결제 및 리텐션을 유도	• 프로모션 코드 사용률 • 첫 결제 후 재구독 비율 • 구독 지속 기간 및 LTV 비교
▪ **장기 충성 사용자 대상 특별 혜택 강화**	최상위 LTV 유저(파워 유저)를 식별해 이탈 방지 및 독점적 경험 제공	• 독점 컨텐츠 참여율 • 파워 유저의 리퍼럴 수 • 파워 유저가 유입 시킨 유저의 LTV 비교

PROJECT 2
공유 오피스 무료체험 사용자 방문 수요 예측 모델링

프로젝트 기간: 2025. 06. 09 ~ 2025. 06. 19
이름: 유윤종
yooyoon97@gmail.com

분석 배경

- 방문 고객의 익일 재방문은 고객 리텐션 및 LTV 증대의 핵심 요소
- 사용자 로그 데이터를 통해 체험 후 행동 패턴을 정밀 분석하고, 모델을 개발함으로써 맞춤형 마케팅·운영 전략 수립을 통해 재방문율 향상 및 수익성 극대화를 도모

분석 목표

1. 재방문 예측 모델을 통해 익일 방문 여부를 높은 정확도로 분류
2. 예측 결과를 바탕으로 재방문을 유도할 수 있는 액션 아이템 도출

사용 데이터

- **공유 오피스 무료체험 사용자의 출입 데이터 테이블**
  - 6,000명의 공유 오피스 사용자의 출입 로그 및 체류 시간 데이터
  - 주요 테이블: 출입 기록 (trial_access_log), 무료 체험 신청 (trial_register), 무료 체험 신청자 결제 여부(trial_payment), 지점 정보(site_area) 등
  - 주요 컬럼: 사용자 ID (user_uuid), 체크인 - 체크아웃 여부 (checkin), 입퇴실 시각 (cdate) 등

분석 과정

- **데이터 전처리**
  - 입실 기록과 퇴실 기록이 매칭되는 경우를 확인하고, 그렇지 않은 로그는 제거
  - 범주형 자료에 Label Encoding 적용 및 날씨 관련 파생변수 생성
  - 원본 데이터에 반응 변수가 존재하지 않았기 때문에 '익일 방문 여부' 라는 이진형 반응 변수를 생성
  - 반응 변수에 데이터 불균형이 존재하여 SMOTE 기법을 적용
- **모델링**
  - 총 4개의 머신 러닝 모델 (Logistic Regression, Random Forest, LightGBM, Stacking) 에 대해 학습 데이터를 통한 모델 적합 수행
  - TimeSeriesSplit을 활용해 시계열 교차검증을 적용하고, Optuna를 사용하여 모델 별 최적의 분류 성능을 확인
  - 3개의 성능 평가 척도 (ROC-AUC, Precision, F1-Score)에 대해 최고의 성능을 보이는 머신 러닝 모델 확인
- **변수 중요도**
  - 각 모델 별 변수 중요도 (Feature Importance)를 확인 하여 모델 성능에 많은 영향을 끼치는 변수를 확인

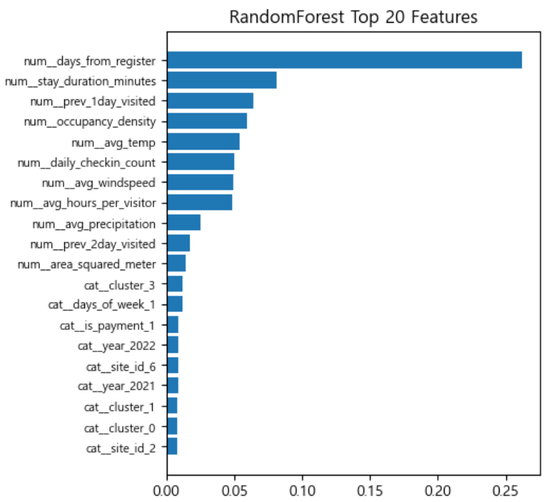
분석 결과

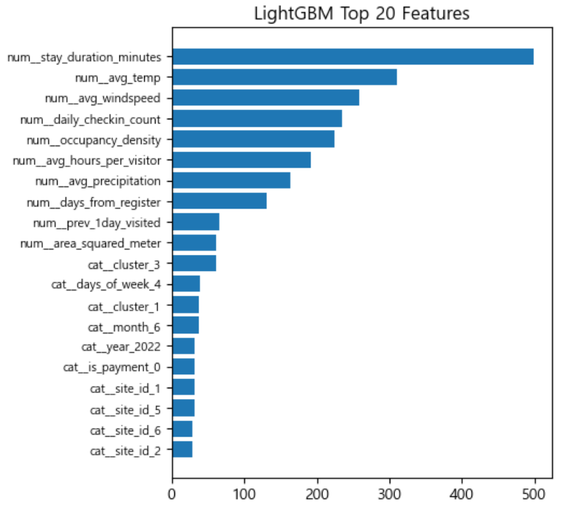
- **각 모델 별 분류 성능 비교 결과**

Method	HPO	ROC-AUC	Class	Precision	Recall	F1-Score
LR	Default	0.7809	0	0.8932	0.6105	0.7253
			1	0.5575	0.8705	0.6797
	Optuna	0.7777	0	0.8725	0.6329	0.7336
			1	0.5621	0.8359	0.6722
RF	Default	0.8088	0	0.7640	0.7945	0.7790
			1	0.6078	0.5648	0.5855
	Optuna	0.8077	0	0.8786	0.6767	0.7646
			1	0.5926	0.8342	0.6930
LGBM	Default	0.8216	0	0.8049	0.7673	0.7856
			1	0.6188	0.6701	0.6434
	Optuna	0.8222	0	0.8819	0.6835	0.7702
			1	0.5988	0.8377	0.6983
Stacking	Default	0.7467	0	0.6436	0.9620	0.7713
			1	0.4507	0.0553	0.0985
	Optuna	0.8209	0	0.7878	0.7770	0.7824
			1	0.6138	0.6287	0.6212

- 최고 성능 모델은 Optuna 튜닝을 거친 LightGBM으로, 타 모델 대비 ROC-AUC가 우수하며 각 Class의 Precision과 Recall도 타 모델 대비 준수하여 재방문자를 놓치지 않고 올바르게 찾아냈음을 의미함.

- Feature Importance Plot





인사이트

- 두 변수 중요도 그래프 비교 결과, 공통적으로 체류시간이 중요하다고 판단되는 변수였으며, LGBM 모델에선 날씨 변수가 모델 성능에 큰 영향을 미쳤음
- 스태킹 모델은 LightGBM 모델 대비 성능 향상이 크지 않았기 때문에 관리 편의성을 고려해 LightGBM을 중심으로 한 예측 모델 운용 고려

추가 개선 방향

- 사용자 세그먼트 별 피드백 구축: 체험 이후 만족도 설문·NPS 점수 등을 모델에 피드백하여 예측 정확도 지속 개선
- 모델 운영 모니터링 & 재학습 주기 단축: 예측 성능이 떨어지는 기간(예: 계절별 변동) 식별 후 주기적 훈련 체계 구축