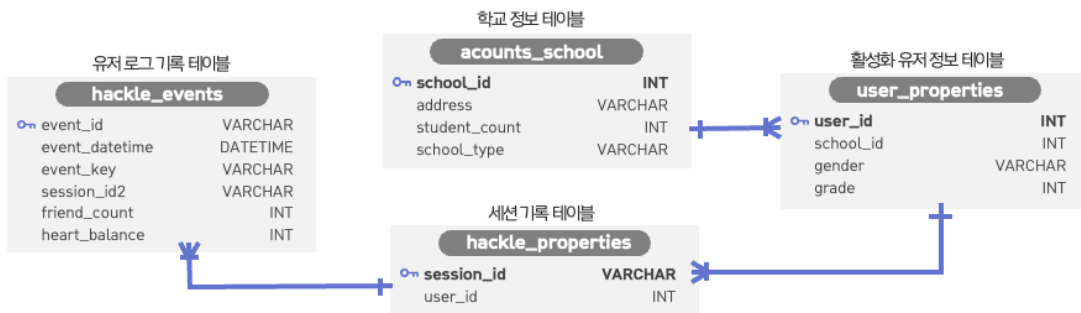


분석배경

- 본 서비스는 10대 청소년을 대상으로 한 '익명 투표' 모바일 어플리케이션으로, 학교 중심의 폐쇄형 서비스로 학교 내에서 서비스가 활성화되지 않는다면 해당 학교 가입 유저의 서비스 이탈 가능성이 매우 높은 구조적 한계를 가지고 있음
- '학교 기반 폐쇄형 커뮤니티'라는 태생적 제약을 극복하기 위해 학교별 서비스 사용행태를 추적하고 확대방안을 모색해야 함

활용 데이터

- 분석 대상 : 총 4,113개 학교 / 유저 98,789명



분석 과정

EDA 단계

- 변수별 분포, 상관관계 확인
- 로그 변환, 원저화로 왜도나 극단값 보정

클러스터링 기법 선정

- 정규성·구형성 가정 기각에 의한
- 분산 기반 클러스터링 기법(DBSCAN)선택

스케일링 기법 선정

- k-distance를 기준으로 RobustScaler 선택

DBSCAN

- 실루엣 점수와 DBI 점수가 나뉘고,
- 노이즈 비율이 지나치게 높아짐

K-means

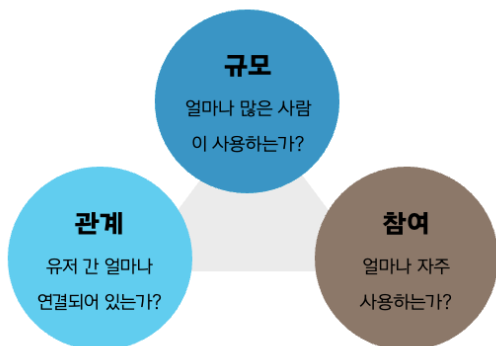
- 다양한 차원 축소 기법 실험
- t-SNE를 통해 안정적인 실루엣 점수 달성

클러스터별 EDA

- 클러스터별 퍼져 분석을 통해
- 클러스터 4가 고활성 클러스터임을 확인

지표 설계

우리 서비스의 활성화 수준을 다각도로 측정하기 위해 규모, 관계, 참여라는 세 가지 핵심 기준을 설정하고, 이를 바탕으로 세부 지표를 정의함



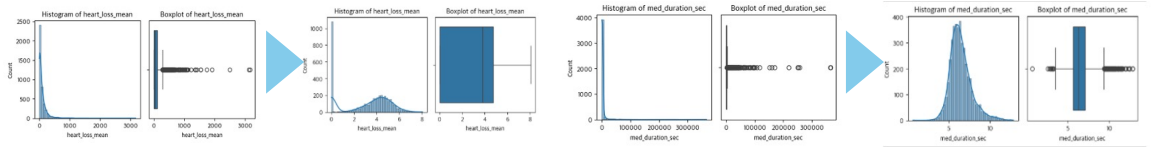
[지표 설계 기준]

구분	지표명	정의
규모	활동 유저 수	기간내 활동 상태의 유저 수
참여	메인 이벤트 참여 수	기간 내 '투표'와 연관된 활동 참여 수 (예. 질문 생성, 질문 열람, 질문 공유 등)
	체류 시간	서비스 이용시간 총합
	하트 사용량	하트 소모 수 (하트 : 결제를 통해 구매하는 포인트)
관계	네트워크 연결 밀도	유저 간 서로 친구로 연결되어 있는 정도 (밀도 = 실제 친구 연결(쌍) 수 / (n(n-1)/2), n=같은 학교 내 유저 수)

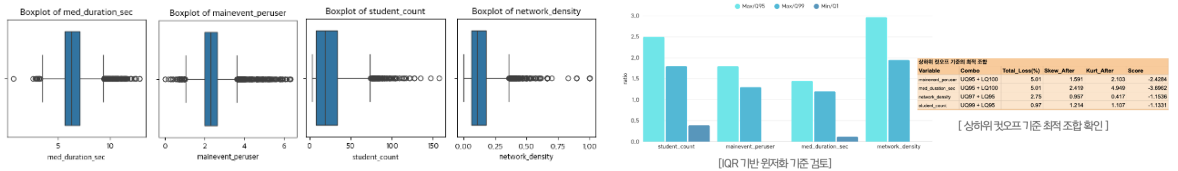
[주요 지표 정의]

클러스터 분석

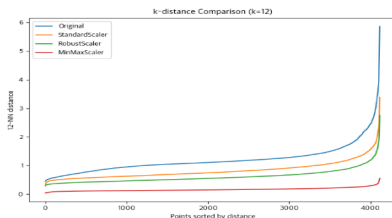
- 하트 사용량, 1인당 메인 이벤트 발생 횟수, 총 체류시간 지표의 심각한 비대칭과 긴 꼬리 형태 확인 후 로그 변환



- IQR 기준 이상치 발견된 지표들의 상위 1%, 5%, 하위 1% 비율 확인 후, 컷오프 최적 조합 찾아 원저화 진행



- 현재 데이터가 skew & kurtosis가 잔존해있고, PCA 결과 비구형을 보이고 있어 DBSCAN을 baseline 모델로 채택
- K-distance를 통해 RobusterScaler 선정 후 GridSearch를 통해 파라미터 최적화 조합 확인



[K-distance 결과]

eps	min_samples	n_clusters	noise_ratio	silhouette	davies_bouldin
0.55	30	3	0.542913	-0.041549	2.324665
0.5	20	3	0.547532	-0.042796	2.413115
0.55	32	3	0.570143	-0.063591	2.281209
0.5	24	3	0.635303	-0.087434	2.31016
0.45	28	3	0.866035	-0.211708	2.357461
0.45	30	3	0.877462	-0.218264	2.320611

[GridSearch 결과]

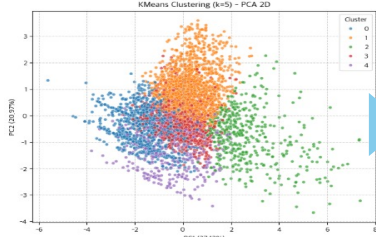
결론

- 실루엣 점수가 음수이고, 노이즈 비율이 지나치게 높으며 DBI 또한 불리한 수준으로, 본 결과는 기각하는 것이 타당
- 클러스터별 최소 샘플 수 제약을 반영했음에도 불구하고 품질이 떨어진다는 점에서 DBSCAN이 현재 데이터 분포 구조에 적합하지 않다는 증거로 볼 수 있음

- 탐색적 분석에서 활용한 DBSCAN의 결과의 신뢰성과 해석력이 부족하여, 군집의 응집력과 분리를 직관적으로 설명 가능한 분산 클러스터링 기법인 Kmeans 진행

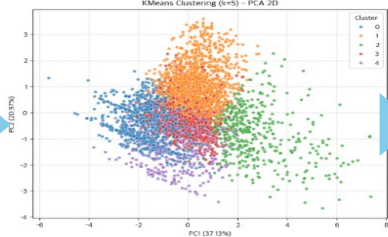
Kmeans + StandardScaler

K=5, Silhouette=0.231,
DBI=1.35, Target Score=0.095



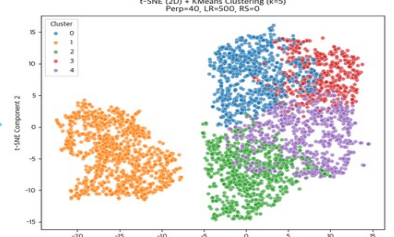
PCA + Kmeans + StandardScaler

K=5, Silhouette=0.244,
DBI=1.303, Target Score=0.114



t-SNE + Kmeans + StandardScaler

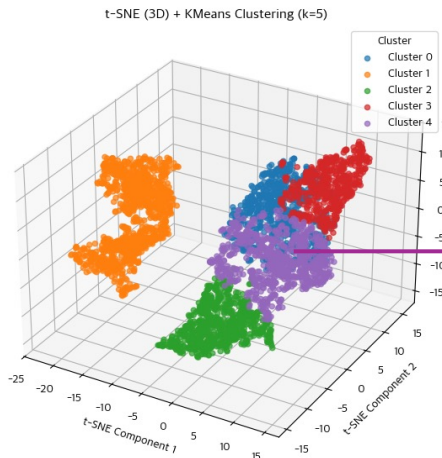
K=5, mean_Silhouette=0.407,
DBI=0.924



결론

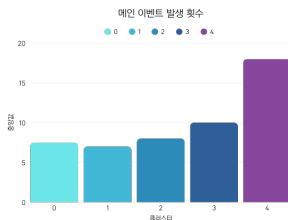
- t-SNE로 차원 축소를 진행한 결과, 안정적인 실루엣 점수와 DBI를 확인
- 다양한 random_state, perplexity, learning_rate 조합에도 안정적인 성능이 유지됨

최종 클러스터

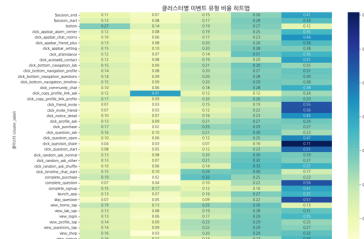


Cluster 4

활동 유저 수는 중간 수준(평균 18명)이지만
활동성과 연결성이 압도적으로 높고, 하트 사용량도 가장 높은
우리 서비스 활성화 수준이 높은 클러스터로 확인



[클러스터별 메인이벤트 발생 횟수]



[클러스터별 메인이벤트 로그 비율]

- Cluster4는 타 클러스터대비 '친구초대', '질문 공유' 이벤트 로그의 비율이 높음
- 즉, 서비스고활성화 학교는 SNS의 소셜 기능을 적극적으로 활용하고 있음

A/B 테스트 설계

SNS 이용에 있어 우리 서비스가 고활성화된 학교는 그렇지 않은 학교 대비 '친구 초대', '질문 공유'와 같은 소셜 기능을 적극적으로 활용하고 있고, 이것이 메인 이벤트의 발생 빈도를 높이고 유저 확대에 기여하고 있는 것으로 확인되어 타 클러스터에도 적용하고자 A/B테스트 설계함

구분	항목	설명
01. 실험 배경 및 가설	관찰된 현상	- 우리 서비스가 고활성화된 학교는 다른 학교 대비 평일, 일과시간에 서비스 이용량이 많다 - 우리 서비스가 고활성화된 학교는 저활성화된 학교 대비 '친구초대'와 '질문 공유' 이벤트 이용 비율이 높다
	목표	클러스터 3의 서비스 활성화 평균 점수 10% 개선
	가설	평일 일과시간에 친구 초대 및 질문 공를 유도하는 푸시 알림을 보내면, 서비스 활성화 지표가 개선될 것이다.
02. 실험 계획	타겟 유저	클러스터 3 - 초기 유저 규모가 충분히 확보되고, 활성화 지표가 중간 수준으로 유지되고 있어 새로운 기능이나 알림에 대한 반응 가능성이 있는 집단임
	실험 방법	학교별 무작위 50:50 비율 분배, 2주 진행 - A그룹(대조군): 기존의 서비스와 알림 정책 유지 - B그룹(실험군): 평일 일과시간에 친구 초대 및 질문 공유를 유도하는 푸시 알림 발송
03. 성과 판단 지표	성과 지표	그룹별 활성화 종합 점수 평균 비교 - 목표: A안 대비 B안의 활성화 종합 점수 평균 20% 이상 개선 성과 검증 방법: t-test를 통한 A그룹과 B그룹의 활성화 점수 평균 비교
	보조 지표	투표 이벤트 참여율 - 알림을 통해 유도된 '질문 공유'를 통해 공유된 질문에 대한 투표 수 DAU - 테스트 기간 내 1일 활성화 사용자 수가 증가했는지 측정
	가드레일 지표	앱 삭제율 - 앱 삭제율이 10% 이상 증가 시, 테스트로 인한 피로도가 증가했다고 판단