

유저 활동일수와 결제율 향상 요인: 네트워크·행동 패턴·모델링 종합 분석

- 데이터분석 6기 1팀
- 백나린, 장위나, 최가림, 황보유리
- 25.07.10 - 25.08.12

목차

Contents

01 프로젝트 개요

02 EDA

03 메인 분석

1. 개인 단위

- a. 활동 일수 영향 요인 종합 분석
- b. 결제 영향 요인 종합 분석
- c. 결제 전 행동 패턴 분석

2. 학교 단위

- a. 활동 일수 영향 요인 종합 분석
- b. 결제 영향 요인 종합 분석

3. 유저 네트워크 영향

- a. 결제 전염성
- b. 투표 전염성
- c. 인플루언서 영향력 분석

4. 핵심 세그먼트 분석

04 종합 인사이트 & 전략 제안

05 부록

프로젝트 개요

- 서비스 소개
- 분석 배경 및 문제 정의
- 분석 방향 확장 과정
- 분석 설계
- 데이터 소개

분석 대상 서비스 소개

10대 청소년을 대상으로 한, “익명 투표” 모바일 애플리케이션

서비스

- 10대를 위한 익명 투표 어플 - 질문에 부합하는 친구에게 투표
- 학교 중심의 반 폐쇄형 서비스 → 학교 단위의 활성 유저 확보가 매우 중요.

투표 기능 설명

1. 친구가 보기로 제시된 질문에 **익명 투표**하며, 질문은 건너뛸 수 있음.
하루 무료 투표는 10회로 제한되며, **포인트를 사용하면 추가 투표 가능.**
2. 투표에서 **지목된 친구**는 **받은 질문만 확인 가능**하며, 투표자는 익명.
포인트 사용 시 투표자의 초성 확인 가능.

그 외 기능

- 출석체크
- 받은 투표에 답장(ask)
- 타임라인: 친구들이 받은 투표들을 확인할 수 있는 탭



2 분석 배경 및 문제 정의

리텐셜과 결제 양측 모두 매우 저조한 상황

I. 개요

분석 배경

- 앱의 MAU 및 결제량 기하급수적으로 감소
- 리텐션 저조
- 학교별 사회적 관계망 구조 차이



문제 정의

- 초반에는 유저의 결제 유발 요인 및 사전 행동 분석을 목표로 삼았으나, 데이터 특성에 따라 리텐션 및 네트워크 관점을 추가하여 분석 범위 확장
- **결제율은 유저 수에 강하게 의존**하는 패턴, 활동 일수와 결제를 동시에 향상시키는 전략 필요
- 결제 감소의 근본 원인은 **유저 리텐션이 짧아 결제 기회가 유지되지 않는 것**
- **사회 관계망**이 리텐션 및 결제에 미치는 영향 분석 필요

분석 목표

**관계망과 행동 패턴을 중심으로
유저 활동일수 및 결제를 동시에 향상시키는 전략 도출**

3 분석 설계

프로젝트 분석 단위, 범위, 방법

I. 개요

유저 단위 분석

- 유저 특성이 활동일수·결제에 미치는 영향
- 결제 직전 행동 패턴
- 핵심 세그먼트(활동일수·결제가 모두 높은 집단) 특성 비교

학교 단위 분석

- 학교별 특성과 네트워크 구조가 활동일수·결제에 미치는 영향
- 핵심 세그먼트(활동일수·결제가 모두 높은 집단) 특성 비교

유저 네트워크 영향 분석

- 투표·결제 행동의 친구 전염력
- 인플루언서의 학급 내 영향력

분석 범위·방법

- **변수 영향력 분석:**
 - 네트워크 지표(Degree Centrality, Clustering Coefficient, Modularity) 산출
 - Pearson 상관계수, Mann-Whitney U 검정, Logistic Regression(오즈비 해석), Random Forest(변수 중요도)로 활동일수·결제 영향 요인 도출
- **행동 패턴 분석:**
 - PrefixSpan 기반 시퀀스 마이닝으로 결제 직전 이벤트 시퀀스 추출
 - 핵심 이벤트 별 전환율, 등장 순서, 체류 시간 산출 및 Mann-Whitney U 검정으로 이벤트 간 체류 시간 비교
- **네트워크 분석:**
 - 친구의 투표·결제 행동 전염성(5일 이내 감염률) 측정
 - 대응표본 T검정으로 친구 수·투표 중심성이 높은 유저 등장 전후의 학급 활동성 변화 비교
- **핵심 세그먼트 분석:**
 - 독립표본 T검정으로 활동·결제 모두 높은 그룹과 다른 그룹의 특성 차이 검정

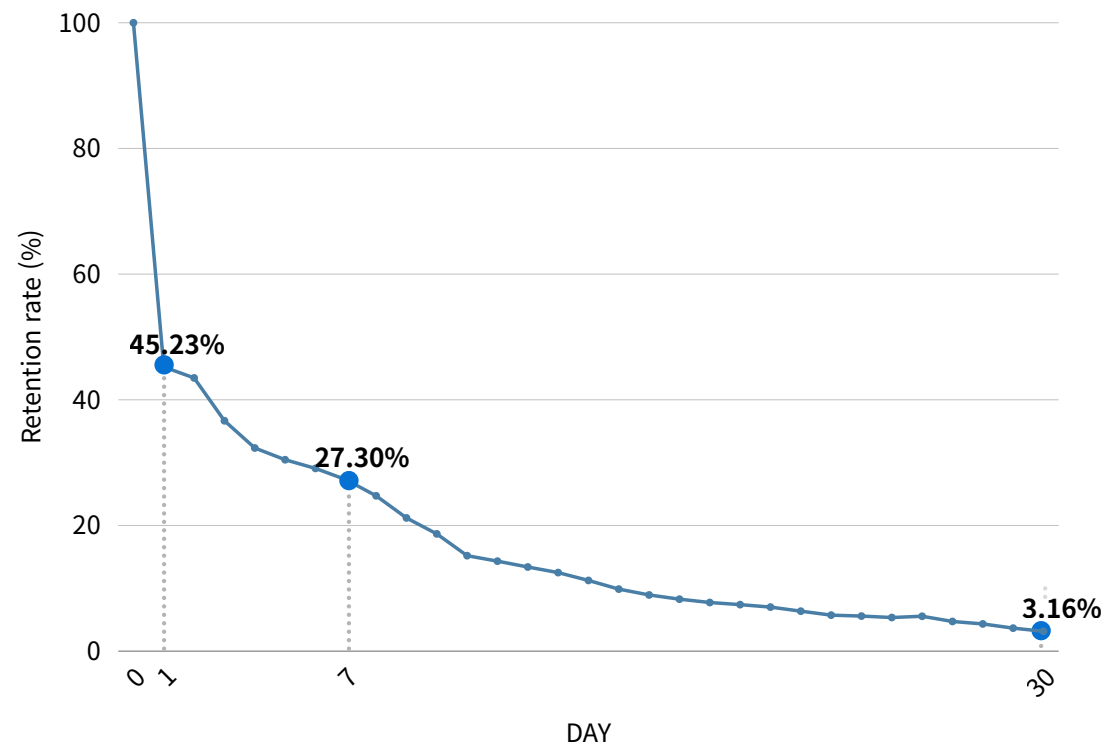
데이터 EDA

서비스 현황과 이용분포

- 서비스 현황
- 서비스 이용 분포
- 학교 네트워크 분포

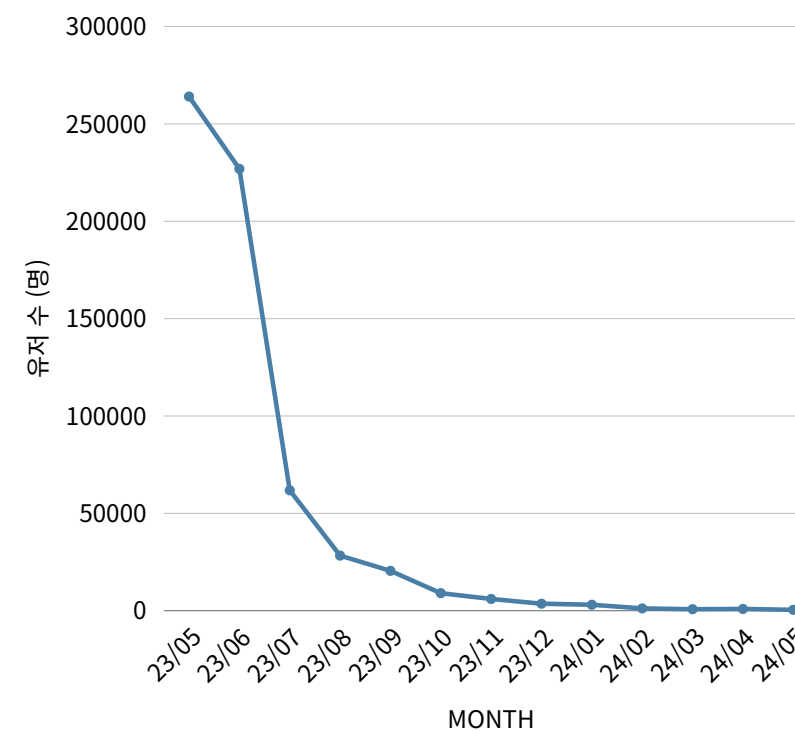
리텐션이 저조하며, 2023년 6월 이후 앱 방문자와 결제량이 동반 급락
→ 신규 유입 감소와 기존 이용자 이탈이 동시에 발생하며 서비스 전반의 지속성이 약화됨

30일 리텐션 곡선(Retention Curve)

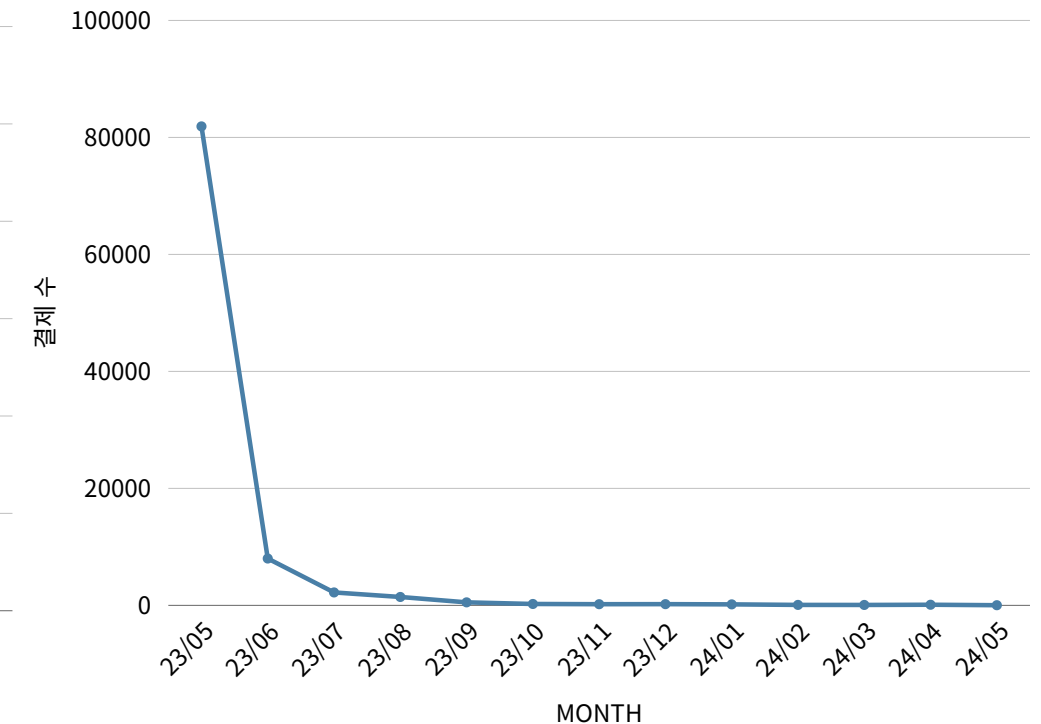


1일차 리텐션 45.23%에서 7일차 27.30%로 빠르게 하락,
30일차에는 3.16%만 잔존
→ 장기 이용자 비중 매우 낮음

월별 방문자 추이(MAU)



월별 결제량 추이



- 2023년 6월 이후 월간 활성 이용자 수가 급격히 감소, 이후 완만한 하락세 지속
- 방문자 감소와 비슷한 시기에 월별 결제량도 동반 급락, 이후 회복 없이 저조한 수준 유지

3 학교 네트워크 지표 분포

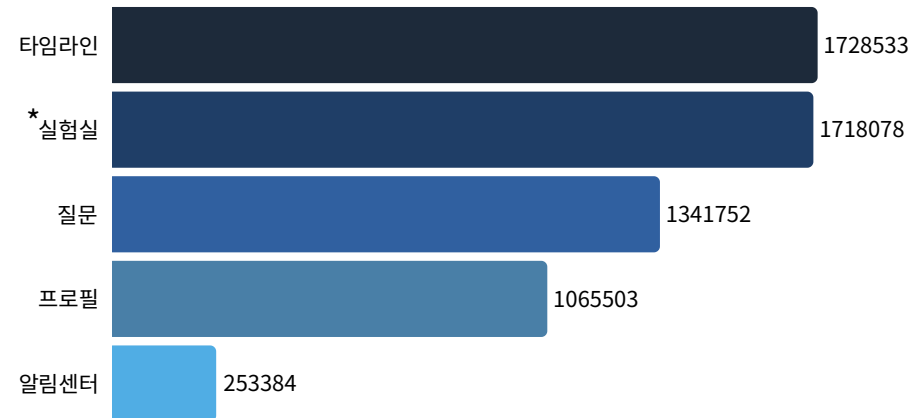
SNA분석을 통한 네트워크 분포 파악

II. EDA

유저 이벤트 이용 분포

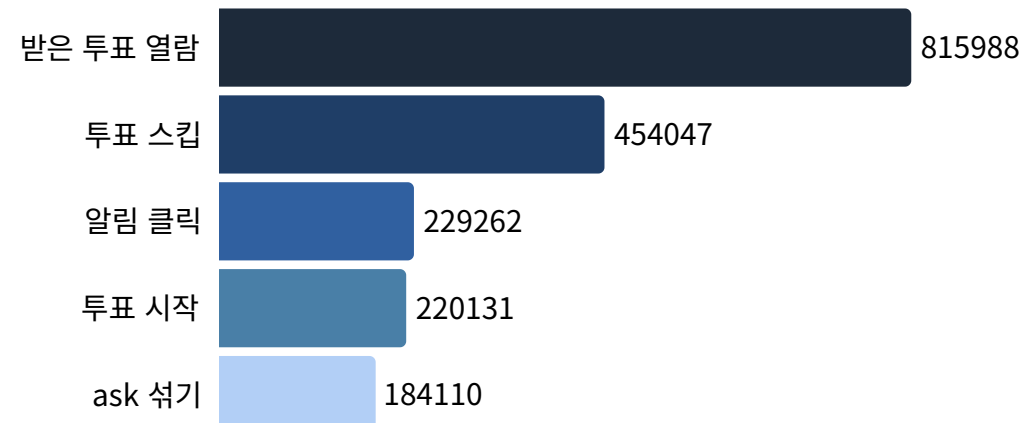
타임라인, 실험실, 질문탭 방문 빈도가 높으며
지목 받은 투표 열람 기능을 가장 많이 이용하는 것으로 나타남

화면 단위 이벤트 이용 분포



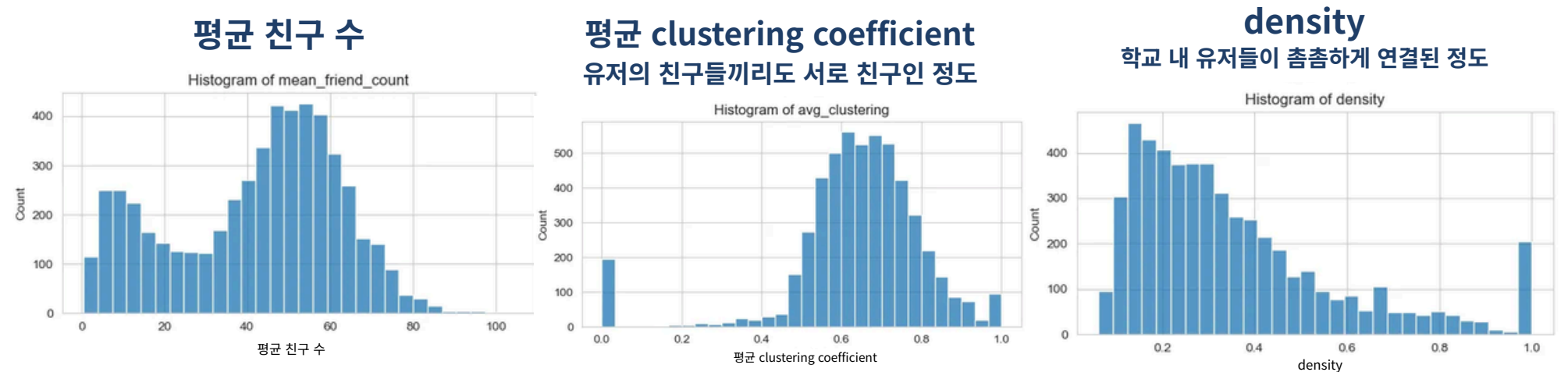
*실험실: 어플의 실험적 기능을 이용하는 탭으로 추정

행동 단위 이벤트 이용 분포

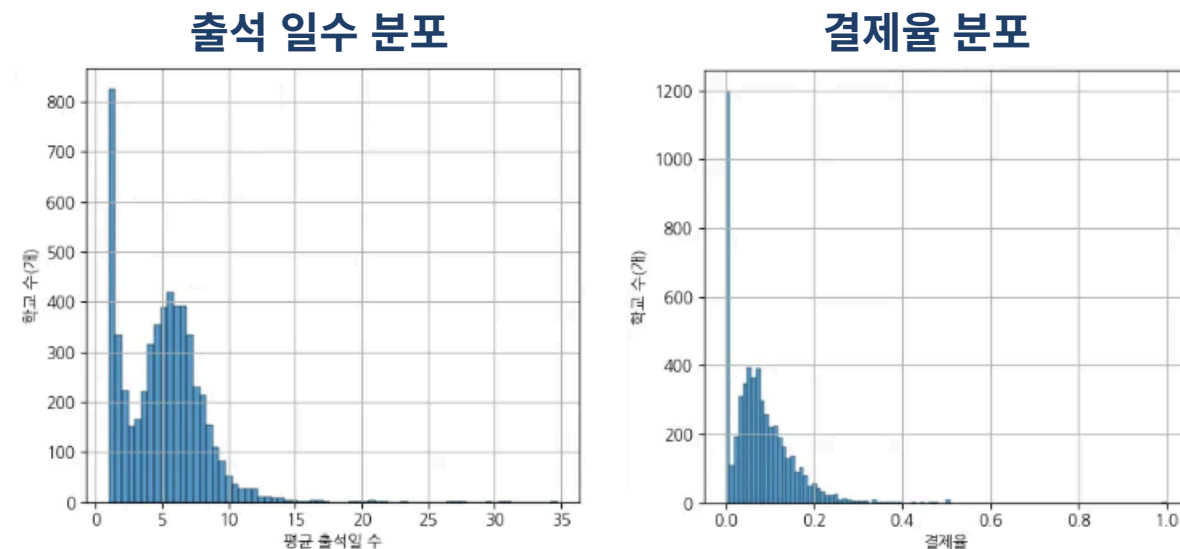


학교 단위 지표 분포

학교 네트워크는 구조적으로 다양한 특징을 보임
평균적으로 친구 수 50명 내외의, 친구들 간 응집성은 높고 전체 유저 간 연결 밀도는 낮은 구조가 나타남



학교의 평균 출석일수는 0일과 7일 근처에서 가장 많았고,
결제율은 대부분 0 ~ 0.2 구간에 분포함



메인 분석

■ 유저 단위

- a. 활동일수 영향 요인 종합 분석
- b. 결제 영향 요인 종합 분석
- c. 결제 전 행동 패턴 분석

■ 학교 단위

- a. 활동일수 영향 요인 종합 분석
- b. 결제 영향 요인 종합 분석

■ 유저 네트워크 영향

- a. 결제 전염성
- b. 투표 전염성
- c. 인플루언서 영향력 분석

■ 핵심 세그먼트 분석

개별 유저의 활동일수에 영향을 미치는 주요 요인은 무엇이며, 그 영향력의 크기와 방향은 어떻게 될까?

친구 수, 학급·학교 ID, 학교 유저 수가 공통적으로 높은 영향력을 보였다

● 분석 방법

- 개인의 활동일수가 7일 이상(분포 상위 25%)인지 여부를 종속변수로 설정
- 상관·그룹 간 차이 검정(Mann-Whitney U)·로지스틱 회귀·랜덤포레스트를 종합 비교

● 랜덤 포레스트 모델 개요

- 모델 성능: accuracy 0.66, precision 0.71, recall 0.71, **ROC AUC 0.79**
- 샘플링 방식: SMOTE(Synthetic Minority Over-sampling Technique) 적용, 종속변수 비율 불균형 보정

● 변수별 종합 분석 결과

변수	상관계수	집단 차이 (p-value < 0.05)	로지스틱 회귀 (Odds rate)	랜덤포레스트 (Featue importance)	종합 해석
학급 id		o		1위 (0.28)	랜덤포레스트에서 변수 중요도 1, 3위 → 개인이 속한 네트워크 자체가 활동일수에 강하게 작용
학교 id		o		3위 (0.19)	
친구 수	0.2	o	1.01	2위 (0.25)	상관·집단 차이 분석 모두에서 유의하고 랜덤포레스트 변수 중요도 2위 → 친구수는 활동일수의 일관된 핵심 변수
학교 유저 수		o	0.999	0.19	랜덤포레스트에서는 중간 정도의 중요도를 보이고, OR은 1에 매우 가까움 → 단독 영향 보다, 다른 변수와 조합될 때(예: 학급 id, 친구 수 등) 간접적인 영향이 있을 수 있음
투표 열람 횟수		o	11.54	0.01	로지스틱 회귀에서 OR>1 → 참여 활동이 곧 활동일수 증가로 연결
결제 여부		o	2.2	0.01	로지스틱 회귀에서 OR>1 → 소비 활동이 곧 활동일수 증가로 연결
차단 횟수		o	1.47	0.01	로지스틱 회귀에서 OR>1 → 사회적 상호작용의 갈등 지표도 활동일수와 일정한 관련을 보임

※상관계수, 그룹 간 차이 검정, 로지스틱 회귀 결과는 유의수준 5%(p<0.05)에서 통계적으로 유의한 변수만 표기

1 유저 단위: 결제 영향 요인 종합 분석

개별 유저의 결제에 영향을 미치는 주요 요인은 무엇이며, 그 영향력의 크기와 방향은 어떻게 될까?

활동일수, 학급 id, 학교 유저 수가 높은 영향력을 보였다

- 분석 방법

○ 개인의 결제 여부를 종속변수로 설정

○ 상관·그룹 간 차이 검정(Mann-Whitney U)·로지스틱 회귀·랜덤포레스트를 종합 비교
- 랜덤 포레스트 모델 개요

○ 모델 성능: accuracy 0.81, precision 0.84, recall 0.84, **ROC AUC 0.90**

○ 샘플링 방식: SMOTE(Synthetic Minority Over-sampling Technique) 적용, 종속변수 비율 불균형 보정

• 변수별 종합 분석 결과

변수	상관계수	집단 차이 (p-value < 0.05)	로지스틱 회귀 (Odds rate)	랜덤포레스트 (Featue importance)	종합 해석
학급id		o	1.00	1위 (0.30)	랜덤포레스트에서 변수 중요도 1위 → 속한 학급 네트워크 특성이 결제 여부에 가장 큰 영향
학교 유저 수		o	0.99	2위 (0.26)	랜덤포레스트에서 변수 중요도 2위 → 학교 규모가 결제 가능성을 높이는 핵심 요인
활동일수		o	1.04	3위 (0.20)	랜덤포레스트에서 변수 중요도 3위 → 개인의 지속적 활동이 결제 가능성을 높이는 핵심 요인
친구 수		o	1.01	0.18	랜덤포레스트에서 변수 중요도 상위권 → 관계망의 크기가 결제 행동과 관련됨
성별 (남성=1)		o	1.61	0.01	로지스틱 회귀에서 OR>1 → 남학생 비율이 높은 경우 결제 확률이 높음
학교 종류 (고등학교=1)		o	1.59	0.00	로지스틱 회귀에서 OR>1 → 고등학교에서 결제 확률이 높음
푸시알림 (ON=1)		o	1.31	0.01	로지스틱 회귀에서 OR>1 → 푸시알림을 받는 경우 결제 확률이 높음

※상관계수, 그룹 간 차이 검정, 로지스틱 회귀 결과는 유의수준 5%(p<0.05)에서 통계적으로 유의한 변수만 표기

1 유저 단위: 결제 직전 행동 패턴 도출

결제 직전의 어떤 행동이 결제 전환율을 높이는 핵심 요인인가?

III. 메인분석

출석체크·투표 완료는 결제 전환율을 크게 높이는 핵심 진입점으로 나타났으며, 질문탭은 직전 집중도는 낮지만 등장 비율이 높아 지속적·간접적 전환 기여도가 큼

데이터 범위 한계

로그 데이터(2023-07-18~2023-08-04) 기준 분석

- 전체 데이터 기준 유저 결제율: 8.7%
- 로그 데이터 기준 세션별 결제율: 0.20%

분석 설계

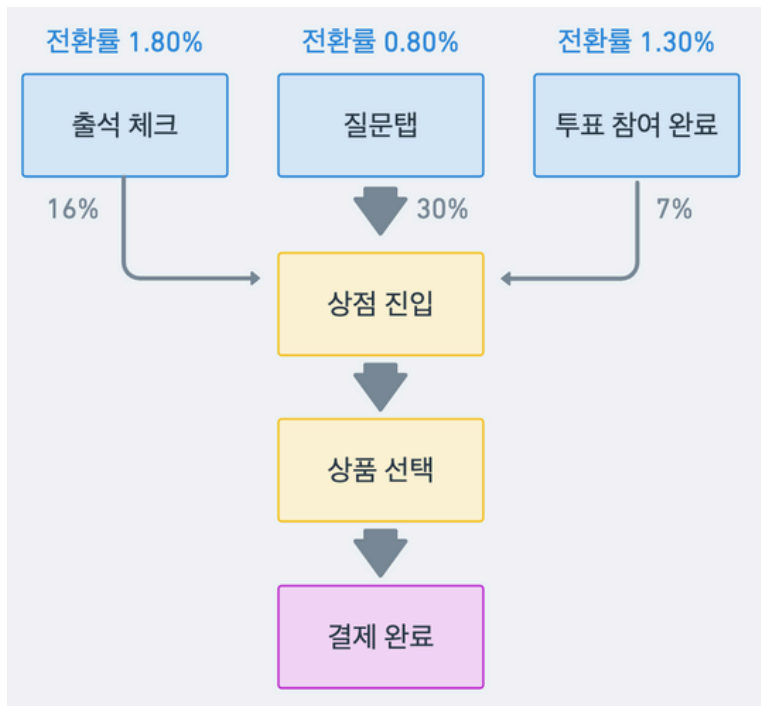
이벤트별 전환율, 결제 세션 상점 진입 전 이벤트 등장 단계, 등장 비율, 체류 시간 비교

- 결제 전환율 = 특정 이벤트가 발생한 세션 중 결제가 완료된 세션 비율
 - 기준선: 전체 세션별 전환율(0.20%)
- 결제 세션 상점 진입 전 이벤트 등장 단계: 결제 세션에서 상점 진입 직전 3단계 중 해당 이벤트가 등장한 순서를 확인 (1단계 = 상점 진입 직전)
- 체류 시간: 결제 직전(결제 세션)과 상점 방문 직전(비결제 세션) 시퀀스의 평균 체류 시간 차이를 Mann-Whitney U 검정으로 비교

이벤트별 종합 분석 결과

행동 (이벤트)	결제 전환율 (%)	상점 진입 전 등장 단계	상점 진입 직전 등장 비율	체류 시간 (초) (결제세션 / 비결제세션)	해석
출석체크	1.80% (기준선 대비 9배)	1단계 집중	0.16	4.7 / 4.9	전환율 ↑, 결제로 이어지는 핵심 진입점
투표 참여 완료	1.30% (기준선 대비 6배)	1단계 집중	0.07	7.9 / 7.2	전환율 ↑, 결제로 이어지는 핵심 진입점
질문탭	0.80%	분산	0.30	2.2 / 1.9 *	가장 자주 등장하는 앱의 핵심 기능, 결제자 체류 시간 ↑
알림 상세 확인	0.80%	분산	0.04	25.0 / 26.4	간접 영향 가능
지목 받은 투표 확인	0.60%	1단계 집중	0.09	10.3 / 11.9	전환율 ↓, 체류 시간 ↑
상점 진입	9.20% (기준선 대비 16배)	(필수)	-	3.9 / 7.8 *	비결제자는 2배 길게 체류
상품 클릭	100%	(필수)	-	14.8 / NaN	구매 단계 이벤트

*표시된 행동은 결제자와 비결제자 간의 결제 직전 체류 시간에 통계적으로 유의미한 차이가 있는 항목



2 학교 단위: 활동일수 영향 요인 종합 분석

III. 메인분석

학교의 평균 활동일수에 영향을 미치는 주요 요인은 무엇이며, 그 영향력의 크기와 방향은 어떻게 될까?

결제율, 평균 친구 수가 모든 분석에서 일관되게 높은 영향력을 보였으며 **density**도 변수 중요도 상위권에 위치했다

- 분석 방법

○ 학교 평균 활동일수가 7일 이상인(분포 상위 25%)인지 여부를 종속변수로 설정

○ 상관·그룹 간 차이 검정(Mann-Whitney U)·로지스틱 회귀·랜덤포레스트를 종합 비교
- 랜덤 포레스트 모델 개요

○ 모델 성능: accuracy 0.87, precision 0.91, recall 0.84, **ROC AUC 0.94**

○ 샘플링 방식: SMOTE(Synthetic Minority Over-sampling Technique) 적용, 종속변수 비율 불균형 보정

• 변수별 종합 분석 결과

변수	상관계수	집단 차이 (p-value < 0.05)	로지스틱 회귀 (Odds rate)	랜덤포레스트 (Featue importance)	종합 해석
결제율	0.59	o	1.45	1위 (0.14)	상관·집단 차이·회귀 모두에서 유의하며 랜덤포레스트 변수 중요도 1위 → 결제율이 높은 학교일수록 활동일수도 높음
평균 친구 수	0.65	o	2.61	2위 (0.11)	랜덤포레스트 변수 중요도 2위, 로지스틱 회귀 OR=2.61 → 네트워크 연결성이 높은 학교가 더 활발하게 활동
신고 수	0.57	o		3위 (0.10)	랜덤포레스트 변수 중요도 3위 → 활동성이 높은 학교에서는 부정·갈등 상황의 발생 빈도도 함께 높을 가능성
density		o		0.09	랜덤포레스트 변수 중요도 4위지만 방향성은 뚜렷하지 않음 → 밀도의 효과는 네트워크 구조 특성에 따라 달라질 수 있음
학교 유저 수	0.48	o	0.52	0.08	상관분석에서는 양의 상관이나 회귀에서는 OR<1 → 규모가 크다고 항상 활동일수가 높지 않으며 관계의 질이 더 중요할 수 있음
평균 clustering coefficient		o	1.35	0.08	로지스틱 회귀에서 OR>1 → 친구의 친구끼리도 많이 연결될 수록 활동일수가 증가하는 경향
modularity		o	1.22	0.06	로지스틱 회귀에서 OR>1 → 적정 수준의 네트워크 분절성은 활동일수를 높이는 요인임
학교 종류 (고등학교=1)		o	0.58	0.05	로지스틱 회귀에서 OR<1 → 중학교가 평균적으로 더 높은 활동일수를 보이는 경향

※상관계수, 그룹 간 차이 검정, 로지스틱 회귀 결과는 유의수준 5%(p<0.05)에서 통계적으로 유의한 변수만 표기

2 학교 단위: 결제율 영향 요인 종합 분석

III. 메인분석

학교의 결제율에 영향을 미치는 주요 요인은 무엇이며, 그 영향력의 크기와 방향은 어떻게 될까?

활동일수, 남성 비율, 평균 친구 수가 모든 분석에서 일관되게 높은 중요도를 보였으며 clustering도 모델링 상위권에 위치함

• 분석 방법

- 학교 평균 결제율이 0.1% 이상(분포 상위 25%)인지 여부를 종속변수로 설정
- 상관·그룹 간 차이 검정(Mann-Whitney U)·로지스틱 회귀·랜덤포레스트를 종합 비교

• 랜덤 포레스트 모델 개요

- 모델 성능: accuracy 0.84, precision 0.85, recall 0.83, **ROC AUC 0.91**
- 샘플링 방식: SMOTE(Synthetic Minority Over-sampling Technique) 적용, 종속변수 비율 불균형 보정

• 변수별 종합 분석 결과

변수	상관계수	집단 차이 (p-value < 0.05)	로지스틱 회귀 (Odds rate)	랜덤포레스트 (Featue importance)	종합 해석
활동일수	0.59	o	2.03	1위 (0.19)	랜덤포레스트에서 변수 중요도 1위, 로지스틱 회귀 OR=2.03 → 활발히 활동하는 학교일수록 결제율도 높음
남성 비율		o	1.81	2위 (0.15)	랜덤포레스트에서 변수 중요도 2위, 로지스틱 회귀에서 OR=1.81 → 남학생 비율이 높은 학교에서 결제율이 높음
평균 친구 수	0.48	o	1.67	3위 (0.09)	상관·집단 차이 분석에서 유의하며 랜덤포레스트 변수 중요도 3위, OR=1.67 → 네트워크 연결성이 결제율 상승에 핵심 역할
평균 clustering coefficient		o	1.2	0.07	로지스틱 회귀에서 OR>1 → 친구의 친구끼리도 많이 연결될 수록 결제율 높음
density		o	1.55	0.07	로지스틱 회귀에서 OR=1.54 → 네트워크 밀도가 높은 학교에서 결제율이 높음
학교 유저 수	0.32	o		0.07	상관분석에서 양의 상관이나 랜덤포레스트에서는 하위권 → 규모 자체보다는 관계의 질과 구조가 결제에 더 큰 영향
푸시ON비율		o	1.32	0.07	로지스틱 회귀에서 OR>1 → 푸시알림을 많이 수신하는 학교가 결제율이 높음
학교 종류(고등학교=1)		o	2.53	0.04	로지스틱 회귀에서 OR=2.52 → 고등학교에서 결제율이 높음

※상관계수, 그룹 간 차이 검정, 로지스틱 회귀 결과는 유의수준 5%(p<0.05)에서 통계적으로 유의한 변수만 표기

3 유저의 네트워크 영향력: 전염성

내가 결제(투표)하면 친구들도 결제(투표)할까?

III. 메인분석

분석 방법 소개

결제 전염성 기준

1. 결제한 친구의 비율 비교

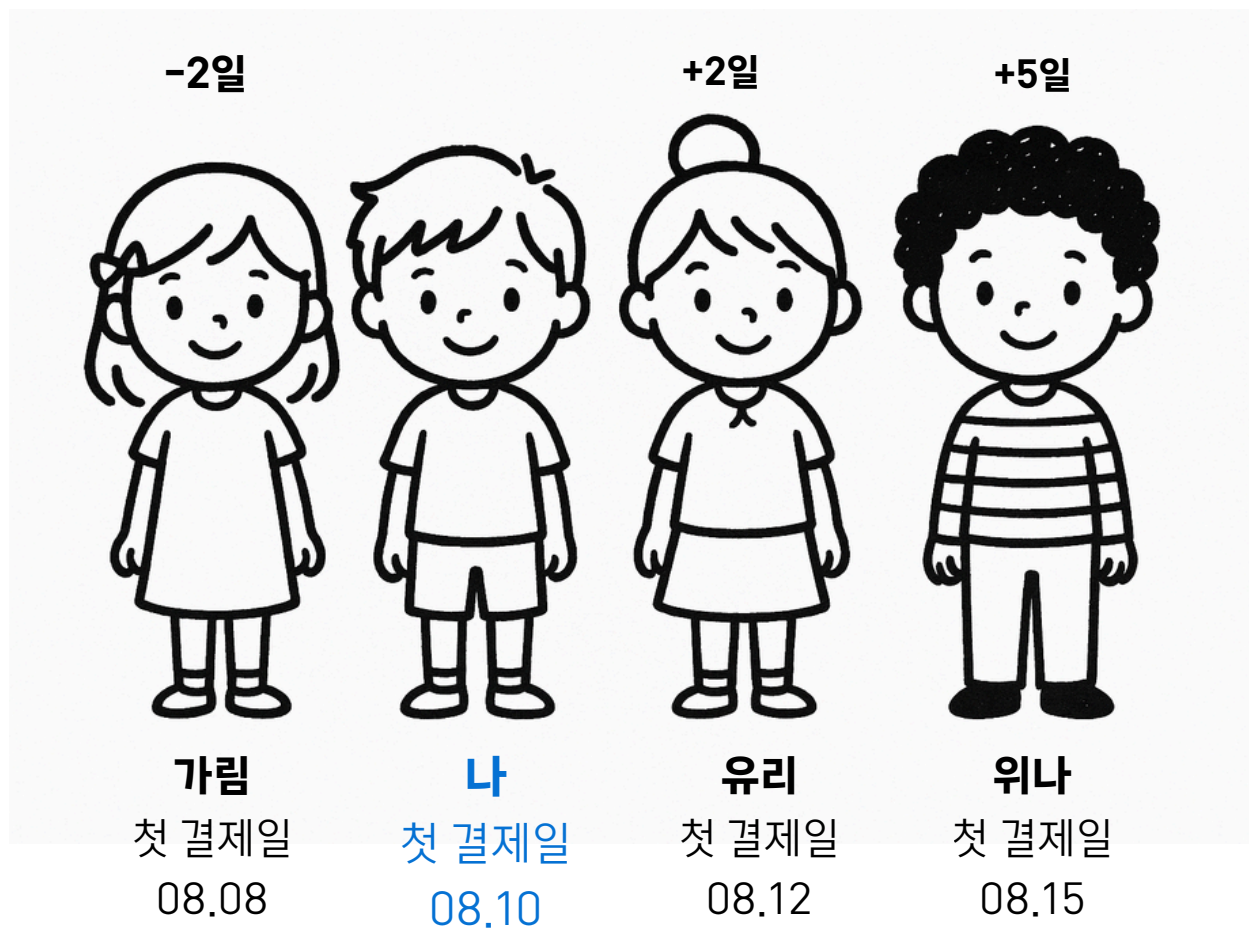
- 비결제자의 결제한 친구 비율과 결제자의 결제한 친구 비율을 비교
- 예) 나의 친구 중 결제한 친구의 비율 구하기
 - 나의 친구 가림, 유리, 위나 모두가 결제를 했으므로 비율은 3명/3명 = 1

2. 친구 결제까지의 누적 비율 계산

- 나를 기준으로 유리의 확산 소요일은 2일, 위나는 5일, 가림은 확산되지 않음.
- 유저의 확산 소요일을 집계하여 누적 비율 계산(중복 포함)

3. 감염률

- 누적 비율 확인 결과, 결제행동과 투표행동 모두 절반의 확산 소요일이 2-3일임을 확인
- 확산 소요일이 5일 이내인 경우를 감염으로 정의

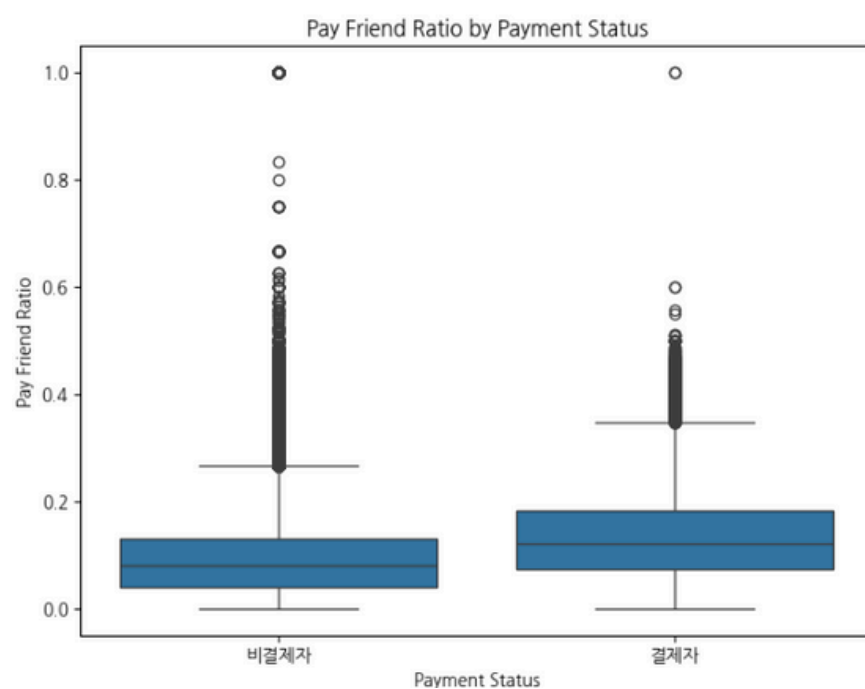


3 유저의 네트워크 영향력: 결제 전염성

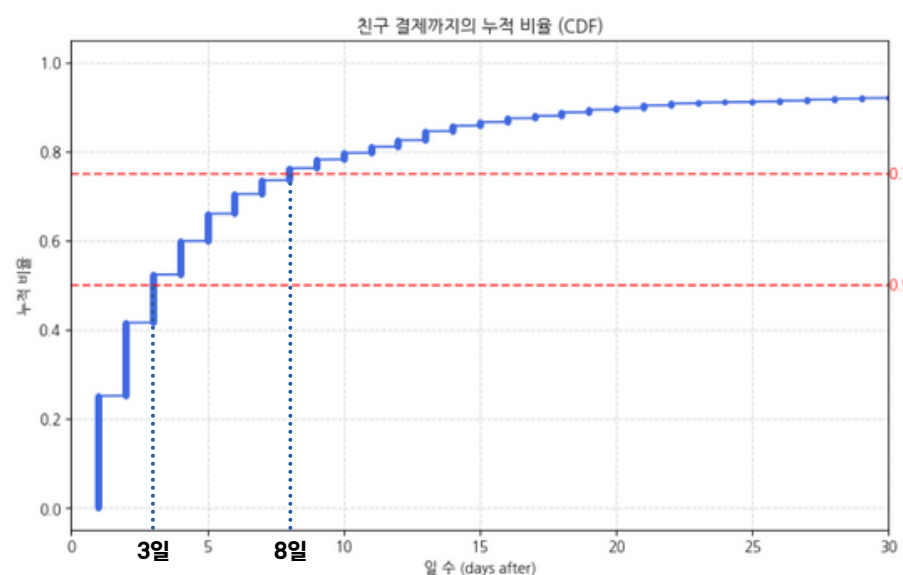
내가 결제하면 친구들도 결제할까?

III. 메인분석

전체 친구의 60%가 최초 결제 후 5일 이내에 결제,
결제가 짧은 기간 내에 네트워크 전반으로 빠르게 확산되는 것으로 나타남

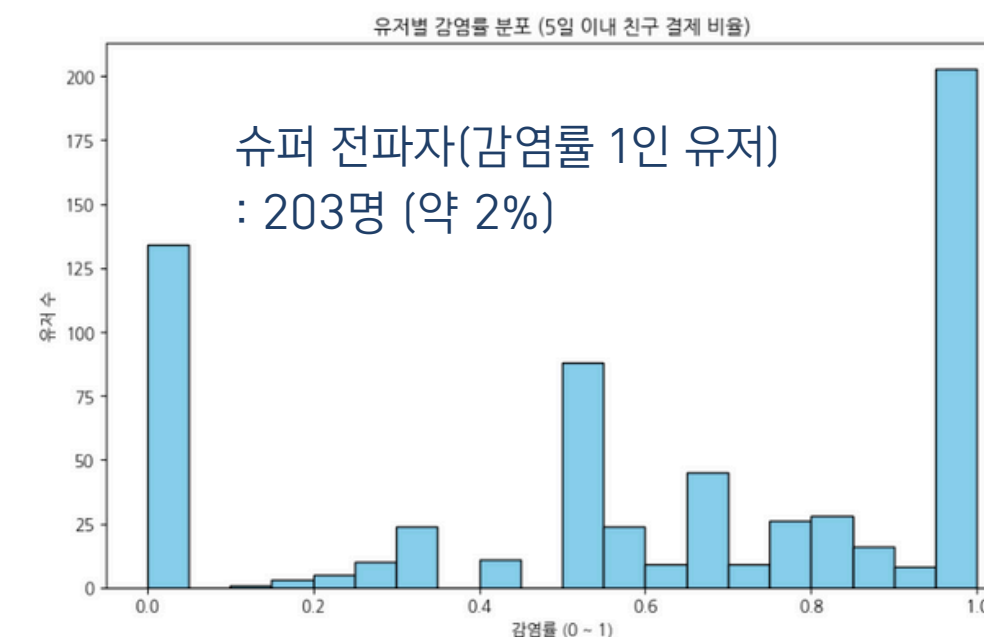


비결제자의 친구 중 결제이력이 있는 평균 친구 비율
: **9.25%**
결제자의 친구 중 결제이력이 있는 평균 친구 비율
: **13.27%**
P-value: 0.000
coef: 5.8632



평균 확산 소요일: **12.3일 후**
상위 50% 확산 소요일: **3일 후**
상위 75% 확산 소요일: **8일 후**
그래프가 거의 수평이 되는 시점: **30.0일 후**
최초 결제일로부터 30일 이내에 결제한 친구 비율
: 92.01%

감염률 = (유저 최초 결제일 이후 5일 내 결제한 친구 수) / (전체 친구 수)



실제 친구 네트워크에서 결제 유저의 친구 중 평균 감염률: **13.07%**

유의성 검정 (Permutation Test)

- 유의성 검정을 위해 1000번의 무작위 샘플링에서 나온 평균 감염률 분포와 비교
- p-value = 0.0000 → 통계적으로 유의미하게 높은 수준
- 즉, 결제는 단순한 우연이 아닌 네트워크 기반 전염성 효과가 존재할 가능성이 높다고 해석

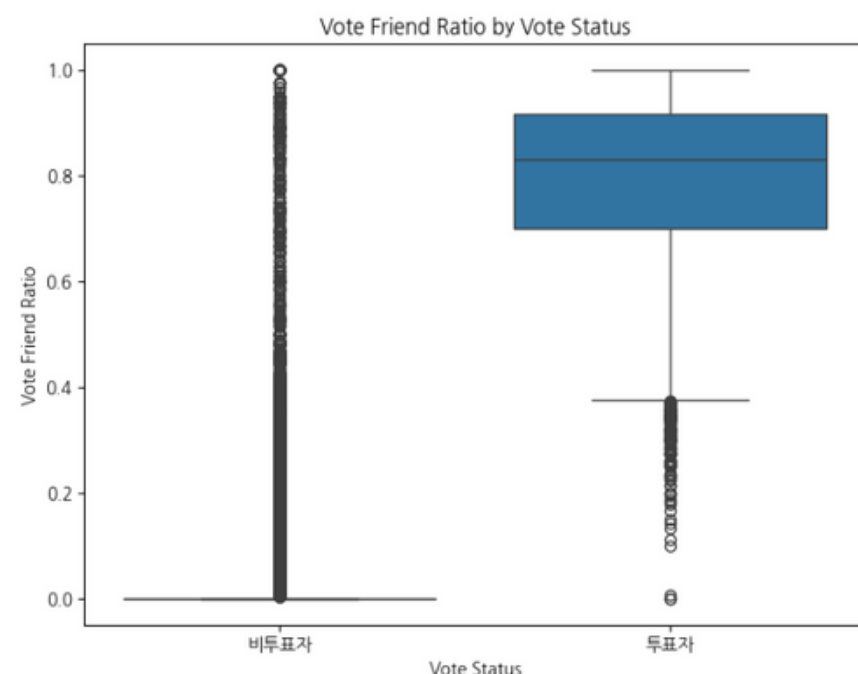
※ 메모리 용량 문제로 인해 전체 데이터 중 10,000건을 무작위로 샘플링하여 분석 진행

3 유저의 네트워크 영향력: 투표 전염성

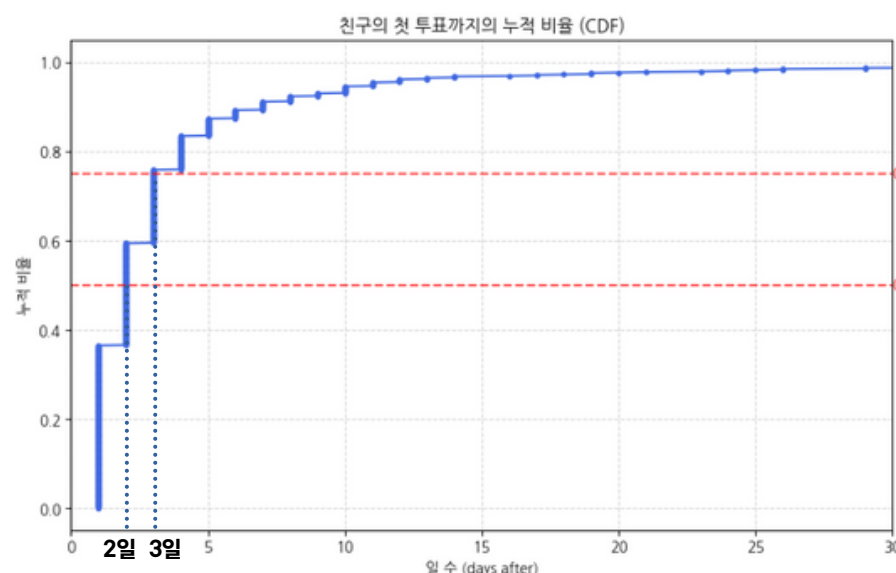
내가 투표하면 친구들도 투표할까?

III. 메인분석

투표 전염성은 평균 감염률이 약 78%이며,
절반 이상이 전파되는 시점이 2일 이내로 매우 빠르게 전염

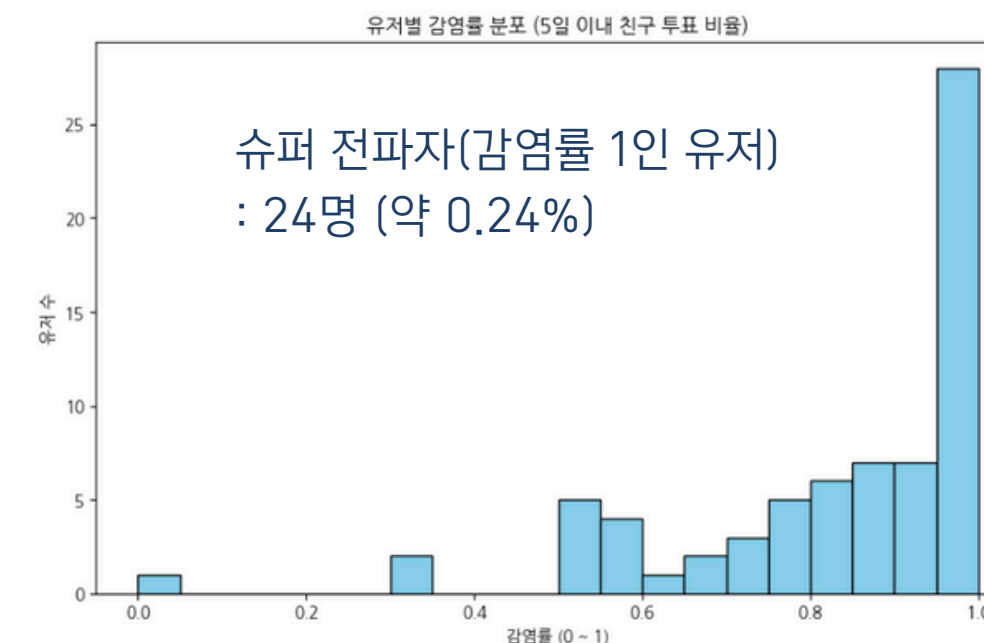


비투표자의 친구 중 투표경험이 있는 평균 친구 비율
: **0.187%**
투표자의 친구 중 투표경험이 있는 평균 친구 비율
: **78.77%**
P-value: 0.000
coef: 16.547



평균 확산 소요일: **4.19일 후**
상위 50% 확산 소요일: **2일 후**
상위 75% 확산 소요일: **3일 후**
그래프가 거의 수평이 되는 시점: **16.0일 후**
최초 투표일로부터 16일 이내에 투표한 친구 비율
: **96.89%**

감염률 = (유저 최초 투표일 이후 5일 내 투표한 친구 수) / (전체 친구 수)



실제 친구 네트워크에서 투표한 유저의 친구 중 평균 감염률: **77.71%**

유의성 검정 (Permutation Test)

- 유의성 검정을 위해 1000번의 무작위 샘플링에서 나온 평균 감염률 분포와 비교
- p-value = 0.0000 → 통계적으로 유의미하게 높은 수준
- 즉, 투표는 단순한 우연이 아닌 네트워크 기반 전염성 효과가 상당한 수준!

※메모리 용량 문제로 인해 전체 데이터 중 10,000건을 무작위로 샘플링하여 분석 진행

3 유저의 네트워크 영향력: 인플루언서 영향력 분석

인플루언서는 누구일까? 영향력이 있을까?

III. 메인분석

친구 수와 투표활동을 기준으로 인플루언서를 선정한 결과, 인플루언서의 존재 여부에 따라 학급의 접속일수·투표·결제에서 유의미한 차이가 확인됨

인플루언서 기준 선정

accounts_user (유저 테이블)

friend_id_list 컬럼 => 친구 수

accounts_userquestionrecord (투표 기록 테이블)

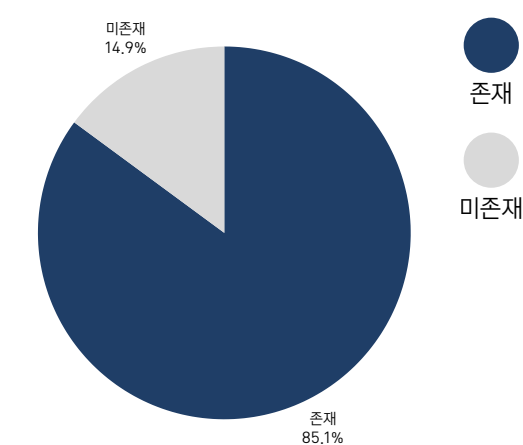
사회연결망 분석을 사용해 투표네트워크 생성 후 다음 지표들 계산

- In-Degree : 이 유저가 다른 사람들에게 투표를 받은 총 횟수
- Out-Degree : 이 유저가 다른 사람들에게 투표를 한 총 횟수
- PageRank :
 - 원리: 각 유저가 가진 영향력을 자신이 투표한 수만큼 N분의 1로 나누어 전달하는 과정을 네트워크 전체가 안정될 때까지 반복적으로 재계산
 - 결과: 단순히 표를 많이 받은 유저가 아닌, 영향력 있는 소수로부터 집중적인 지지를 받는 네트워크 내 '여론 주도층'을 식별

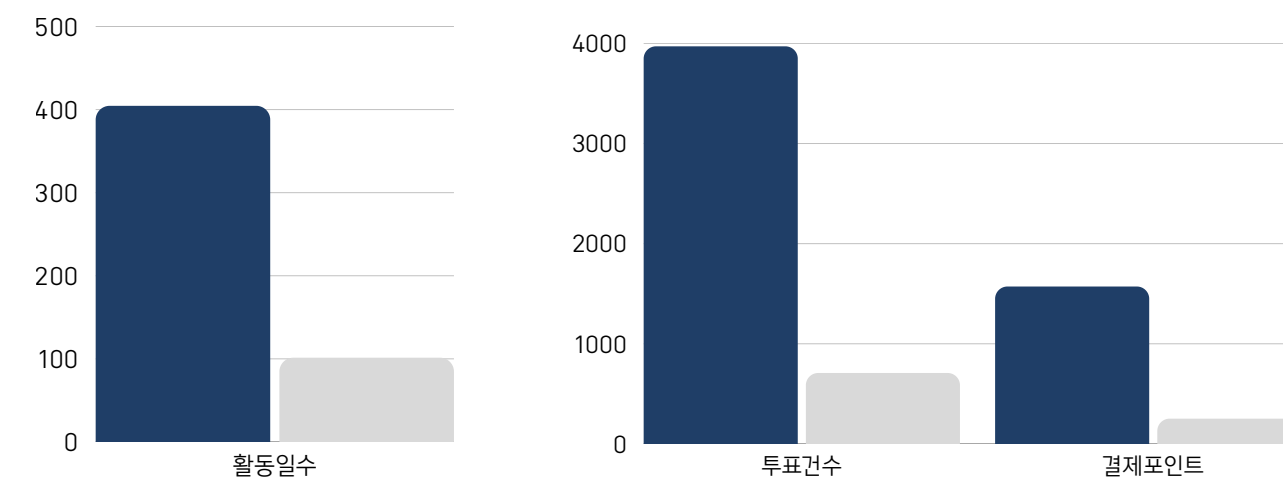
4가지 지표를 바탕으로 데이터 표준화 후 클러스터링을 위한 최적의 k 찾기 (Elbow Method)

k-means 클러스터링 진행 후 가장 활동적인 클러스터를 인플루언서 집단으로 정의

인플루언서 보유 학급 비율



인플루언서 유무에 따른 학급 내 활동량 평균 차이 (독립표본 T검정 완료)



3 유저의 네트워크 영향력: 인플루언서 영향력 분석

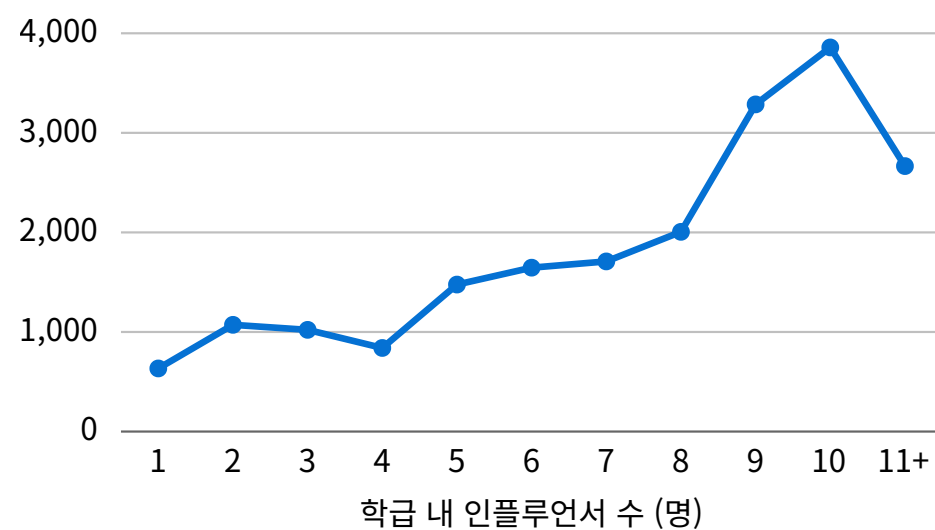
인플루언서가 학급에 미치는 영향력은 어느정도일까?

III. 메인분석

학급 내 인플루언서의 수가 증가할수록 결제 포인트의 양, 투표 건수의 양, 활동일수 모두 증가하는 추세를 보임

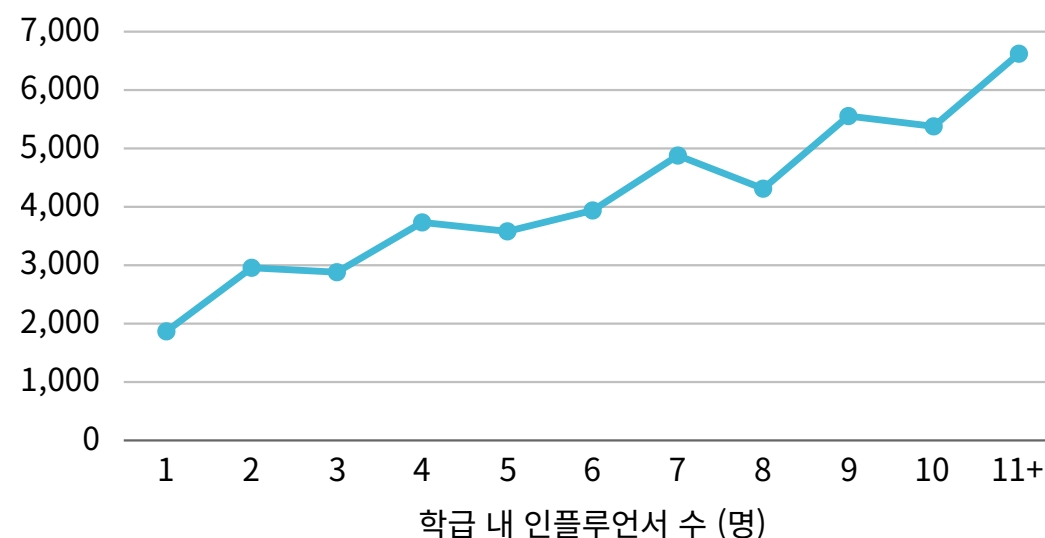
학급 내 인플루언서 수에 따른 인플루언서 등장 전후 학급 내 평균 활동량 차이

학급 내 인플루언서 수 별 결제 포인트



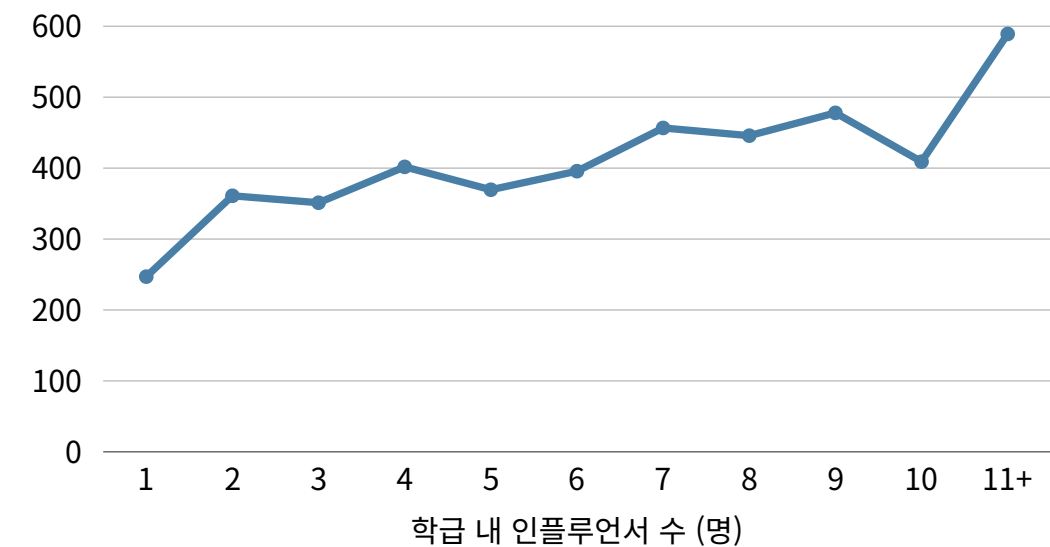
학급 내 인플루언서 수가 증가할 수록
학급 당 **결제 포인트**가 늘어나는 양상

학급 내 인플루언서 수 별 투표수



학급 내 인플루언서 수가 증가할 수록
학급 당 **투표수**가 증가하는 양상

학급 내 인플루언서 수 별 학급 활동일수



학급 내 인플루언서 수가 증가할 수록
학급의 **활동일**이 늘어나는 양상

인플루언서 등장 이전과 이후를 비교한 **대응 표본 T-검정** 결과,
학생들의 **평균 결제 포인트 양, 평균 투표 수, 평균 활동일수**가
통계적으로 유의미하게 **증가**한 것으로 나타남

3 유저의 네트워크 영향력: 인플루언서 영향력 분석

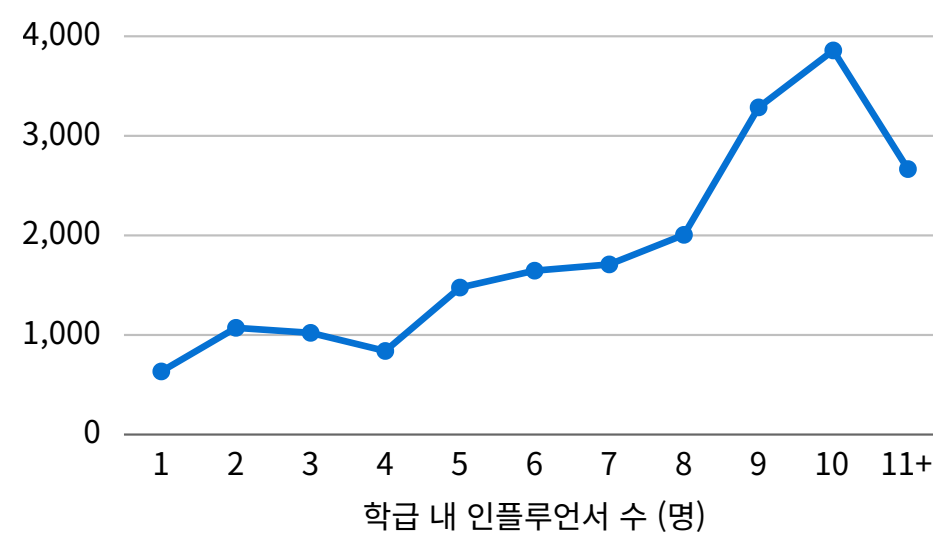
인플루언서가 학급에 미치는 영향력은 어느정도일까?

III. 메인분석

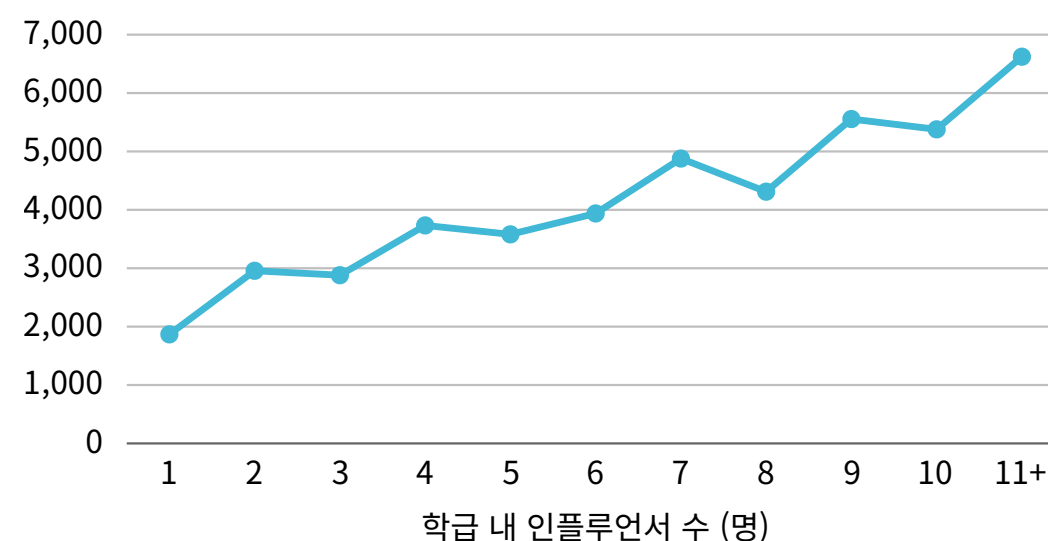
학급 내 인플루언서의 수가 증가할수록 결제 포인트의 양, 투표 건수의 양, 활동일수 모두 증가하는 추세를 보임

학급 내 인플루언서 수에 따른 인플루언서 등장 전후 학급 내 평균 활동량 차이

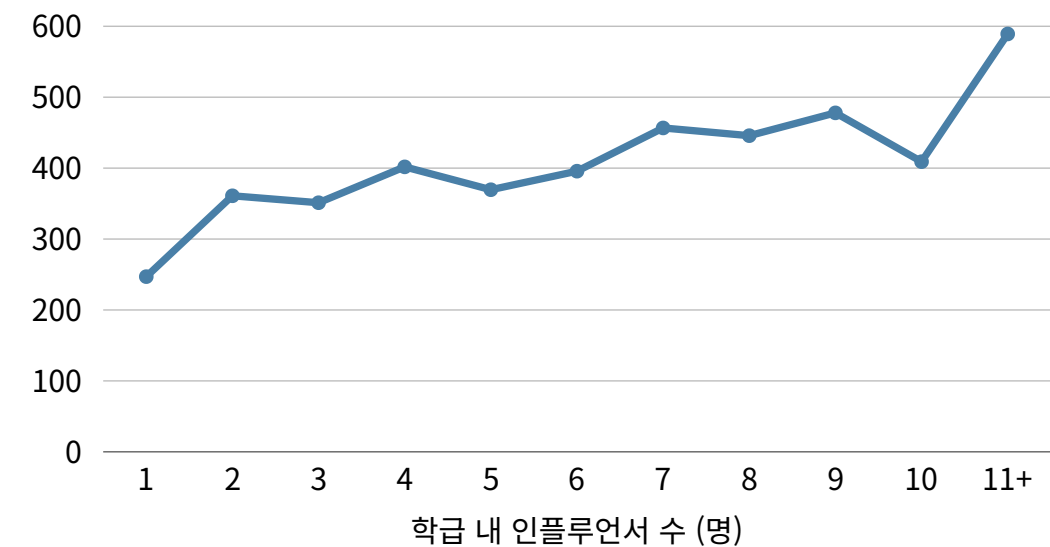
학급 내 인플루언서 수 별 결제 포인트



학급 내 인플루언서 수 별 투표수



학급 내 인플루언서 수 별 학급 활동일수



인플루언서 등장 이전과 이후를 비교한 **대응 표본 T-검정** 결과,
학생들의 **평균 결제 포인트 양, 평균 투표 수, 평균 활동일수**가
통계적으로 유의미하게 **증가**한 것으로 나타남

4 핵심 세그먼트 분석 - 개인 및 학교 단위

III. 메인분석

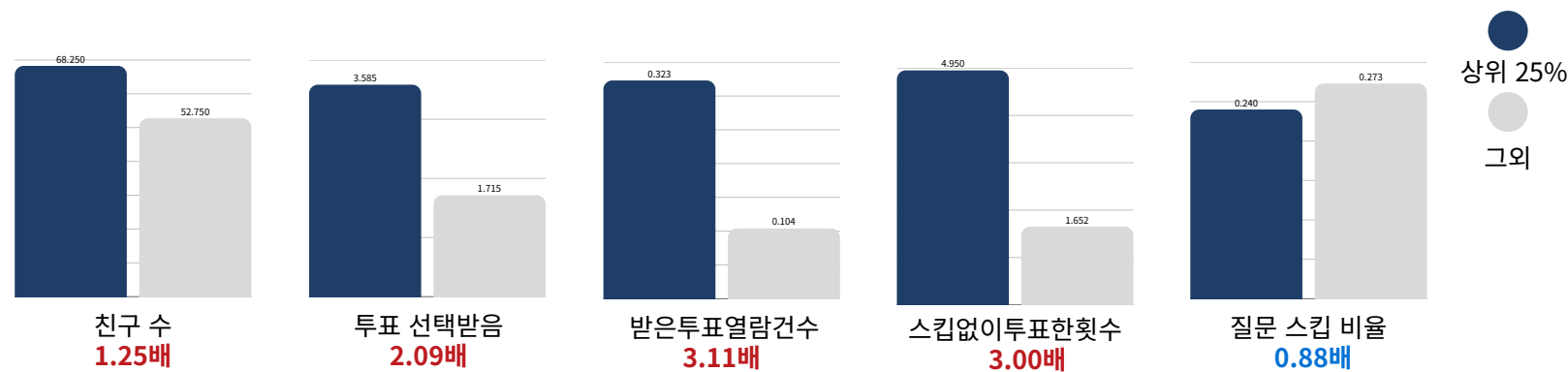
활동일수와 결제 모두에서 상위 25%에 속하는 유저 및 학교의 특징은 무엇일까?

상위 유저들은 그외 학생들에 비해 친구 수·투표 관련 지표·차단 관련 지표에서 유의미하게 높고, 질문 스킵 비율은 낮음

상위 학교는 촘촘하지만 분절된 네트워크를 가지며, 남성 비율과 차단·신고 활동이 높게 나타남

*핵심 세그먼트와 그 외 유저들 평균 활동량 변수 차이

*핵심 세그먼트 : 활동일수 및 결제횟수 상위 25% 유저

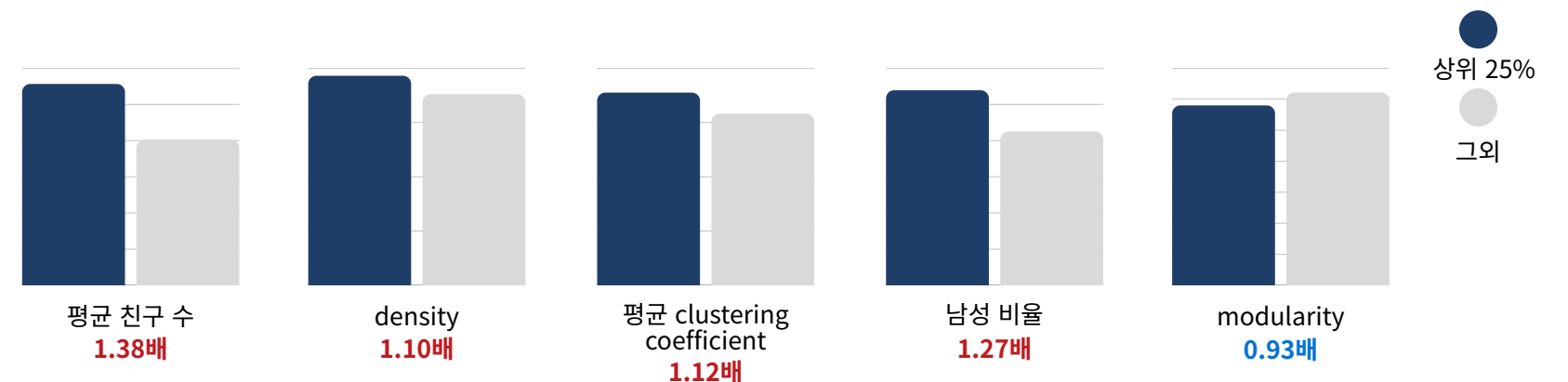


독립표본 T-검정 결과, 핵심 세그먼트는 친구 수·투표 관련 지표, 차단 및 신고 관련 지표(2배 이상), 알림ON(1.04배), 고등학교 비율(1.15배)에서 유의미하게 높고, 질문 스킵 비율은 낮았다.

이는 활동일수와 결제횟수가 모두 높은 유저가 더 많은 친구와 연결되어 있고 서비스 핵심 기능을 더 활발하고 몰입도 있게 이용하며 투표에서 더 많이 선택되고 이를 확인하며, 부정적 상호작용에서도 뚜렷한 행동 차이를 보임을 시사한다.

*핵심 세그먼트와 그 외 학교들 평균 활동량 변수 차이

*핵심 세그먼트 : 활동일수 및 결제 율상위 25% 학교



독립표본 T-검정 결과, 핵심 세그먼트는 평균 친구 수, density, 평균 clustering coefficient, 남성 비율, 차단·신고 관련 지표(약 2배), 학교 유저 수(1.08배)에서 유의미하게 높고, modularity는 낮았다.

이는 활동일수와 결제율이 모두 높은 학교의 유저들이 더 촘촘하게 연결되어있지만 전체적으로는 분절된 네트워크를 형성하며, 남성 비율이 높고, 부정적 상호작용에서도 뚜렷한 행동 차이를 보임을 시사한다.

종합 인사이드 & 전략 제안

“친구 수와 네트워크 구조가 활동성과 결제를 좌우한다”



유저 친구 수는 활동성과 결제의 핵심 변수

- 모든 분석 수준(개인 / 학교 단위, 상관 관계 / odds rate / 모델 주요 변수)에서 일관된 주요 변수로 나타남
- 친구 수가 많은 유저, 평균 친구 수가 많은 학교일수록 활동 일수와 결제율 모두 높게 나타남



결제와 활동은 서로 영향을 주는 ‘선순환’ 구조

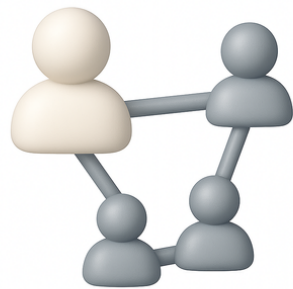
- 상관분석, 회귀, 모델링 모두에서 서로가 예측 변수로 등장함
- 활동 일수가 높으면 결제율 또한 높아짐
- 즉, 결제와 활동 일수는 상호 관계를 가짐



개인이 속한 네트워크는 개인의 활동성과 결제에 큰 영향을 줌

- [전염성] 투표/결제를 완료한 유저의 친구 78%가 5일 이내에 투표, 13%가 5일 이내에 결제
- [인플루언서] 인플루언서가 많은 학급일수록 앱 접속 일수, 투표량, 결제량이 유의미하게 증가
- 학급 id, 학교 id, 학교 유저 수 등 개인의 네트워크와 관련된 지표들이 모델링의 핵심 변수로 등장

“친구 수와 네트워크 구조가 활동성과 결제를 좌우한다”



네트워크 구조는 학교의 활동성과 결제에 중요한 영향

- 평균 친구 수가 많을 수록, 친구의 친구끼리도 많이 연결될 수록, 밀도가 높을 수록 활동 일수와 결제가 높아지는 경향 확인됨
- 네트워크 분절성 역시 활동성과 결제 각각을 예측하는 데는 긍정적이지만, 두 지표가 모두 높은 학교에서는 오히려 낮게 나타나 복합 조건에서는 낮은 분절성이 유리할 수 있음을 시사함
- 학교 유저 수는 학교 결제 예측에서는 양의 영향, 학교 활동일수 분석에서 음의 영향.
이는 학교의 규모보다 관계의 질이 핵심이라는 것을 시사



개인의 투표 참여도 또한 활동 지속성을 예측하는 중요한 변수

- 투표 열람 수, 투표 완료는 활동 일수의 예측력 높은 행동
- 시권설 마이닝에서도 투표 완료 > 상점 진입 > 결제로 이어지는 흐름이 뚜렷하게 나타남



출석 체크, 푸시 알림, 성별, 학교 종류도 결제에 유의한 영향

- 출석 체크는 등장 시 결제 전환율이 가장 높은 행동
- 푸시 알림 수신 비율이 높을수록 결제율 상승
- 고등학교, 남학생 비율이 높은 학교일 수록 결제율 높음

활동일수와 결제여부 예측 모델링

[모델링 주요 기여 변수]
친구수, 활동일수, 결제 경험,
학급/학교 ID(유저), 네트워크 구조(학교)

결제 직전의 행동패턴 분석

[결제 직전 주요 행동패턴]
전환율 ↑ 체류시간을 바탕으로
출석 체크, 투표 완료, 상점 진입이
결제 직전 주요 기능임을 파악

친구들의 활동을 유도하는 전염성 분석

[친구 전염성 주요 지표]
투표에 **78%**의 전염성 영향
결제에 **13%** 전염성 영향
→ 사회적 연결을 통한 확산 경로 존재

학급의 활동을 주도하는 인플루언서 분석

[인플루언서 주요 지표]
인플루언서 ↑ 하면
학급의 활동일수·투표량·결제량 ↑

활동일수, 결제 상위 25%
핵심 세그먼트 특징

[세그먼트 주요 지표]
개인 : 투표 활동을 활발히 주고 받음
학교 : 친구수가 많고 남성의 비율이 높음
네트워크의 끈끈함과 밀도가 높으며 분절도가 낮음

사회적 연결 강화

근거

친구 수는 활동성과 결제의 핵심 변수

친구의 친구끼리도 많이 연결된 네트워크일수록
접속일수와 결제율 높음

친구의 활동이 나의 활동을 자극하고,
나의 활동도 친구에게 전염



실행방안

친구 초대 이벤트

친구 3명을 초대 시 수박바 지급

공동 미션

친구 3명과 함께 일주일 출석 시 보상 지급

사회적 동기 부여

주간 활동 랭킹(Top5)시스템 추가
업적 기반 배지 제공

재방문 습관 형성

근거

자주 오는 유저가 더 오래 머물고 더 많이 전환함
→ 일단 꾸준히 앱에 오게 만드는 것이 핵심

체류시간·재방문 증대를 위해 앱 기능 다양화 필요
참여형 콘텐츠 확대 필요

친구의 활동이 나의 활동을 자극하고,
나의 활동도 친구에게 전염



실행방안

연속 출석/투표 보상제 운영

활동 일수 기반 레벨 및 보상 부여

프로필에 유저 level 추가

참여형 콘텐츠 확대

타임라인 상호작용(하트, 댓글) 추가
글로벌 투표 기능 추가

핵심 인플루언서 육성

근거

인플루언서(친구 수, 투표 중심성 높은 유저) 존재 여부에 따라
학급의 접속 일수 · 투표 · 결제에서
유의미한 차이 발생

인플루언서가 많을수록 접속일수 · 투표 · 결제 증가



실행방안

인플루언서 조기 포착

친구 수, 투표 상호작용 등 특정 패턴으로
인플루언서 예상 유저 식별

활동 강화 인센티브

활동이 많은 유저에게 쿠폰·포인트 제공

우선 노출

친구 추천 탭에 우선 노출

행동 패턴 기반 전환 최적화

근거

[출석체크], [투표 완료] 는 결제 직전에 자주 등장하며
결제 전환율이 가장 높은 이벤트

비결제자는 결제자 보다 2배 긴
상점 체류시간을 가지고
이탈하는 패턴이 도출 됨



실행방안

출석·투표 완료 보상

3일 연속 이벤트 완료 시 포인트 30% 할인 쿠폰 제공

상점 진입 유도 CTA

‘출석 보상 쿠폰 받기’ 배너 → 상점 이동

리마인드 푸시

비결제·긴 시간 상점 체류자 대상
‘망설이셨죠? 오늘만 할인 쿠폰!’ 발송

감사합니다

- 데이터분석 6기 1팀
- 백나린, 장위나, 최가림, 황보유리