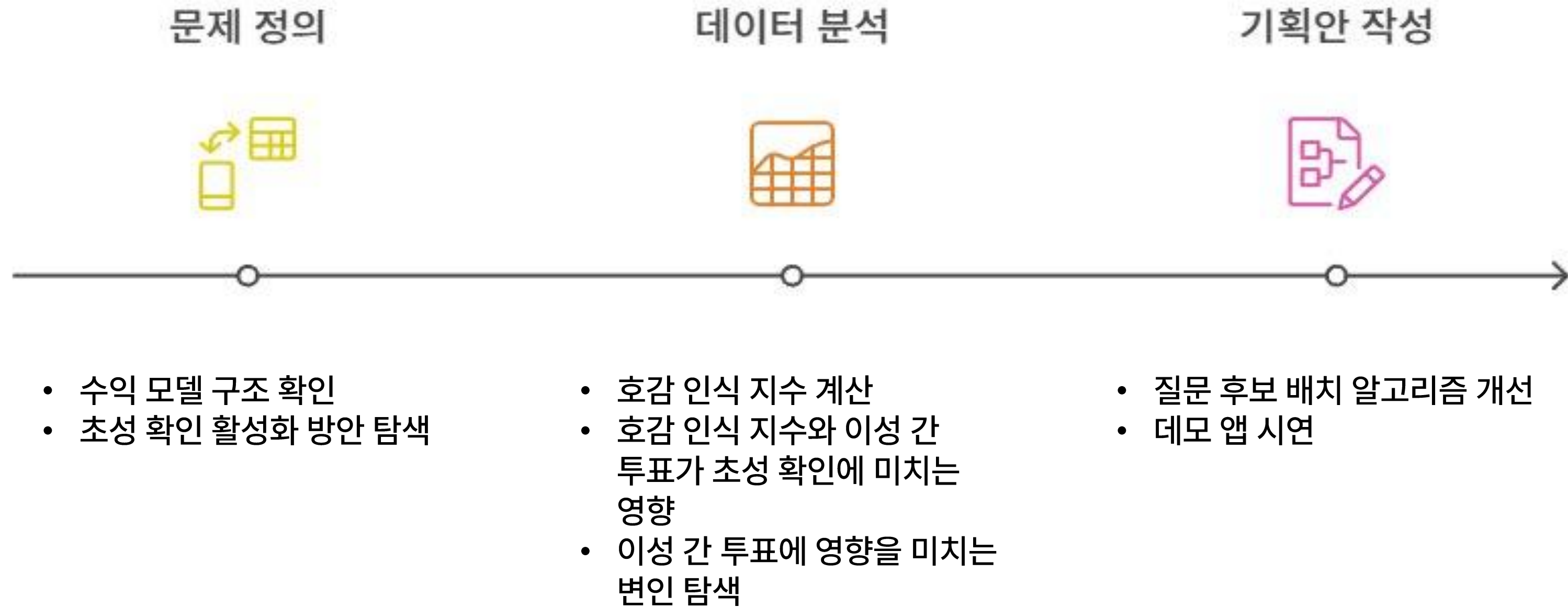


수익성 개선을 위한 질문 후보 배치 전략

3팀 | 김명균, 권기범, 남선우, 이유진, 임자은, 조성학

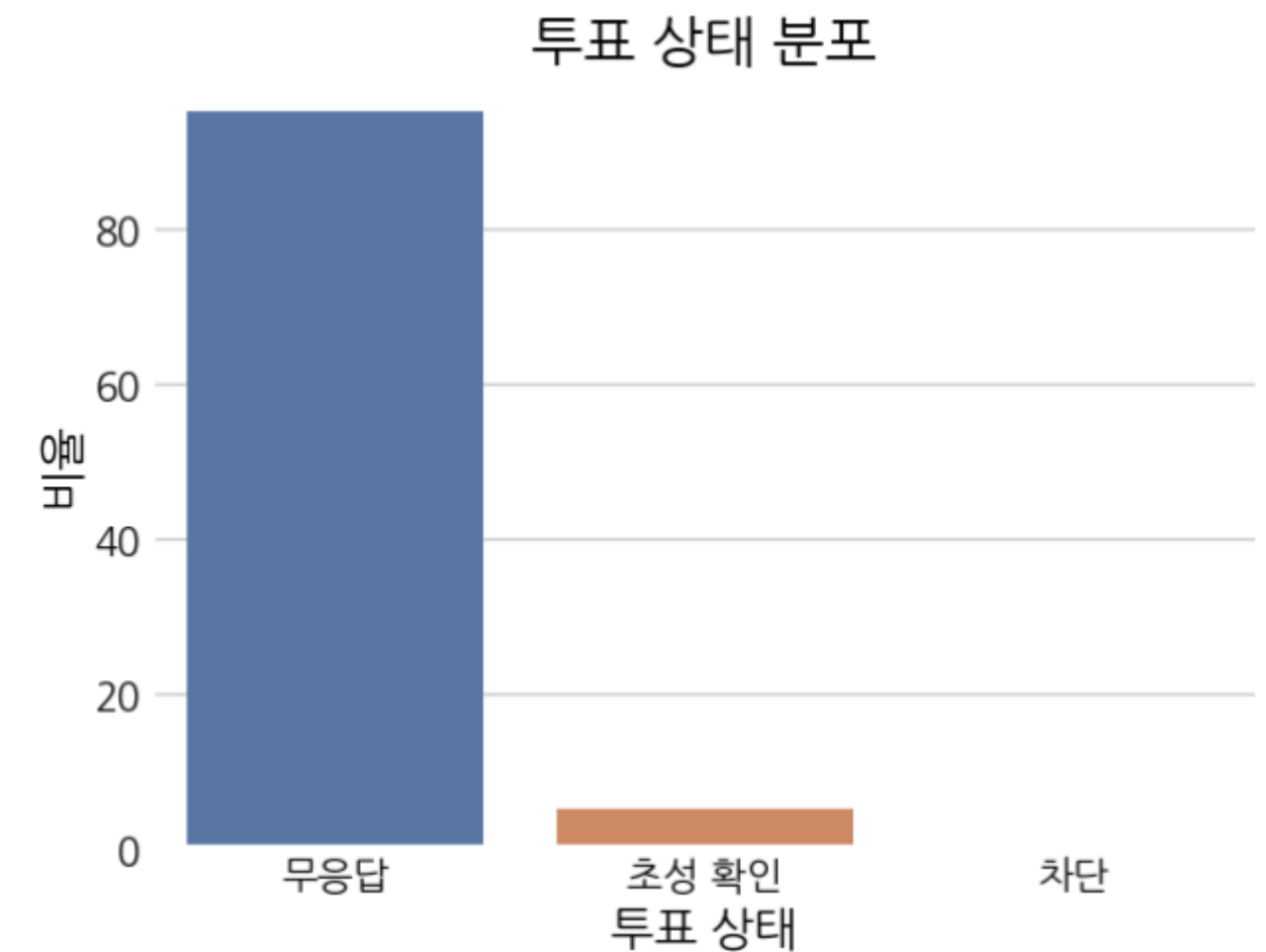
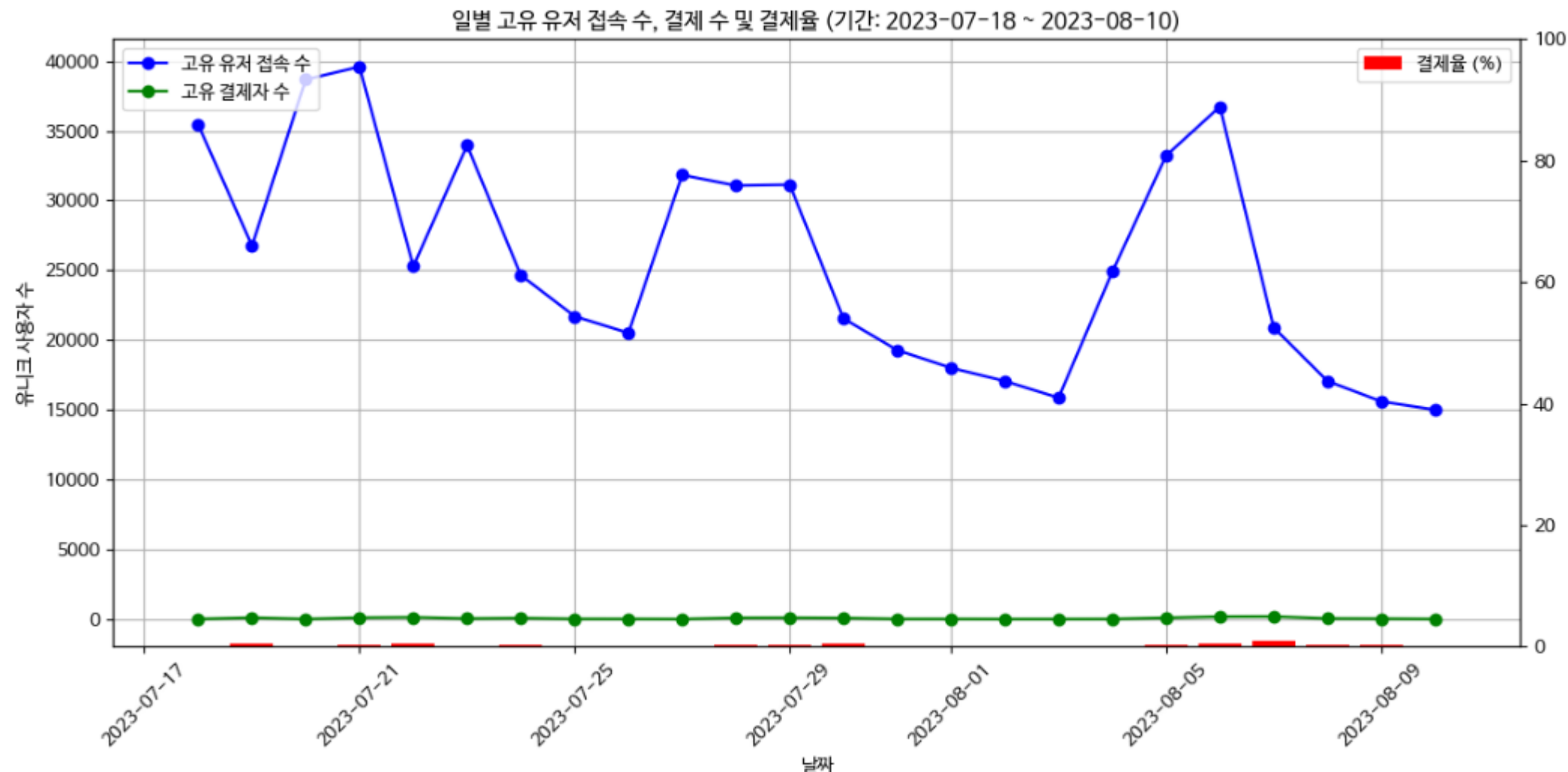
프로젝트 개요



배경 및 목적

목적: 수익 모델에서 가장 중요한 "초성 확인" 활성화 유도

- 일별로 접속하는 고유 유저 수에 비해 결제율이 1% 미만으로 매우 낮음
- 현재 H 서비스의 수익 모델에서 가장 수익성에 기여할 수 있는 기능은 "초성 확인"
 - 200 포인트 이상의 높은 포인트 소비는 모두 "초성 확인"에서 발생
- 전체 투표에서 초성 확인이 이루어진 비율은 5%에 불과



7개의 테이블을 분석에 사용

테이블 이름	내용
accounts_user	•유저의 기본 정보 및 서비스 이용 기록 저장 •'gender', 'friend_id_list' 컬럼을 이용해 이성 친구 비율 계산
accounts_userquestionrecord	•유저가 참여한 투표 내역 저장 •투표자/득표자 정보, 질문 정보, 응답 결과 포함 •'user_id', 'chosen_user_id', 'status', 'question_id' 컬럼 사용
accounts_pointhistory	•유저의 포인트 사용 및 획득 내역 저장 •답장 및 초성 확인에 사용되는 포인트를 확인
polls_usercandidate	•각 투표에 나온 후보자 정보 저장 •후보자의 성별을 유저 테이블로 식별한 후, 관련 정보를 'question_piece_id' 컬럼을 이용해 투표 기록 테이블과 병합
polls_question	•질문 고유 아이디와 내용 저장 •'id'를 기준으로 'question_text'를 투표 기록 테이블과 병합
hackle_events	•유저의 활동 정보 저장(23.07.18 ~ 23.08.10) •하나의 시간, 이벤트, 세션에 대해 동일 유저의 로그는 하나만 남게끔 처리
hackle_properties	•유저의 속성 정보 저장 •'user_id' 기준으로 유저 테이블과 병합해 유저 테이블에 'session_id' 정보 추가

초성 확인 기능의 특징

부끄럽지만, 내 마음을 표현하고 싶어!

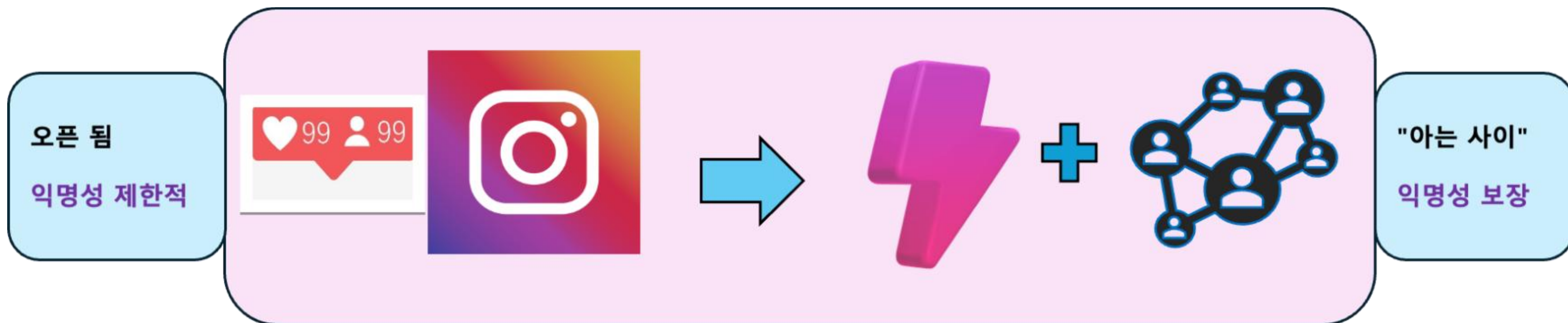
요즘 10대는 익명 플랫폼을 활용해 감정을 간접적으로 표현하고 반응을 살핍니다.



이성 간 투표에서 초성 확인 비율이 높은 이유

Z 세대 (요즘 10대)의 커뮤니케이션 특징

- Z 세대는 다른 연령층 보다 인터넷 기반의 익명 소통 및 커뮤니케이션에 익숙하다
- 익명 플랫폼을 통해 감정을 간접적으로 표현하고 상대방의 반응을 살피는 경향이 높다
- 직접적인 감정 표현을 부담스러워 하며, 익명성을 활용한 관심 표현과 피드백 확인이 중요해진 세대



초성 확인이 많이 발생하는 이유

자신을 향한 관심을 확인하려는 심리

- 이성에게 투표를 받으면 단순한 호기심이 아닌 호감 표현일 가능성을 고려
- 익명 SNS에서는 특정 숫자-알파벳 조합 (예: 'o99','o59')으로 암묵적인 호감 표현이 확산하고 있음

이성 간 투표에는 감정적 의미를 내포할 가능성이 높음

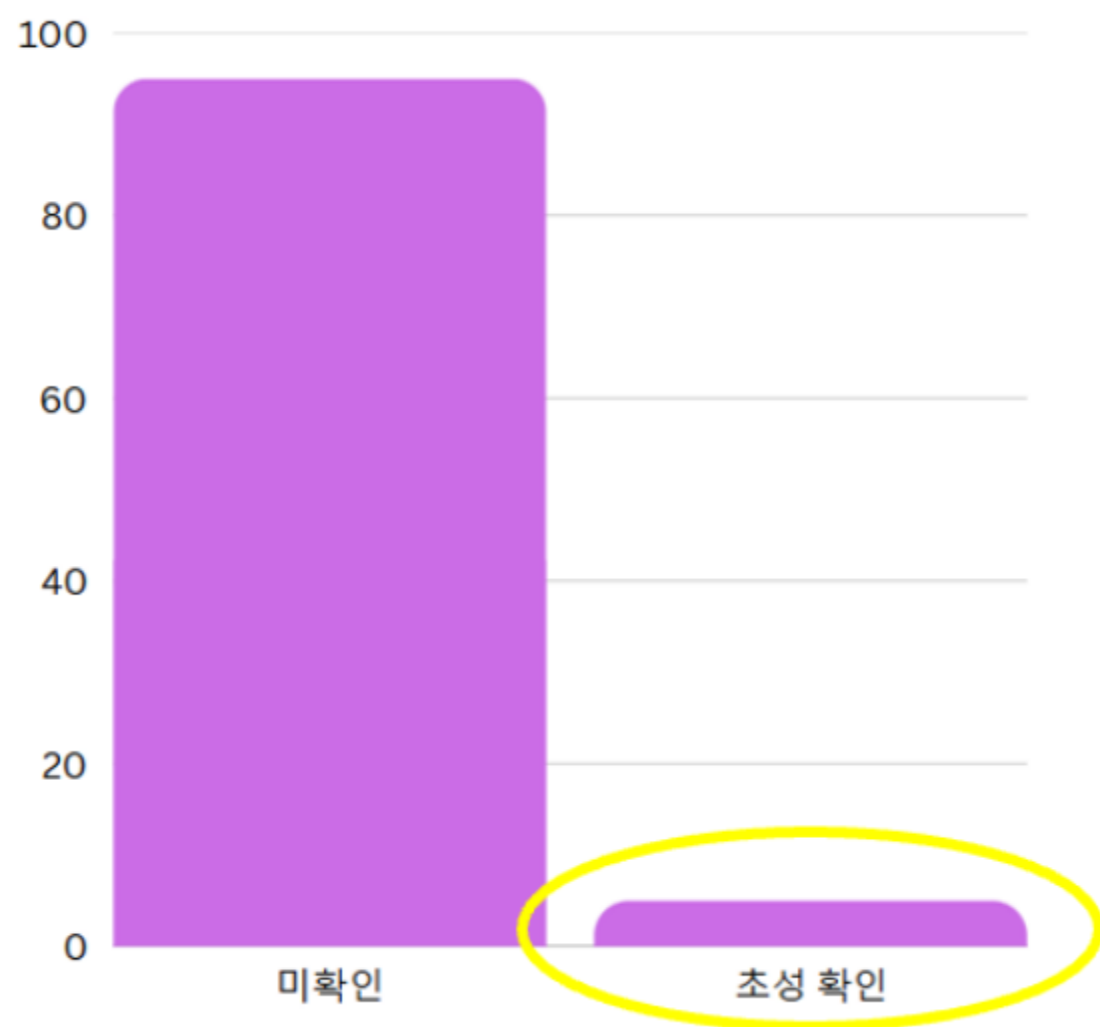
- 이성에게 받은 투표는 썸이나 관심의 신호로 해석될 가능성이 큼
- 특히 연애 관련 질문은 관계 형성의 시작점으로 작용

익명성이 만든 불확실성을 해소하려는 본능적 반응

- Z세대는 FOMO 심리 - 내 주변일은 놓치지 않고 알고 싶어하는 심리가 강함
- 투표를 보낸 상대방을 직접 확인할 수 없기에, 초성 확인 기능이 단서 제공 역할

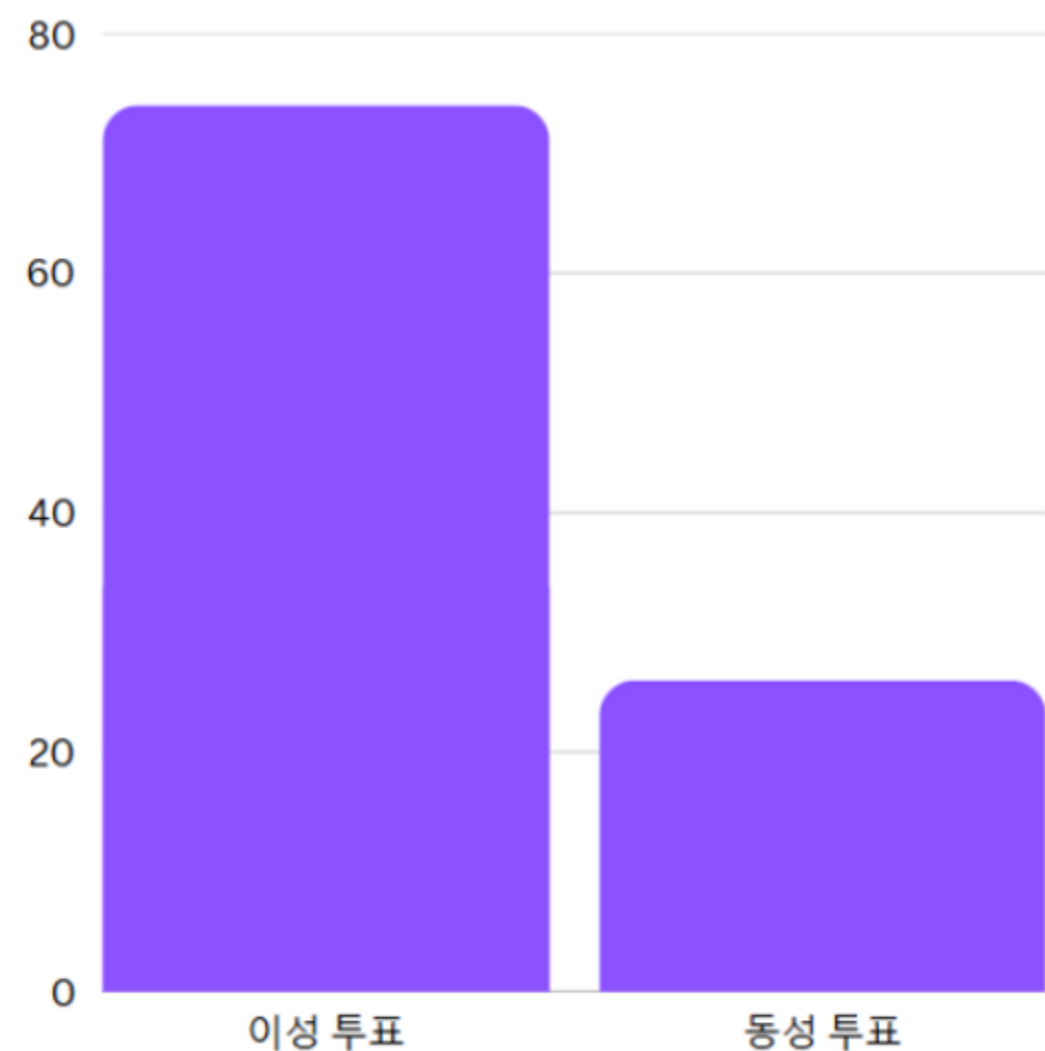
초성 확인 기능 사용 실태

전체 투표 중 초성 확인 비율



- 전체 투표 중 초성 확인이 발생한 비율: 5%

초성 확인 투표 중 이성 투표 비율



- 이성에게서 온 투표가 → 초성 확인으로 연결되는 비율이 압도적으로 높았음

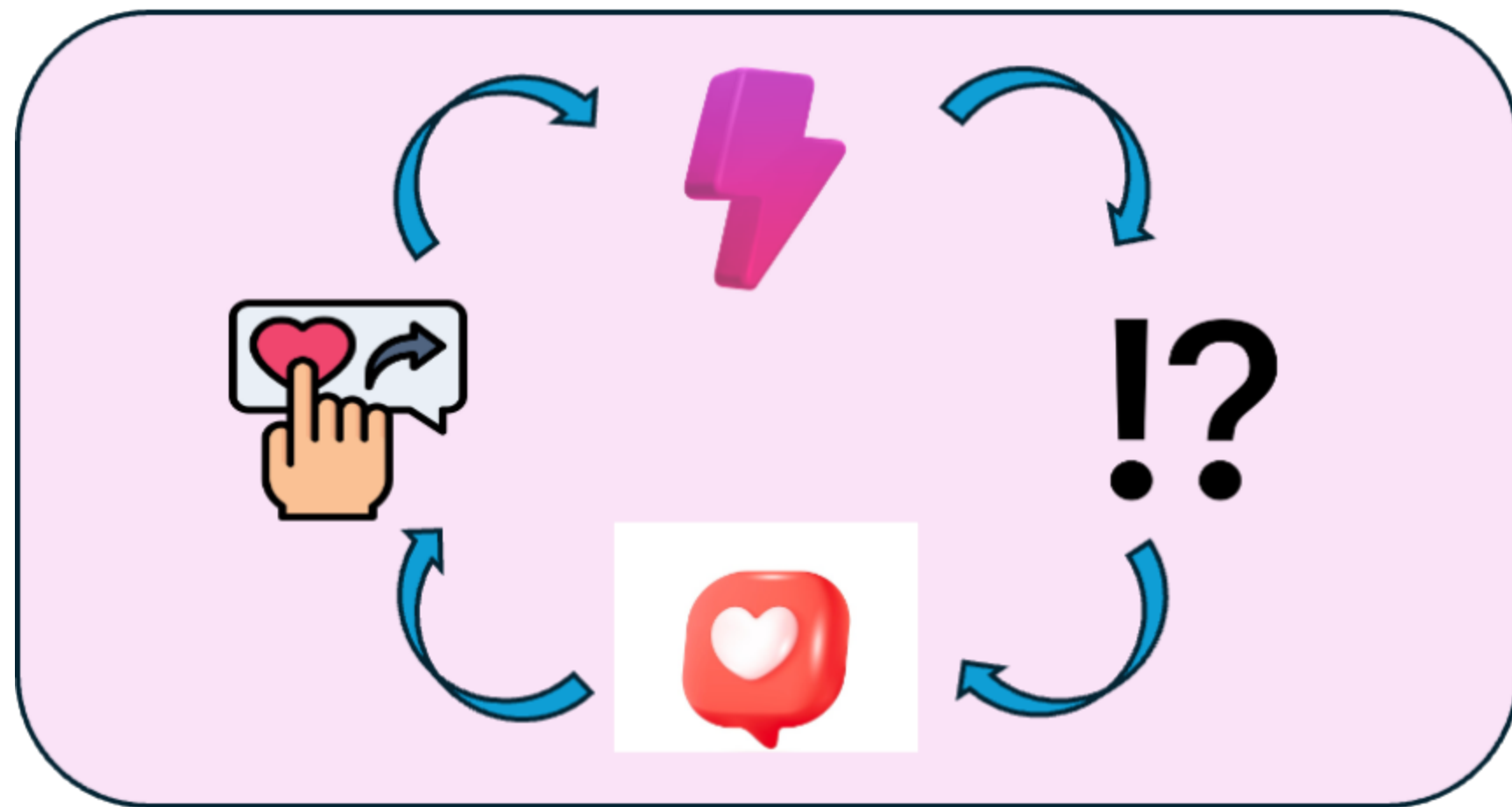
초성 확인 기능이 많이 사용된 질문

투표자가 간접적으로 호감을 표현할 수 있는 종류의 질문이 압도적으로 많음

질문 내용	언급 횟수
아직 잘 모르지만, 더 알아가 보고 싶은 사람은?	279
내 생일에 꼭 연락 왔으면 좋겠는 사람은?	231
사귀면 정말 잘해줄 것 같은 사람은?	228
여기서 가장 예쁜 사람은?	202
누가 봐도 좋아할 것 같은 호감인 사람은?	197

결론: 초성 확인 기능

- 초성 확인 기능은 단순한 호기심이 아닌 Z세대의 관계 탐색 도구
- 익명 환경에서 관계의 진정성을 탐색하려는 과정에서 필연적으로 발생하는 행동 패턴
- 이성 간 투표는 감정적 의미를 내포할 가능성이 크며, 초성 확인 기능을 통해 신호 해석이 이루어짐
- 결과적으로 초성 확인 기능은 앱의 참여율과 수익성을 견인하는 핵심 요소로 작용



이성에게 호감 질문을 받았을 때, 초성 확인 여부에 미치는 영향을 검증

- 호감 질문과 초성 확인의 관계
 - 귀무가설: 호감 질문을 받아도 초성 확인 확률은 변하지 않는다
 - 대립가설: 호감 질문을 받으면 초성 확인 확률이 달라진다
- 이성 간 투표와 초성 확인
 - 귀무가설: 이성에게 투표를 받아도 초성 확인 확률은 변하지 않는다
 - 대립가설: 이성에게 투표를 받으면 초성 확인 확률이 달라진다
- 호감 질문과 이성 간 투표의 상호작용 여부
 - 귀무가설: 호감 질문과 이성 투표간 상호작용이 없다
 - 대립가설: 호감 질문과 이성 투표 간 상호작용이 있을 것이다

질문 라벨링 기준 및 방법

질문을 이성 간 감정적 호감을 내포하는 지 여부에 따라 라벨링

- 호감 인식 지수

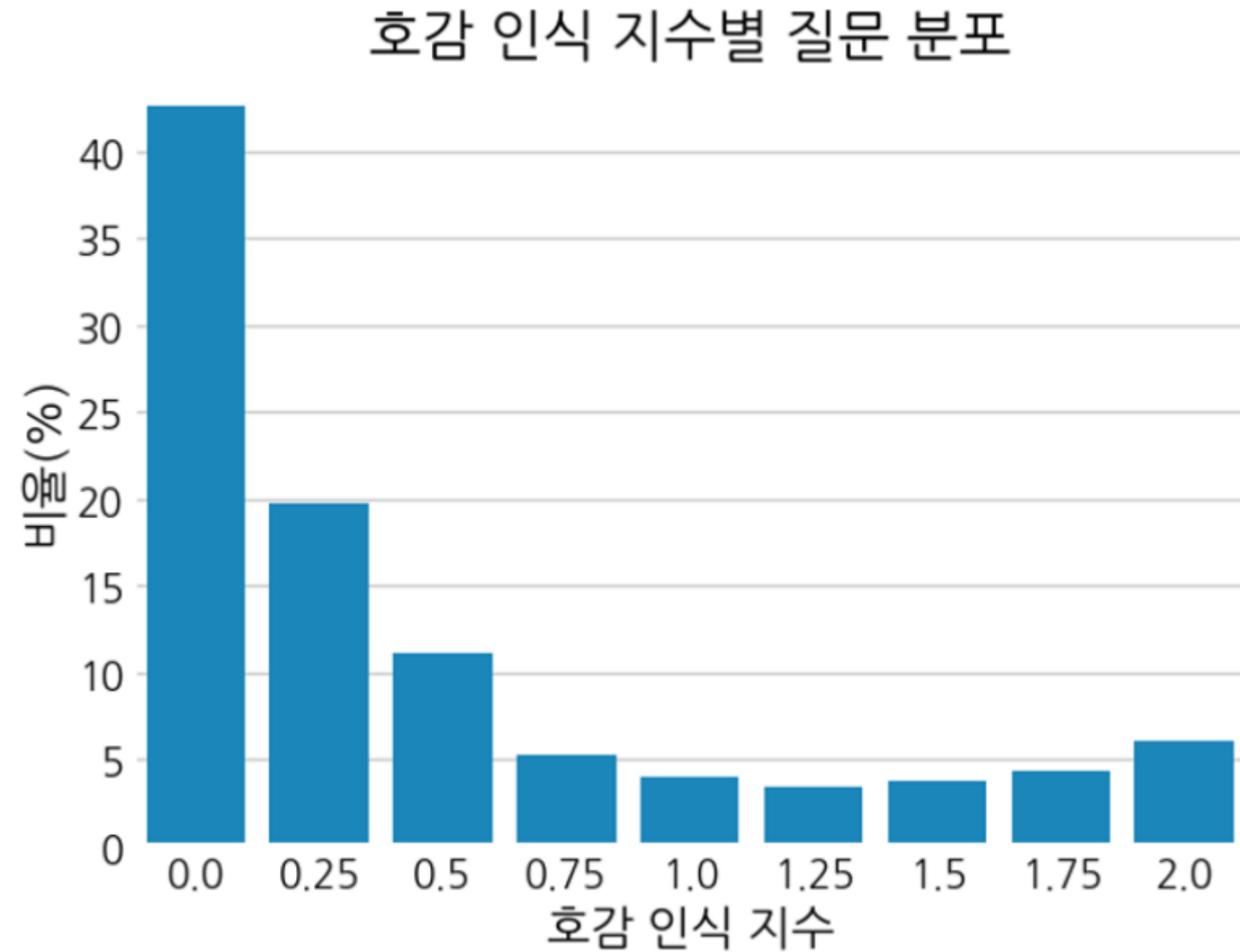
호감 인식 분류	기준	호감 인식 점수
No Romantic	이성에게 투표 받았을 때 투표대상자의 입장에서 이성적인 호감으로 보기는 어려움	0
Neutral	이성에게 투표 받았을 때 투표대상자의 입장에서 이성적인 호감으로 생각할 수도 있으나 확신을 주진 않음	1
Romantic	이성에게 투표 받았을 때 투표대상자의 입장에서 이성적인 호감으로 보임	2

최소한의 객관성을 위해 팀원의 과반수인 4명을 투입하여 각자 1차 라벨링 진행 후,
1차 라벨링 결과를 평균을 내어 최종 라벨링

라벨링 결과

호감 인식 지수별 질문 분포

- 호감을 전혀 인식하지 못하는 질문(0)
 - 42.5% (가장 높은 비율)
- 호감을 곧바로 인식할 수 있는 질문(2)
 - 6%
- 애매하게 호감을 표현한 질문(1.25)
 - 3% (가장 낮은 비율)



라벨별 특징

호감 인식 지수 특징 분석 과정

- 워드 클라우드를 활용하여 라벨별 특징 분석
- 라벨러 의견이 일치한 질문 중심으로 특징 도출
- 공통 단어 제거
 - 모든 라벨(0, 1, 2)에서 TOP20 빈도 단어 불용어 처리

불용어로 처리한 단어	0점	1점	2점
친구는?	TOP 6	TOP 5	TOP 5
친구	TOP 10	TOP 8	TOP 10
같은	TOP 1	TOP 3	TOP 14
수	TOP 19	TOP 16	TOP 11
것	TOP 2	TOP 4	TOP 4
사람	TOP 4	TOP 9	TOP 4
사람은?	TOP 3	TOP 1	TOP 2
가장	TOP 4	TOP 2	TOP 7

호감 인식 지수 0 (No Romantic) 특징



키워드	순위
잘	1
제일	2
할	3
많이	4
있을	5
때	6
좋아할	7
먹을	8
많을	9
나중에	10

- **예측 키워드가 많이 나타남**
 - 예: 24시간 폰만 할 것 같은 사람은?
- **장소 키워드 많이 안보임**
- **정도를 나타내는 키워드가 많이 나타남**
 - 예: 갤러리에 자기 사진이 가장 많을 것 같은 친구는?
- **특성에 관련된 키워드가 나타남**
 - 예: 무채색이 가장 잘 어울리는 사람은?

호감 인식 지수 2 (Romantic) 특징



키워드	순위
싶은	1
같이	2
가고	3
하고	4
있는	5
하는	6
보고	7
지금	8
내	9
있다	10

- 장소 키워드가 많이 나타난다
 - 예: 같이 롯데월드 가고 싶은 친구!
- 데이트 관련 키워드가 많이 나타난다
 - 예: 봄날 같이 피크닉 떠나고 싶은 사람은?
- 투표대상자와 투표자만의 시간을 갖고 싶어하는 뜻을 내포하는 키워드가 나타난다.
- 투표대상자에게 욕망을 표현하는 키워드가 나타난다
- 직접적으로 호감을 표하는 단어(예: 사랑해) 보다는 애둘러서 감정을 표현하는 단어가 많이 나타났다.

호감 인식 지수 1 (Neutral) 특징

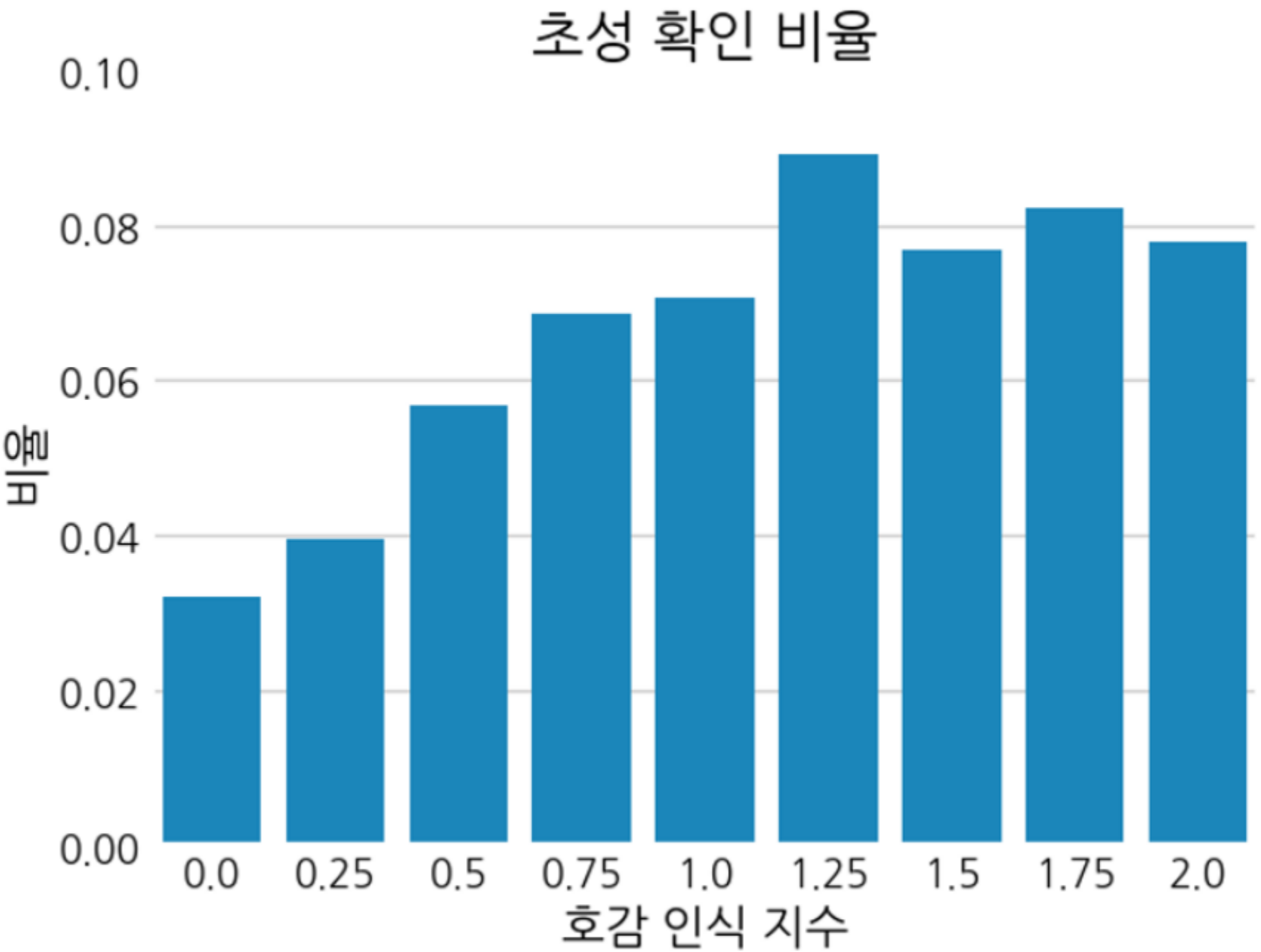


키워드	순위
싶은	1
같이	2
제일	3
잘	4
있는	5
친해지고	6
때	7
먼저	8
내가	9
좋은	10

- 호감 인식 지수 2에서 나타난 키워드와 호감 인식 지수 0에서 나타난 키워드가 혼합
 - 예: 제일 나랑 잘 맞을 것 같은 사람, 같이 대화하면 재밌는 친구는?
- 키워드 "친해지고" 등장
 - 예: 아직 어색하지만 친해지고 싶은 친구
- 외모에 대한 키워드가 다수 등장
 - 예: 꽃이 가장 잘 어울리는 사람은?
- 상상과 관련된 키워드가 등장
 - 예: 내가 전화하면 바로 달려올 것 같은 친구?

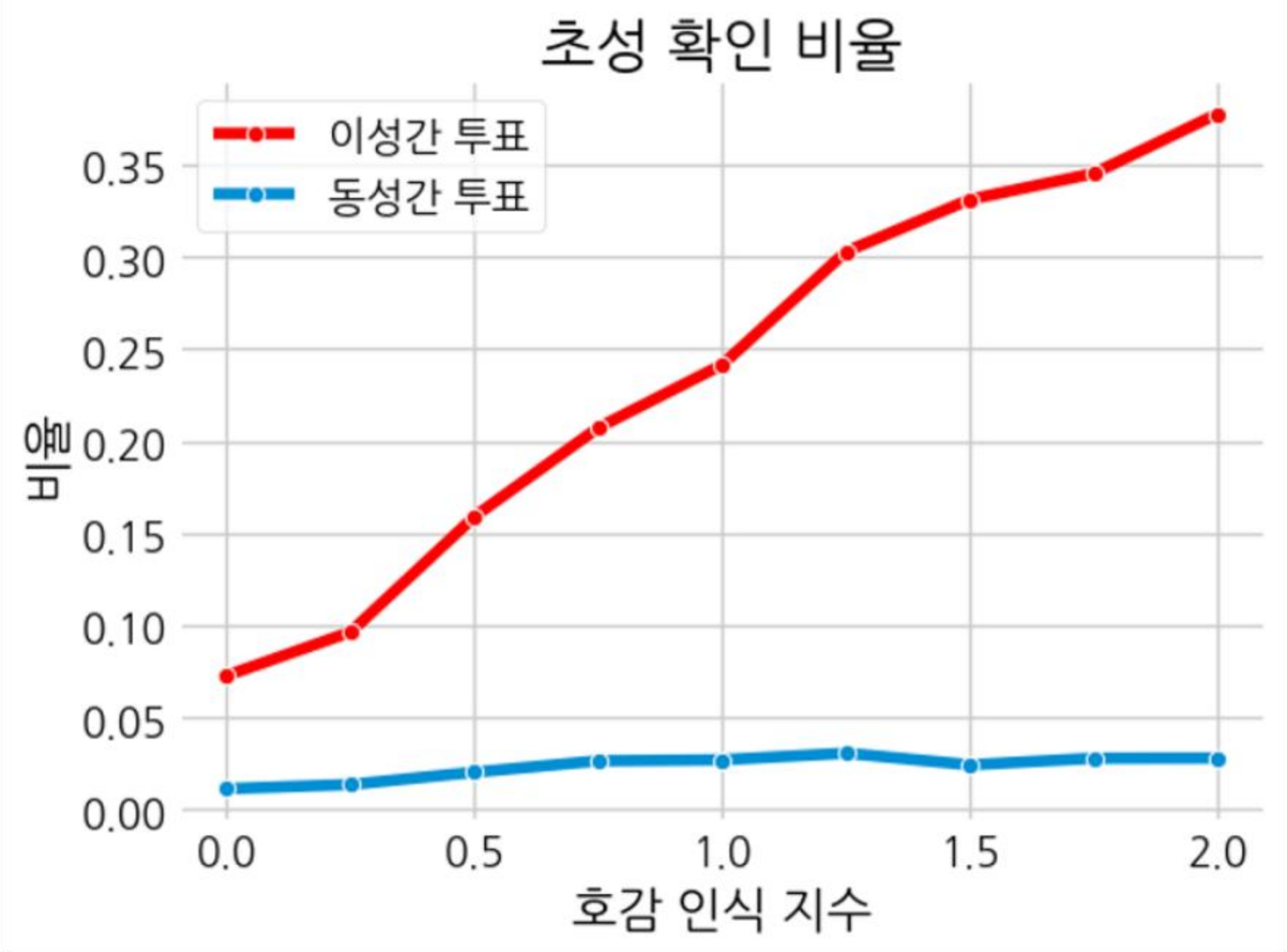
호감 인식 지수와 이성 간 투표가 초성 확인에 미치는 영향

호감 인식 지수별 초성 확인 비율



호감 인식 지수와 이성 간 투표가 초성 확인에 미치는 영향

호감 인식 지수에 따른 이성, 동성 초성 확인 비율 비교



호감 인식 지수와 이성 간 투표가 초성 확인에 미치는 영향

로지스틱 회귀분석

“호감 인식 지수와 이성 투표 여부가 초성 확인에 어떤 영향을 미치는가?”

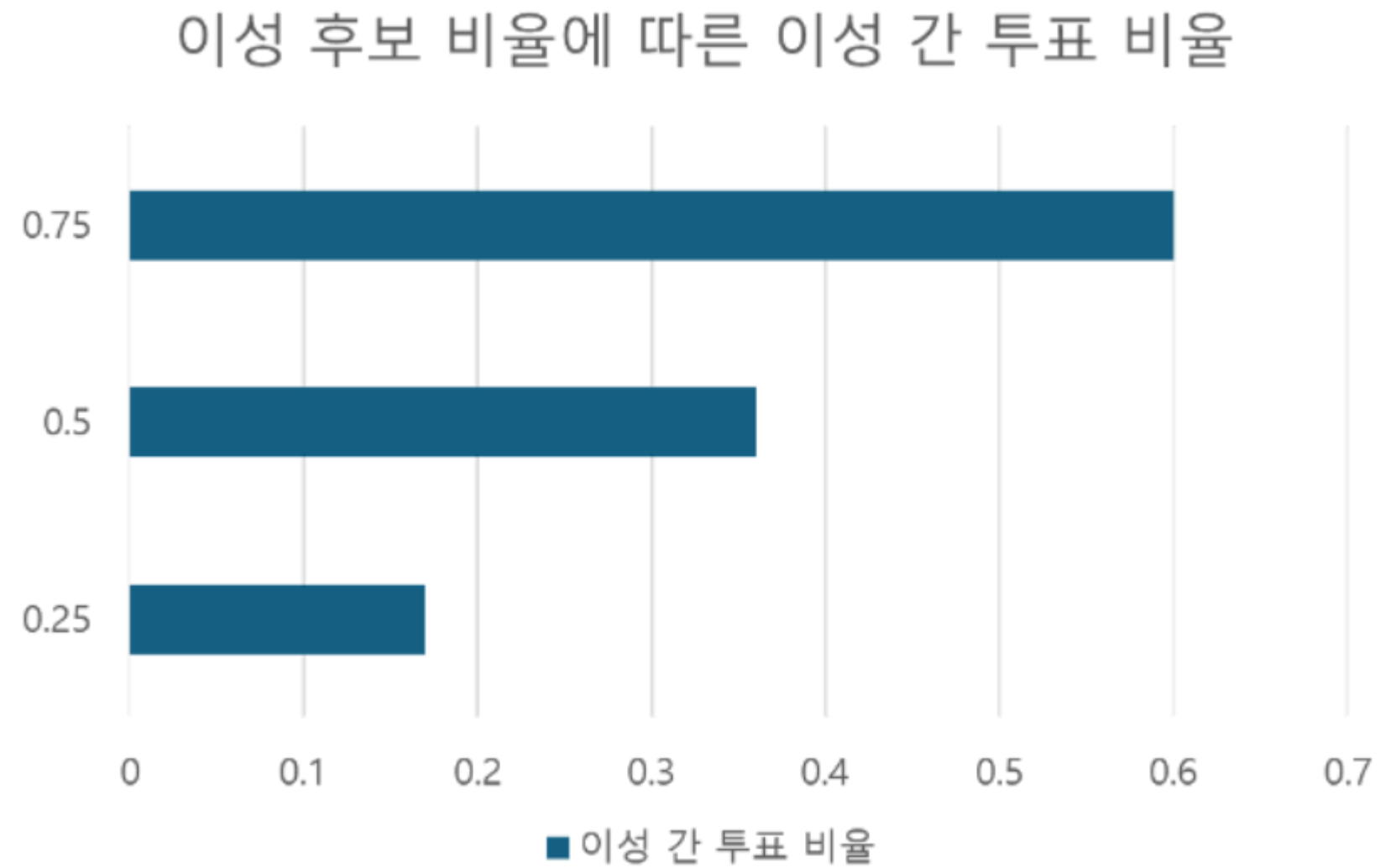
$$odds\ ratio = e^{\beta_1} = \exp(\beta_1)$$

독립 변수	계수	Z-statistic	P-value	오즈비
호감 인식 지수	0.48	43.64	$P < .01$	1.61
이성 간 투표 여부	1.88	135.17	$P < .01$	6.54
호감 인식 지수 * 이성 간 투표 여부	0.63	46.70	$P < .01$	1.88

이성 간 투표를 이끌어내는 변인

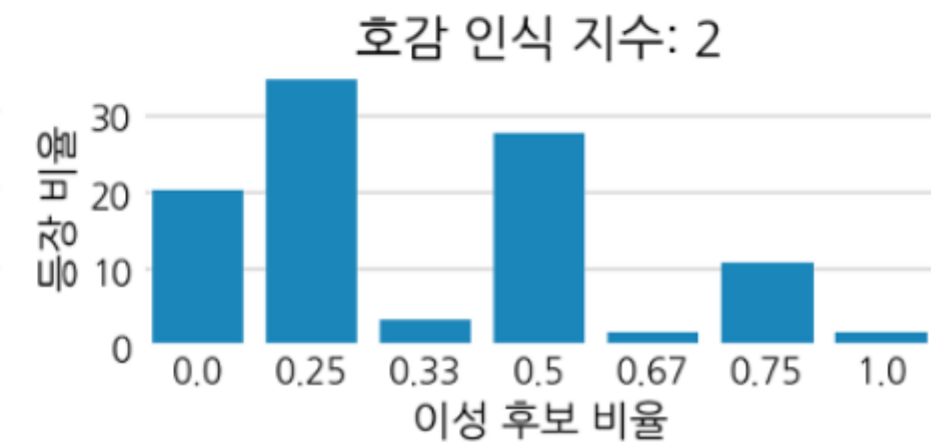
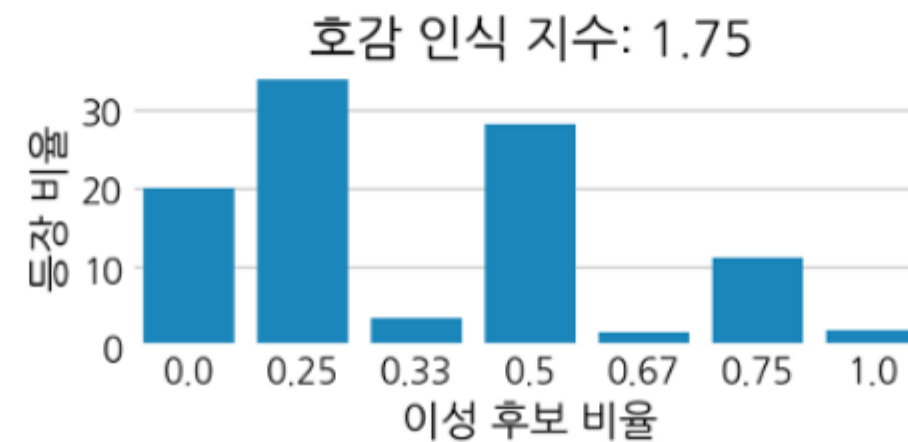
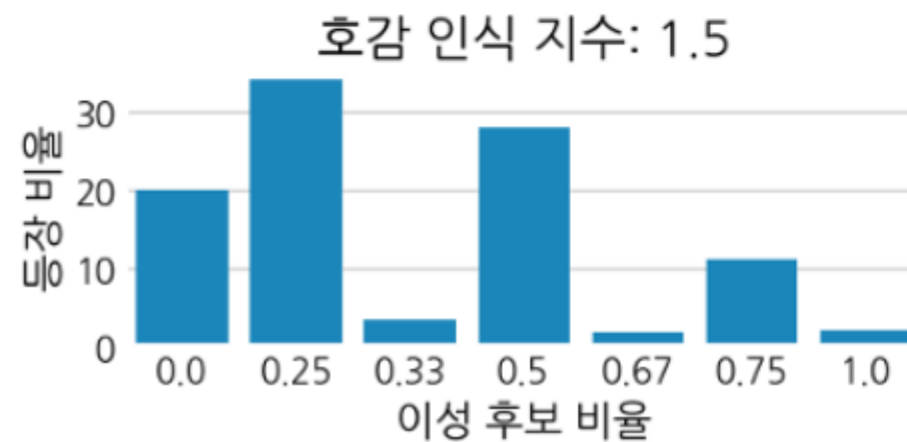
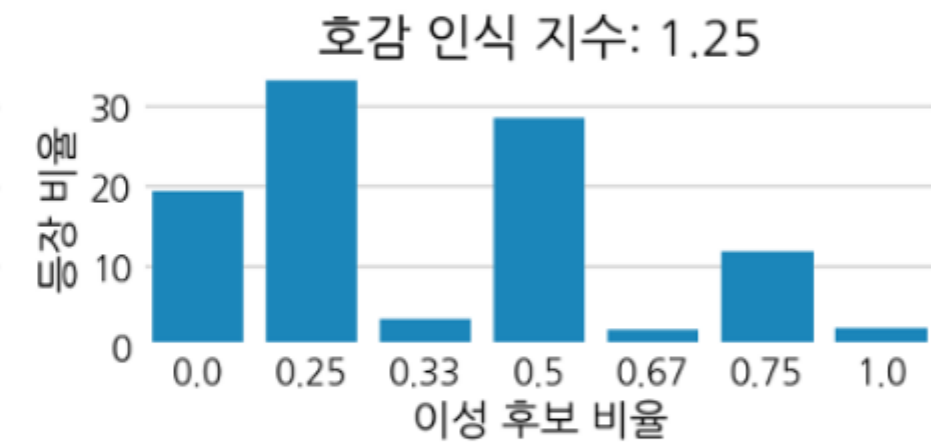
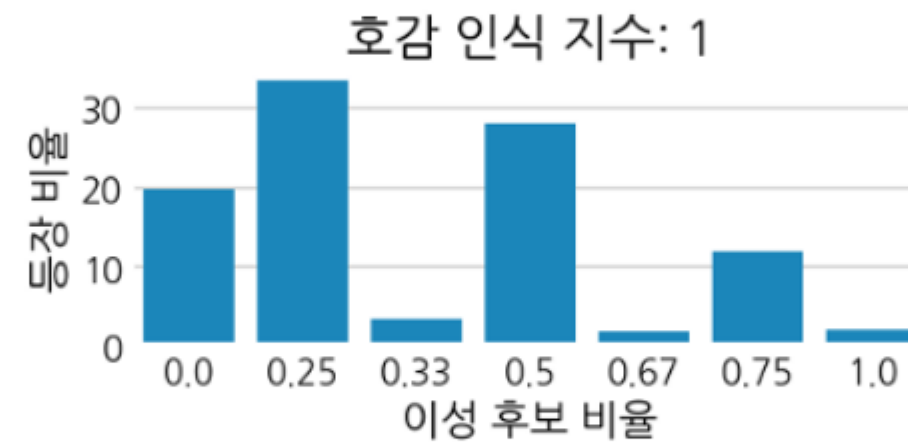
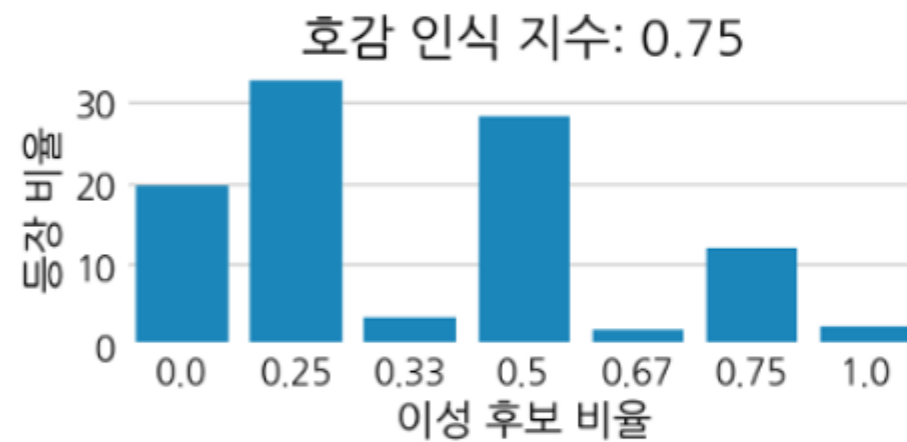
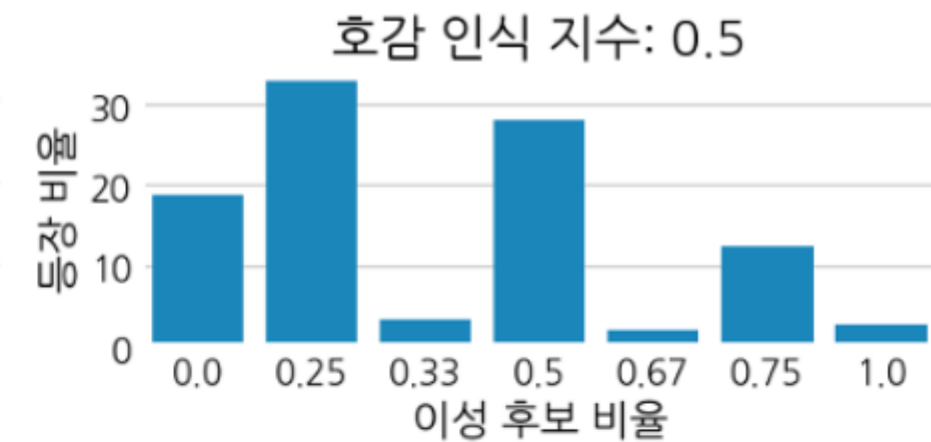
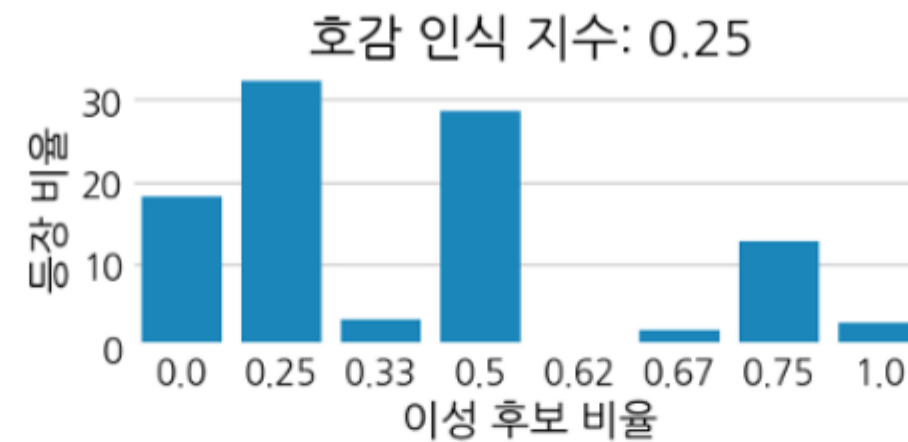
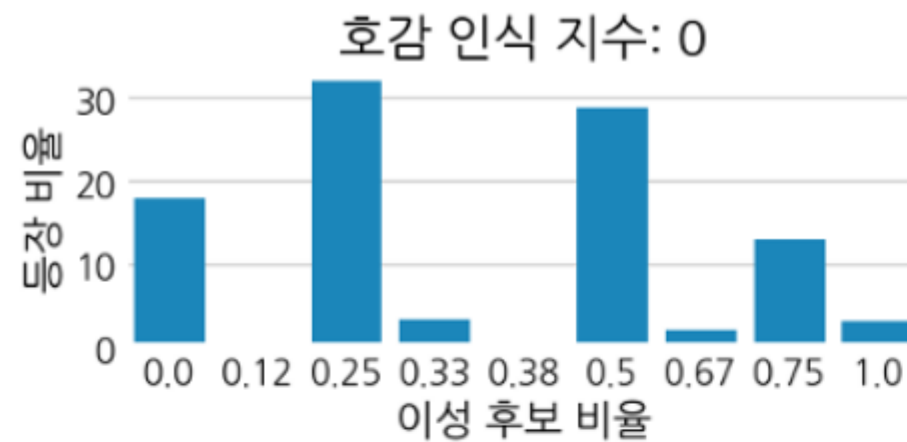
이성 후보 수가 많이 배치될 때 이성 간 투표가 많이 발생

유형	이성 후보 수(비율)
동성 간 투표	1명(29%)
이성 간 투표	2명(56%)



호감 인식 지수와 이성 후보 비율 분포

모든 질문에 동일한 이성 후보 비율 분포



이성 후보 배치 알고리즘 개선안 기획서

배경

- 수익 모델에서 가장 중요하지만 발생 빈도가 낮은 '초성 확인'
- 수익 개선을 위해서는 유저들이 초성 확인 사용 횟수를 늘려야 함

현황

- 호감을 느낄 수 있는 질문에서 이성을 투표했을 때 초성 확인 비율이 가장 높음(37%)
- 현재 H 서비스에서는 질문의 특성을 고려하지 않은 채 랜덤하게 후보를 배치

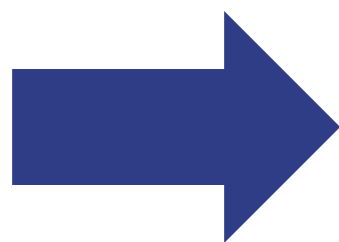
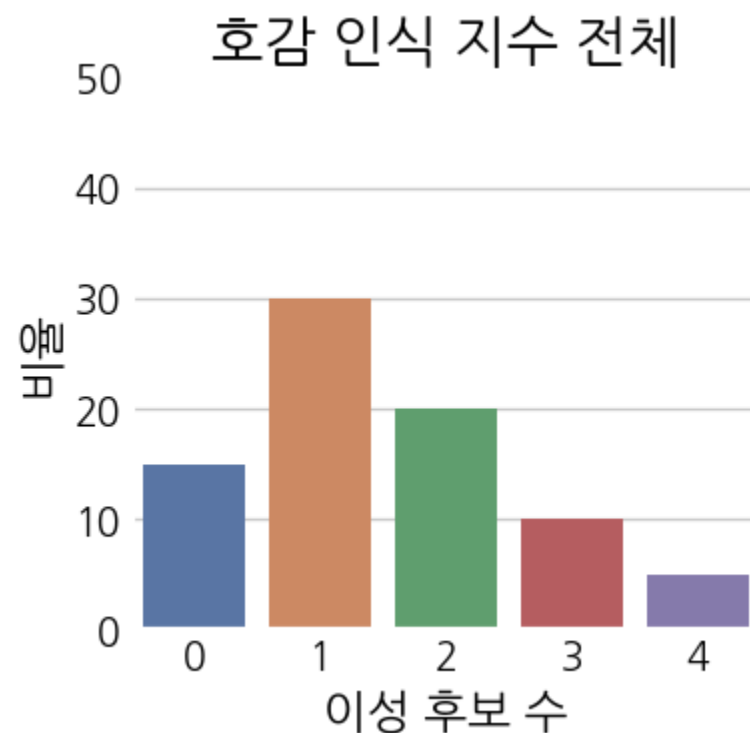
이성 후보 배치 알고리즘 개선안 기획서

컨셉

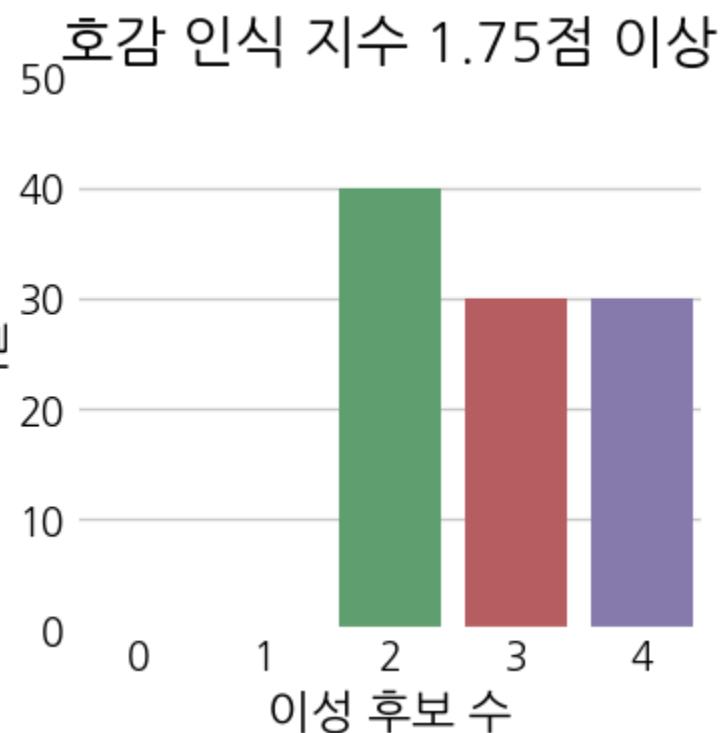
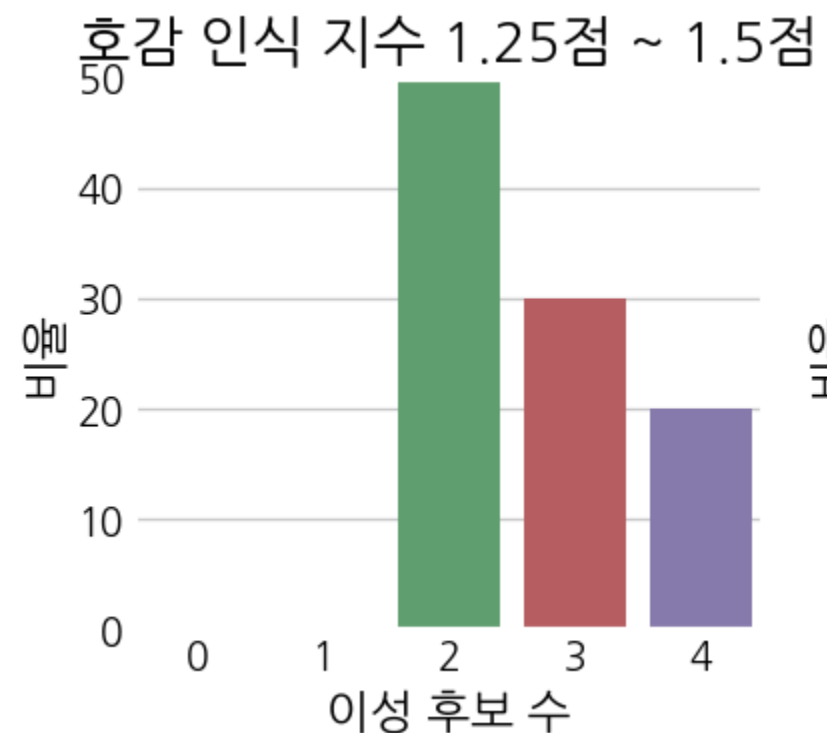
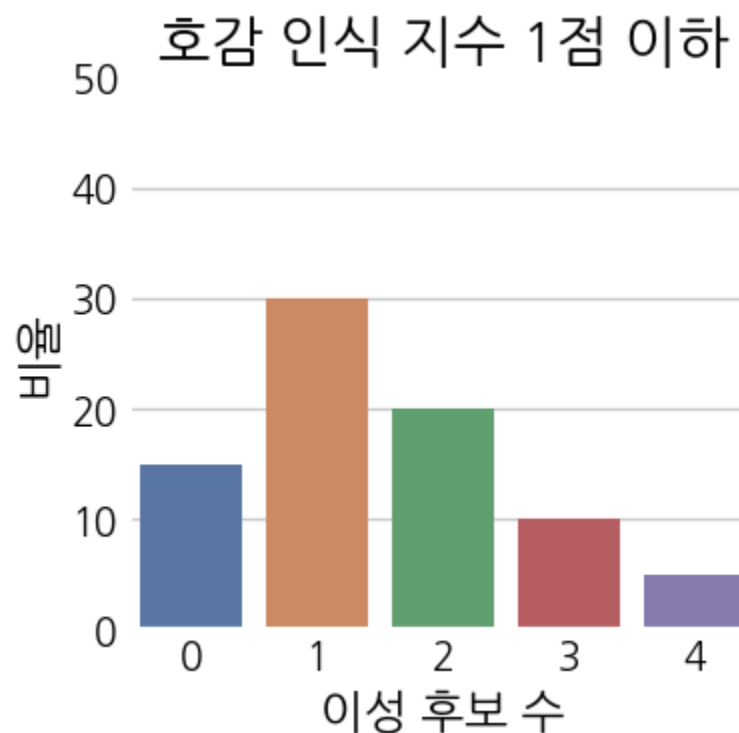
"설레는 질문, 설레는 후보"

- 호감 인식 지수에 따른 이성 후보 수 배치 알고리즘 개발

BEFORE



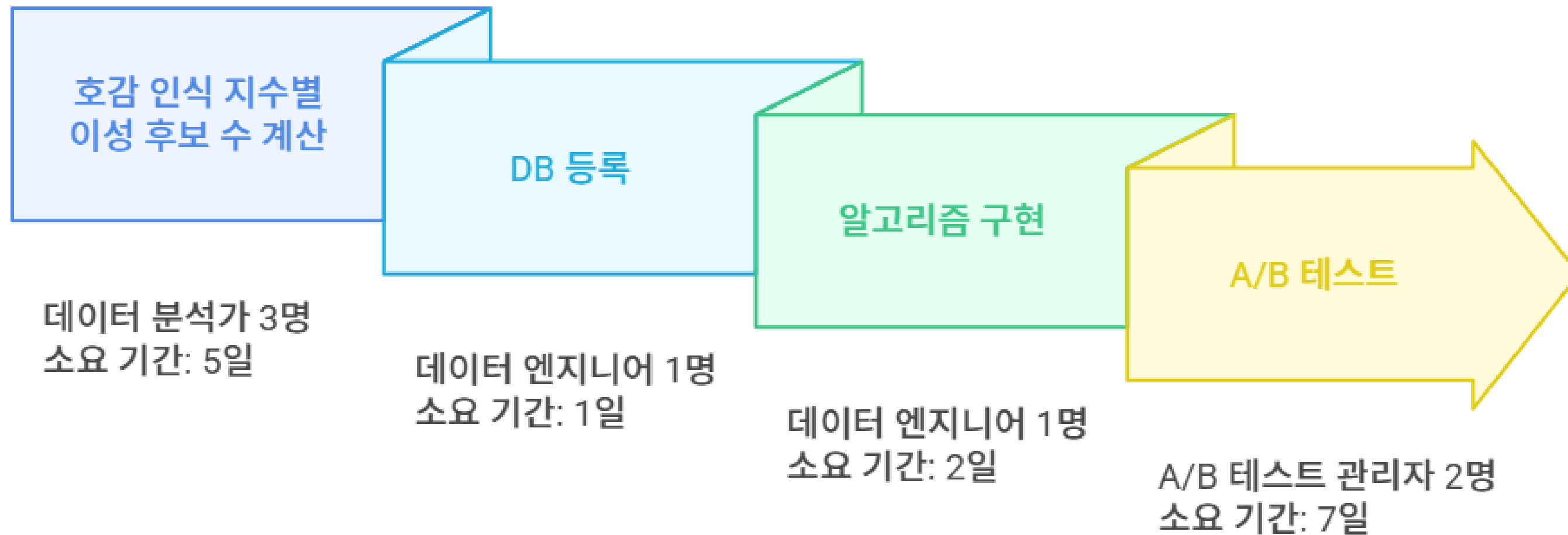
AFTER



이성 후보 배치 알고리즘 개선안 기획서

단계별 실행 과정

- 총 필요 인원: 6명 (데이터 분석가 3명, 데이터 엔지니어 1명, A/B 테스트 관리자 2명)
- 총 소요 기간: 15일



이성 후보 배치 알고리즘 개선안 기획서

호감 인식 지수별 적정 이성 후보 수 계산

- 초성 확인 비율이 가장 급격하게 증가하는 구간: 호감 인식 지수 1.25점
- 이성을 투표했을 때 평균 이성 후보 수: 2명
 - 호감 인식 지수가 1.25점 이상일 때 이성 후보 수를 최소 2명 이상으로 설정
- 고려사항
 - 유저의 이성 친구 수를 고려한 이성 후보 배치 필요(최소 이성 친구 수: 4명)
 - 이성 후보 수와 관련이 있는 변인 탐색 필요(예. 유저의 이성 친구 비율)
 - 유저들이 질문에 투표하면서 후보 배치에 작위적인 느낌이 들지 않도록 설계

이성 후보 배치 알고리즘 개선안 기획서

DB 등록 & 알고리즘 구현

- DB 등록: polls_question 테이블에 '호감 인식 지수', '적정 이성 후보 수' 컬럼 추가

polls_question	
id	integer
question_text	integer
created_at	timestamp



polls_question	
id	integer
question_text	integer
created_at	timestamp
attraction_index	float
opposite_gender_candidate_num	integer

- 알고리즘 구현: 기존 임의 배치 → 호감 인식 지수에 따른 적정 이성 후보 수 계산 후 임의 배치
 - 데모 앱에서 구체적으로 설명

이성 후보 배치 알고리즘 개선안 기획서

A/B 테스트

- 연구 가설
 - 호감 인식 지수 1.25점 이상의 질문에 이성 후보 수를 최소 2명 이상 나오게 설정하면 그렇지 않을 때보다 초성 확인이 더 발생할 것이다.

- 변인 설정

독립변인	통제변인	성공 지표	보조 지표	가드레일 지표
1.25점 이상 질문에 최소 이성 후보 수 적용 여부	유저 성별, 이성 친구 비율	초성 확인 횟수 10% 이상 증가	이성 간 투표 횟수	유저 이탈률

- 조건 별 투표 횟수: 1,000회
- 연구 기간: 14일

이성 후보 배치 알고리즘 개선안 기획서

리스크 요인

- 호감 인식 지수가 높은 질문에서 동성 선택을 제한할 가능성
 - 호감 인식 지수가 높은 질문에서 동성을 선택하는 경우
 - 남학생: 마음에 드는 이성 없어서 장난으로 동성을 선택
 - 여학생: 남학생보다 호감 표현이 관대함 (예. "같이 데이트 가자", "같이 손잡자" 등)
- 여학생의 경우, 호감 인식 지수가 높은 질문에서 이성 후보 비율을 높게 설정하면 투표자가 원하는 대상을 선택할 수 없을 수 있음

이성 후보 배치 알고리즘

STEP1

사용자 성별 확인

- 사용자의 성별을 DB에서 조회하여 이성 친구와 동성 친구 분류

STEP2

친구 데이터 로딩

- 친구 데이터를 성별 필드를 기준으로 이성 친구와 동성 친구로 분리

STEP3

점수에 따른 이성 후보 배치

- (A) Score ≥ 1.25
 - 이성 친구를 최소 2명 이상 배치
 - 이성 친구가 2명 이하이면 가능한 모든 이성 친구를 배치한 후, 남은 자리를 동성 친구로 채움
 - 이성 친구가 4명 이상이라면 2~4명을 랜덤으로 선정하여 배치한 후, 남은 자리를 동성 친구로 채움
- (B) Score < 1.25
 - 동성/이성 구분 없이 랜덤으로 4명의 친구 배치

데모 앱 소개

기능

- 로그인 기능
- 친구 목록 조회 기능
- 질문 투표 목록 확인 기능

설정

- 유저: 아이돌/배우, 코드잇 DA 2기 교육생, 테스트 계정 등 약 80여명의 가상 유저 생성
- 가정: 알고리즘 시연을 위해 동성 36명, 이성 4명의 친구가 존재하는 극단적인 상황 연출
- 남,녀 ID를 통한 이성 배치 알고리즘 시연 진행 (남성 유저 ID: 0, 여성 유저 ID: 100)

데모 앱 아키텍처

Frontend



- 로그인 화면 UI
- 친구목록 UI
- 질문 투표 UI

Backend

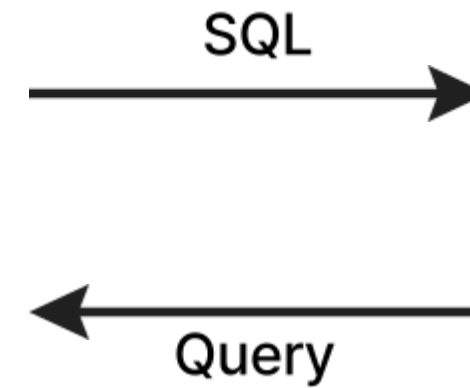
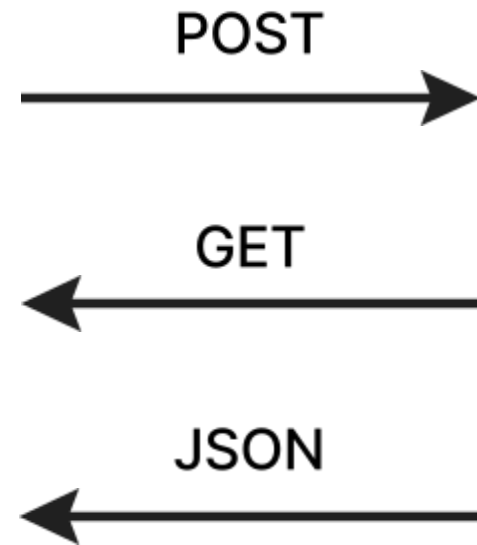


- POST /login (로그인 처리)
- GET /friends (친구 목록)
- GET /vote/questions (투표)
- BigQuery (데이터 조회)

Data Storage



- user_table (사용자 정보)
- friends_table (친구 관계)
- sample_question_table (질문)
- 프로필 사진 (클라우드 스토리지)



목표 및 기대 효과

목표

- 초성 확인 10% 증가
- 기존 전체 초성 확인 횟수: 66,633건
- 예상 증가량: 6,663건

기대 효과

- 초성 확인 증가를 통한 수익 증가
- 1,000 포인트 당 2,900원 = 1포인트 당 2.9원
- 예상 수익 = 초성 확인 증가량 * (300 ~ 1,000) * 2.9
 - 예: $6,663 * (300 \sim 1,000) * 2.9 = 5,796,810\text{원} \sim 19,322,700\text{원}$ 의 수익 발생

감사합니다.

Q&A

- 초성 확인 10% 증가
- 기존 전체 초성 확인 횟수: 66,633건
- 기존 호감 인식 지수 1.25점 이상 질문에서 이성 후보자가 0~1명일 때: 5,788건
- 기존 호감 인식 지수 1.25점 이상 질문에서 이성 후보자가 2명 이상일때: 12,136건
- 예상 증가량: $12,136 - 5,788 = 6,348$ 건 → 약 10% 증가

