

Thinkgood

씽굿 AI 서포터즈

# 씽굿 추천 시스템 제안 및 설계

사용자 행동 데이터 기반 개인화 추천

---

| TEAM 씽굿 AI 서포터즈

김승현 · 박현서 · 신경은 · 한재희

# — 목차

01

## 문제 정의

제한된 데이터 환경에서도 개인화가 가능했는가

데이터 구조 분석 · 로그 부재 파악

한재희

02

## 프로젝트 요약

현재 데이터로 구현한 개인화 AI 추천 시스템

전처리 · 모델링 · 세그먼트 전략

김승현

03

## DB 설계 방향

추천 성능을 확장하기 위한 데이터 구조 제안

멤버 · 공모전 · 접수 테이블 개선안

박현서

04

## 단계별 적용 전략

이미 구현된 AI를 서비스로 확장하는 방법

데이터 입력 구조 · A/B 테스트 · 확장성

신경은

SECTION 01

# 문제 정의

# 공모전 추천, 정말 개인화일까요?

대부분의 경쟁사들은  
"추천 공모전"이라는 메뉴를 운영하고 있지만,  
실상을 들여다보면 진정한 의미의 개인화와는 거리가 멍니다.

Ad

## 광고비 기반 상단 노출

사용자의 취향보다 광고비를 지불한 공모전이 우선적으로 노출되는 구조입니다.



## 운영자 픽(Pick) & 단순 정렬

운영자가 수동으로 선정한 추천이거나, 조회수/인기순 등 모두에게 동일한 정렬 결과를 보여줍니다.



## 단순 필터링 리스트

"나에게 맞춰진 추천"이 아닌, 사용자가 설정한 카테고리 조건에 맞는 공모전을 나열하는 수준입니다.

"사용자는 '나를 위한 추천'이 아닌 '**모두를 위한 정렬**'을 보고 있습니다."



## 과정(Process)이 기록되는 구조

행동 하나하나가 데이터로 남음

- 노출 (Impression)
- 클릭 (Click)
- 30분 시청 (Duration)
- 이탈 (Leave)
- 연속 재생 (Chain)

✓ 취향 즉시 파악 가능



## 결과(Result)만 남는 구조

고민은 길지만 데이터는 없음

VS

- 탐색 (비회원)
- 고민 (기록 없음)
- 접수 (Application)

\*1년에 1~2회 발생 (저빈도)

⚠ 학습 재료 부족

우리는 이 문제를 해결하기 위해

**‘지금 가능한 것’으로  
개인화 추천 AI를 구현하고  
‘다음에 필요한 것’을 고려해  
확장 구조까지 설계했습니다.**

# 접근 방식: AI 실행 전략을 두 단계로 나누었습니다.

모델 성능보다 중요한 것은 '올바른 질문'과 '현실적인 데이터 설계'입니다

Pragmatic AI Strategy



핵심 질문 : “완벽한 모델을 만들 수 있는가” (X)  
“공모전 도메인에 AI 도입이 가능한가” (O)



## PHASE 1. 검증 (VERIFICATION)

**"지금 있는 데이터로  
개인화(AI)는 어디까지 가능한가?"**

### → 제한된 신호의 최대 활용

접수 로그, 관심사 태그 같은 제한된 신호만으로도  
사용자군(Segment)별 추천 전략을 다르게 설계합니다.



NEXT

## PHASE 2. 설계 (BLUEPRINT)

**"더 좋은 추천을 위해서는  
어떤 로그/DB가 필요한가?"**

### → 과정 로그의 구조화

노출/클릭/체류 데이터를 담을 DB 구조를 설계하고,  
비회원 탐색 로그를 수집할 파이프라인을 정의합니다.

### Key Goal

**"개인화처럼 동작하는 경험"**

Rule-based + AI Hybrid Strategy

### Key Goal

**"추천 품질을 단계적으로 높일 수 있는 데이터 구조 설계"**

Feature Store & Quality Guardrails

# Phase 2가 필요한 이유:

우리는 빙산의 일각만 보고 추천을 진행하고 있습니다.

## 과정 데이터 (Process)

MISSING

- ▶ 노출, 클릭, 스크롤, 체류
- ∞ 빈도 매우 높음 (실시간 누적)
- 💡 취향/의도를 파악할 수 있는 **진짜 단서**

"이 데이터를 잡아야 비회원 탐색 단계부터 개인화가 가능  
해집니다."

## 결과 데이터 (Result)

CURRENT

- 📌 접수, 수상
- 📊 빈도 낮음 (1년에 1~2회)
- ❗ 학습 데이터 부족 (**Cold Start**)

### 안전지향적 수렴

"실패하지 않기 위해  
인기순/광고만 노출"

Depth: Shallow

Depth: Medium

Depth: Deep

# 추천 = 인프라 돈이 되는 구조 만들기

단순 기능 도입이 아닌, 데이터가 가치로 전환되는 3단계 로드맵

 Infra Roadmap



 운영 효율성(단기)을 시작으로 비즈니스 확장(장기)까지 이어지는 선순환 구조

SECTION 02

# 프로젝트 요약

# 추천시스템 설계 방향

데이터 환경 기반 추천 시스템 접근 전략



## 데이터 환경

현재 운영 DB 중 3가지 사용



### 회원(Member) 테이블

사용자 정보 및 관심사



### 공모전(Contest) 테이블

추천 대상 아이템 정보



### 접수(Log) 테이블

유일한 사용자 행동 신호 (Action)

핵심

**LIMITATION** 사용자 대부분이 1회 또는 소수의 접수 이력만 보유하여,  
성능 중심의 단일 추천 AI 모델 적용에 한계가 존재합니다.



## 프로젝트 목표

현재 데이터 환경에서도 운영 가능할

AI 추천 시스템을 구현하고

단일 개인화 모델이 아닌

사용자 세그먼트별 목적 중심 추천 전략에  
집중했습니다.



그룹별 전략



운영 적용성



콜드스타트 보완

# 전처리 및 최종 변수 구성

추천 시스템 적용을 위한 데이터 구조 정비 프로세스



## Member Table

USER DEMOGRAPHICS

- ✓ 사용자 식별 기준(PK) 확립  
고유 식별자 생성으로 데이터 정합성 확보
- ▼ 추천 비대상 제거  
직원, 주최사 등 추천 불필요 계정 필터링
- ◎ 핵심 변수 선별  
연령, 성별, 관심사 등 주요 Feature 추출

01



## Contest Table

ITEM INFORMATION

- ▢ 공모전 데이터만 선별  
대외활동, 강연 제외하고 순수 공모전 집중
- ◁▶ 텍스트 정보의 정형화  
자격요건(지역/연령) 텍스트를 코드로 변환
- ★ 추천 활용 변수 최적화  
분야, 상금, 마감일 등 핵심 속성 구조화

02



## Log (접수)

INTERACTION DATA

- ▢ 비회원 접수 이력 통합  
학습 데이터 확보를 위해 비회원 정보 확보
- ▷◁ 데이터 구조 재정리  
접수번호 기준 → 참가자 기준으로 관점 전환
- ☰ 1인 1행 정규화  
사용자-공모전 관계를 분석 가능한 형태로 변환

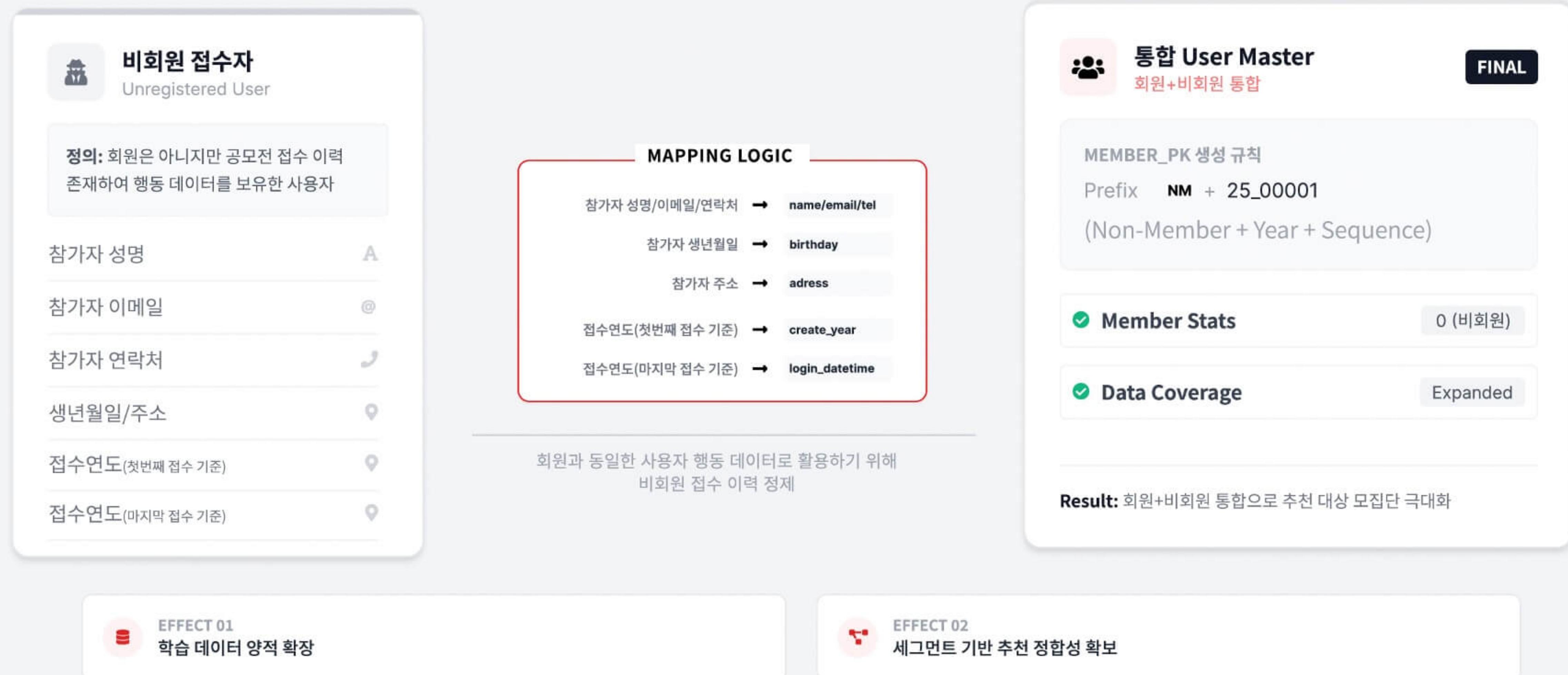
03



“ 전처리는 추천에 바로 사용할 수 있도록 데이터의 형태와 기준을 맞춰주는 과정입니다.

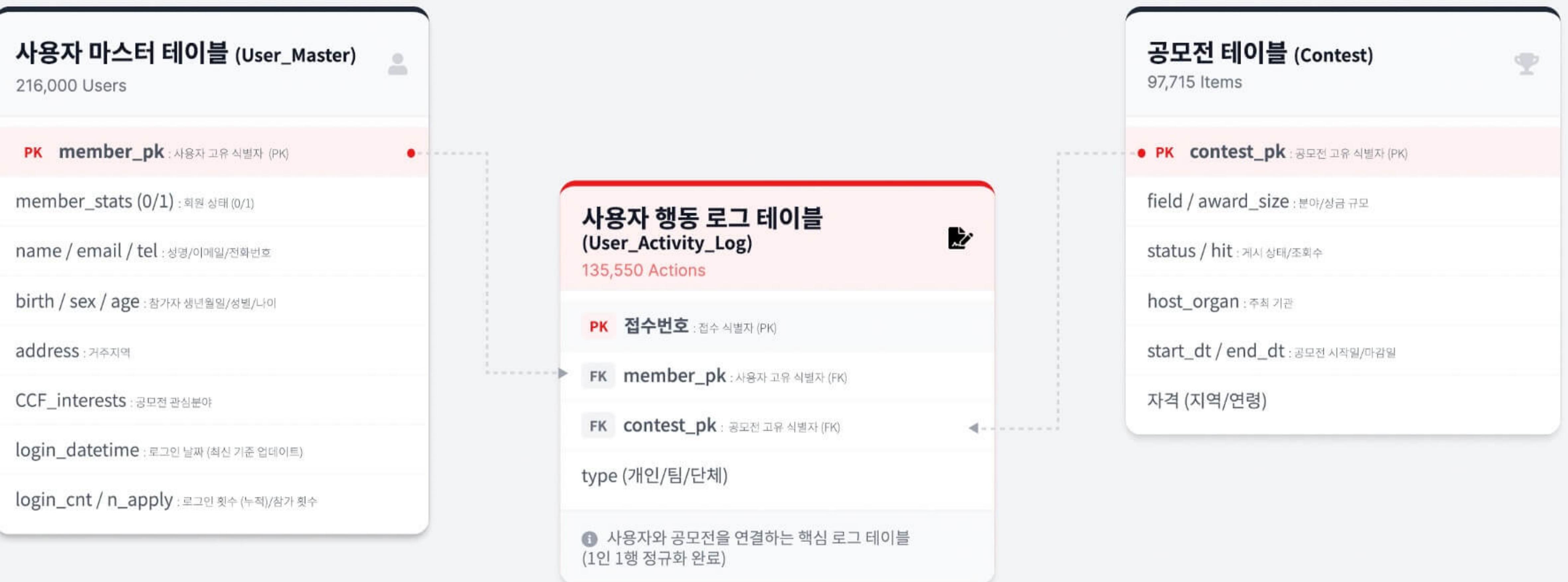
# 비회원 접수 이력 활용

추천 정확도 향상을 위한 데이터 통합 프로세스



# 최종 테이블 구조

추천 알고리즘 구현을 위한 최종 테이블 형태

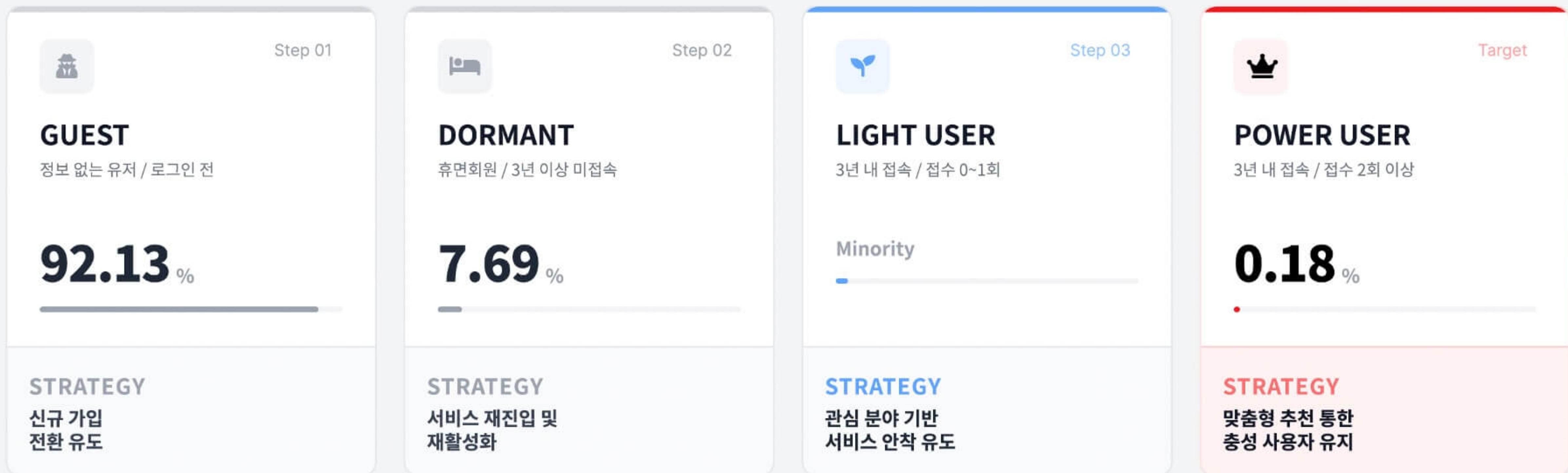


사용자(User) – 행동(Action) – 아이템(Item) **3축을 명확히 연결** 합니다.



# 세그먼트 정의

추천 목적에 따른 4단계 사용자 세그먼트 분류



◎ 목적별 세그먼트 정의로 전환·재활성화·유지 전략 을 구체화합니다.



## PROJECT SUMMARY

# 세그먼트별 AI 추천 모델

초기 진입 사용자와 이탈 우려 사용자를 위한 전략



## GUEST

정보 없는 유저 / 로그인 전

92.13%

### MODEL

#### Popularity + Recency

### LOGIC

사용자 정보가 없으므로,  
전체 사용자 기준 인기 공모전 중심 추천

### ◉ EXPECTATION

서비스 이해도 향상 및 첫 방문 이탈률 감소  
별도 정보 입력 없이 즉시 탐색 가능



## DORMANT

휴면 회원 / 3년 이상 미접속

7.69%

### MODEL

#### Popularity + Rule-based (Easy)

### LOGIC

인기도와 함께 **진입 장벽(낮은 난이도)**을 고려한  
재참여 유도형 추천

### ◉ EXPECTATION

재접수 및 재방문율 개선  
부담 없는 선택지 제공으로 Retention 확보



진입 장벽을 낮춰 '**재진입**'을 먼저 확보합니다.

## PROJECT SUMMARY

# 세그먼트별 AI 추천 모델

사용자 행동 데이터 축적에 따른 정밀 개인화 전략



## LIGHT USER

관심사 기반 탐색 유도



### MODEL

#### Content-based Filtering

(IDF 기반 중요 키워드 추출)



### LOGIC

사용자 관심 분야와

공모전 태그(Tag) 유사도 기반 추천



### EXPECTATION

명시적 관심사 기반 참여 확률 상승



## POWER USER

충성 고객 정밀 타겟팅



### MODEL

#### Item-Profile Hybrid

(Item-CF + Content-based)



### LOGIC

과거 접수 이력과 관심 패턴을 결합한

정밀 개인화 추천 제공



### EXPECTATION

충성 사용자 유지 / 고도화된 실험 가능

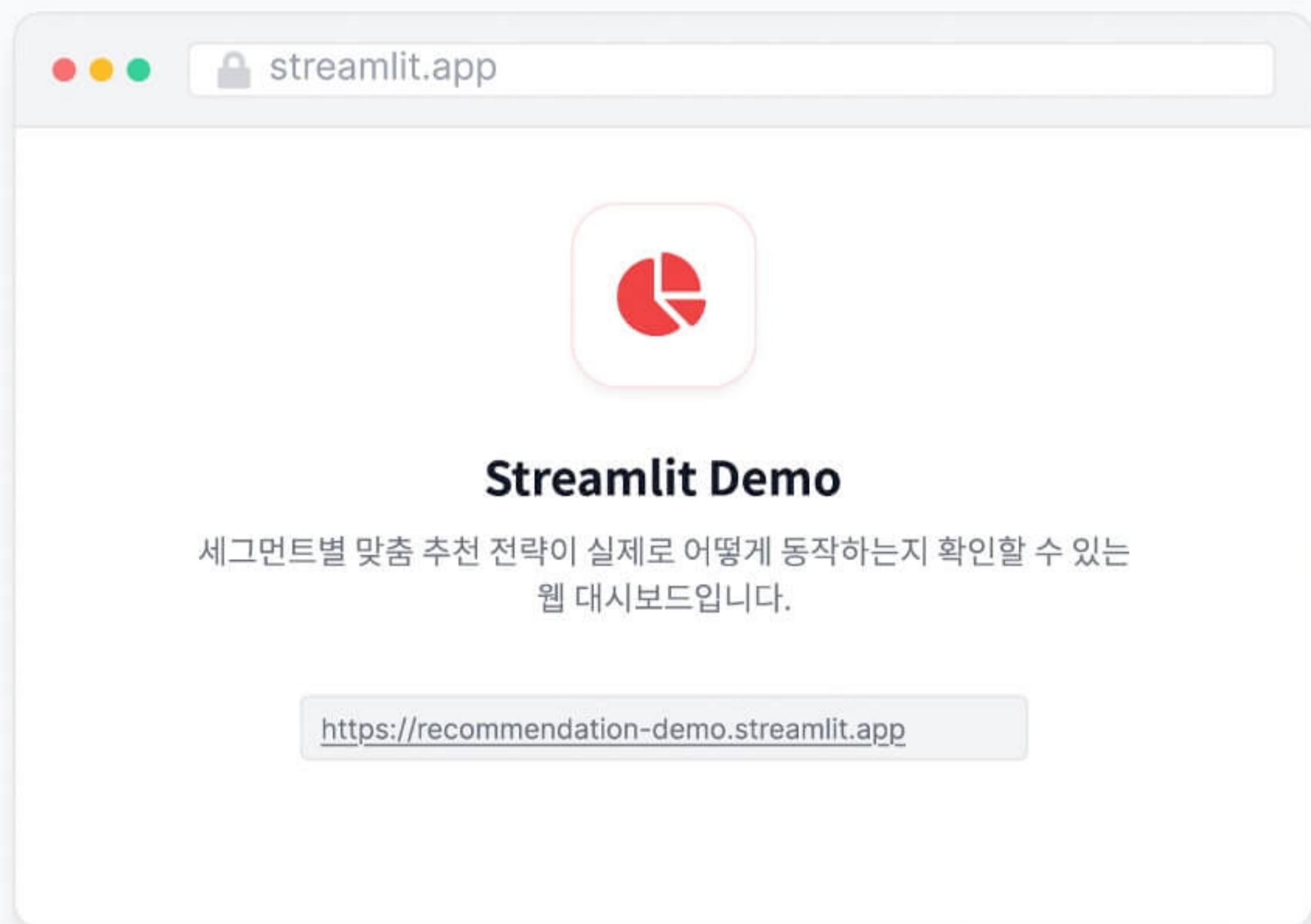


정보가 쌓일수록 '정밀 개인화'로 이동합니다.



# Demo 화면

개인화처럼 동작하는 경험을 보여주는 데모 시스템



SECTION 03

---

# DB 제안 사항

# DB 제안 - 멤버



## **⚠ CURRENT ISSUES**

#### 회원 고유 식별 칼럼 부재

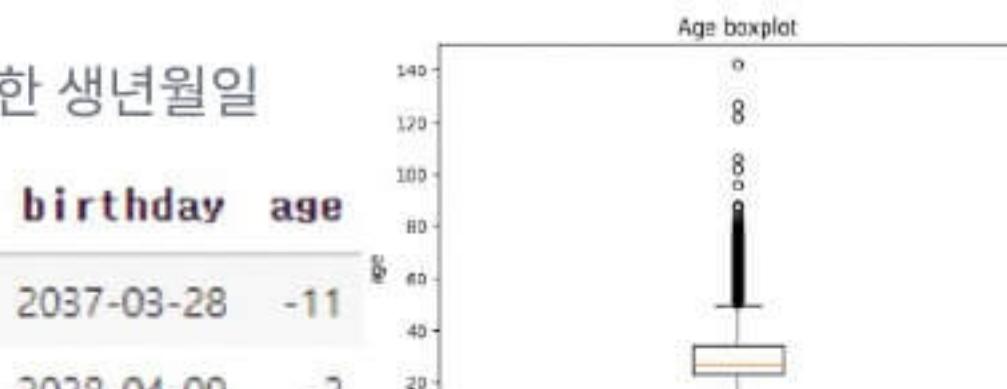
ID 형식이 통일되어 있지 않아 시스템적 식별 불가

```
df_M['member_id'].unique() [126109:126139]  
array(['113925455', '1139276496214185', '113995893',  
'114001571', '114004500306027', '114007066', '  
114034735', '114040880', '1140913412742454',  
'114099884', '114112972', '114151300', '114190'  
'114216688', '114219435', '114223973', '114252',  
'11426650', '114266101', '114286143', '1143028',  
'1143077048210981', '114340063', '114348645',  
'11441949'], dtype='object')  
  
df_M['member_id'].unique() [126109:126139]  
array(['zzzrtrt012', 'zzzsongs000', 'zzzung11', 'zzzuou', 'zzzvltd',  
'zzzx0111', 'zzzxxxx98', 'zzzz1373', 'zzzzcoolcool',  
'zzzzfuuc57894', 'zzzzz543', 'zzzzw', 'zzzzzzz', 'zzzzzzizon',  
'김예설', '김한서HAPPYVK', '내가찢는다', '박서현', '별이래', '보율한보돌씨', '신경식', '안세린',  
'오유진', '윤사월', '이정찬', '자수인', '최소희', '행운을만드는여자', '황자은', '효정'],  
dtype='object')
```

## \* 중복 가입 및 정보 오류

동일 이메일/전화번호 중복 계정, 부정확한 생년월일

M24_05345	tnals74	1	오수 만	000smo@naver.com	5276- 0704	010-
M24_05344	tnals0704	1	오수 만	000smo@naver.com	5276- 0704	010-



## ■ 데이터 혼재 (Mixed Columns)

member\_char1~14 등 여분 칼럼에  
주최기관/거주지 정보 혼재

```
df_M['seubar_char1'].unique()

array(['서울특별시', None, '부산광역시', '경기도', '대전광역시', '전라남도', '충청북도', '경상북도', '인천광역시', '광주광역시', '경상남도', '충청남도', '대구광역시', '전라북도', '강원도', '세종특별자치시', '제주특별자치도'], dtype=object)

df_M['seubar_char14'].unique()

array(['서울특별시', None, '부산광역시', '경기도', '대전광역시', '전라남도', '충청북도', '경상북도', '인천광역시', '광주광역시', '경상남도', '충청남도', '대구광역시', '전라북도', '강원도', '세종특별자치시', '제주특별자치도'], dtype=object)

df_M['seubar_char2'].unique()

array([None, 'KE500', '더학생', ..., '없는것같다', '친업주부', '독동중학교'], dtype=object)

df_M['seubar_char15'].unique()

array([None, '+082079183406206AA936EE2F315FFBD28630072', 'E5D632222AECFA191CEA69A2275A2010E66D1E62']
```

## ✓ PROPOSED SOLUTION



고유 식별 체계 및 인증 도입

고유 회원 번호(UUID 등) 도입 및 이메일/전화번호 기준 1인 1계정 제한,  
본인인증 프로세스 추가

데이터 관리 가이드라인 수립



임시 칼럼 사용을 지양하고 명확한 칼럼 정의 및 입력 규칙(Validation) 적용

## EXPECTED EFFECTS



### ⑥ 데이터 연결 안정성



#### ▶ 처리 속도 향상



개인정보 보호



#### ▶ 분석 데이터 오염 방지

# DB 제안 - 공모전

## ▲ CURRENT ISSUES

### ▣ 높은 결측률 및 낮은 유용성

결측치가 90% 이상인 다수 칼럼 존재로 데이터 품질 저하

standard	95146
cheering	95017
guidance	96468
coach	90379
connect_url	97714
connect_url_nm	97715

- 97715 행 중 결측치 개수

### ☰ 유사 정보 분산

날짜, 상금, 참가 자격 등 유사한 정보가 여러 칼럼에 산재

### ◁ 비정형 텍스트 및 형식 불일치

지역/연령/명칭 등이 단순 텍스트로 저장되어 기계 학습 불가

전주지역에 거주하며 환경에 관심 있는 모든 중고등학생 5년 이내 창업(팀) 놀이터에 관심 있는 안산시민 누구나 만 19세 ~ 만 35세의 개인 및 그룹 만 40세 미만 본 센터에서 상담을 받은 경험이 있는 이용자
강동구에 거주하거나 강동구를 자주 방문하는 주민 강동두내 활동자격1가지이상충족 강원도민 강원도에 거주하는 만 65세 이상(1958년 12월 31일 이전 출생)

## ✓ PROPOSED SOLUTION

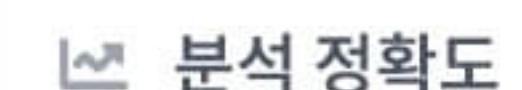
### 데이터 정제 및 구조 통합

사용 불가 칼럼 과감히 폐기(질 우선) 및 시작일/마감일, 상금 등 핵심 정보 통합 관리

### 텍스트 코드화 및 정규화

연령/지역 등 텍스트 데이터를 코드화하고, 후원사/개최사 등  
중복 정보는 별도 테이블로 분리

## EXPECTED EFFECTS



분석 정확도



자원 절약



데이터 일관성



확장성 및 보안

# DB 제안 - 접수

☞ TABLE: LOG (APPLICATION)

## ▲ CURRENT ISSUES

### ▣ 참가자 고유 식별자(PK) 부재

회원/비회원이 혼재되어 개별 참가자 식별이 불가능

### ❖ 팀별 고유 번호 형식 불일치

팀원 간 연결 고리가 약하고 접수 번호 규칙이 제각각임

접수번호	참가자성명	팀명	참가형태
TW07110001	박교리	스마트한 남매	팀 단체
TW07110001-1	박정의	스마트한 남매	단체참가

### ② 접수/참가 기준 모호성

단체 접수 내 개별 참가자 구분 불가, 명칭 혼재

— 참가형태 칼럼의 고유값 —  
['개인참가', '개인', '팀', '팀참가', '단체참가', '단체']

## ✓ PROPOSED SOLUTION



### 고유 PK 생성 및 1인 1행 원칙

회원 테이블과 연결 가능한 고유 PK를 생성하여 '한 행이 한 명의 참가자'를 대변하도록 정규화



### 데이터 표준화 가이드라인

팀 접수 번호 규칙 통일(예: AVCD-1, AVCD-2) 및 참가 형태 명칭 일관화



접수 데이터도 '개인 단위'로 추적 가능해야 합니다.



### Plus Alpha: 행동 로그 축적 필요

단순 접수 이력을 넘어 '방문 로그', '체류 시간' 등 행동 데이터를 축적하여 고도화된 추천 기반 마련

## EXPECTED EFFECTS

↪ 분석 신뢰도 향상

🚫 중복 데이터 차단

SECTION 04

# 단계별 적용 전략

# 그래서, 쌍궁은 지금 무엇부터 시작해야 할까?

≠ CURRENT SITUATION

- AI 모델 성능을 높이기엔 데이터 부족

- ▣ 사이트 전면 개편 = 추천 시스템의 Golden Time

## JAN

- 현재 (Before)

- 공모전 정보 저장 방식**

줄글 위주의 비정형 텍스트로 저장

- 로그 & DB 구조**

추천 관점으로 설계되지 않음 (단순 적재)

"데이터 활용의 한계가 존재함"

## PLAN

- 리뉴얼 (After)

- AI 도입 & DB 전면 개편**

추천 시스템 활용을 고려한 구조 설계

- 지속적 성장 기반 마련**

AI 활용 기업으로 도약 (개인화 추천 기능)

"데이터 기반 의사결정 가능"



🔑 **'지금의 리뉴얼'**이 데이터 설계의 기회입니다.

# AI 성능을 위한 관리자 입력 구조의 전환

관리자 입력 방식의 구조적 변화

추천/분석 시스템 즉시 활용 가능

## TEXT

- Before (비정형)

참가자격 (텍스트 필드)

"만 19세 이상, 서울 거주자 우대하며 기획에 관심 있는 분..."



- ✖ 컴퓨터가 조건을 이해하지 못함
- ✖ 필터링 및 매칭 불가능
- ✖ 분석 시 텍스트 파싱 비용 발생

## DATA

- After (정형화)

조건형/선택형/숫자형 컬럼

```
"age_min": 19,  
"region": "Seoul",  
"category": "Planning/Idea"
```

DEMO REFERENCE

- 🔗 [recommendation-demo.streamlit.app](https://recommendation-demo.streamlit.app)
- 🔗 [huggingface.co/.../ThinkGood\\_AI\\_Recommendation\\_System](https://huggingface.co/.../ThinkGood_AI_Recommendation_System)

‘텍스트 → 코드’로 바꾸면 곧바로 성과가 납니다.

# 데이터 및 AI 기반 운영의 시작

도구의 활용과 데이터 축적으로 시작하는 운영의 변화



## 분석 도구 활용 ANALYTICS TOOLS

### G Google Analytics + AI

이미 사용 중인 도구를 활용하여  
전문 지식이 없어도 핵심 지표 확인 가능



### ⌚ 사용자 흐름 가시화

어떤 공모전에서 이탈이 발생하는지,  
추천 영역이 실제로 클릭되는지 눈으로 확인



## 변화의 출발점 NEW STANDARD

### CHANGE 01

#### 데이터 자산화 시작

정형 데이터와 기본 행동 로그를 축적하여  
분석 가능한 자산으로 전환



### CHANGE 02

#### 운영 판단 기준의 이동

막연한 감(感)이 아닌 ‘숫자’를 보기 시작하며  
객관적 데이터 기반 의사결정 체계 확립



데이터가 **‘운영의 기준’**을 바꿉니다.



# AI 도입과 함께 이뤄져야 하는 것 : A/B Test

서로 다른 두 가지 방식을 동시에 적용해보고, 어떤 방식이 실제 사용자 반응에서 더 효과적인지를 데이터로 검증하는 방법입니다.

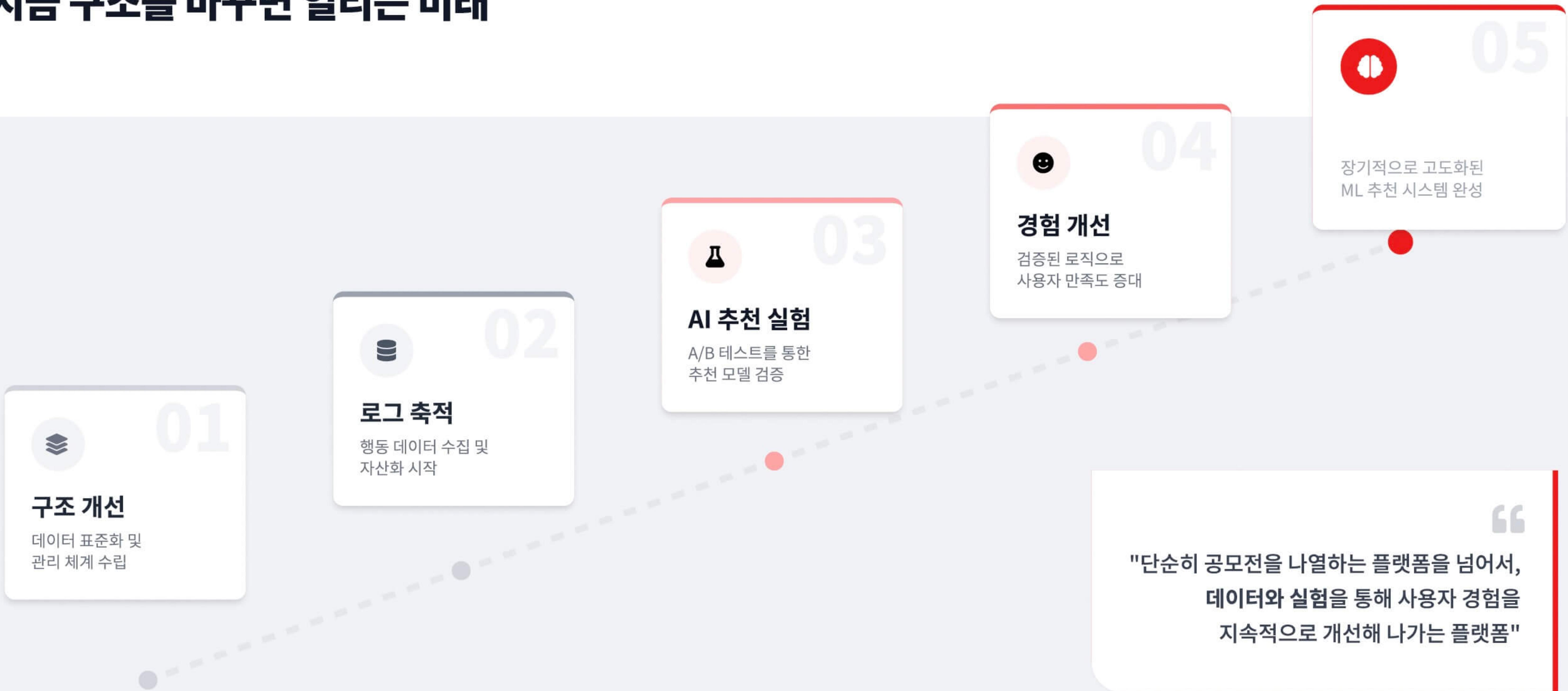


## GROWTH PRINCIPLE

“추천 AI는 한 번 만들어서 계속 쓰는 것이 아닌,  
지속적으로 성능을 검증하며 성장하는 도구 입니다.”

▶ 실험 문화가 추천 성능을 키웁니다.

# 지금 구조를 바꾸면 열리는 미래



▣ 구조를 바꾸면 '**진화하는 플랫폼**'이 됩니다.



# 감사합니다

씽굿 AI 서포터즈 최종 발표를 마칩니다.

**씽굿 서포터즈**

THINKGOOD SUPPORTERS

김승현 박현서 신경은 한재희