

# 실전 이탈 예측과 유의사항

NCSoft 분석모델링팀

장윤제

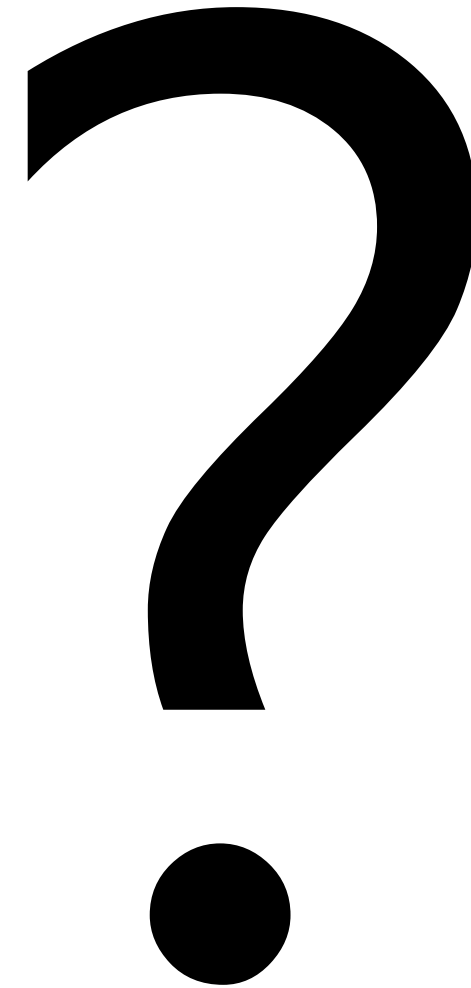
# 목차

1. 이탈예측
2. 예측 모델링 프로세스
3. 실전 고려 사항
4. 결론

# 이탈예측

# 이탈 예측

이탈 예측이란?



# 이탈 예측

이탈 예측이란?

유저



유저

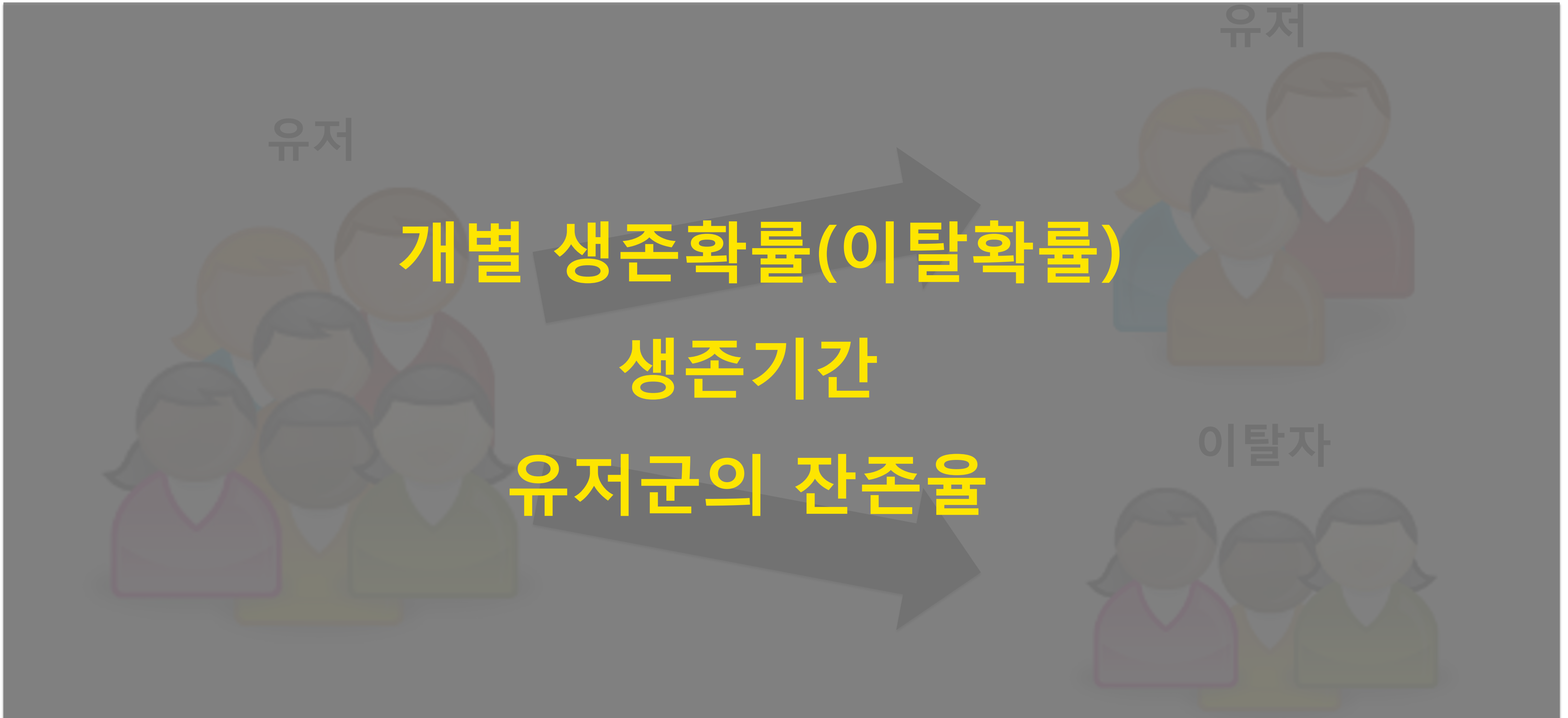


이탈자



# 이탈 예측

이탈 예측이란?



# 이탈 예측

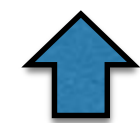
이탈 예측 목적

고객의 이탈을 방지하고

고객 생애 가치 (LTV), 투자 대비 수익 (ROI) 등

정확히 계산하여

**데이터 기반의 의사 결정에 일조**



분석가의 꿈

# 이탈 예측

이탈 예측 중요성

신규 유저 유입 대비 기존 유저 유지의 비용이 작다

또한 모바일 시대에는,

1) 유저 생존기간이 짧음

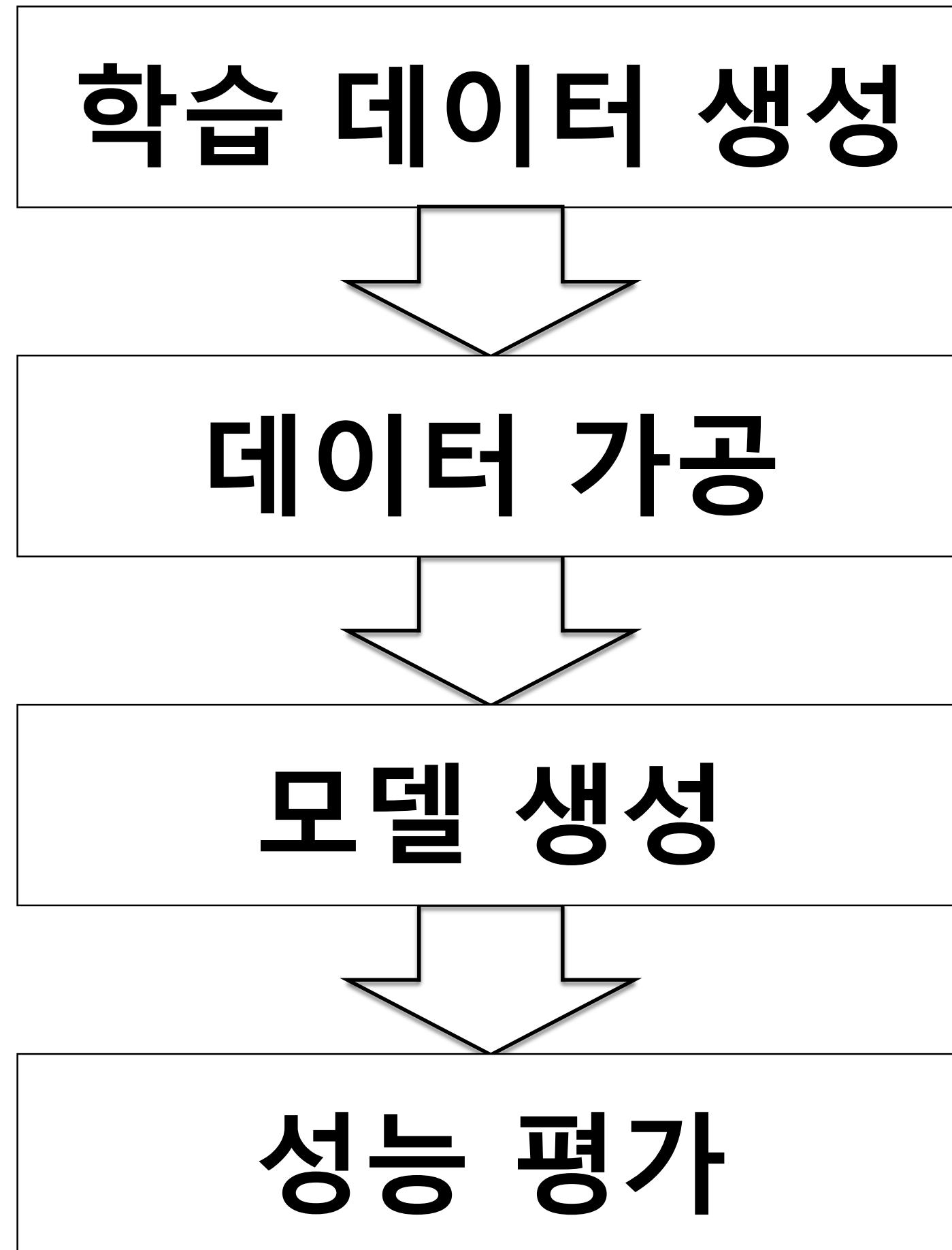
2)이탈 방지를 위한 직접적인 방안이 존재



# 예측 모델링 프로세스 & 실전 고려사항

# 예측 모델링 프로세스

예측 모델링 프로세스



# 예측 모델링 프로세스

실전에서 경험한 문제점



이탈은 어떻게 정의하지?  
누구를 대상으로 정하지?

# 실전고려사항

학습 데이터 생성

학습 대상

학습 데이터 = +

레이블

# 실전고려사항

학습 데이터 생성

학습 대상 ← 누구십니까?

학습 데이터 = +

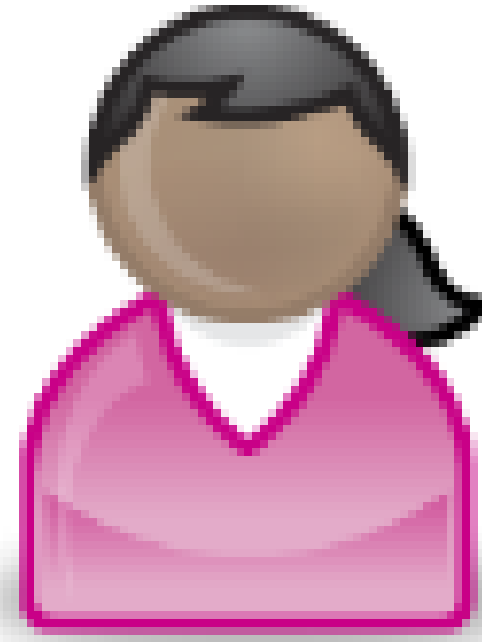
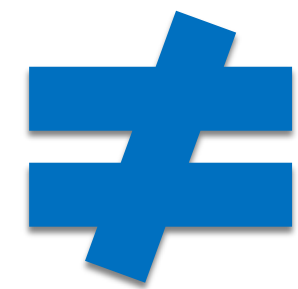
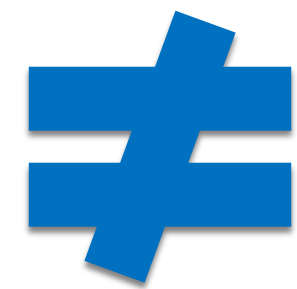
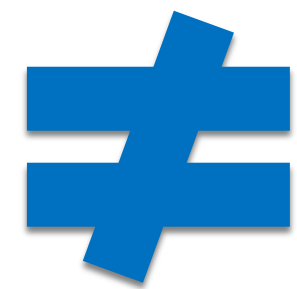
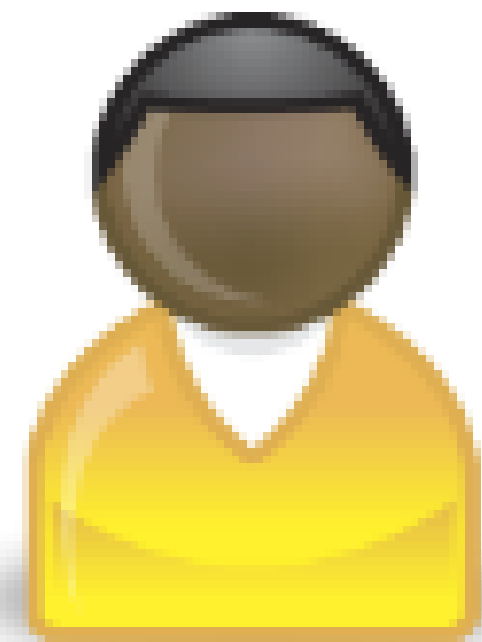
레이블 ← 이탈하셨나요?

# 실전고려사항

학습 데이터 생성 - 대상은 누구인가

전체 유저에 대해 이탈예측을 하자니 ...

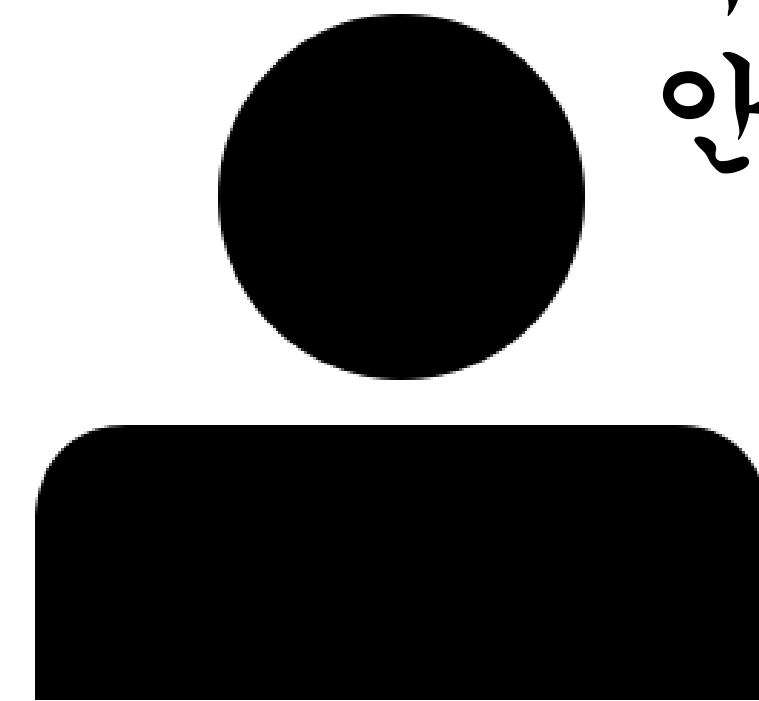
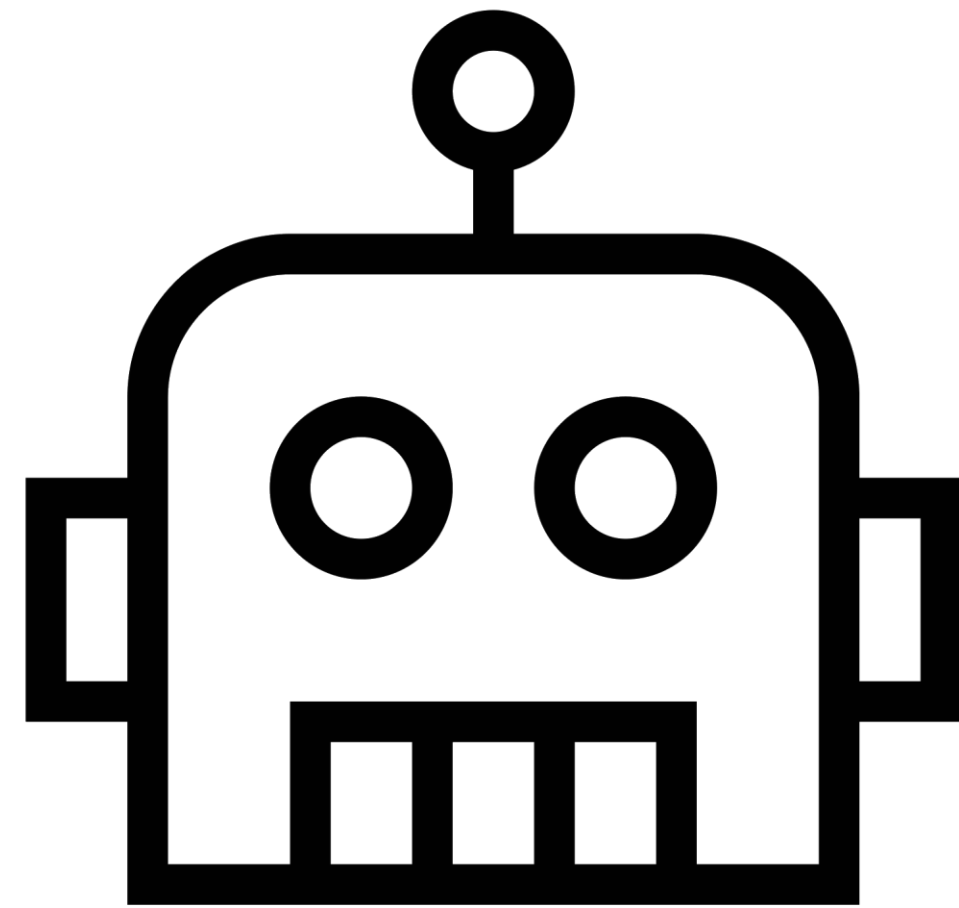
유저마다 성향이 너무 다르다



# 실전고려사항

학습 데이터 생성 - 대상은 누구인가

모든 유저를 대상으로 하는 이탈 예측은 비효율적



우린 참  
안 맞아.....

# 실전고려사항

학습 데이터 생성 - 대상은 누구인가

✓ BOT이 아니고

✓ 악성 유저도 아니면서

✓ 장기간 플레이 할 것 같은 유저



# 실전고려사항

학습 데이터 생성 - 대상은 누구인가

진성유저

유저 클러스터링으로 분류



이탈 정책/목표에  
맞는 유저군

# 실전고려사항

학습 데이터 생성

학습 대상 ← 누구십니까?

학습 데이터 = +

레이블 ← 이탈하셨나요?

# 실전고려사항

학습 데이터 생성 - 이탈의 정의

이탈이란 ?

~~탈퇴?~~

1년 이상 미접속 유저 중 서비스 탈퇴 유저는 0.8%

# 실전고려사항

학습 데이터 생성 - 이탈의 정의

이탈이란 ?

~~탈퇴?~~

**연속 미접속 기간**

1년 이상 미접속 유저 중 서비스 탈퇴 비율은 0.8%

# 실전고려사항

학습 데이터 생성 - 이탈의 정의

## 미접속 기간

짧으면



비이탈자 오탐

*저... 이탈 안했어요... 그냥 주말에 숙취때문에 접속을 못했어요*

길면



이탈자 방치

*유저가 로그인 하지 않은지 4달... 아직도 돌아올 것이라 믿는다*

# 실전고려사항

학습 데이터 생성 - 이탈의 정의

## 적절한 미접속 기간

내부 요인

+

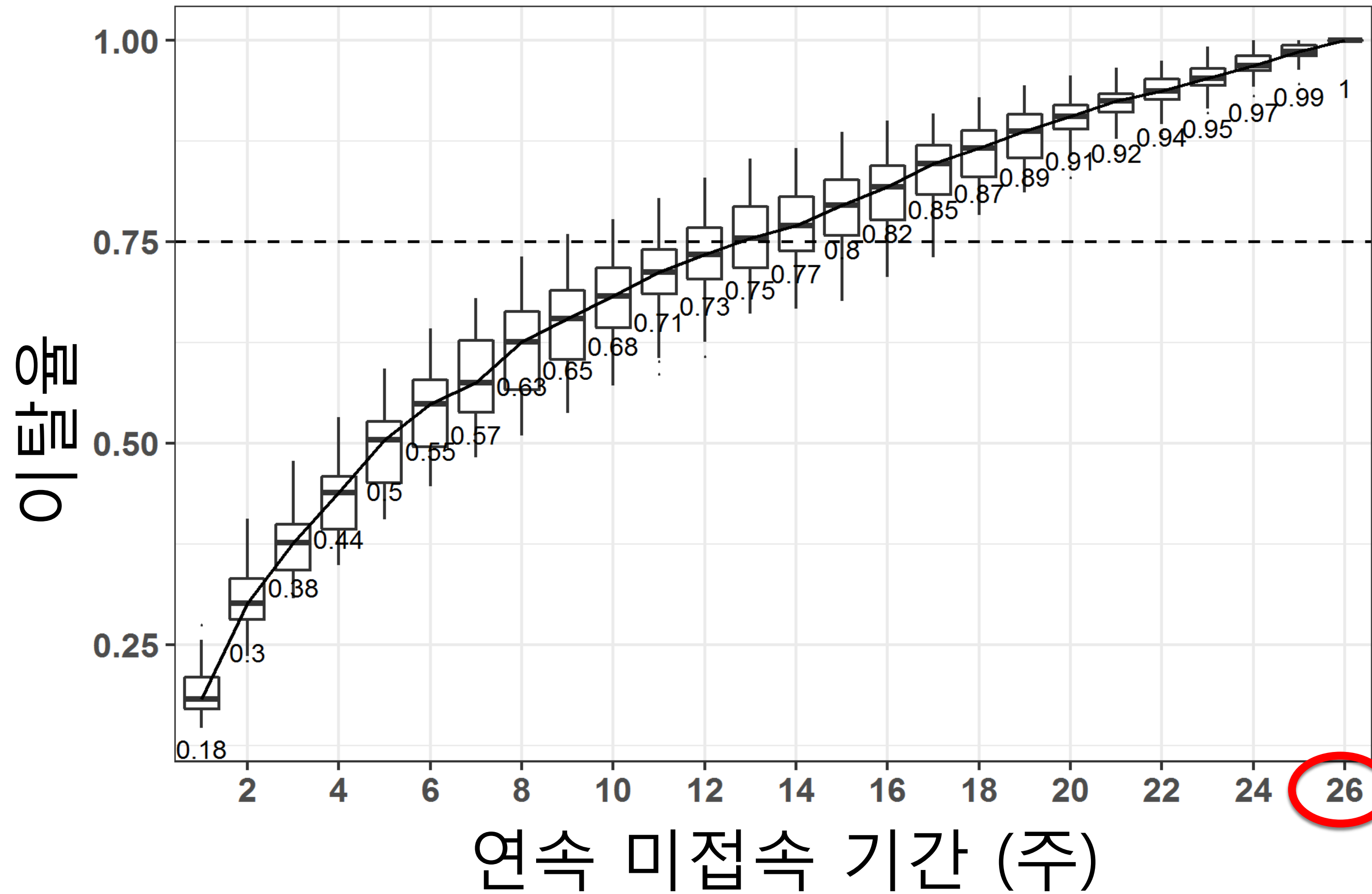
미접속 기간에 따른 복귀/이탈율

- 업데이트 주기
- 향후 사업 계획 등

- 1주 미접속 사용자가 복귀/이탈하는 비율
- 2주 연속 미접속 사용자가 복귀/이탈하는 비율
- 3주 연속 미접속 사용자가 복귀/이탈하는 비율
- ...
- N주 연속 미접속 사용자가 복귀/이탈하는 비율

# 실전고려사항

학습 데이터 생성 - 이탈의 정의

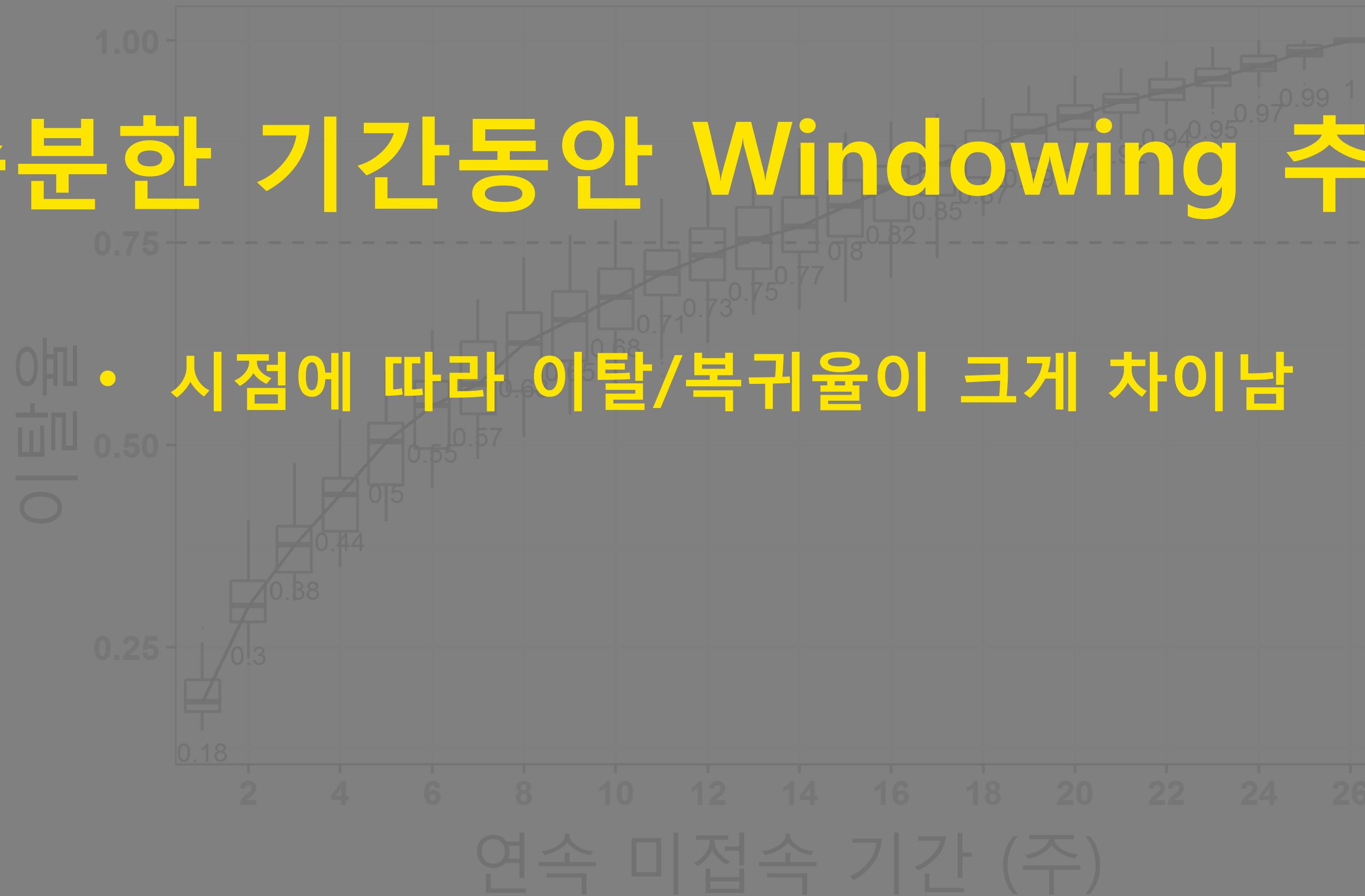


# 실전고려사항

학습 데이터 생성 - 이탈의 정의

## 충분한 기간동안 Windowing 추천

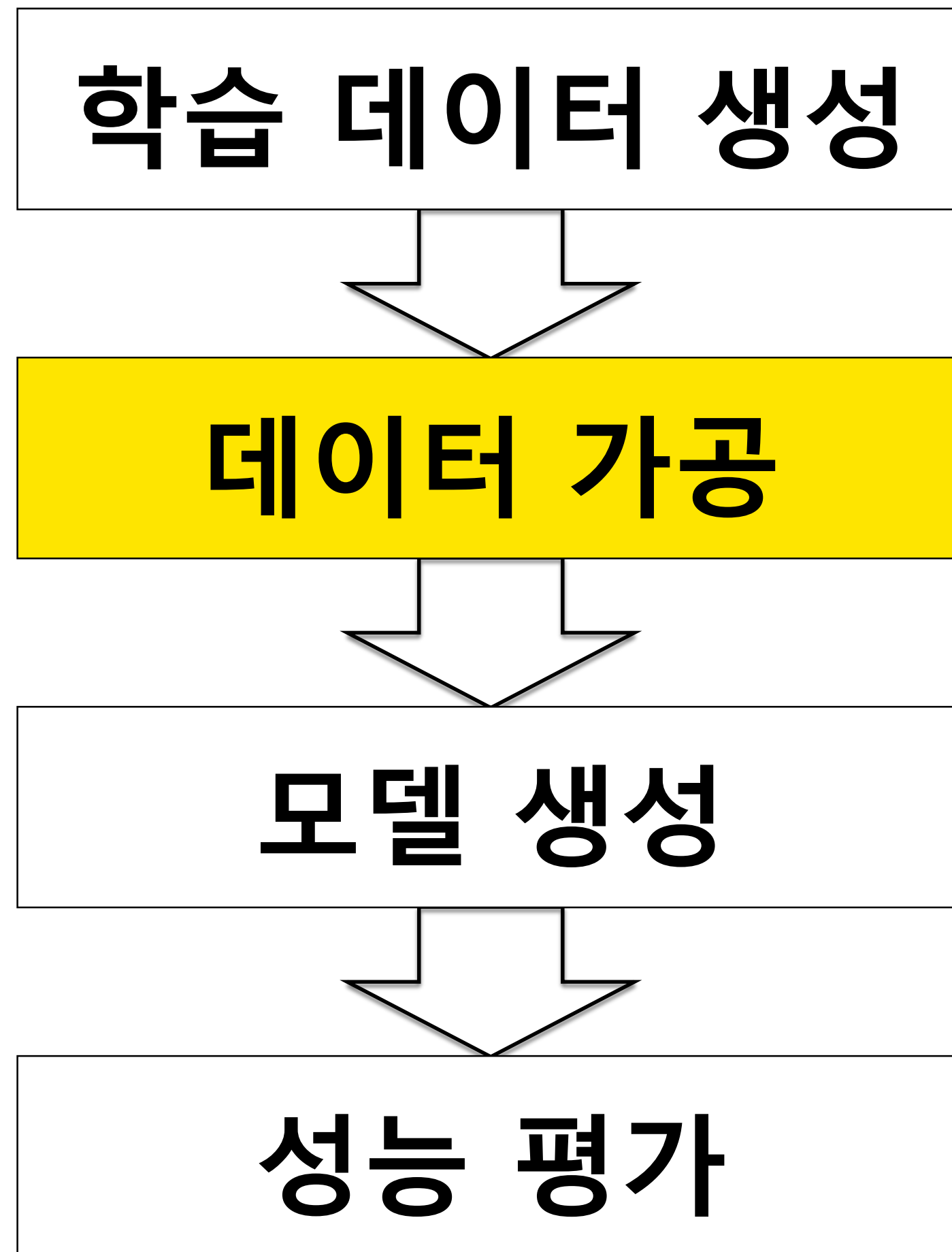
- 시점에 따라 이탈/복귀율이 크게 차이남





# 예측 모델링 프로세스

실전에서 경험한 문제점



이탈은 어떻게 정의하지?  
누구를 대상으로 정하지?

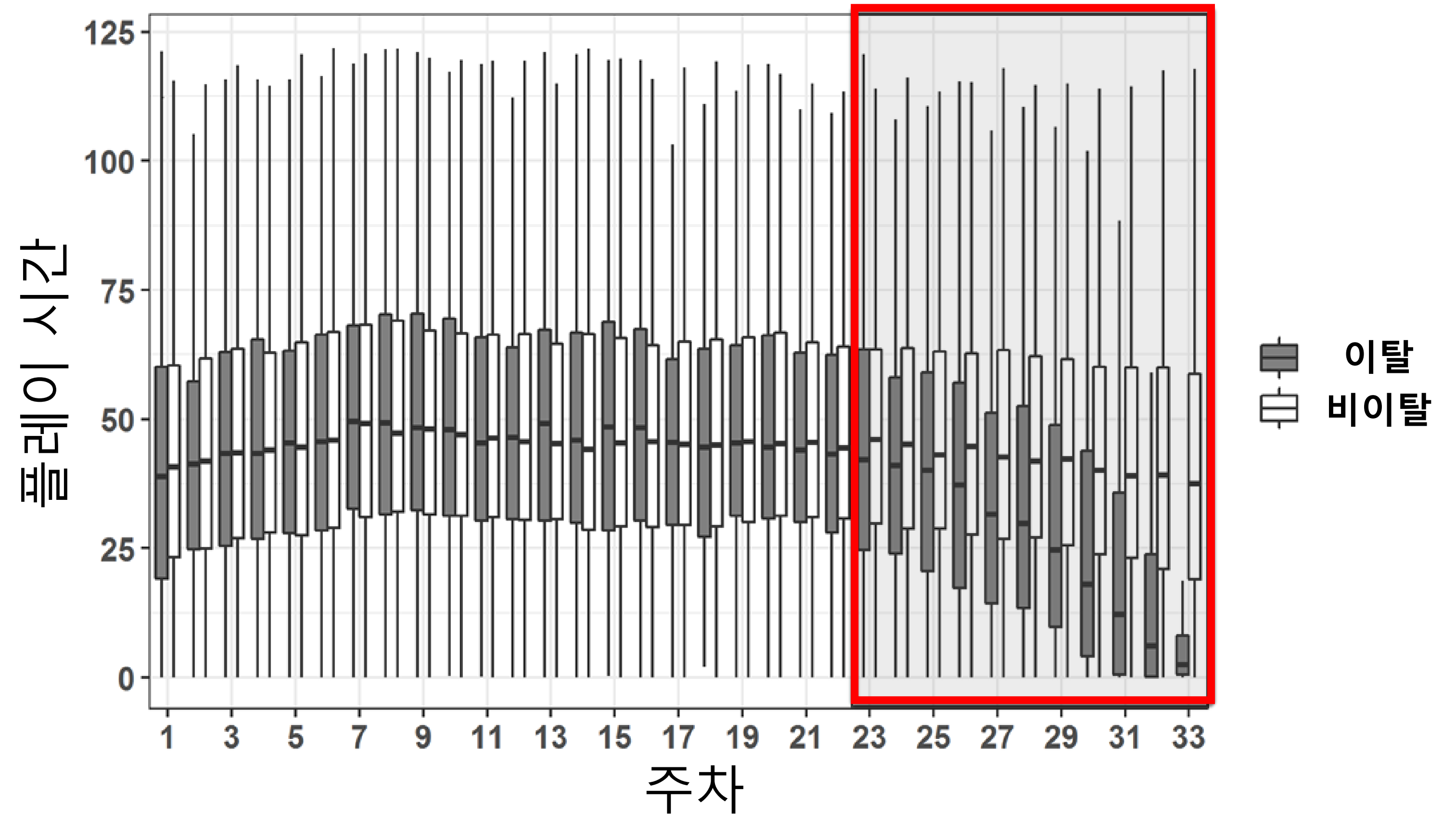


어떤 데이터를  
가공하여 사용하지?

# 실전고려사항

데이터 가공 - 활동 이력

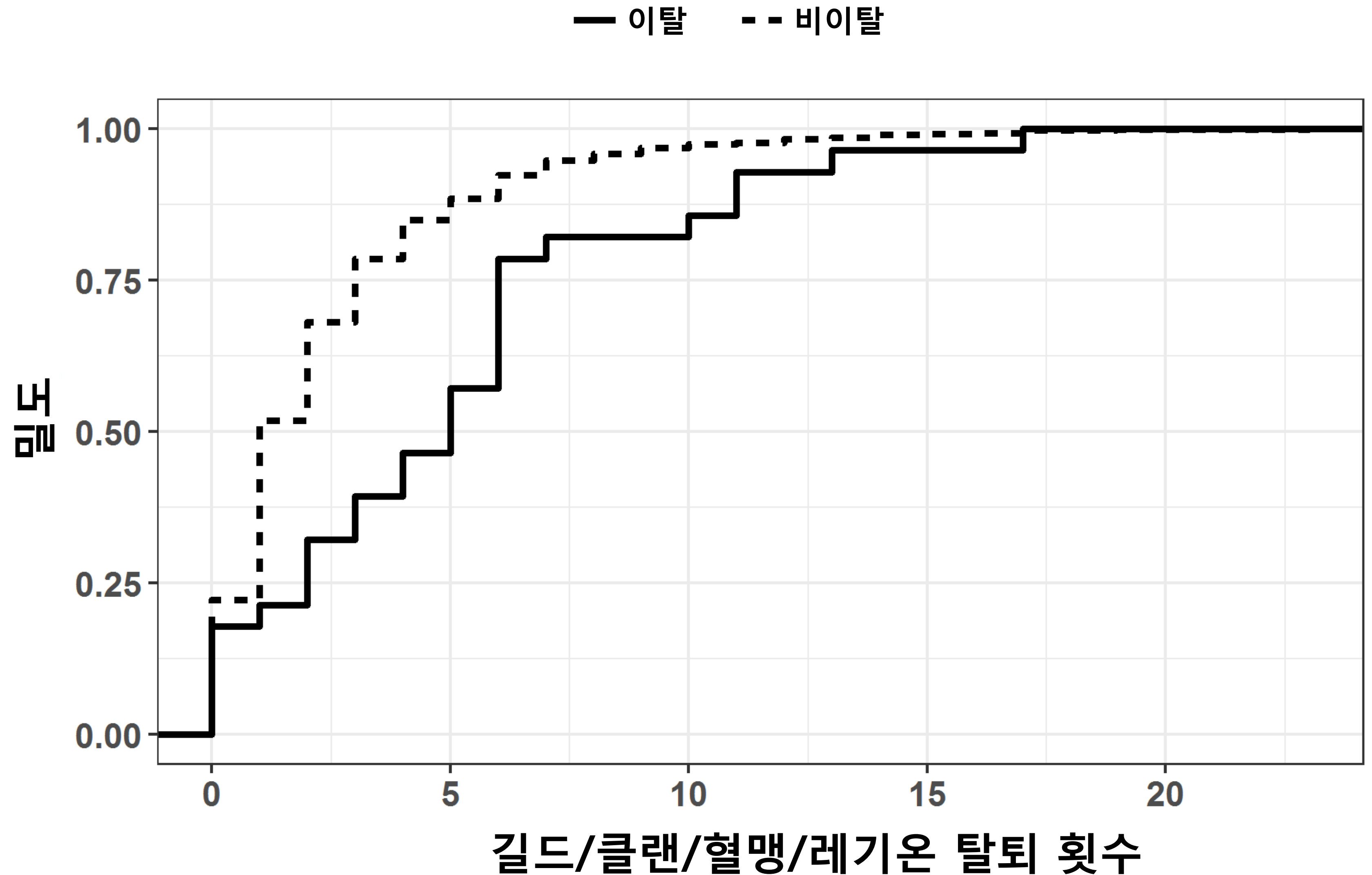
- **활동이력**



# 실전고려사항

데이터 가공 - 사회 활동

- 활동이력
- **사회활동**



# 실전고려사항

데이터 가공

- 활동이력
- 사회활동
- 결제이력 등에 대하여 추세, 변동성 등 계산하여  
변수가공

# 실전고려사항

데이터 가공 - 변수 가공



"Iron Man", Tony Stark (Robert Downey Jr.)

# 실전고려사항

데이터 가공 - 변수 가공

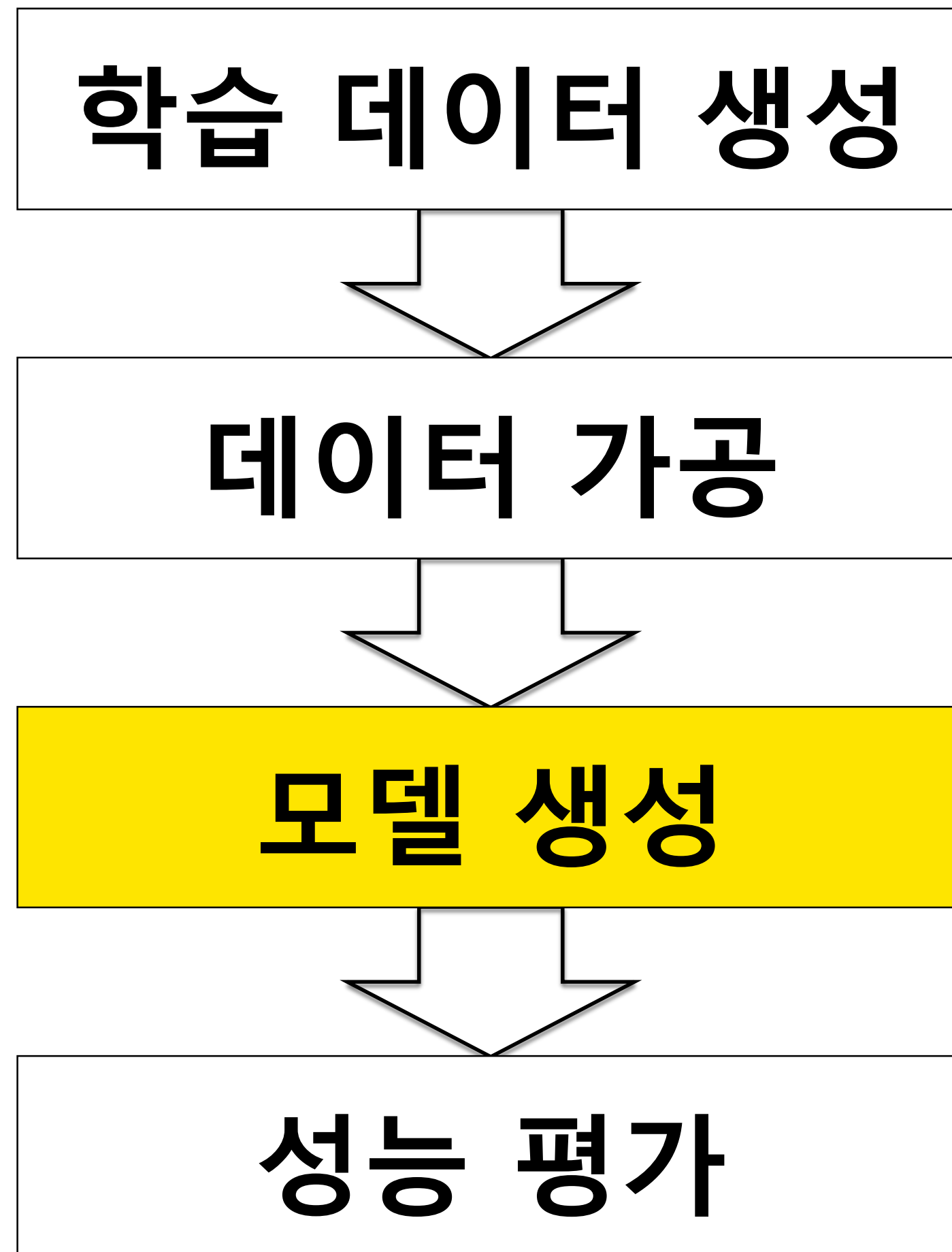


도메인 지식 전문가와 협의

“Iron Man”, Tony Stark (Robert Downey Jr.)

# 예측 모델링 프로세스

실전에서 경험한 문제점



이탈은 어떻게 정의하지?  
누구를 대상으로 정하지?



어떤 데이터를  
가공하여 사용하지?



어떤 알고리즘을 사용하지 ?

# 실전고려사항

모델 생성

예측력 + 해석력

- 이탈의 원인을 모르면, 정확하게 이탈자를 예측해도 이탈을 방지 할수 없다



# 실전고려사항

모델 생성

모델링/예측 목표에 따라,

단순 이탈 확률 VS

생존기간

예측 변수 : 이탈자이냐 아니냐

예측 변수 : 유저의 생존기간

Logistic Regression,  
Random Forest, SVM, XGB 등

생존분석법 – Cox Regression,  
Survival Random Forest 등

# 실전고려사항

모델 생성

모델링/예측 목표에 따라,

단순 이탈 확률 VS

생존기간

예측 변수 : 이탈자이냐 아니냐

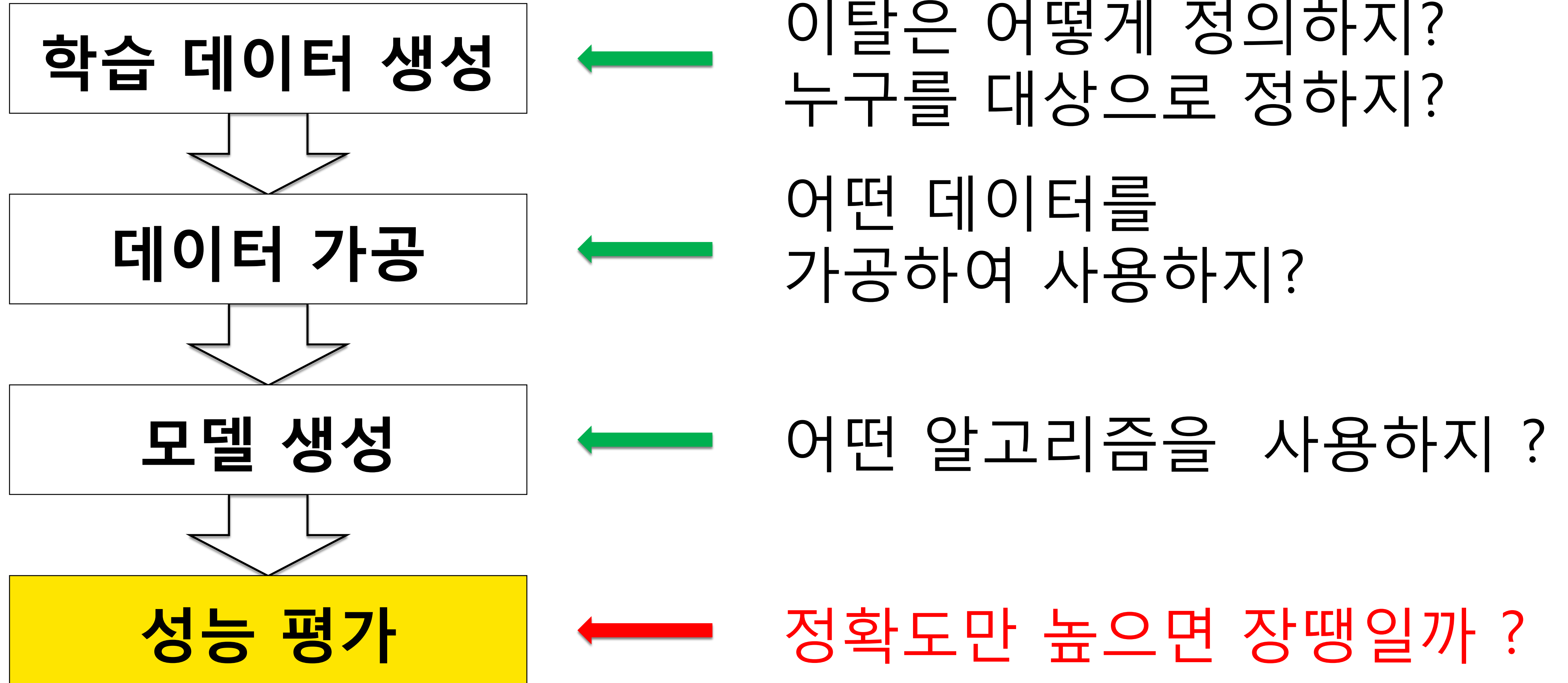
예측 변수 : 유저의 생존기간

Logistic Regression,  
Random Forest, SVM, XGB 등

생존분석법 – Cox Regression,  
Survival Random Forest 등

# 예측 모델링 프로세스

실전에서 경험한 문제점



# 실전고려사항

성능 평가

## Precision

확실히  
이탈 예측 가능한  
유저만 예측할래

## VS

## Recall

최대한 많은 유저들을  
잠재 이탈자로 간주할래

이탈 정책에 부합하도록 성능 측정

# 실전고려사항

성능 평가

내일 이탈 예측 → 3일후 이탈

이탈하긴 했는데... 맞췄다고하긴 좀 그렇네 .....

접속은하지만 게임 내 활동/구매 X

영혼은 이탈했는데, 몸이 남았나 ????

# 실전고려사항

성능 평가

1주후 이탈 예측률 100%, 2주후 이탈 예측률 10%

나..... 잘하고 있는거니?

업데이트 전에는 예측률 100%, 업데이트 후 0%

업데이트마다 모델 새로 만들꺼야?

# 실전고려사항

성능 평가

Time Dependent ROC Curve

IBS (Integrated Brier Score)

다수의 평가 데이터 사용

# 결론



# 결론

- 이탈의 정의는 무엇인가? → 탈퇴 X, 연속 미접속으로 정의
- 이탈 예측의 대상은 누구인가? → 이탈 정책에 부합하는 유저군  
e.g. 장기 진성 유저
- 이탈 예측을 위한 데이터는 어떤 것이 있는가? → 활동, 결제, 매출, 사회활동  
도메인 지식 전문가의 도움
- 어떤 모델링 기법을 사용해야 하는가? → 해석력 + 예측력  
생존기간 (생존분석) 추천
- 성능 평가시 고려사항은 무엇인가? → Precision vs Recall  
이탈 정책에 맞도록 성능 측정

결론

~~정답~~

~~비법~~

~~사랑~~

# 경험담 & 팁

# We are Hiring!

## 주요 업무

- 진성 유저 클러스터링
- 이상재화 증감 탐지 모델링
- 작업장 & BOT 탐지
- 이탈예측
- 데이터 분석 경진 대회
  - 2017년 IEEE CIG 소속 국제 대회 개최
  - 2018년 대회 개최 예정 (6월 경)
- 업무 소개 자료 : <http://blog.ncsoft.com/?p=28900>

# 감사합니다

NCSoft 분석모델링팀

장윤제 (yoonjaej@ncsoft.com)

# Q&A

NCSoft 분석모델링팀

**장윤제 (yoonjaej@ncsoft.com)**

어떻게 표현할  
방법이 없네~

Chun ho

