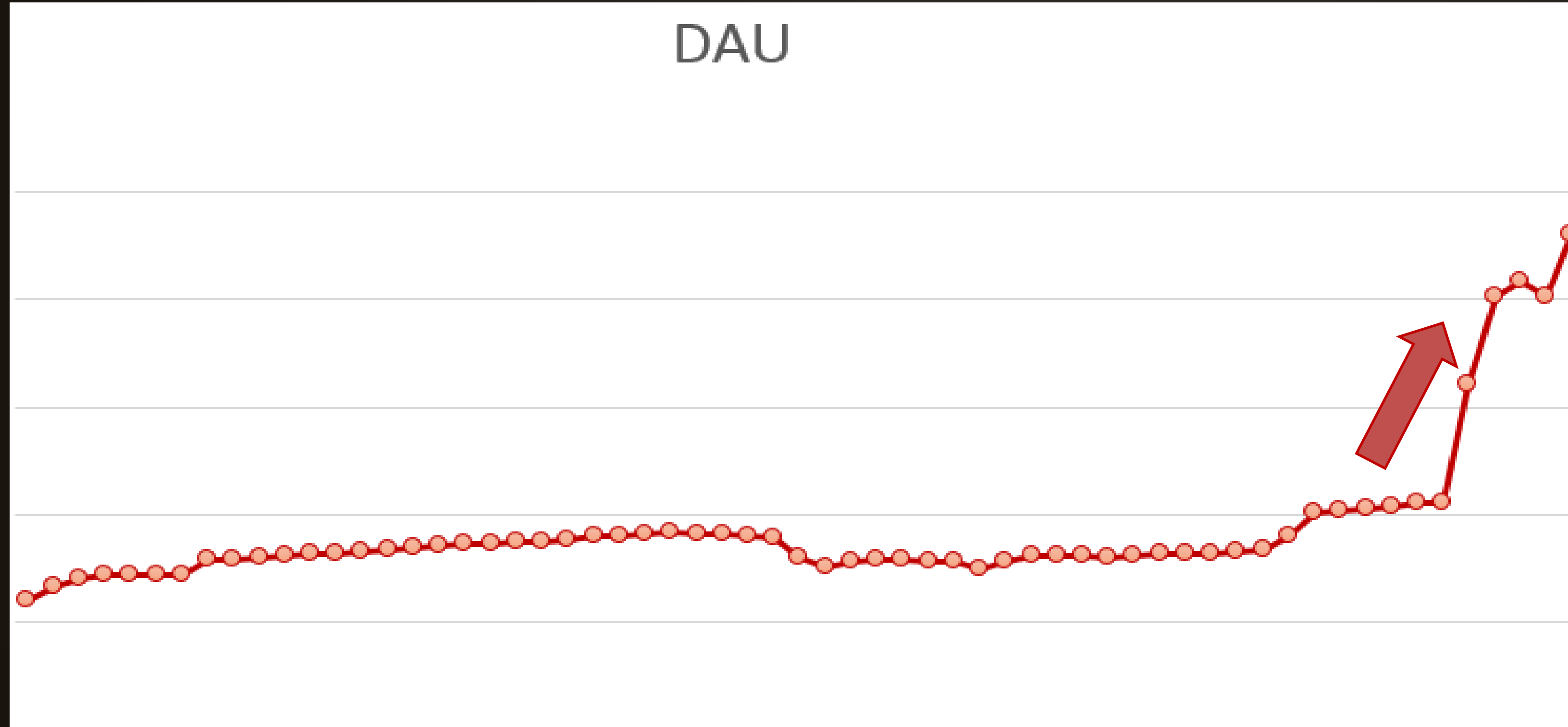


데이터마이닝을 활용한 진성 유저 지표 만들기



엔씨소프트 엄혜민

AU가 높다고 다 좋은걸까?



모바일 게임 업계를 중심으로 진성 유저 파악에 대한 관심 점차 증가



MMO에서도 진성 유저를 파악해 보자

[“우리 서비스의 진성 유저는 어디 있을까?” 탭조이-파이브락스 ...](#)

platum.kr/archives/28209 ▼

2014. 10. 6. - 이어지는 세션에서는 퍼포먼스 마케팅 전문가인 요기요(알지피코리아, RGP Korea) 박지희 부사장이 '진성 유저가 통과하는 마케팅 퍼널(Marketing ... 이 페이지를 15. 5. 28에 방문했습니다.

[진성유저와 진성앱을 찾아라... '앱 디스커버리 세미나' 성황리 ...](#)

platum.kr/archives/28507 ▼

2014. 10. 14. - 이번 세미나에서는 앱 디스커버리에 기반해 게임과 앱 서비스의 진성유저를 찾는 접근방식에 대해 각계 전문가들이 발표 및 토론을 하는 자리였다. 이 페이지를 15. 5. 28에 방문했습니다.

[진성 유저 분석을 통해 진짜 고객을 찾아라! 선데이토즈 임정...](#)

www.inven.co.kr/webzine/news/?news=119810 ▼

2014. 10. 14. - "성공적인 모바일 마케팅을 진행하기 위해서는 진성유저를 찾는 방법을 강구하라." 최근 들어 광고회사가 데이터 분석회사를 사들이거나, 혹은 ...

[온오프믹스 :: App Discovery : 우리 게임/앱의 진성 유저는 ...](#)

onoffmix.com/event/33891 ▼

탭조이(Tapjoy)와 파이브락스(5Rocks)가 선사하는 첫번째 세미나의 테마는 App Discovery! 각 분야 전문가들과 업계분들을 모시고 생각과 고민을 나누어 보고자 ...

[\[PDF\] TAPJOY INSIGHTS](#)

home.tapjoy.com/blog/wp-content/.../Insights_Issue7_04_korean.pdf ▼

제7호. 2013년 3월. TAPJOY. INSIGHTS. 읽을거리. 모바일 플랫폼. 선택. 진성유저 확보. Tapjoy의 PPE 광고로 진성유저 확보하기 ...

[\[한국경제\] 진성유저 확보 '겜포인트' 구글 인기 154위 진입!](#)

blog.ladybugs.co.kr/420 ▼

2013. 9. 23. - 진성유저 확보형 리워드앱 '겜포인트'가 추석 연휴에 이어 상승세를 보이고 있다. 레디벅의 게임 리워드앱 '겜포인트'는 8월 18일 출시되어 설치에서 ... 이 페이지를 15. 5. 28에 방문했습니다.

진성 유저??



- Real User?
- 매출이 높은 유저?
- 게임을 열심히 하는 유저?

신라의 골품제

4두품	관등		골품·공복의 색				
	등급	관등명	진골	6두품	5두품	4두품	
1	1	이별찬					
2	2	이찬					
3	3	잡찬					
4	4	파진찬					
5	5	대아찬					
6	6	아찬		자주색			
7	7	일길찬					
8	8	사찬					
9	9	급별찬					
10	10	대나마		붉은색			
11	11	나마					
12	12	대사				파란색	
13	13	사지					
14	14	길사					
15	15	대오					
16	16	소오					
17	17	조위					
		집 크기					

신라의 골품제

4두품	관등		골품·공복의 색				
	등급	관등명	진골	6두품	5두품	4두품	
1	1	이별찬					
2	2	이찬					
3	3	잡찬					
4	4	파진찬					
5	5	대아찬					
6	6	아찬		자주색			
7	7	일길찬					
8	8	사찬					
9	9	급별찬					
10	10	대나마		붉은색			
11	11	나마					
12	12	대사				파란색	
13	13	사지					
14	14	길사					
15	15	대오					
16	16	소오					
17	17	조위					
		집 크기					

신라의 골품제

5두품	관등		골품·공복의 색				
	등급	관등명	진골	6두품	5두품	4두품	
1	1	이별찬					
2	2	이찬					
3	3	잡찬					
4	4	파진찬					
5	5	대아찬					
6	6	아찬		자주색			
7	7	일길찬					
8	8	사찬					
9	9	급별찬					
10	10	대나마		붉은색			
11	11	나마					
12	12	대사				파란색	
13	13	사지					
14	14	길사					
15	15	대오					
16	16	소오					
17	17	조위					
		집 크기					

진성 유저란?

- 저마다 다른 이미지의 진성 유저
- 유저 개개인이 어떤 행동을 하는지 알지 못함
- 진성 유저의 기준 불명확



데이터마이닝을 이용한
탐사 분석 수행

작업 흐름

플레이 유형 파악 (고객 세그먼테이션)



유형별 태깅 및 진성 유저 그룹 선정



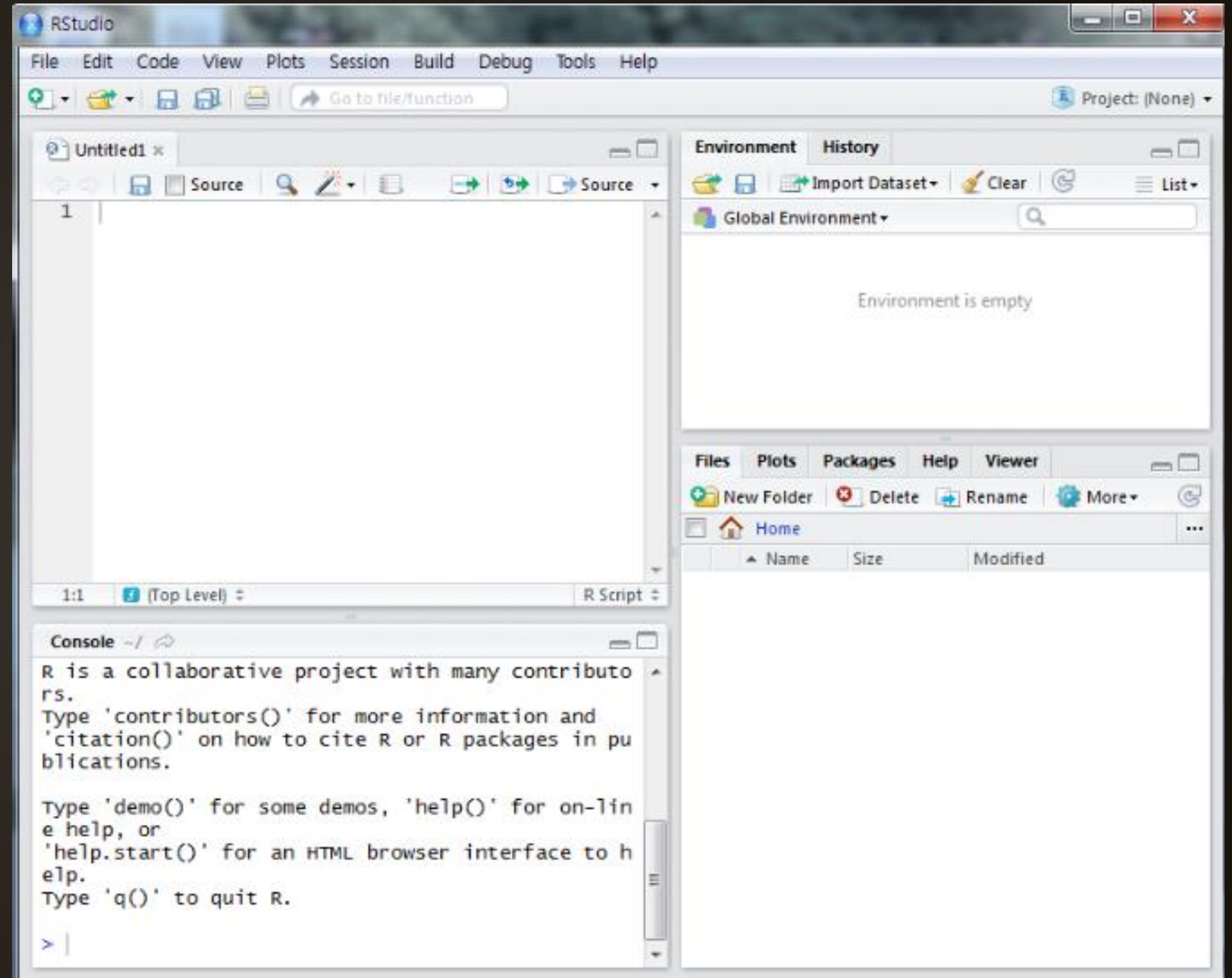
지표 생성 자동화

분류 알고리즘 적용

작업 도구



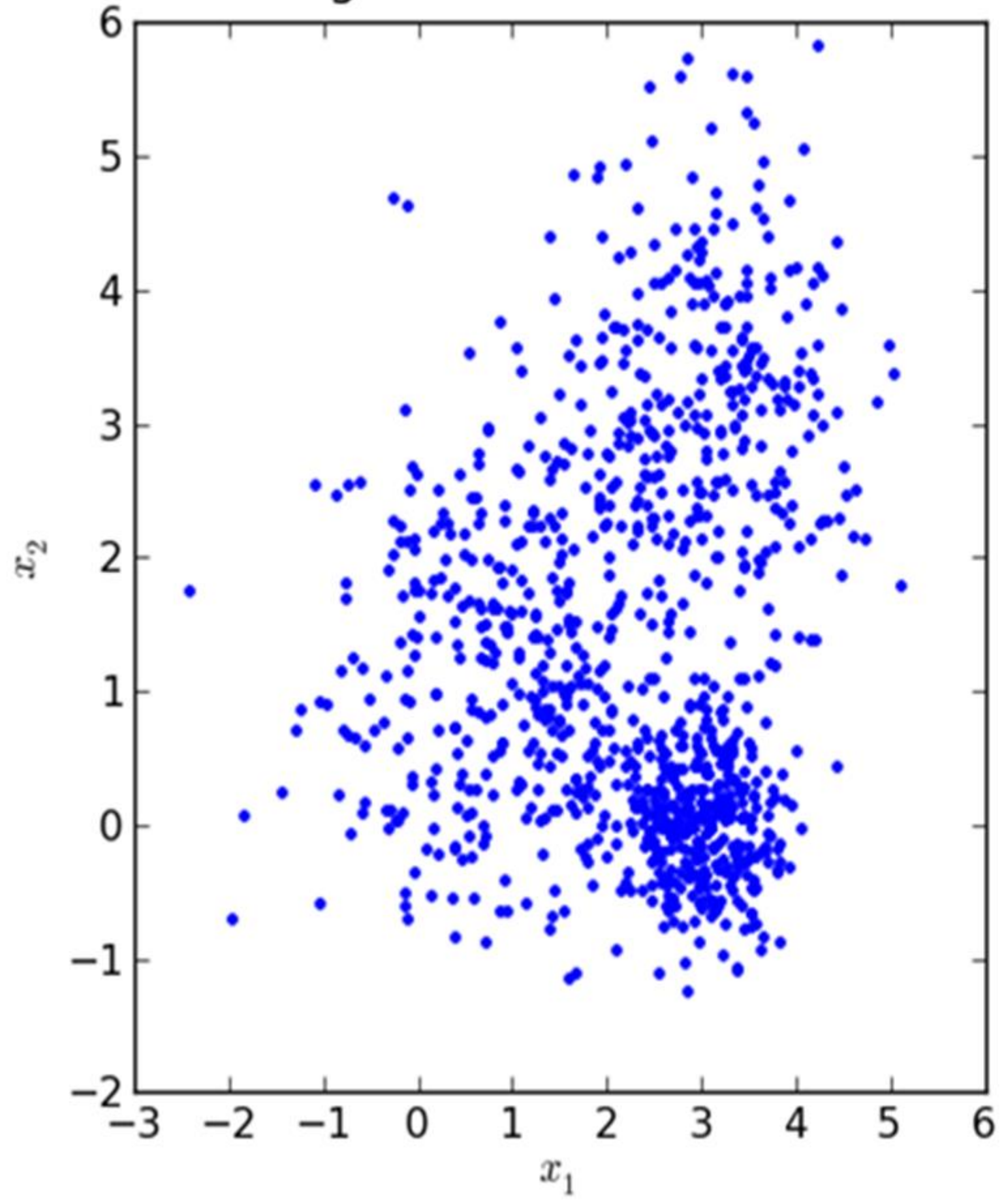
오픈소스 통계 소프트웨어



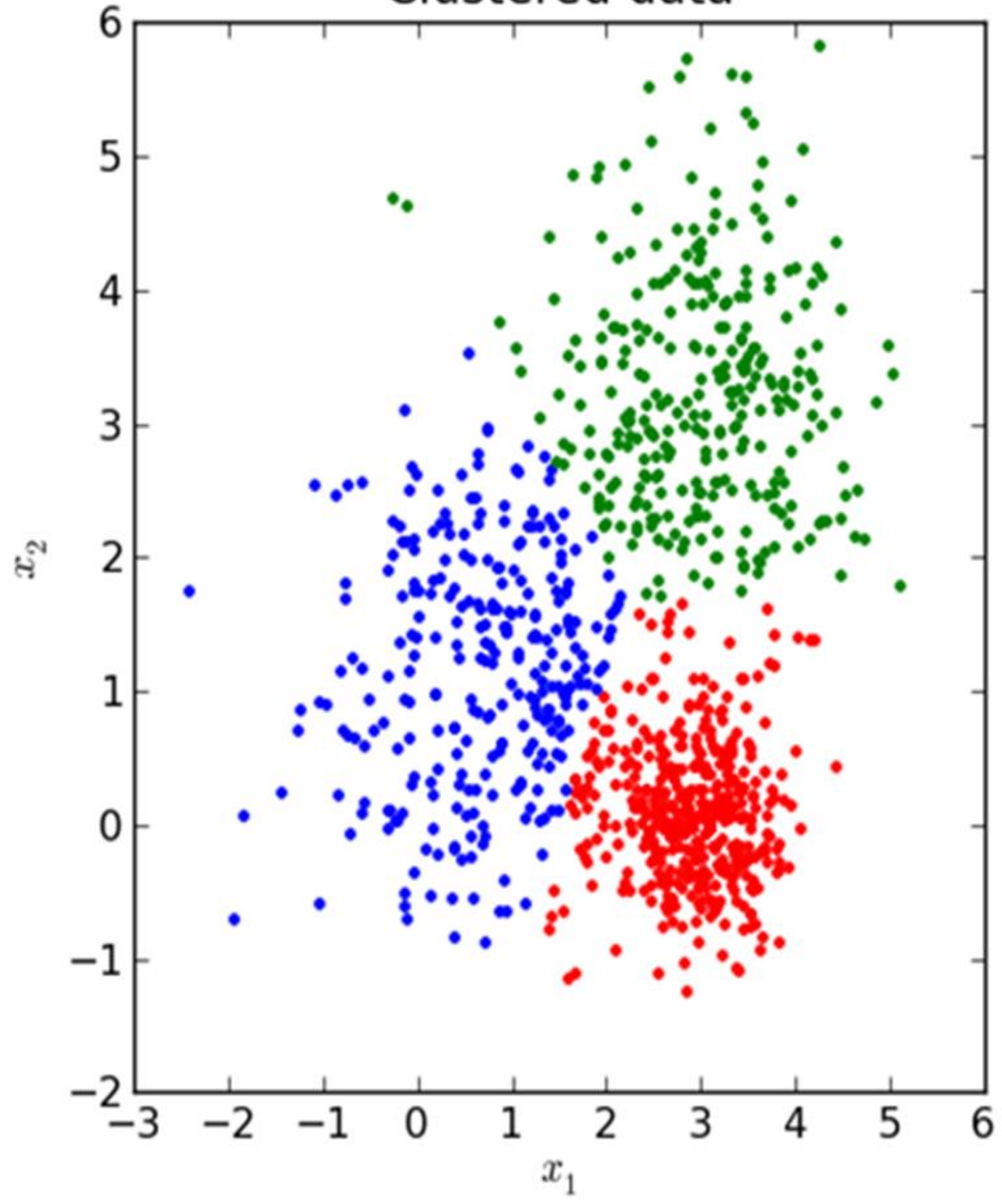
플레이 유형 파악 – 사용 기법

- K-means 클러스터링
 - 유형 => 클러스터(군집)라고 지칭
 - 데이터 비교를 통해 유사한 것들끼리 군집을 묶는 방법
 - 전체 중에 어떤 유형의 집단들이 형성되는지 확인하거나, 잠재된 집단을 발견하고자 할 때 사용

Original unclustered data

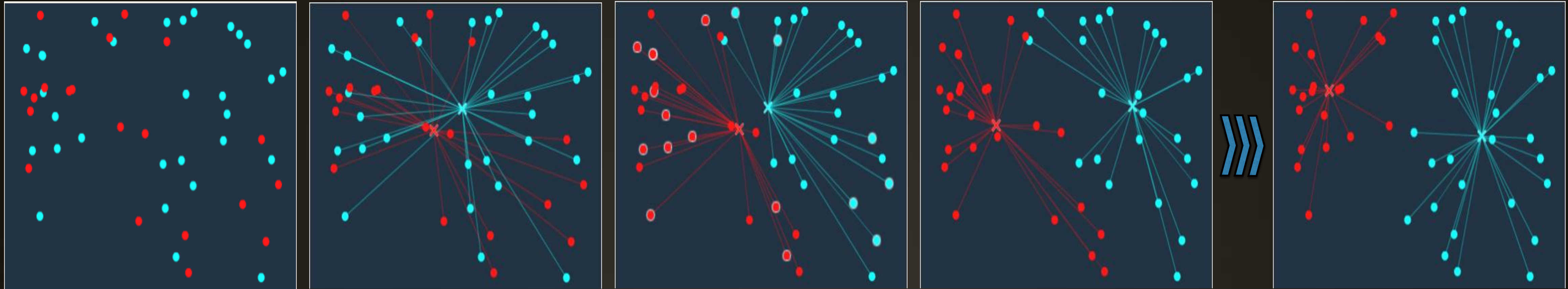


Clustered data



데이터를 이용해
3개 군집으로 분류

50개 데이터를 2개 군집으로 분류하는 경우



1. 각 점에
랜덤하게
클러스터를
분할

2. 클러스터의
중심을
계산
(x표시)

3. 점의
클러스터를
가장 가까운
중심의
클러스터로
변경

2~3 반복
(중심 변경이
더 이상 없을
때까지)

짜잔!

플레이 유형 파악 – Feature 선정

- Feature – 유저를 설명해 주는 데이터
 - 플레이타임, 사냥 관련 횟수, 경제 활동 관련 횟수 등등
 - 도메인 지식을 통해 항목 선정

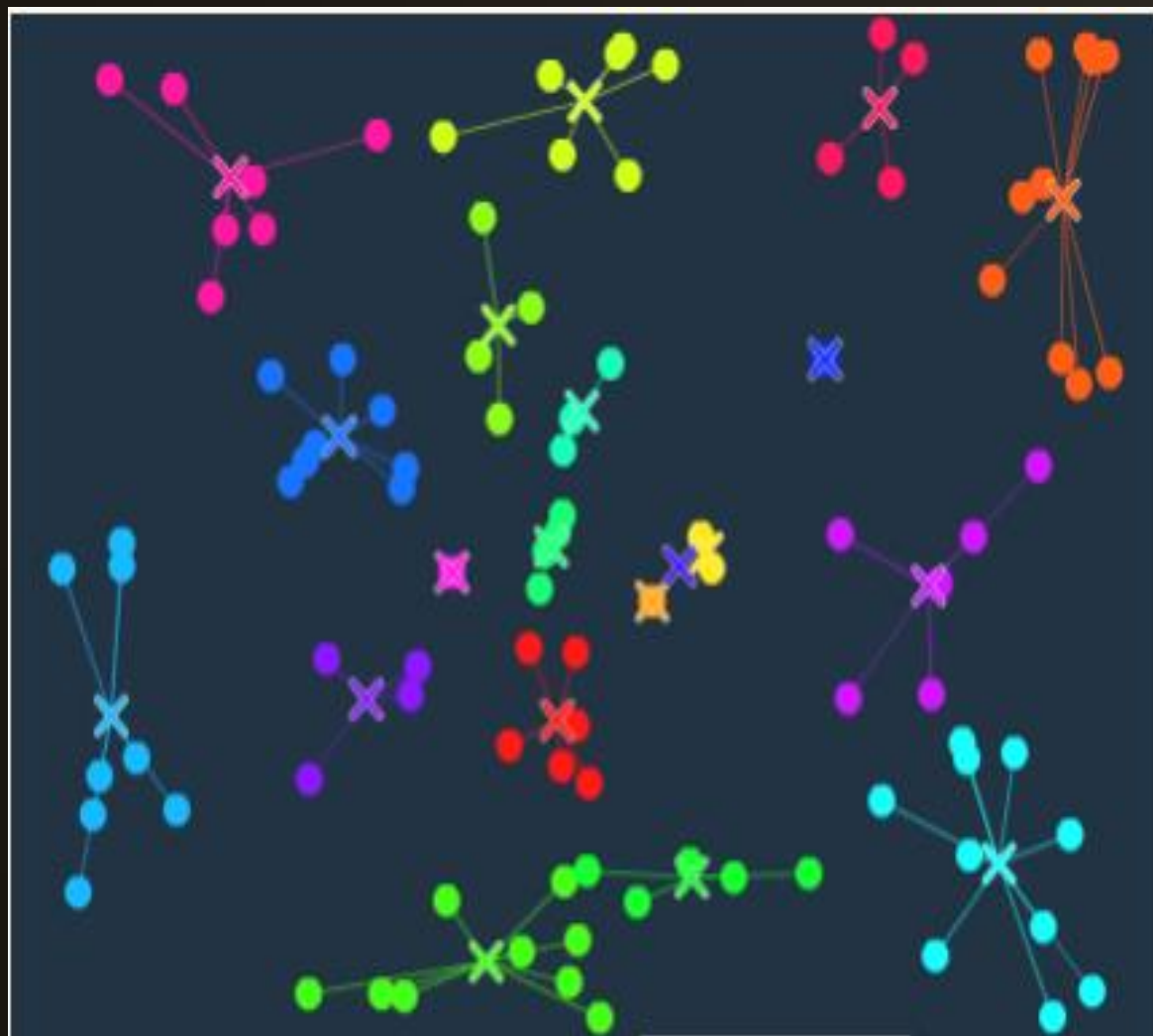
› 게임로그 데이터를 이용하여 집계 (일주일 주기)

유저	플레이타임	사냥	상점거래	등등	등등	등등
A씨	1000	450	10
B씨	47	20	10

› 유형 확인 및 반복 작업을 통해 Feature 조정

플레이 유형 파악 – 유형 개수 선정

- 유형 개수 – 몇 개의 군집으로 분류할 것인가
 - 사용 목적에 따라 개수 조절 가능

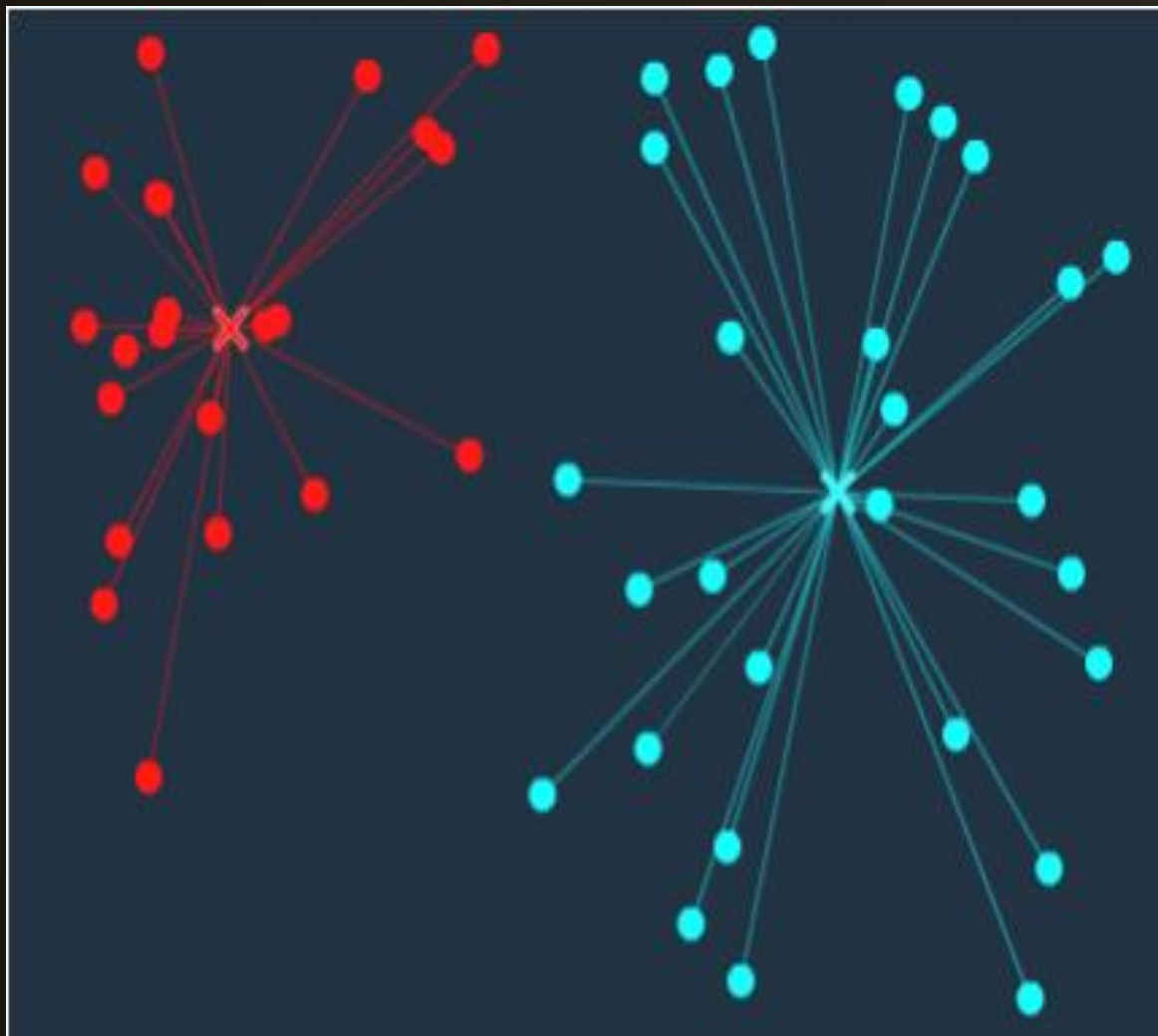


개수가 많으면?

- (장점) 세분화 가능 – 다양한 유형 확인
- (단점) 특징 확인이 번거로움
- (단점) 노이즈. 주별 결과가 항상 같지 않을 수도...

플레이 유형 파악 - 유형 개수 선정

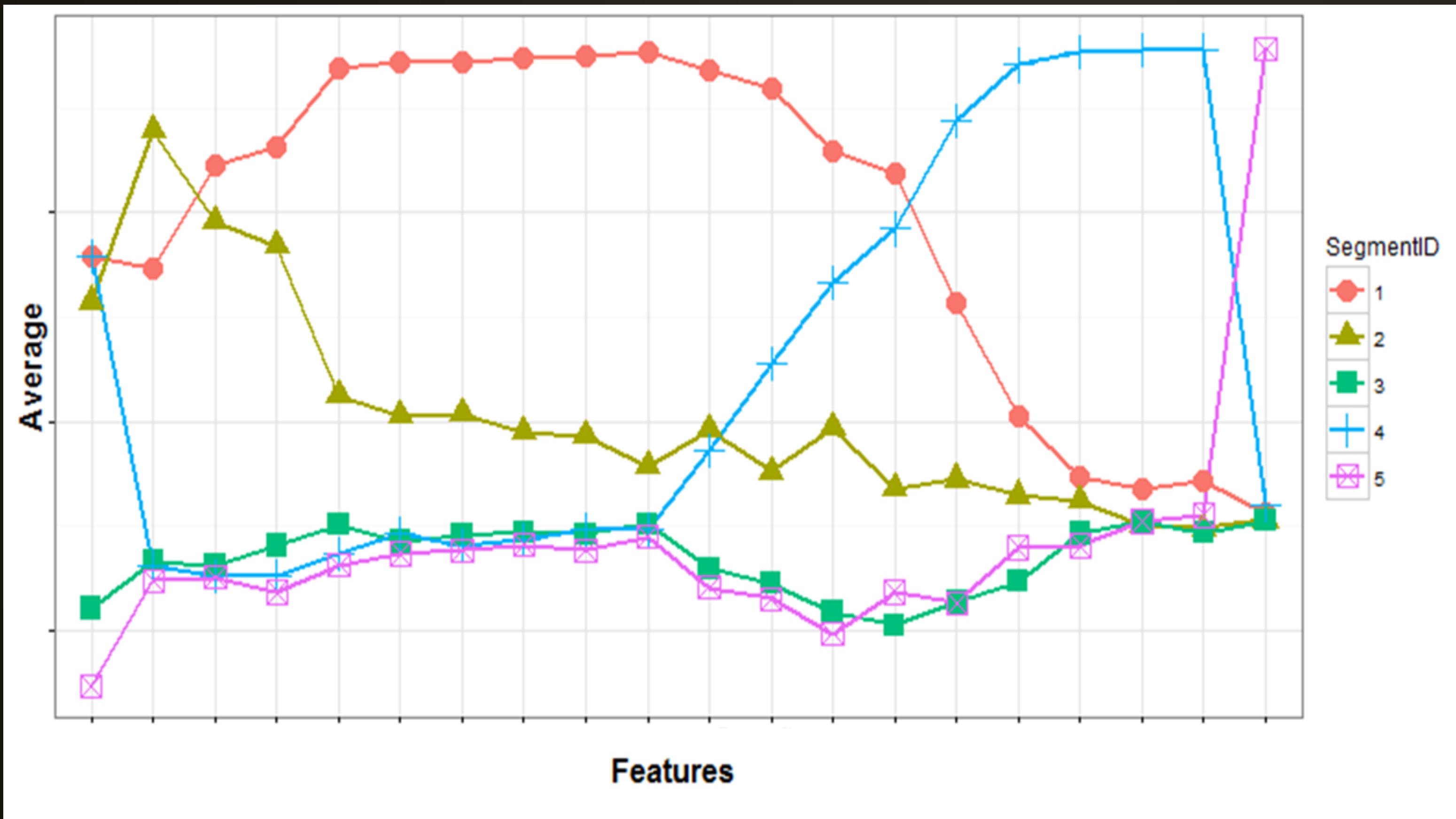
- 유형 개수 - 몇 개의 군집으로 분류할 것인가
- 사용 목적에 따라 개수 조절 가능



개수가 적으면?

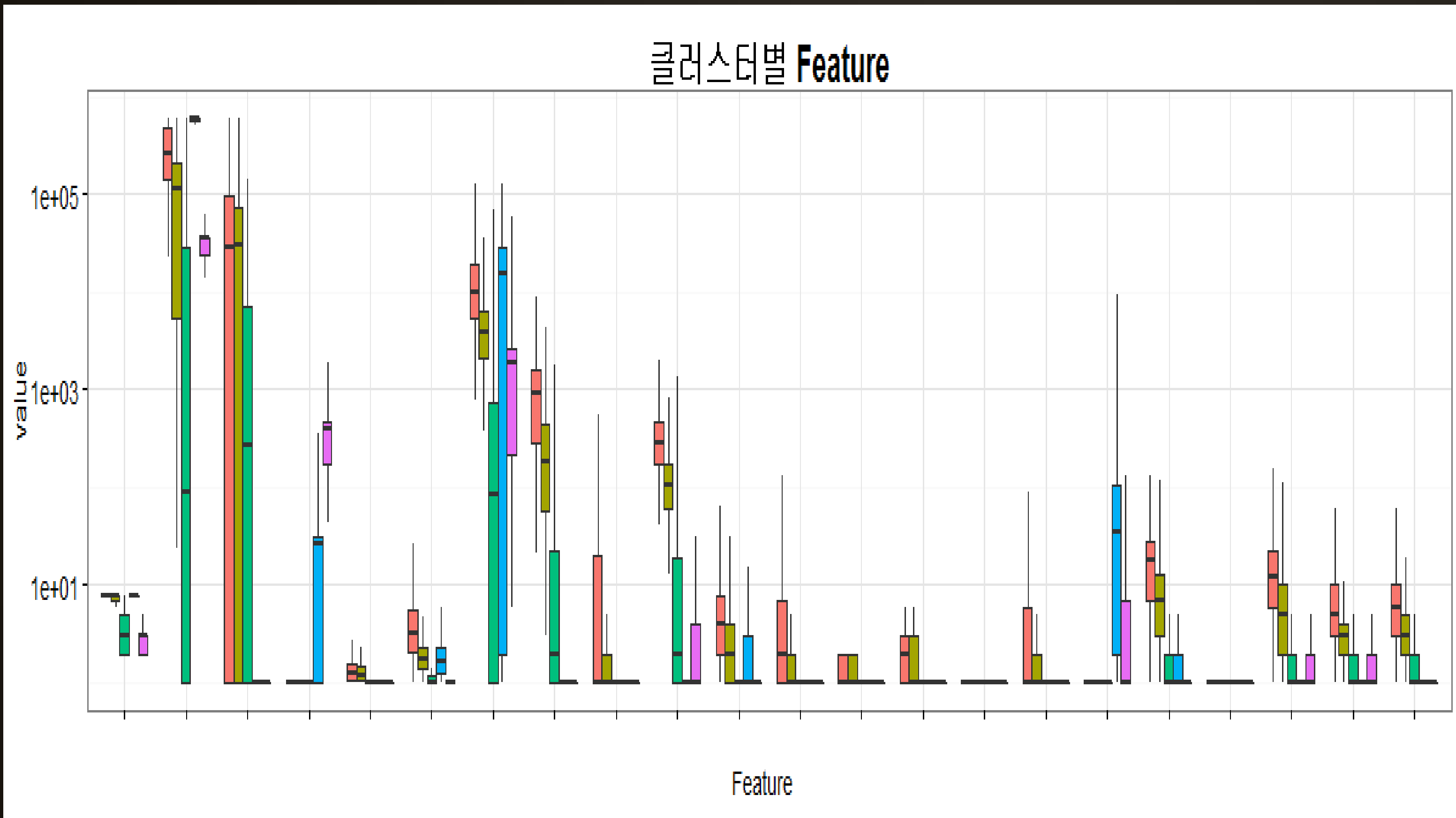
- (장점) 특징 확인 및 결과 유지의 편의성
- (단점) 다양한 유형이 통합되어 있을 수 있음

유형 태깅 및 진성 유저 그룹 선정 - 그룹별 플레이 특징 분석



각 그룹별
Feature별
평균값 확인

유형 태깅 및 진성 유저 그룹 선정 - 그룹별 플레이 특징 분석

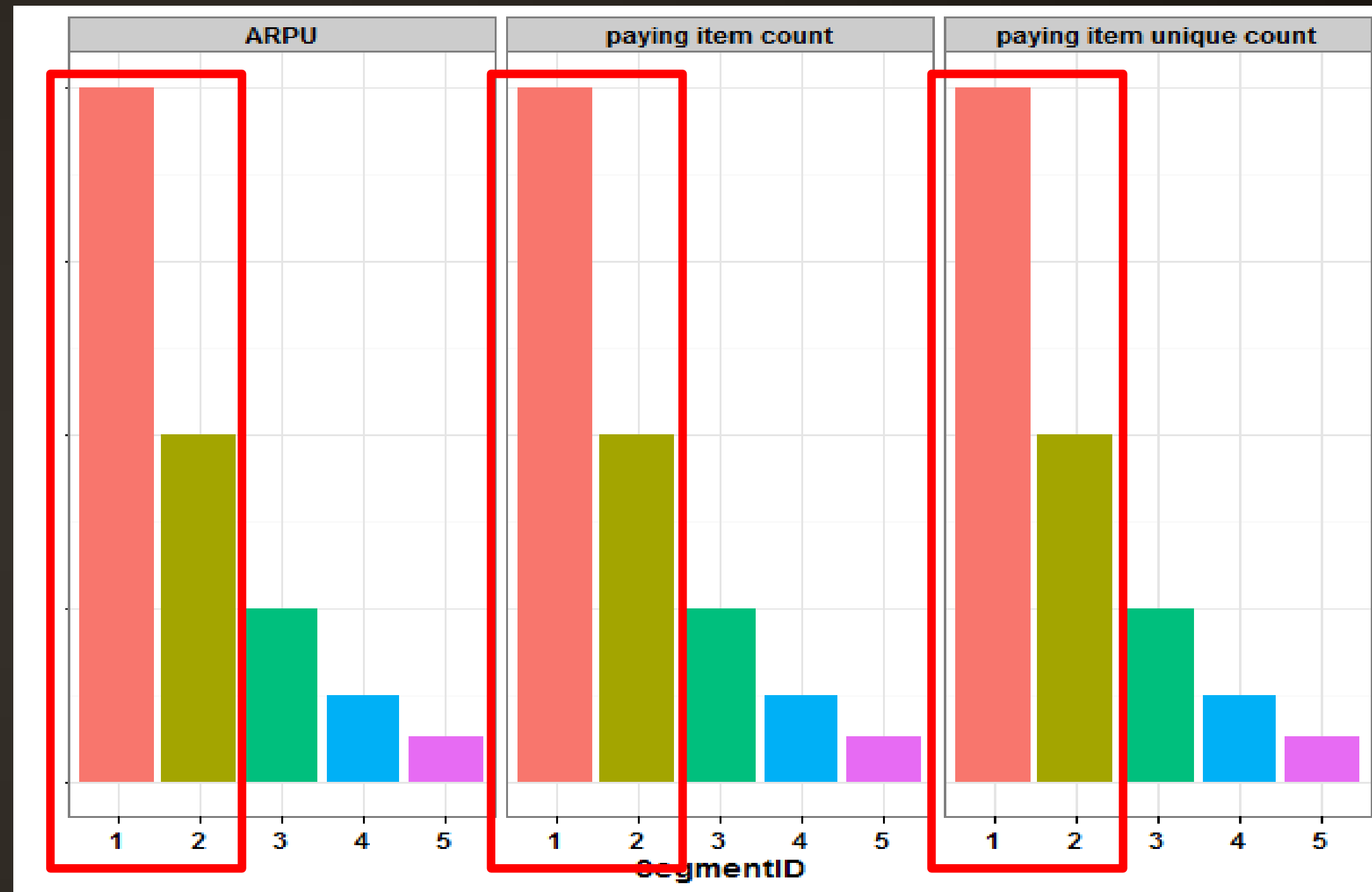
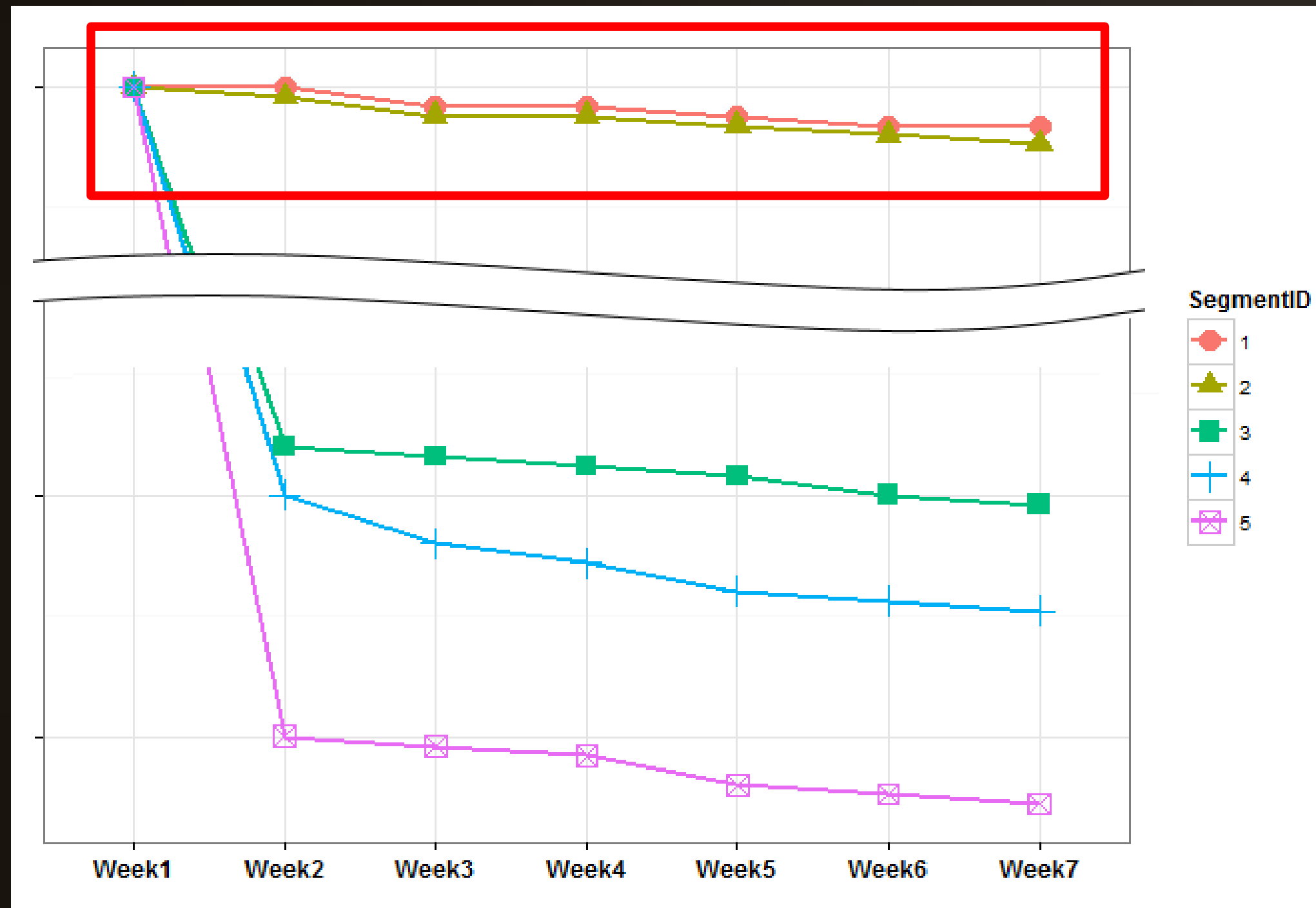


각 그룹별
Feature별
분포 확인

유형 태깅 및 진성 유저 그룹 선정 - 그룹별 플레이 특징

- 1번: 다양한 콘텐츠에 대한 높은 사용량
- 2번: 중반 정도의 고른 수치 분포
- 3번: 전반적인 콘텐츠 이용률 매우 낮음
- 4번: 높은 플레이시간, 상업 활동 위주
- 5번: 무료 이용권으로 접속, 재화 획득 목적

유형 태깅 및 진성 유저 그룹 선정 - 그룹별 잔존율 및 결제



1번, 2번 그룹이 잔존율과 결제 모두 높다!

유형 태깅 및 진성 유저 그룹 선정

- 1번: 열성형
- 2번: 일반형
- 3번: Light형
- 4번: Gold farming형
- 5번: Cherry picker형

지표 생성 자동화



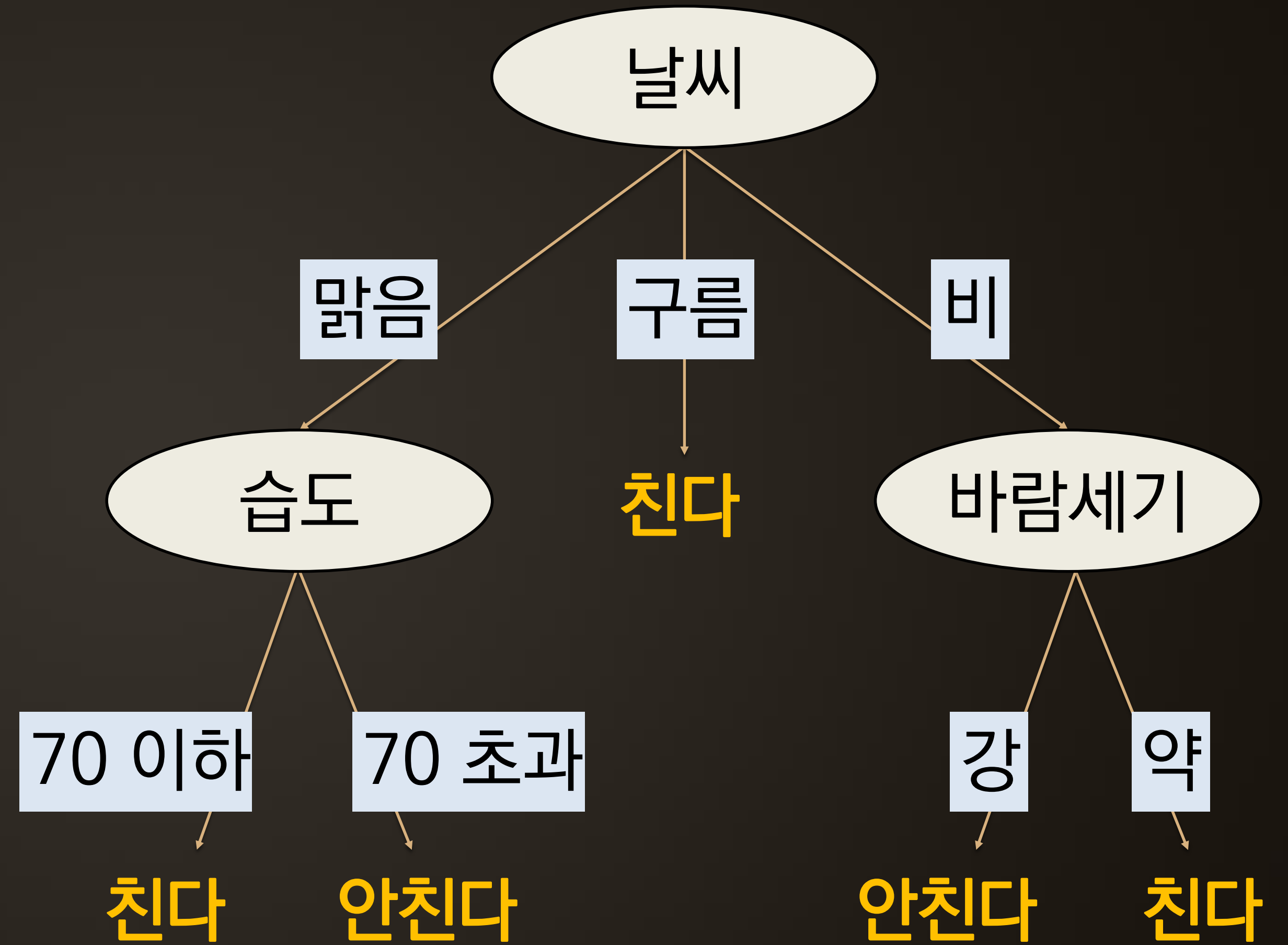
이걸 매번
일일이
돌릴 수는 없고...

지표 생성 자동화 – 사용 기법

- Random Forest
 - 여러 개 Decision Tree를 융합하여 사용하는 분류 알고리즘
 - 높은 정확도 대비 낮은 과적합 (Over Fitting) <- 안정적인 성능 보장

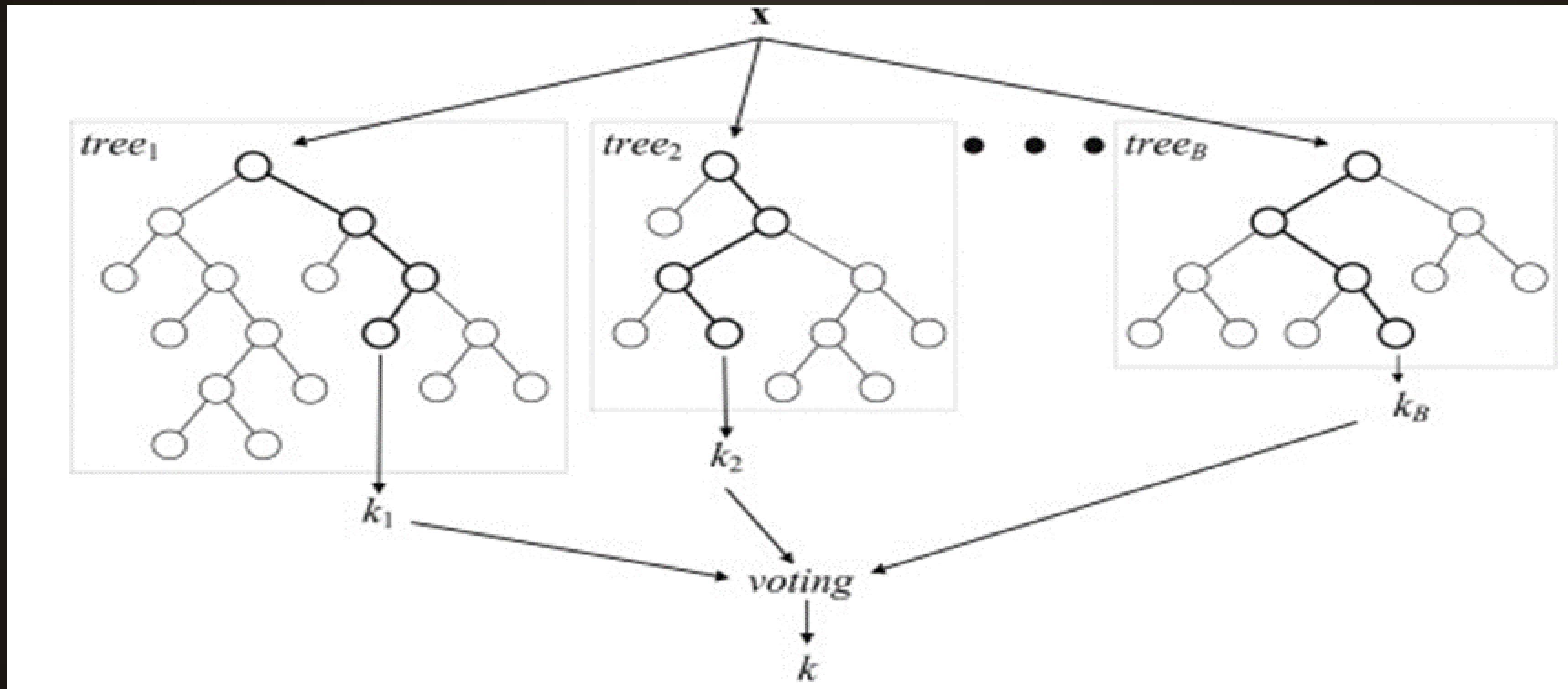
데이터				결과
날씨	기온	습도	바람세기	골프를 치는가
맑음	22	95	약	안친다
맑음	21	70	약	친다
맑음	24	70	강	친다
맑음	29	85	약	안친다
맑음	27	90	강	안친다
구름	27	75	약	친다
구름	28	78	약	친다
구름	22	90	강	친다
구름	18	65	강	친다
비	22	80	강	안친다
비	21	96	약	친다
비	20	80	약	친다
비	18	70	강	안친다
비	24	80	약	친다

골프 예측에 관한 Decision Tree

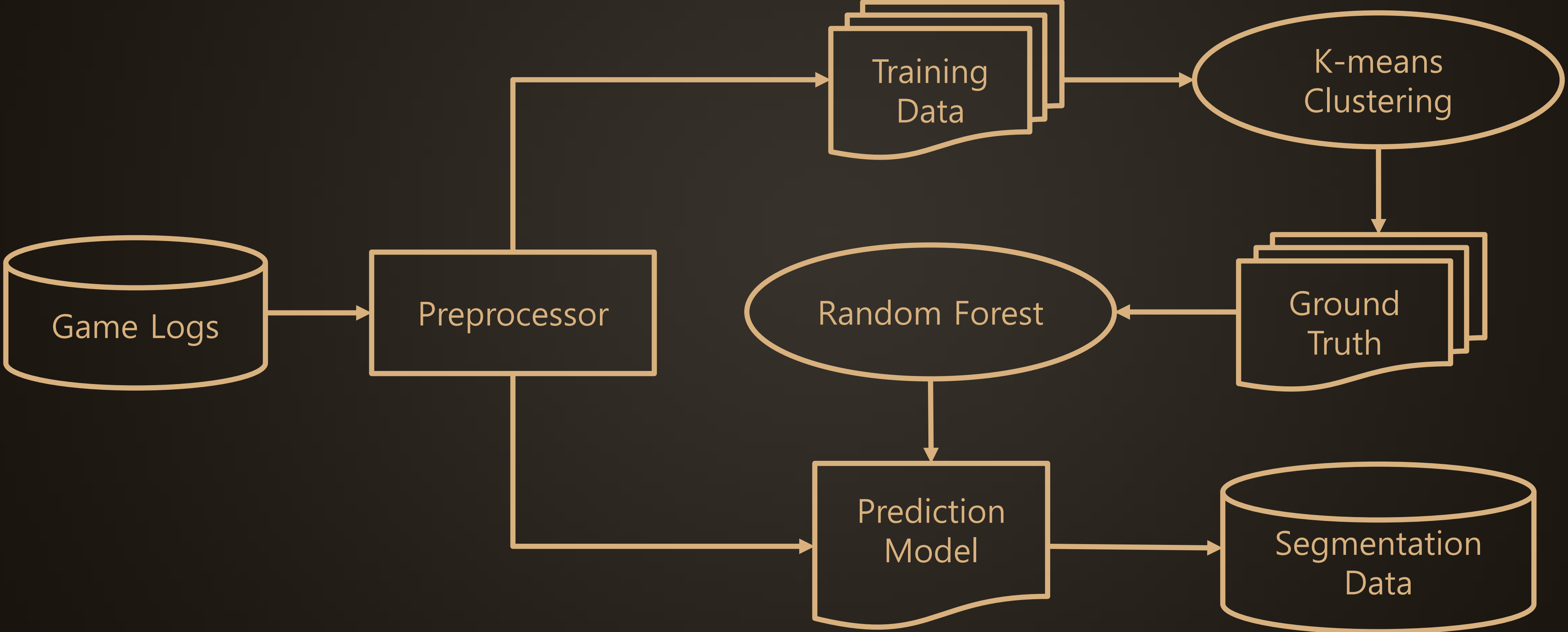


지표 생성 자동화 - 사용 기법

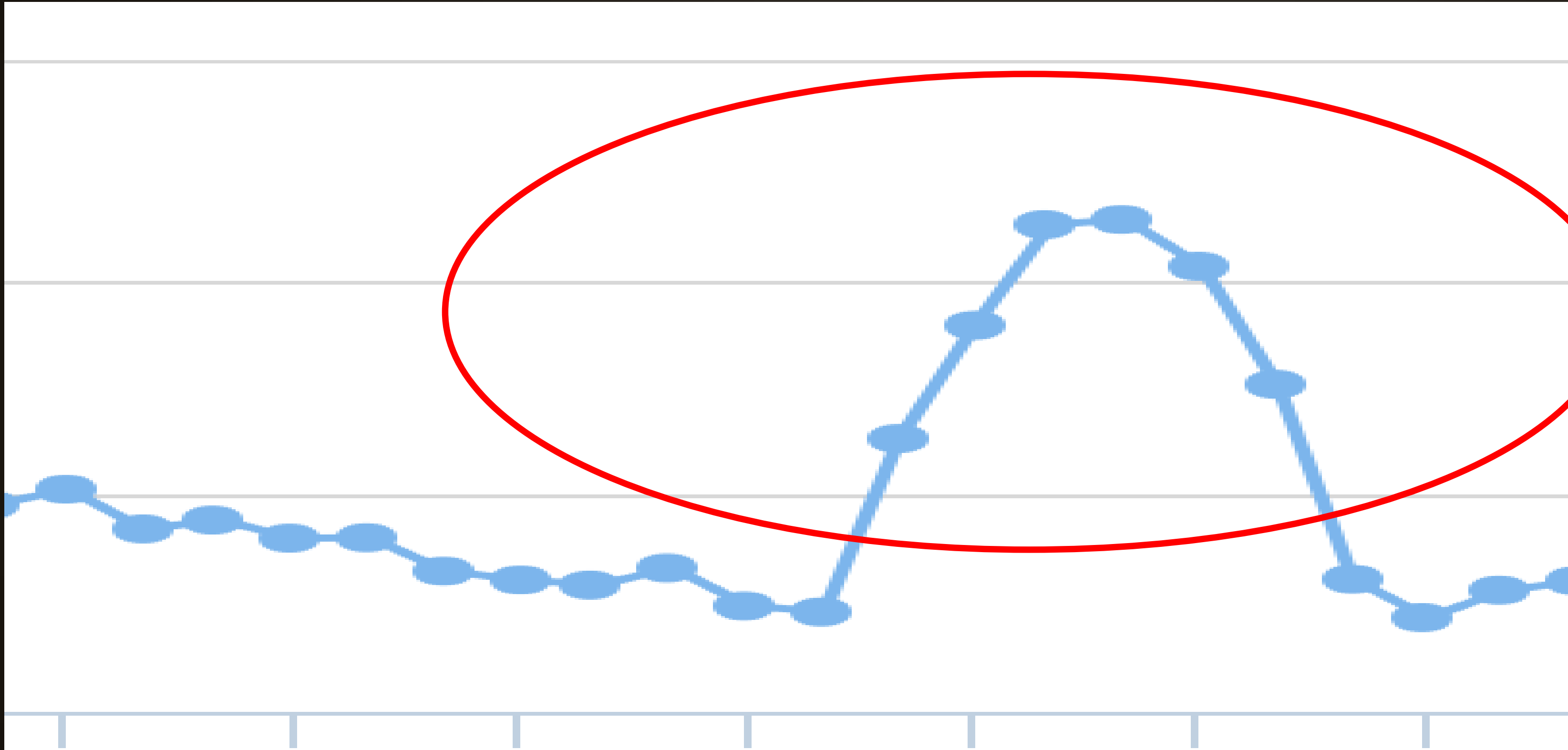
- Random Forest



지표 생성 자동화 - 시스템 구조

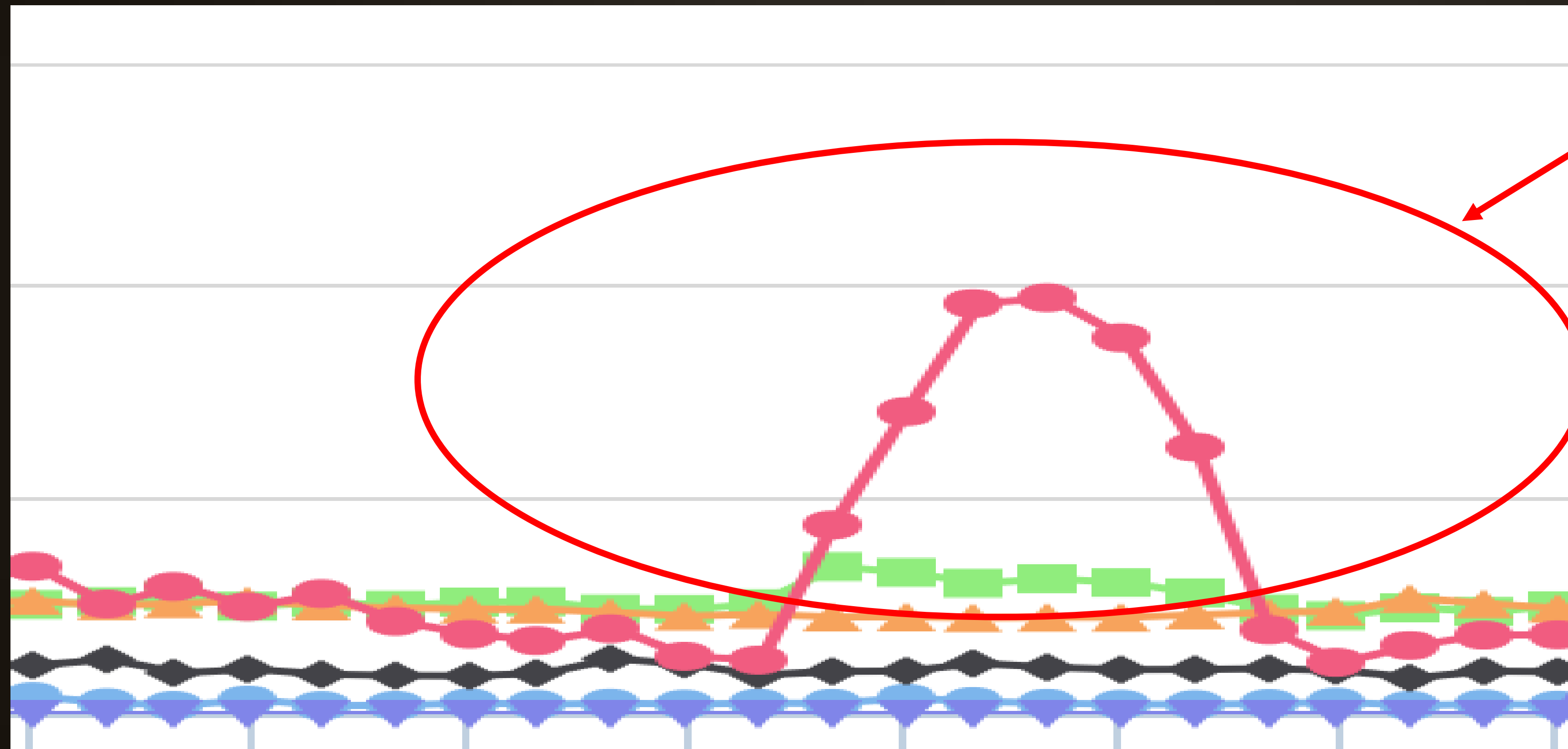


지표 활용 예



AU가 늘었다
줄었다
왜?

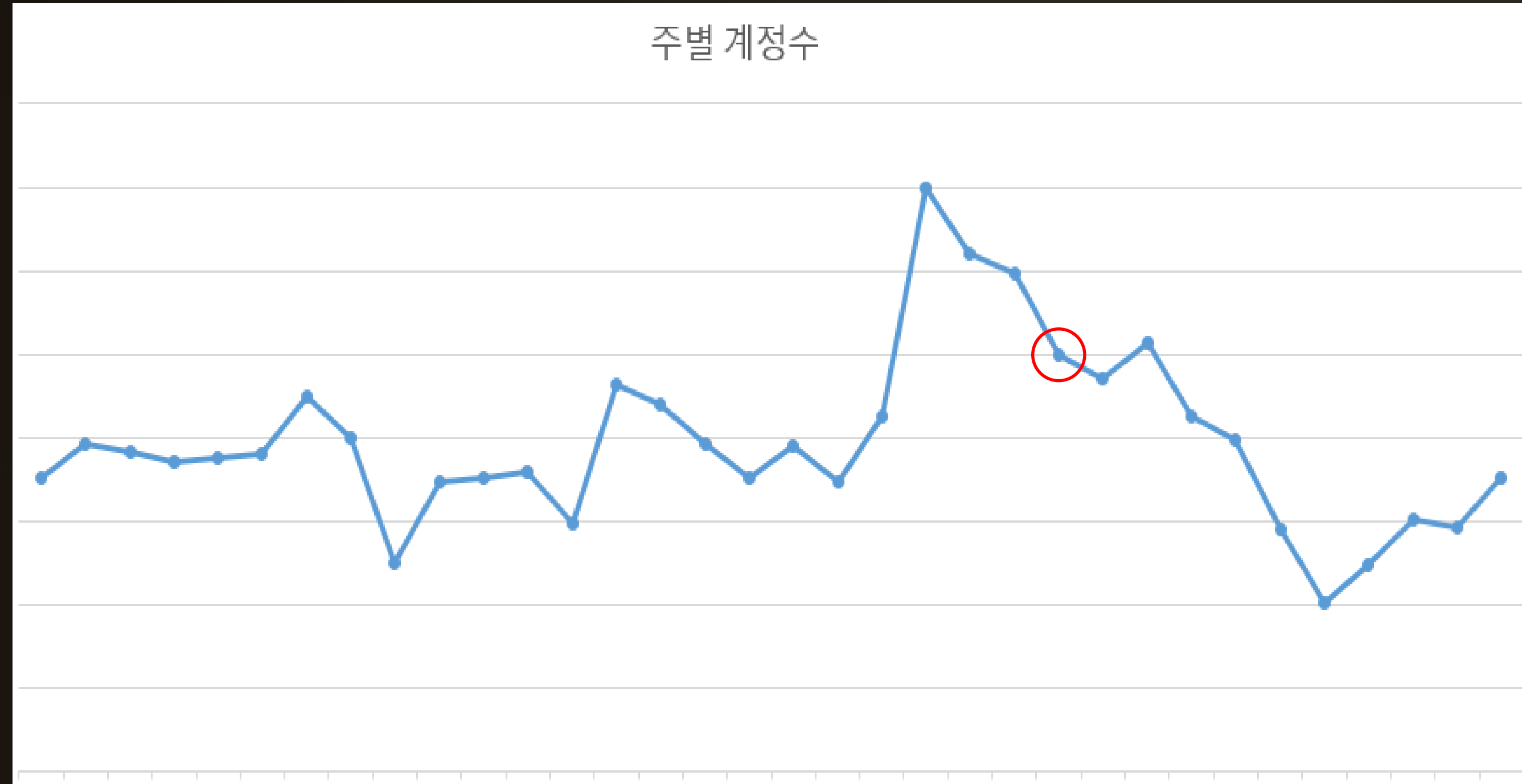
지표 활용 예



Cherry picker

지표 기준선 추정

○ 시점의 AU는 많은 걸까, 적은 걸까?

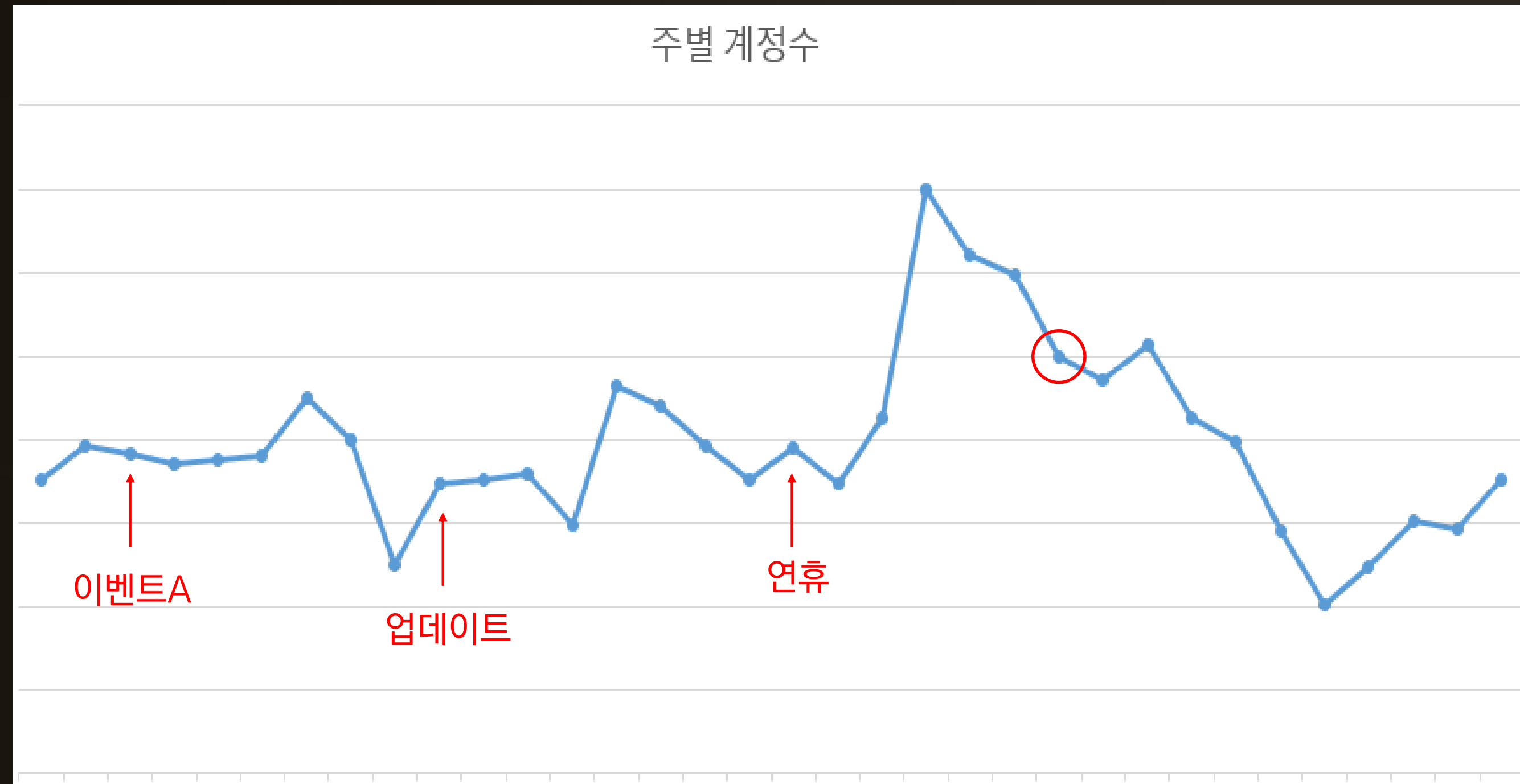


적다
(지난주보다 떨어졌다)

VS

많다
(한 달 전보다 올랐다)

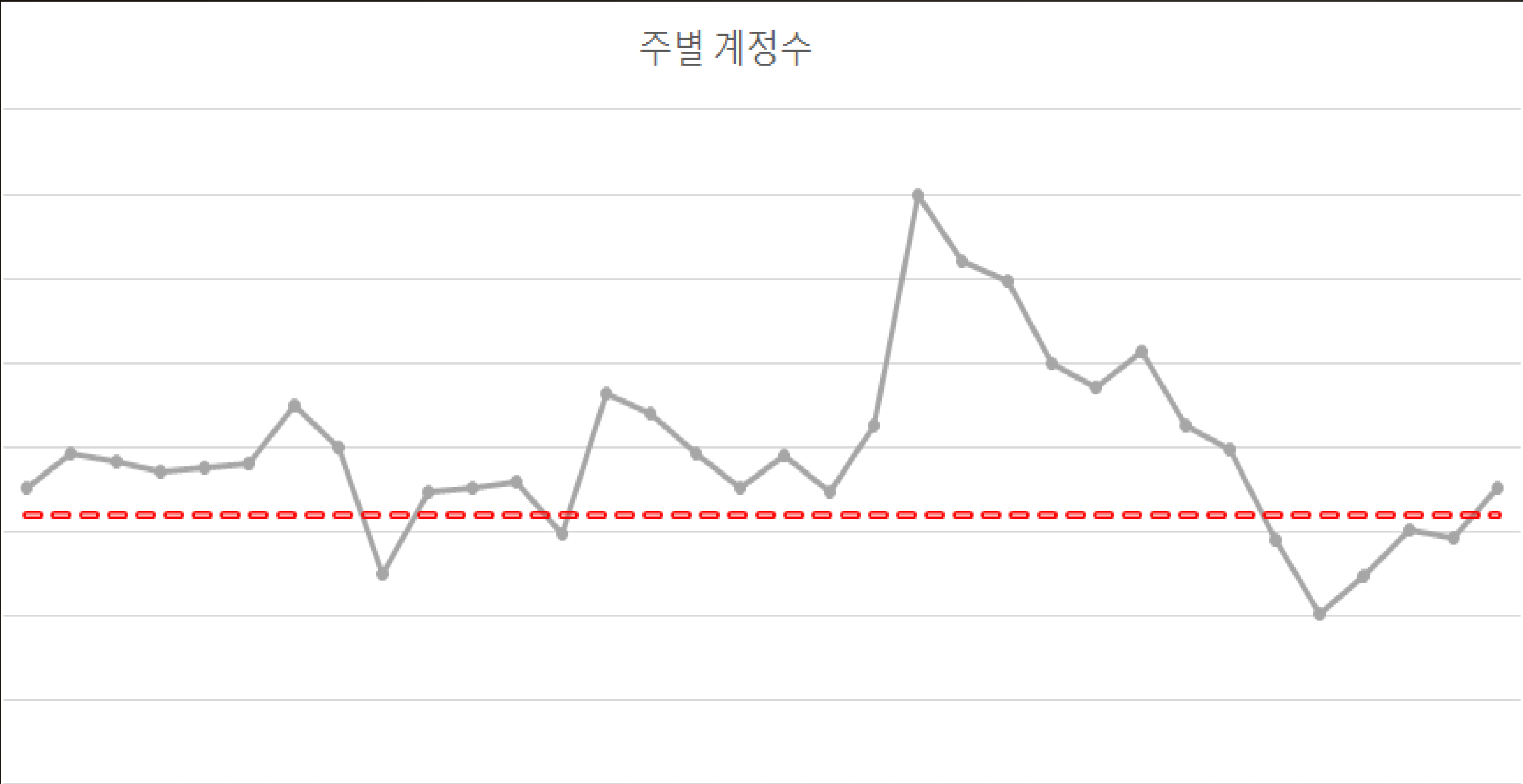
다양한 프로모션 및 이벤트로 인한 AU 변화



계절, 휴일
이벤트, 업데이트...

지표 변화에 영향을 주는
요인들로 인해
판단이 어려울 수 있다

지표 기준선



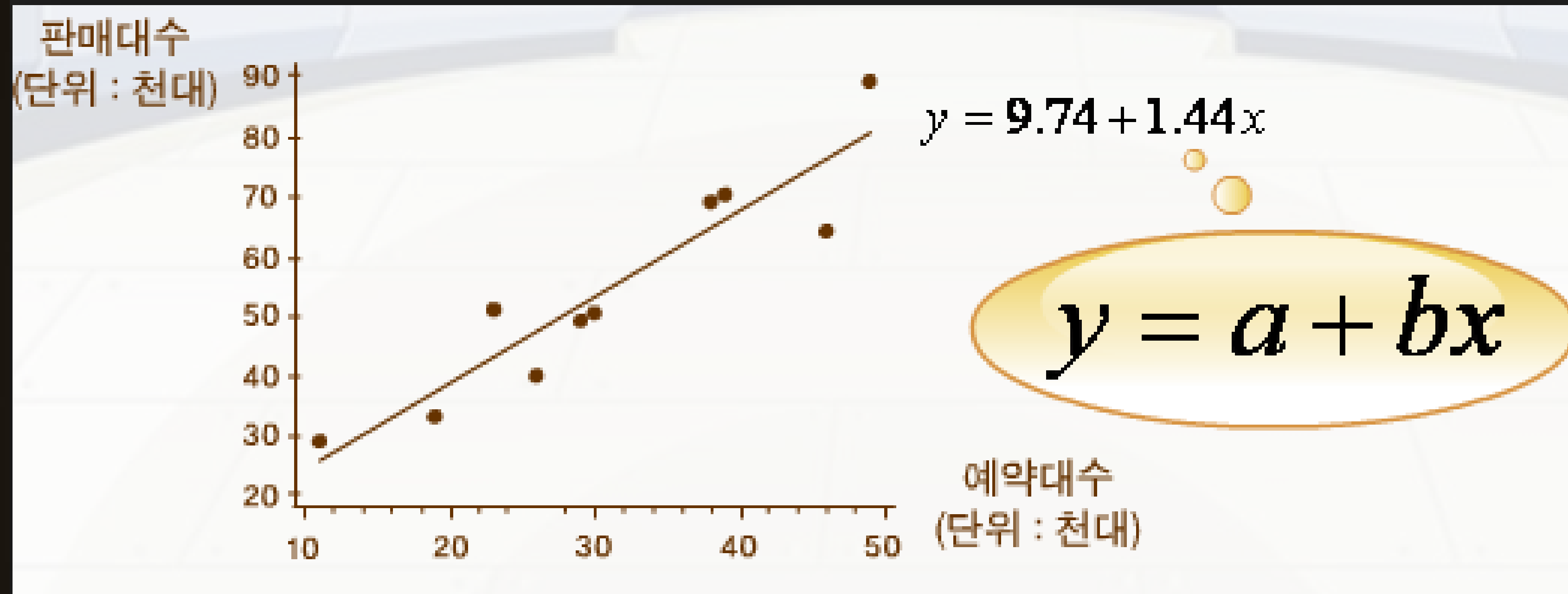
지표 변화에 영향을 주는
요인을 제외한 값을
찾아보자

지표 기준선 – 사용 기법

- 회귀 분석

- 한 종속변수가 다른 독립 변수들에 의해서 어떻게 설명 또는 예측되는지 알아보기 위해 자료를 적절한 함수식으로 표현하여 분석하는 통계적 방법

에어컨 예약 대수가 한 단위 올랐을 때 판매대수는 얼마나 바뀔까?



- a (Intercept): x라는 요인을 통제했을 때의 평균적인 y값
- b (회귀계수): x라는 요인이 한 단위 오를 때 y에 영향을 미치는 값
→ 예약 대수가 한 단위 오를 때 판매대수는 1.44 증가

지표 기준선

- 진성 유저 지표의 각 세그먼트별 AU, 계정 비율, ARPU 추출
- 프로모션, 업데이트, 이벤트 등의 이슈 사항 정리
- 진성 유저 지표의 각 세그먼트별 AU, 계정 비율, ARPU에 대해 회귀 분석
 - 지표 기준선을 찾고, 프로모션/이벤트 등의 영향도를 파악

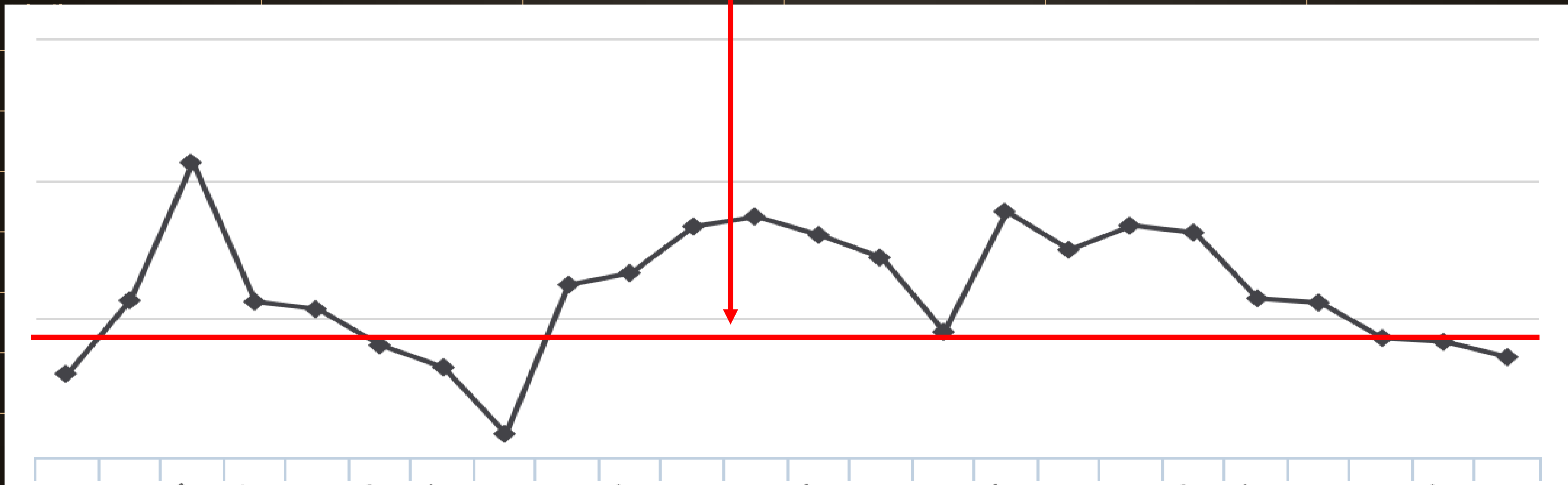
지표 기준선 결과 - 세그먼트별 ARPU

	1번 그룹	2번 그룹	3번 그룹	4번 그룹	5번 그룹
Intercept	1	0.626	0.215	0.197	0.062
이벤트 1	0.41	-	-	-	-
이벤트 2	0.99	0.84	0.85	0.21	-
이벤트 3	-	-	-	-	0.75
이벤트 4	-	-	-	-	-0.56
이벤트 5	-	-	-	-	-
이벤트 6	-	-	-	-	-
이벤트 7	0.54	0.56	0.50	-	-
이벤트 8	-	-	-	-	-0.47

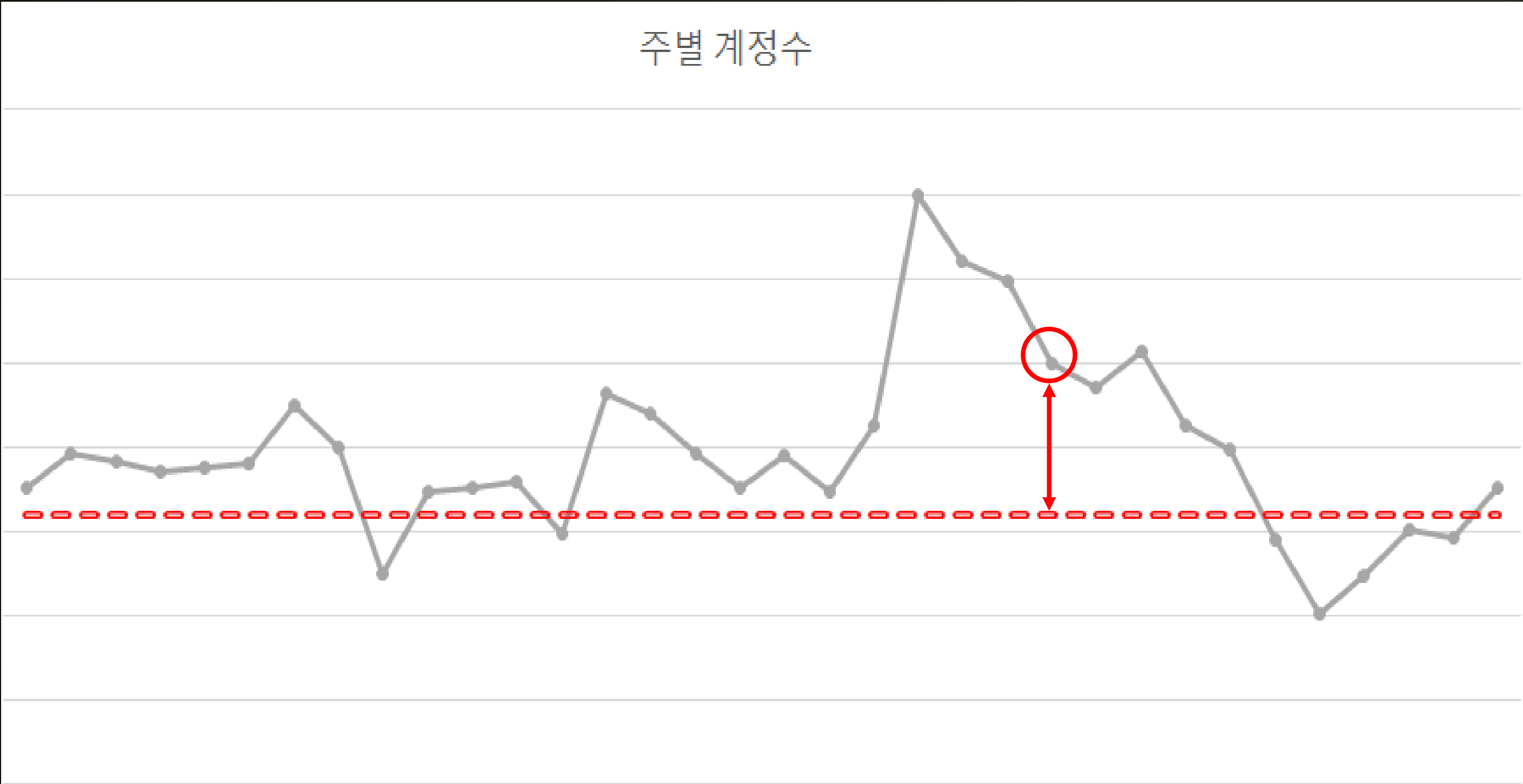
(통계적으로 유의미하지 않은 경우 - 표시)

지표 기준선 결과 - 세그먼트별 ARPU

	1번 그룹	2번 그룹	3번 그룹	4번 그룹	5번 그룹
Intercept	1	0.626	0.215	0.197	0.062



지표 기준선

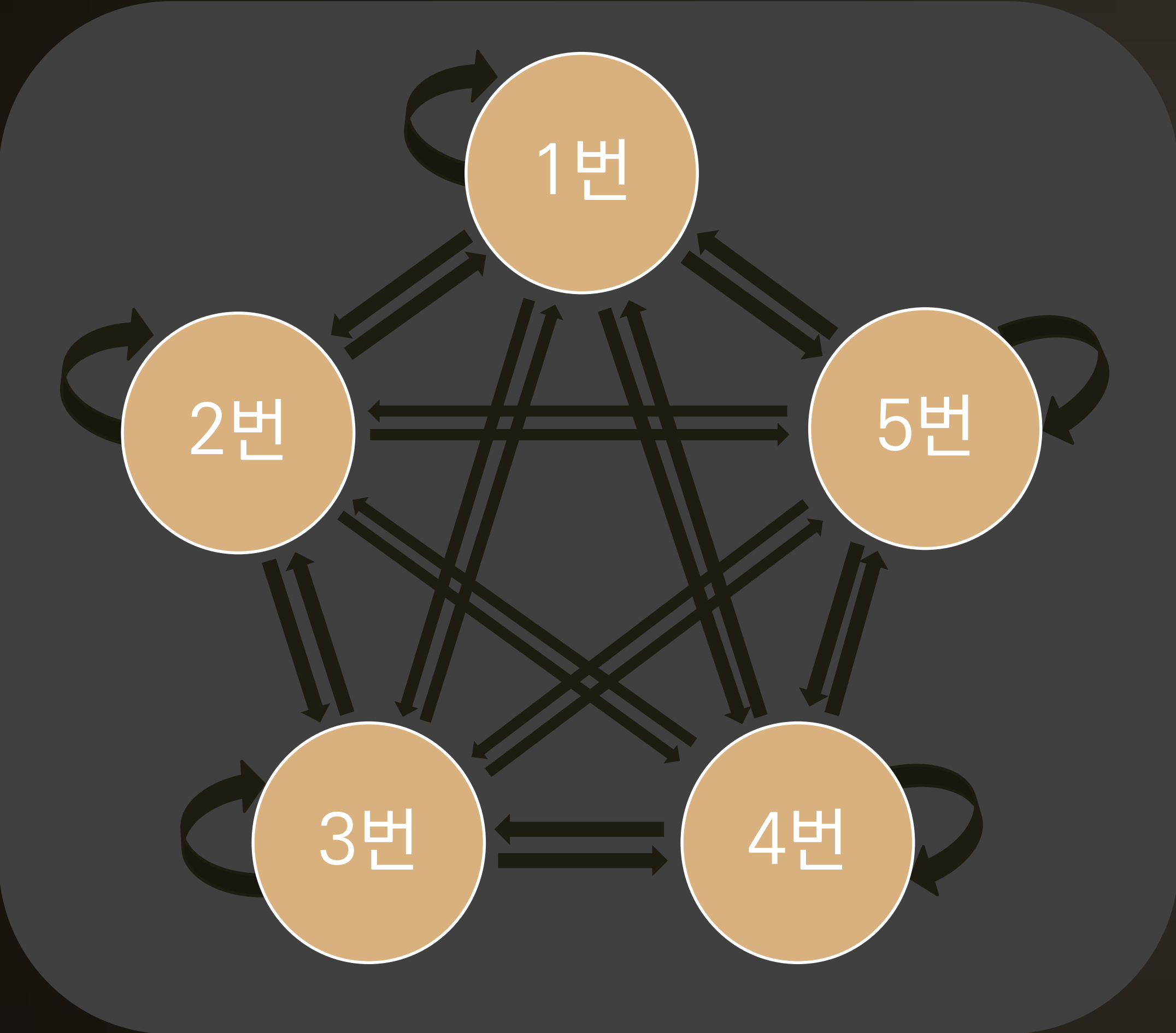


기준선과 비교



유형을 나누어 추적해 보면,
유저의 변화를 파악하는 것도
가능하다

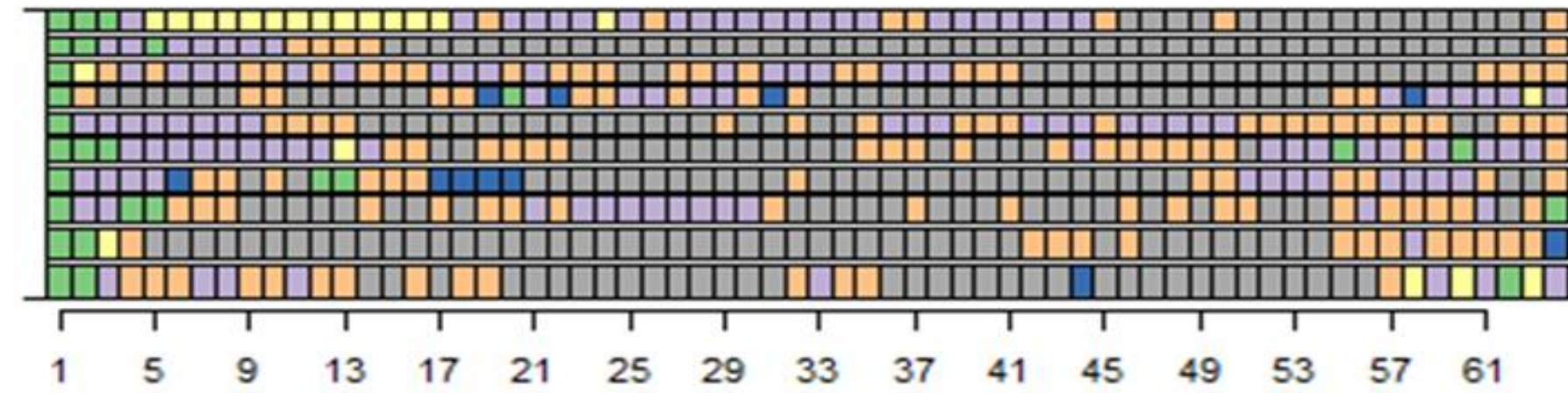
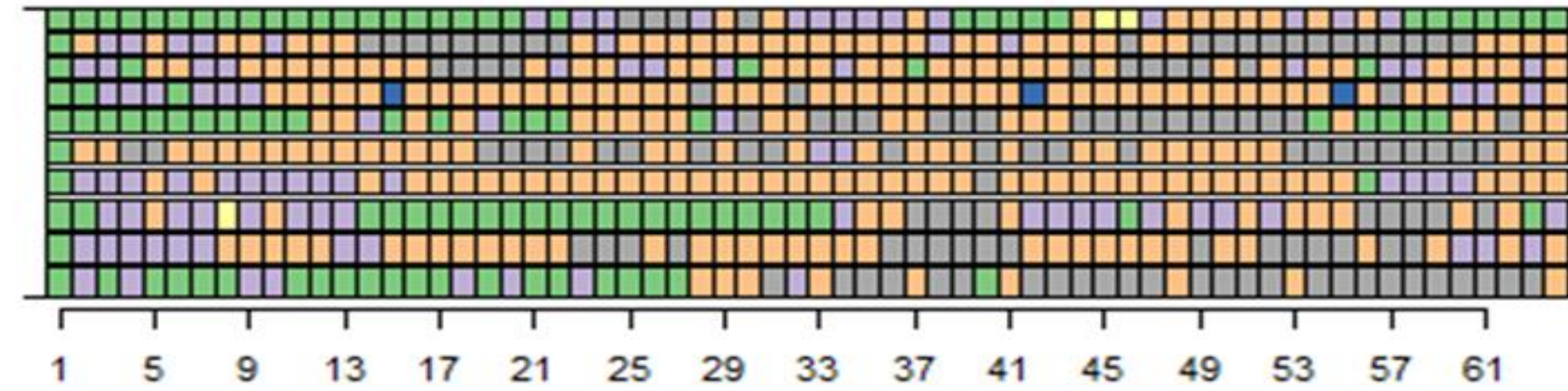
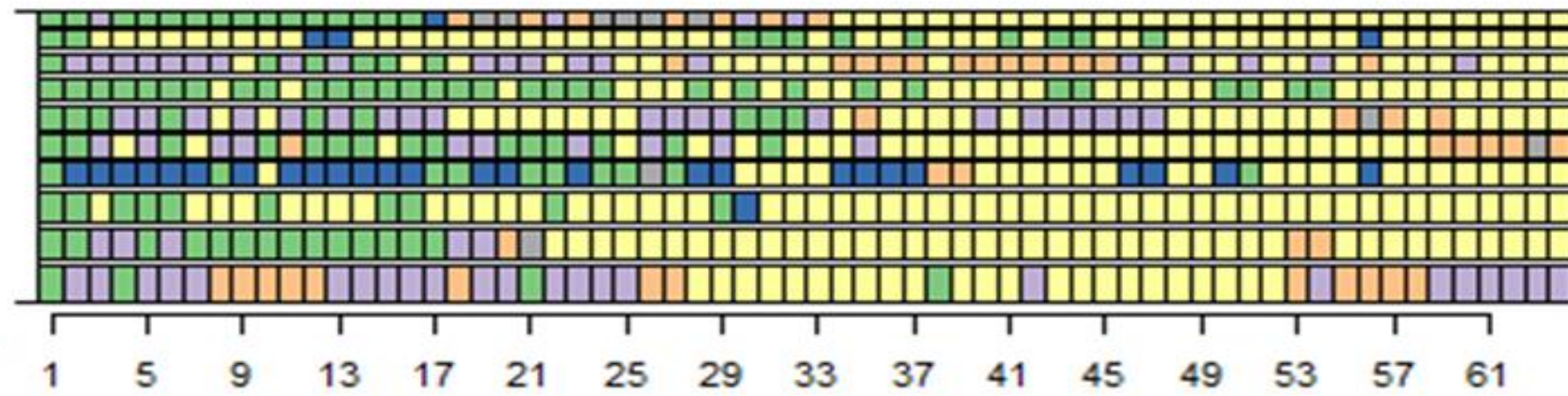
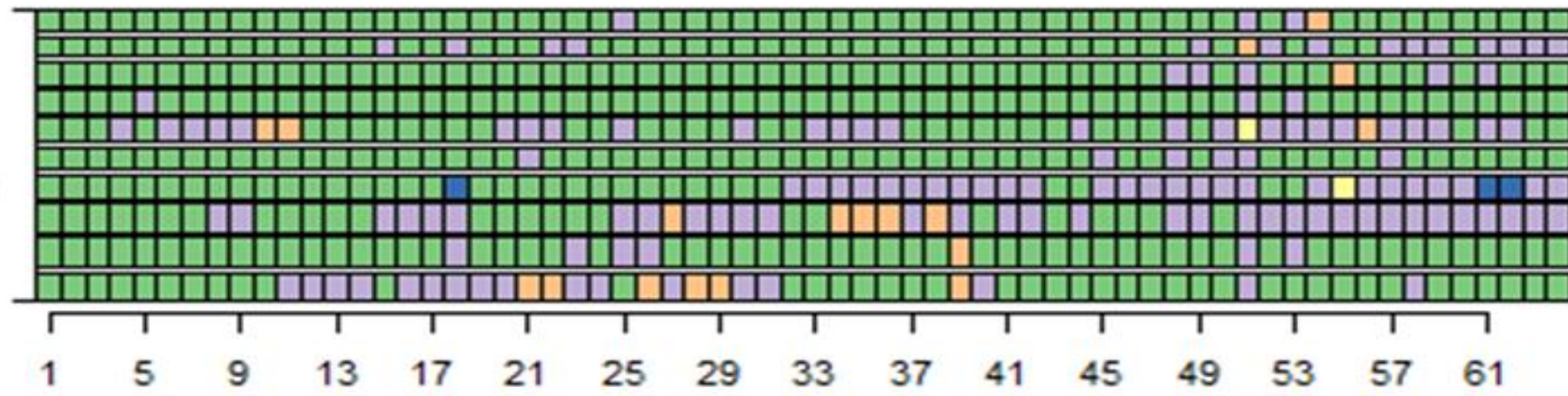
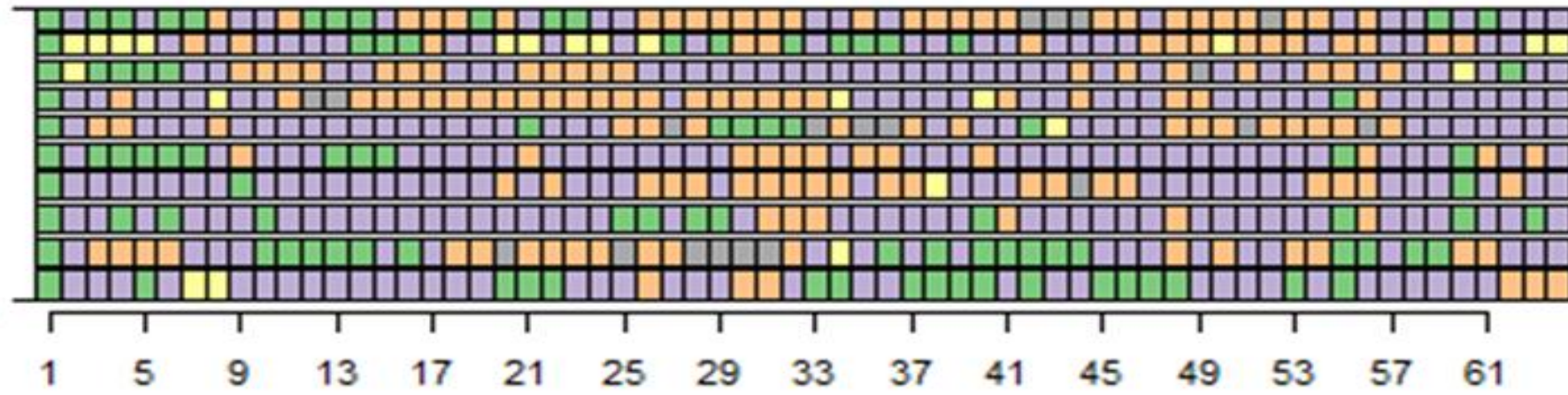
유형별 전이 확률 측정



매주 각 번호에서 이동하는 비율을 추적

유형별 변화 시퀀스 분석

sequence frequency plot



마치며

기존의 AU 지표에서 좀 더 나아가서 유형별로 유저 동향을 파악하면,
현재 게임 세계의 흐름을 감지할 수 있다

유형별 기준선을 찾고, 외부 영향도를 파악할 수 있다

세그먼트별로 어디서 들어오고 나가는지,
유입/유출을 추적하면, 유저의 변화를 파악할 수 있다

작업은 R을 이용하면 생각보다 간단하다 (데이터만 준비하자)

이제는 지표를 활용할 때다



