

회귀 모델 기반의 게임 재화 이상 증감 탐지 룰 개발

엔씨소프트 분석모델링팀

엄혜민

Overview

오늘 소개 드리고자 하는 내용

- 운영자가 쉽게 파악하기 힘든 버그로 인한 **게임 재화 이상 증감 현상**을
- **회귀 분석 기법**을 이용하여
- **조기에 자동으로 탐지**하는 기술에 대해



이야기 해 보려 합니다.

목차

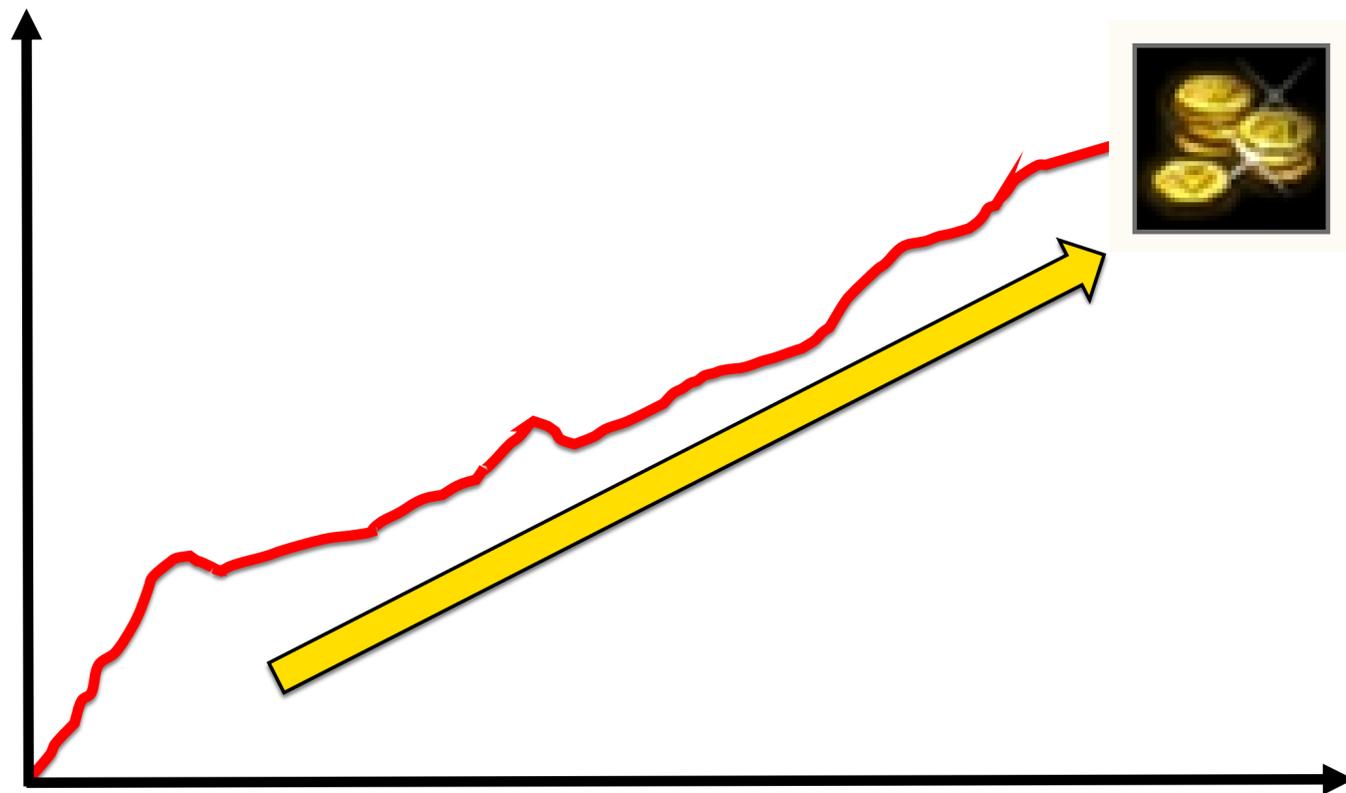
1. 게임 재화 이상 탐지 왜 필요할까?
2. 일반적으로 이야기하는 이상 데이터란?
3. 게임 재화에서 나타날 수 있는 이상 데이터는?
4. 구체적인 탐지 방법
5. 이상 탐지 프로세스 적용
6. 마무리

게임 재화 이상 탐지는 왜 필요할까?

게임 재화 이상 탐지는 왜 필요할까?

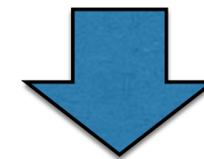
게임 재화 증감 모니터링

- 게임 경제 관점에서 게임 내에 재화량을 모니터링하는 것이 중요함



게임 내 재화가 계속 늘어만 간다면?

게임 머니 인플레이션 발생



화폐 가치 하락

신규 유저와 기존 유저간 빈부 격차

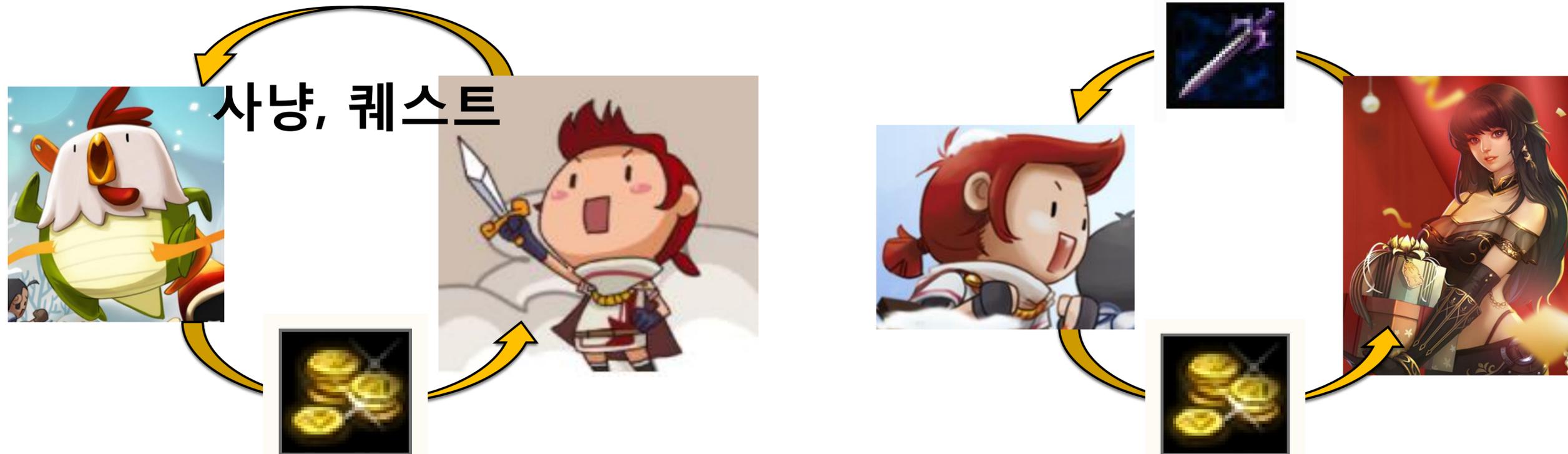
...

게임 재화 이상 탐지는 왜 필요할까?

게임 재화 증감 모니터링

- 유저들이 적절히 재화를 모으고 있는지, 또 적절히 소비하고 있는지를 체크하며 게임 환경 관리 필요

적절한 재화의 생산과 소비



게임 재화 이상 탐지는 왜 필요할까?

게임 재화 증감 모니터링

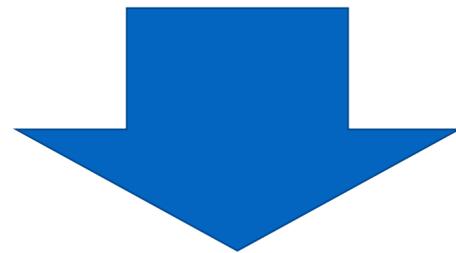
- 그런데 간혹 게임 내 버그나 어뷰징으로 인해 게임 재화가 급격히 증가하는 경우, 게임 서비스에 치명적인 영향을 줄 수 있음



게임 재화 이상 탐지는 왜 필요할까?

게임 재화 증감 모니터링

- 실제 게임 내 재화 복사 버그가 뒤늦게 발견되어, 서버 롤백까지 진행되었던 사례
 - 초기 : 게임 플레이 과정 중에 의도치 않게 간헐적으로 발생
 - 후반 : 해당 행위를 하는 경우가 늘어나게 되면서 재화가 급증
 - 게임 경제 안정 및 선량한 이용자들을 위한 해결책으로 서버 롤백까지 진행되는 문제

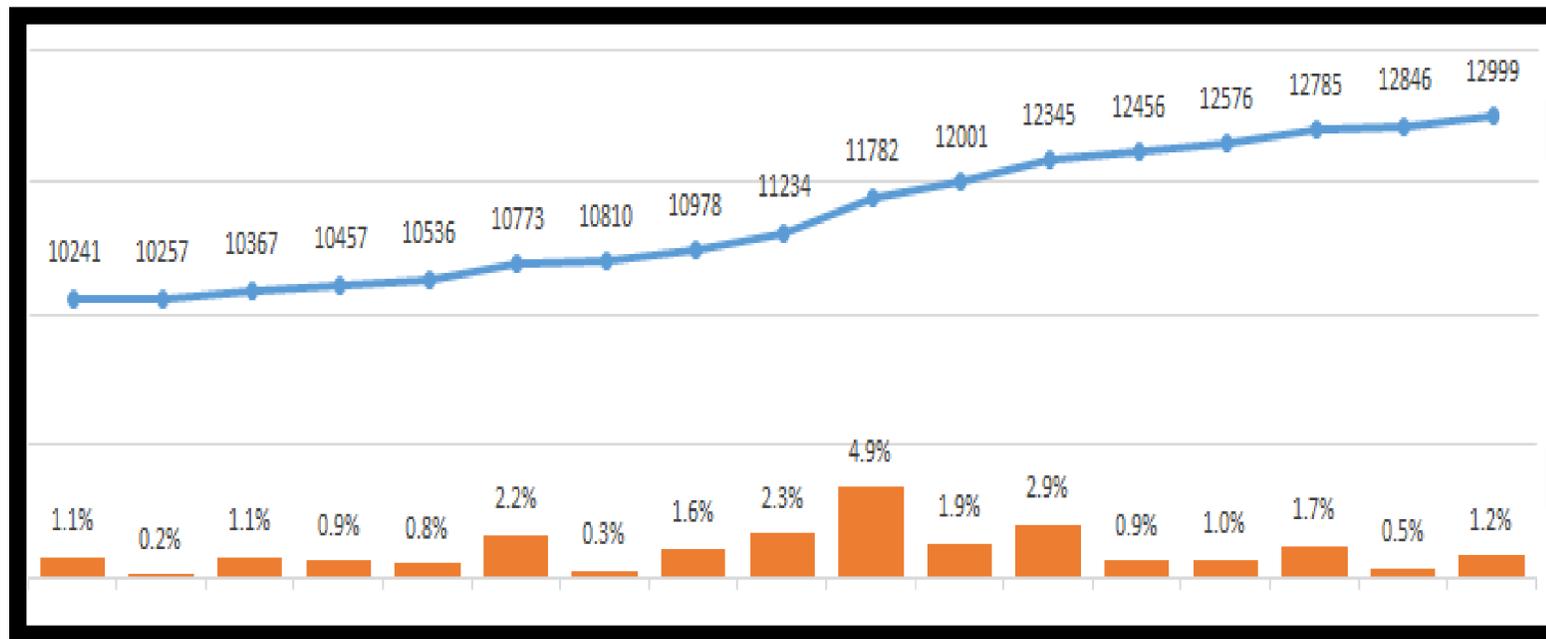


재화 이상 증가 현상을 **조기에** 감지하는 것이 매우 중요!

게임 재화 이상 탐지는 왜 필요할까?

게임 재화 증감 모니터링

- QA를 통해 모든 문제를 찾기는 현실적으로 불가능
- 대개의 경우 주요 게임 내 지표를 모니터링 하면서 급격한 변동이 발생할 경우 문제를 탐사해 나가는 방식을 취함



재화 증가 추이는 어떤지

전일 혹은 전주 대비
변화율은 얼마나 되는지

게임 재화 이상 탐지는 왜 필요할까?

게임 재화 증감 모니터링

- 비정상적의 경우는 확률상 매우 드물게 발생함
- 때문에 운영자가 실시간으로 다양한 지표를 지속적으로 확인하는 것은 비효율적인 면이 큼

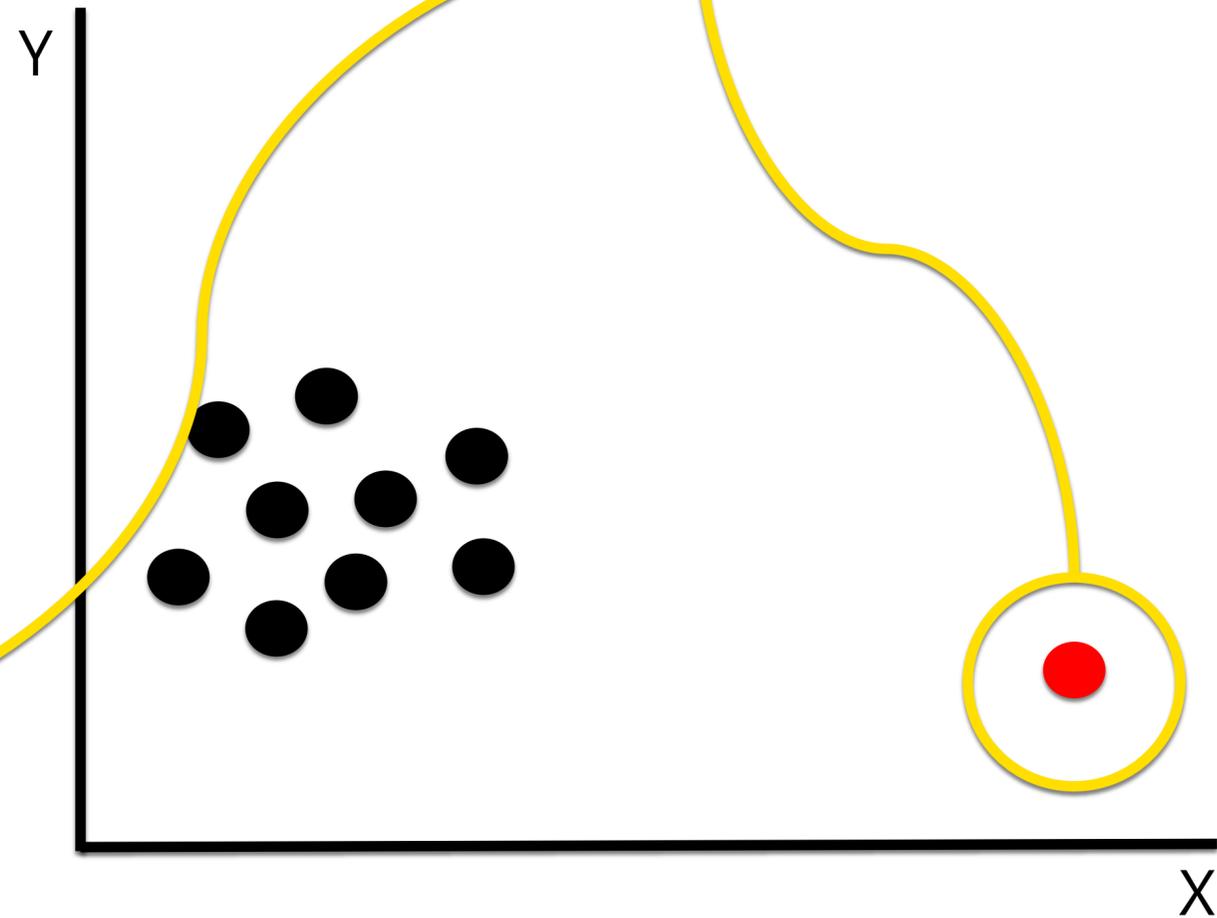
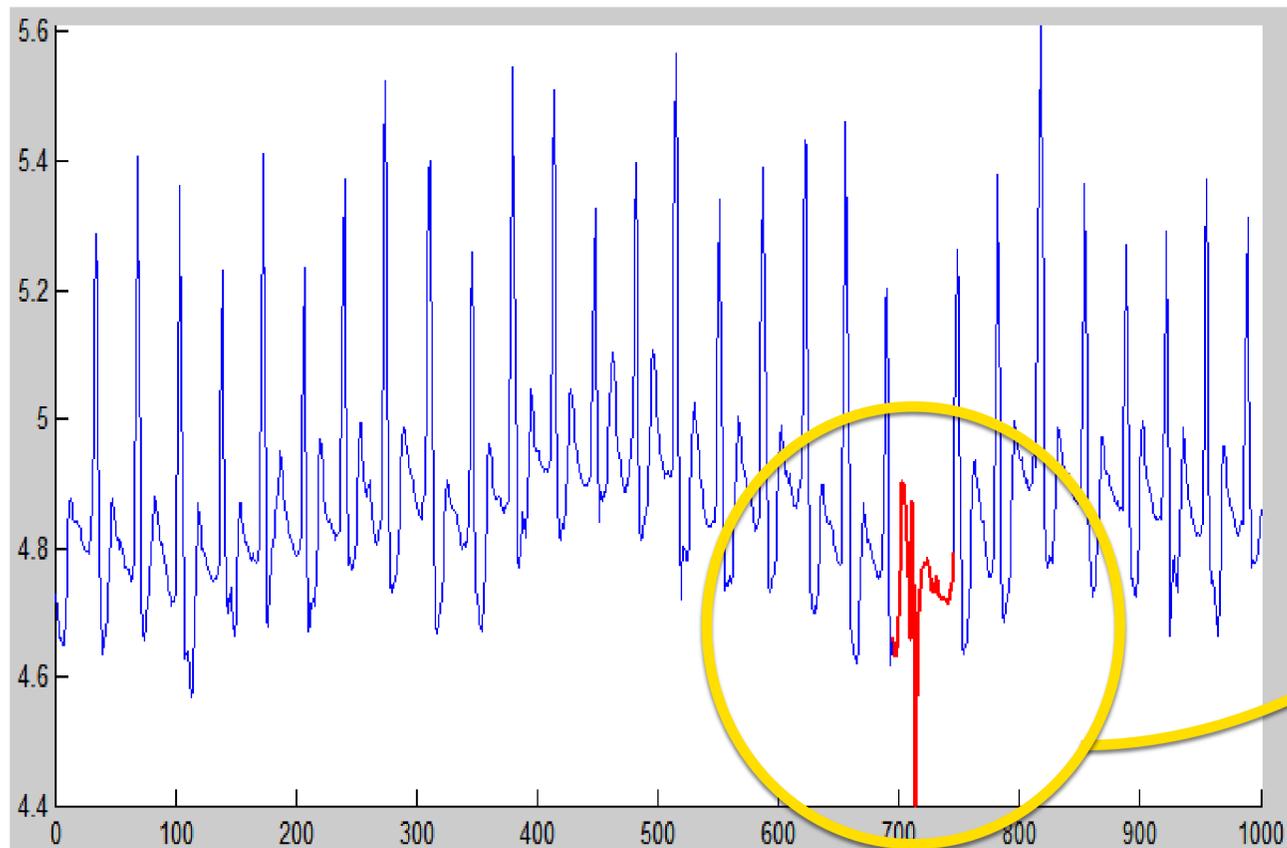


일반적으로 이야기하는 이상 데이터란?

이상 데이터란 무엇인가?

일반적인 이상 데이터

- 기존 데이터의 분포에서 크게 벗어나는 데이터를 일반적으로 **이상 데이터** (Outlier) 라고 함

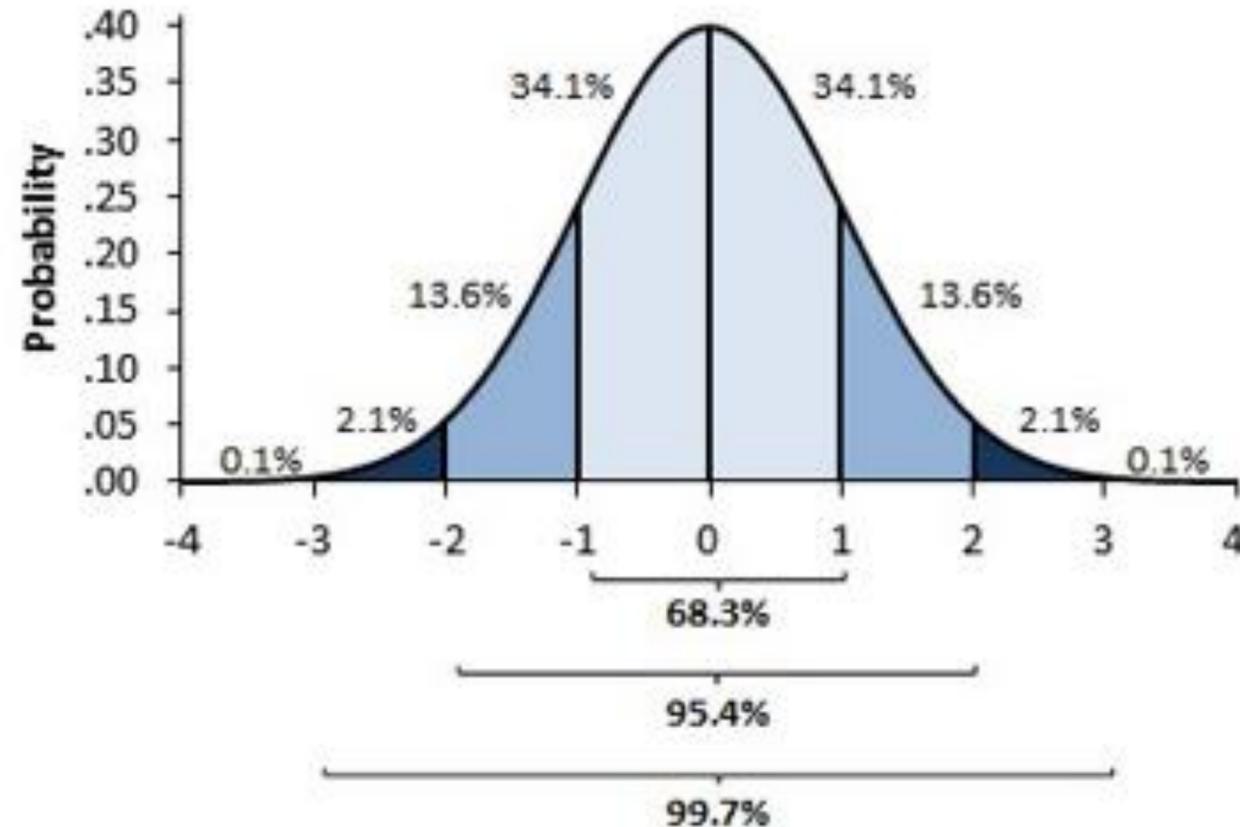


이상 데이터란 무엇인가?

일반적인 이상 데이터 탐지 방법

1. 기존 데이터 분포와 비교해 크게 벗어나는 데이터를 찾는 방법

➤ 평균, 분산을 이용한 탐지



전체 데이터의 평균으로부터
얼마나 떨어져 있는가

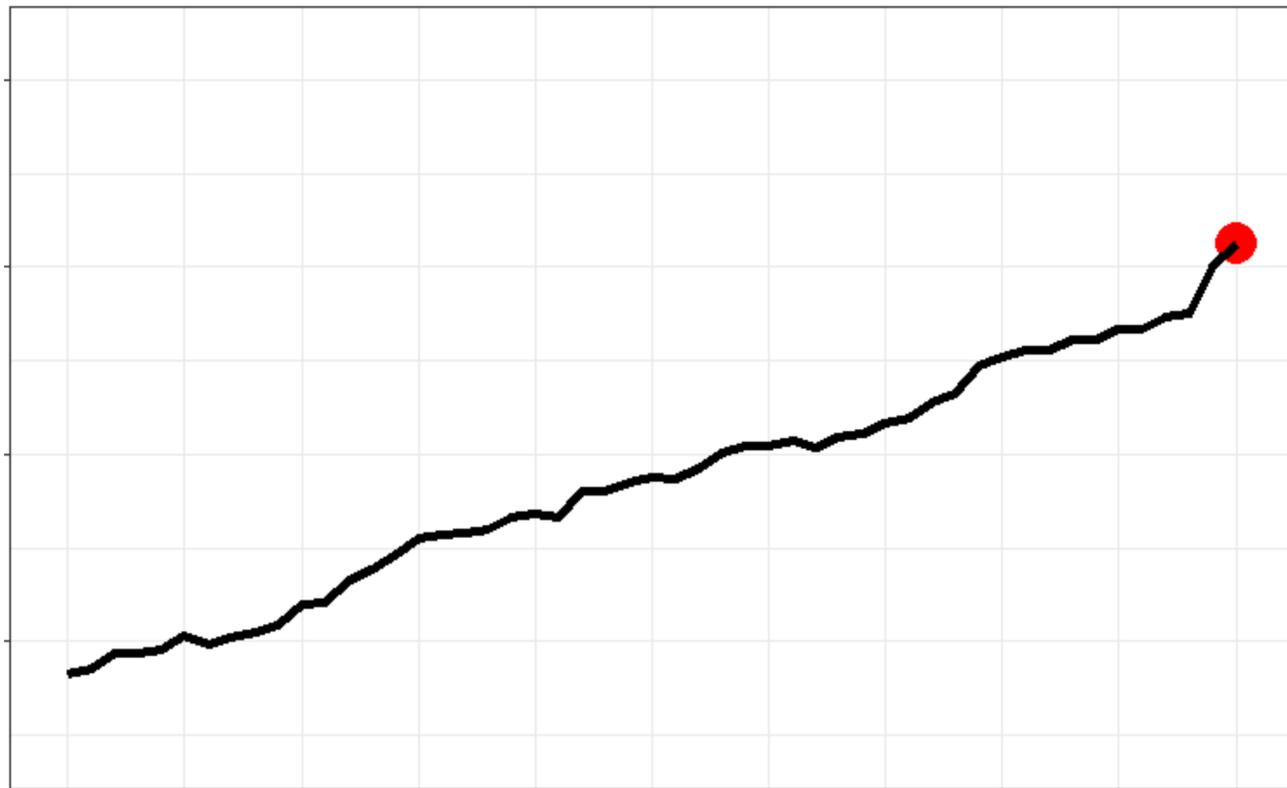
보통 (평균 \pm 2 혹은 3 * 표준편차)를
벗어나는 데이터를 이상치로 탐지

이상 데이터란 무엇인가?

일반적인 이상 데이터 탐지 방법

2. 추세가 있는 데이터에서 이상 데이터를 찾는 방법

➤ 이동 평균, 분산을 이용한 탐지



데이터가 상승 추세 혹은 하락 추세인 경우

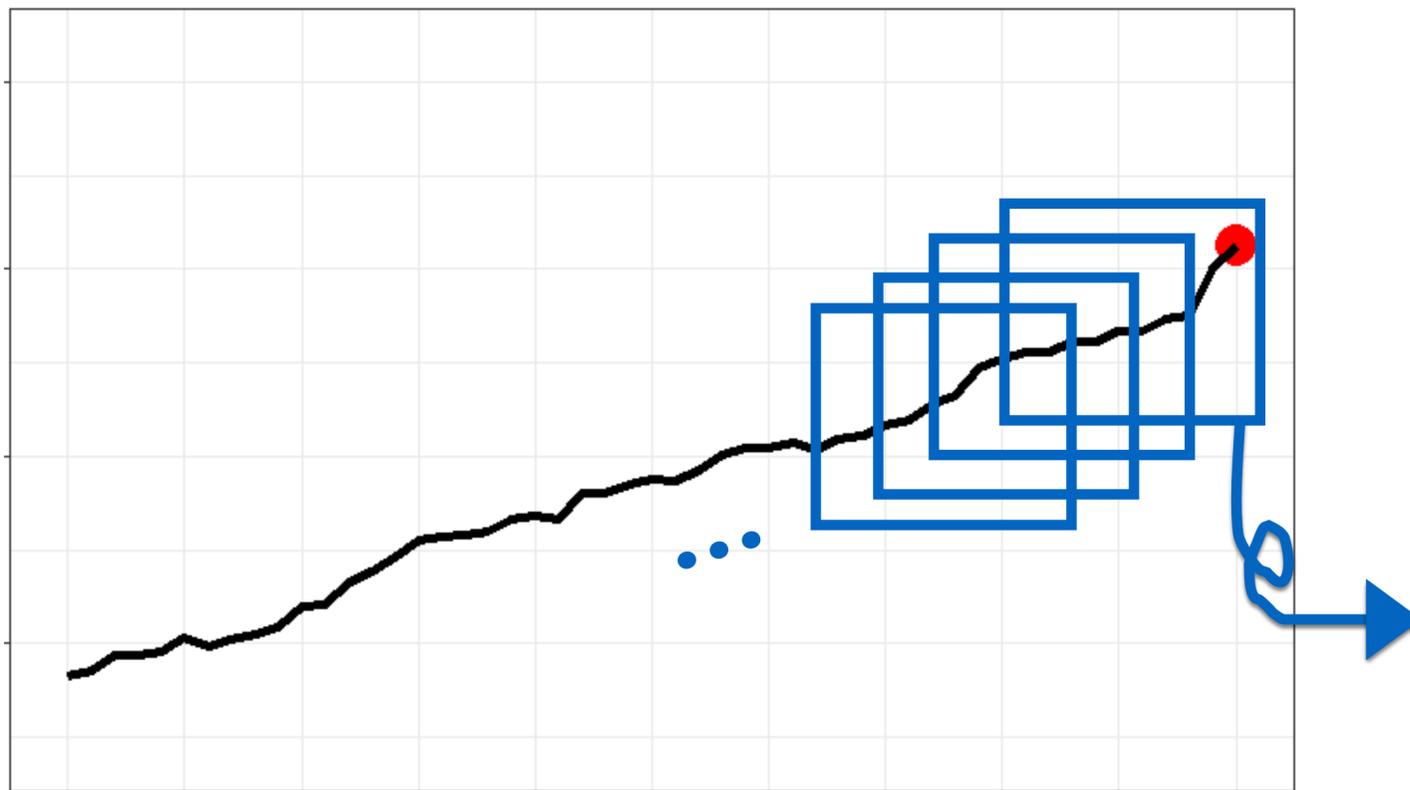
단순히 과거 데이터 분포만 본다면
시간이 갈수록 잘못 탐지하는 경우가 늘어남

이상 데이터란 무엇인가?

일반적인 이상 데이터 탐지 방법

2. 추세가 있는 데이터에서 이상 데이터를 찾는 방법

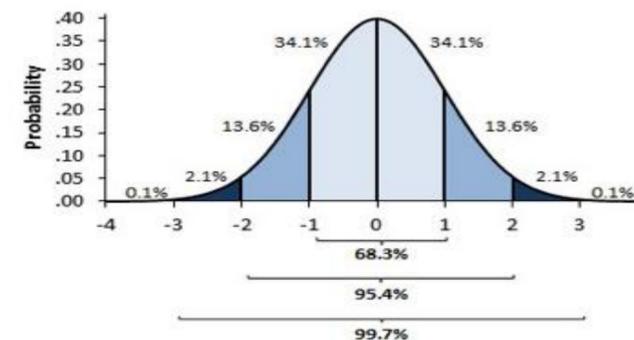
➤ 이동 평균, 분산을 이용한 탐지



탐지하는 시점부터 과거 N번의 데이터로

평균과 분산을 구해

분포에서 크게 벗어나는 값 탐지

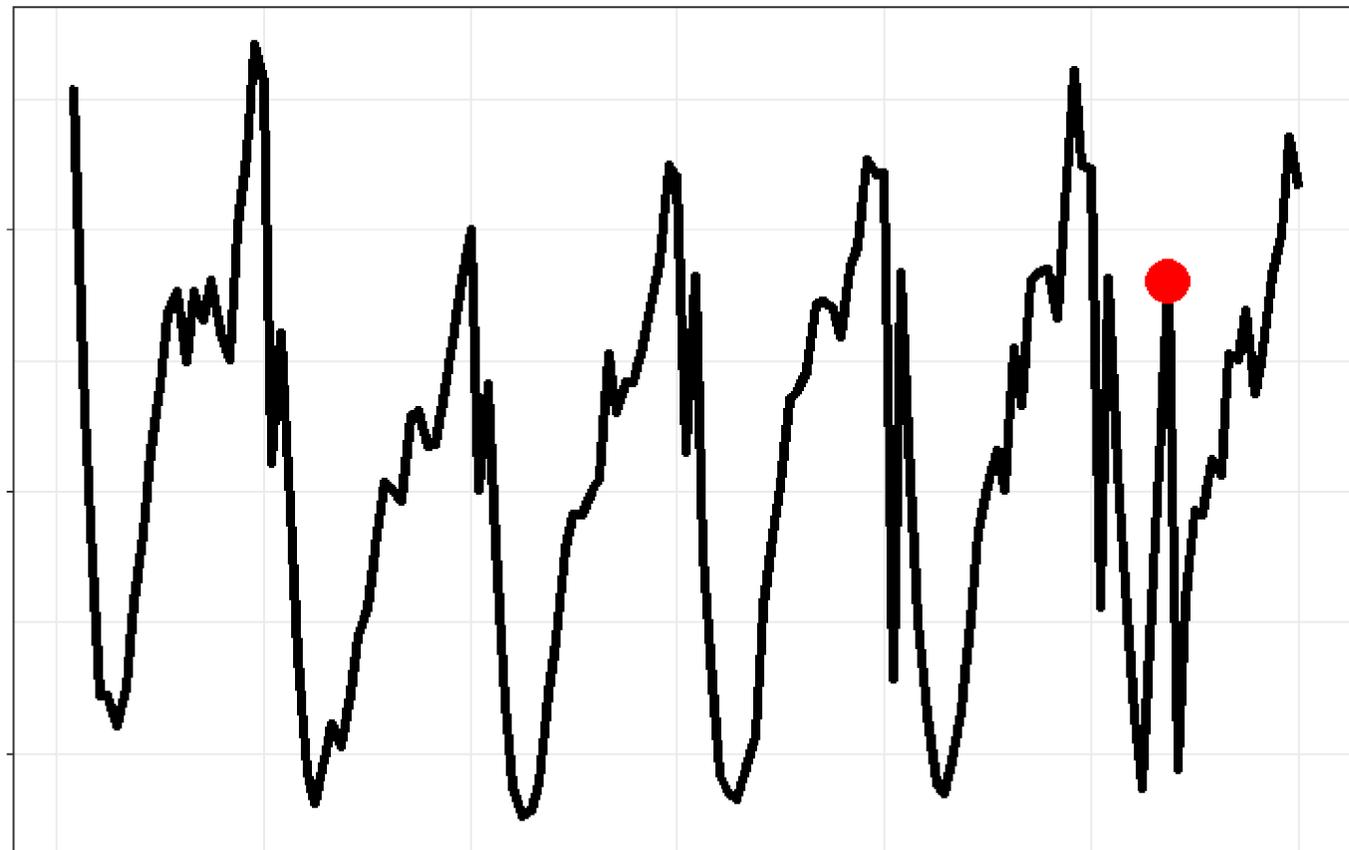


이상 데이터란 무엇인가?

일반적인 이상 데이터 탐지 방법

3. 주기성이 있는 데이터에서 이상 데이터를 찾는 방법

➤ 시계열 분석 (STL)을 이용한 탐지



주기성을 갖는 데이터인 경우

전체 분포 혹은 과거 N 시점의 분포로 봐도

이상치가 정상 범위에 포함되어

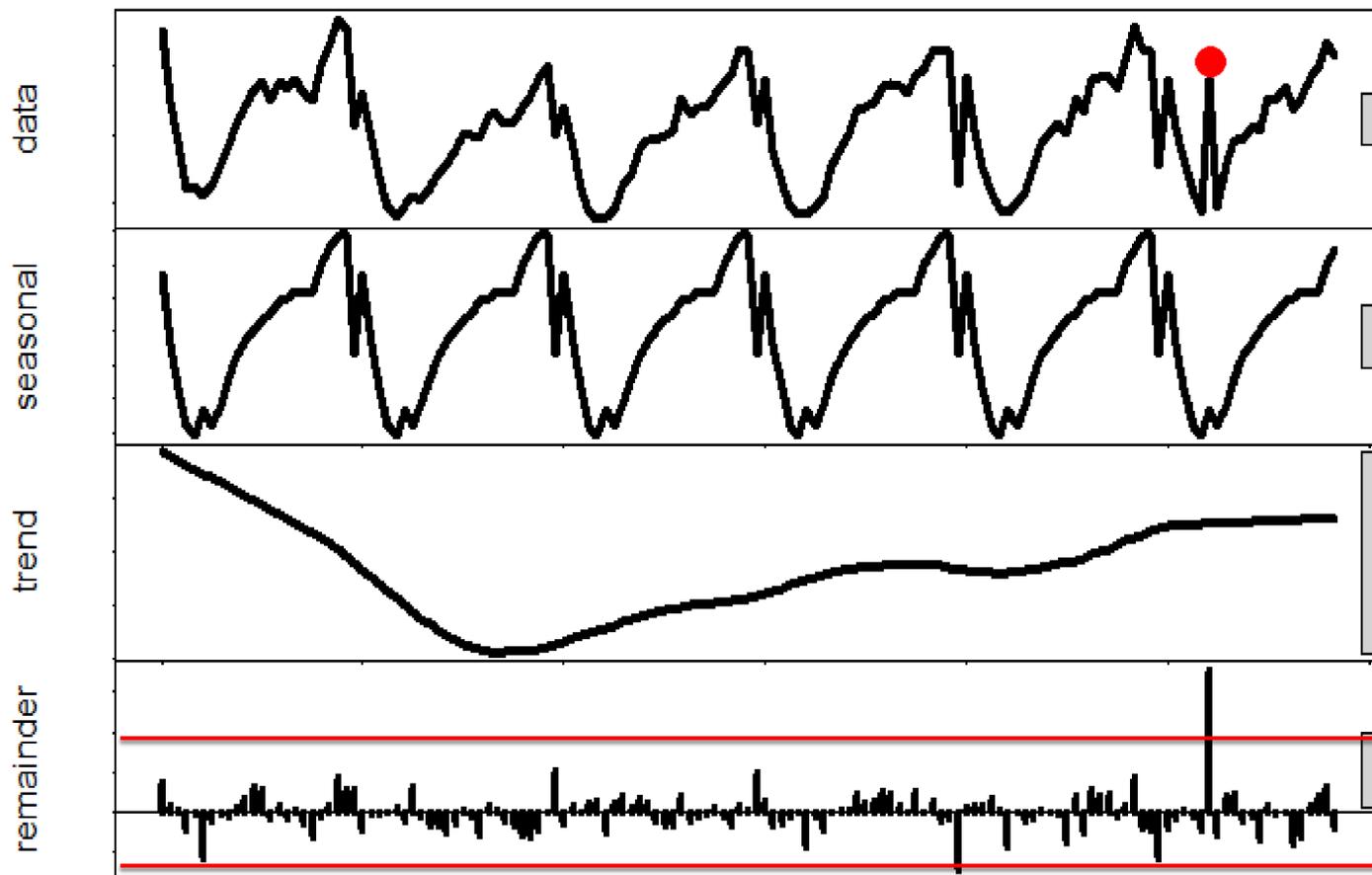
제대로 탐지 못할 수 있음

이상 데이터란 무엇인가?

일반적인 이상 데이터 탐지 방법

3. 주기성이 있는 데이터에서 이상 데이터를 찾는 방법

➤ 시계열 분석 (STL)을 이용한 탐지



시계열 분석 방법 (STL 이용)

- 데이터를 계절성 (Seasonal), 장기적인 변동 패턴 (Trend), 나머지 (Remainder)로 분해

Remainder의 평균, 표준편차 대비 크게

벗어나는 시점의 데이터 탐지

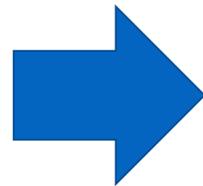
게임 재화에서 나타날 수 있는 이상 데이터는?

게임 재화에서 나타날 수 있는 이상 데이터는?

시간에 따른 변동폭을 고려해야 하는 데이터

- 새벽 시간대와 같이 사용자가 많지 않은 시간대에 어뷰징이 일어나는 경우

시간	CASE
04.24 01시	1000
04.24 02시	500
04.24 03시	800
04.24 04시	700
04.24 05시	800
04.24 06시	650
04.24 07시	300



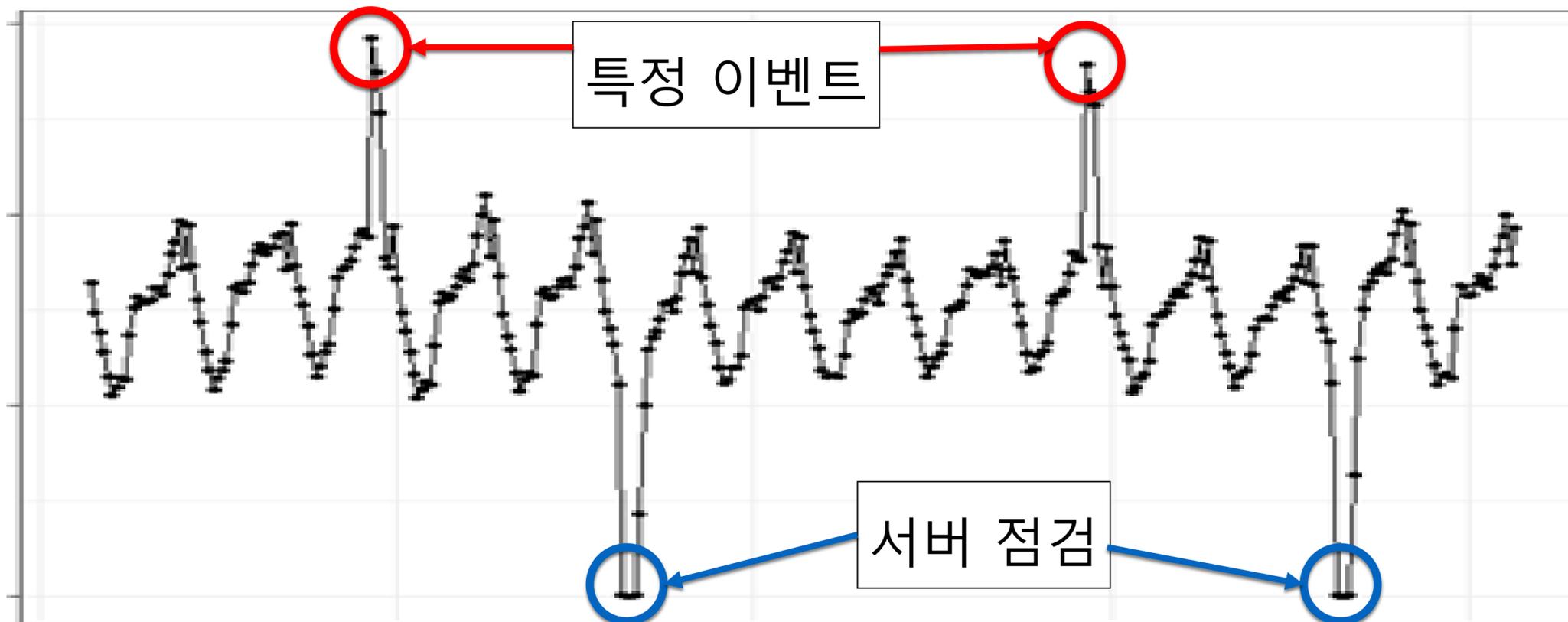
새벽 시간대는 재화 증감이 적기 때문에
버그나 어뷰징으로 인해 잘못 증가되더라도
저녁 시간대의 증감량과 비슷할 수 있음

시간대에 따른 주기성 고려가 필요

게임 재화에서 나타날 수 있는 이상 데이터는?

재화에 영향을 주는 추가적인 요인도 고려해야 하는 데이터

- 고려해야 할 주기성: 24시간, 일주일, 월, 계절 등등
- 위의 주기성 외에도 여러 가지 이벤트들에 의한 영향도 고려해야 함

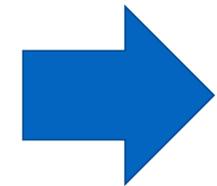


게임 재화에서 나타날 수 있는 이상 데이터는?

재화 이상 데이터를 찾기 위해서는

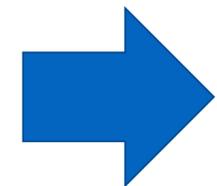


재화 증감에 영향을 주는 요인들을 모두 고려해서



주기성, 이벤트

'정상 상태'를 정의 (모델링) 해야 함



이렇게 정의된 정상 상태에서 크게 벗어난 경우를 이상으로 판단 가능

지금까지

1. 게임 재화 이상 탐지 왜 필요할까?
2. 일반적으로 이야기하는 이상 데이터란?
3. 게임 재화에서 나타날 수 있는 이상 데이터는?
4. 구체적인 탐지 방법
5. 이상 탐지 프로세스 적용
6. 마무리

이제 반 왔습니다



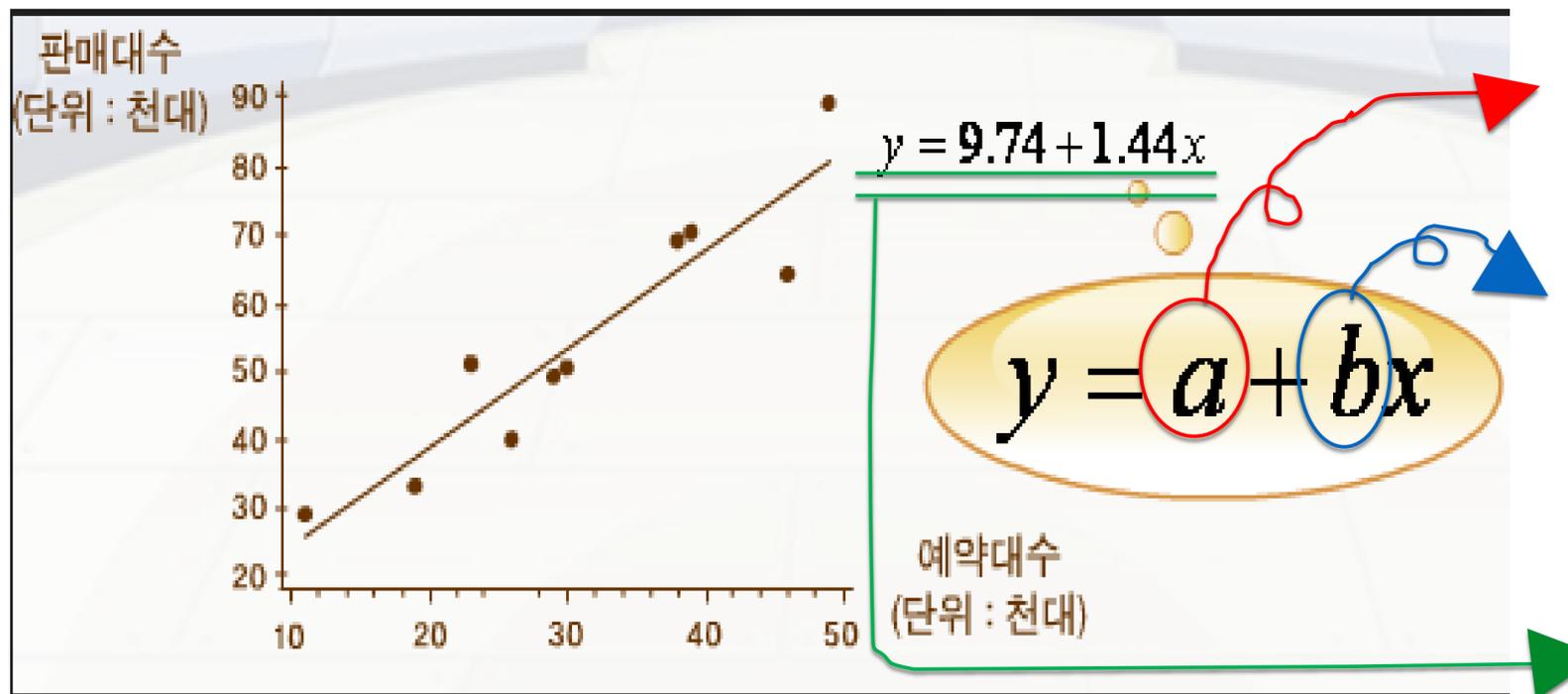
구체적인 탐지 방법

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

재화 증감량을 예측하는 회귀 모델 생성

➤ 회귀 모델 - 어떤 자료에 대해서 그 값에 영향을 주는 조건을 고려하여 구한 평균

예) 에어컨 예약 대수가 한 단위 올랐을 때 판매대수는 얼마나 바뀔까?



Intercept: x라는 요인을 통제했을 때의 평균적인 y값

회귀계수: x라는 요인이 한 단위 오른 때 y에 영향을 미치는 값

예약 대수가 한 단위 오를 때 판매 대수는 1.44 증가

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

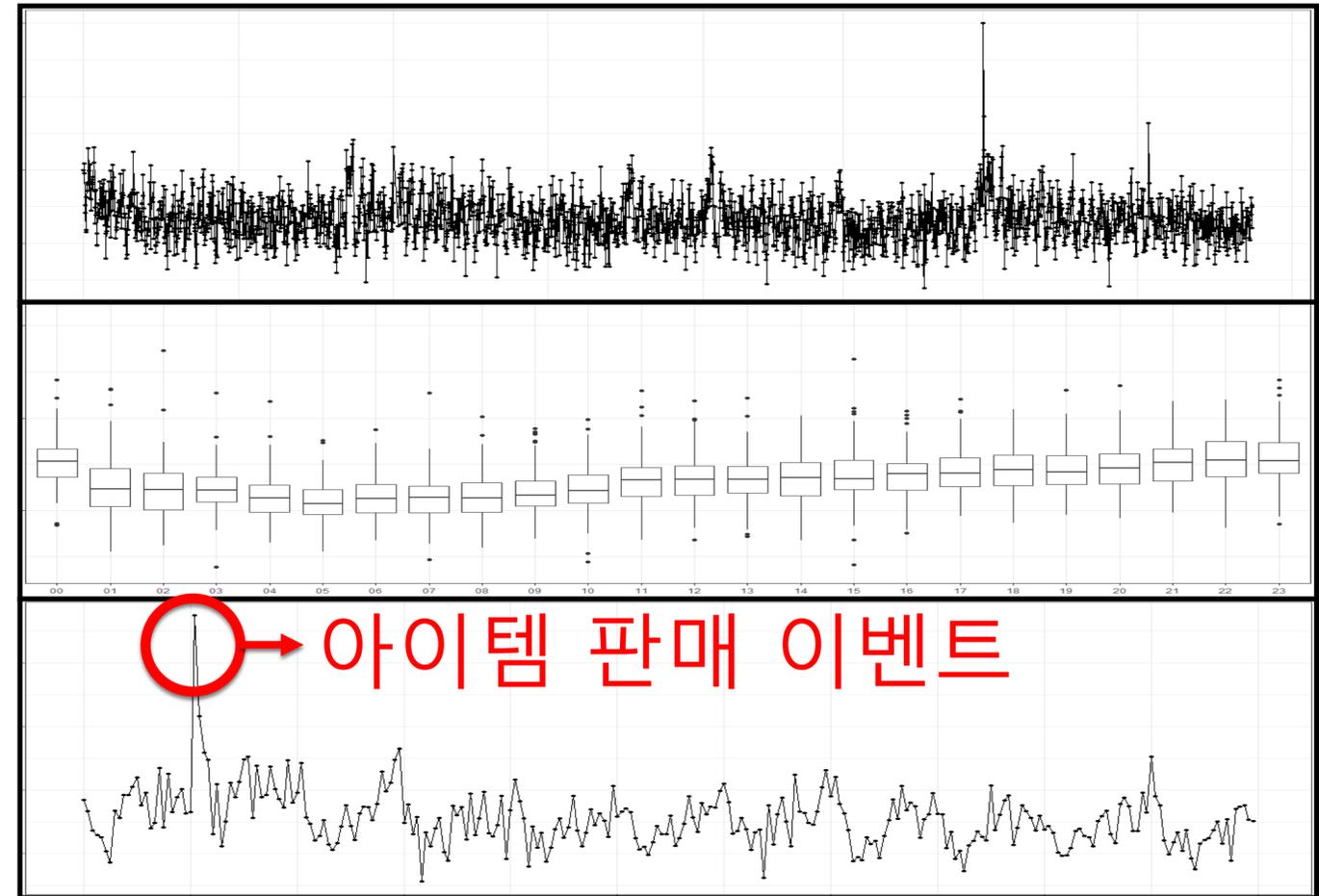
재화 증감량을 예측하는 회귀 모델 생성

➤ 우선은 데이터 탐색부터

전체적인 증감 추이는 어떤지

시간대별이나 요일별 변동은 없는지

특정 이벤트에 따른 변화는 없는지



어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

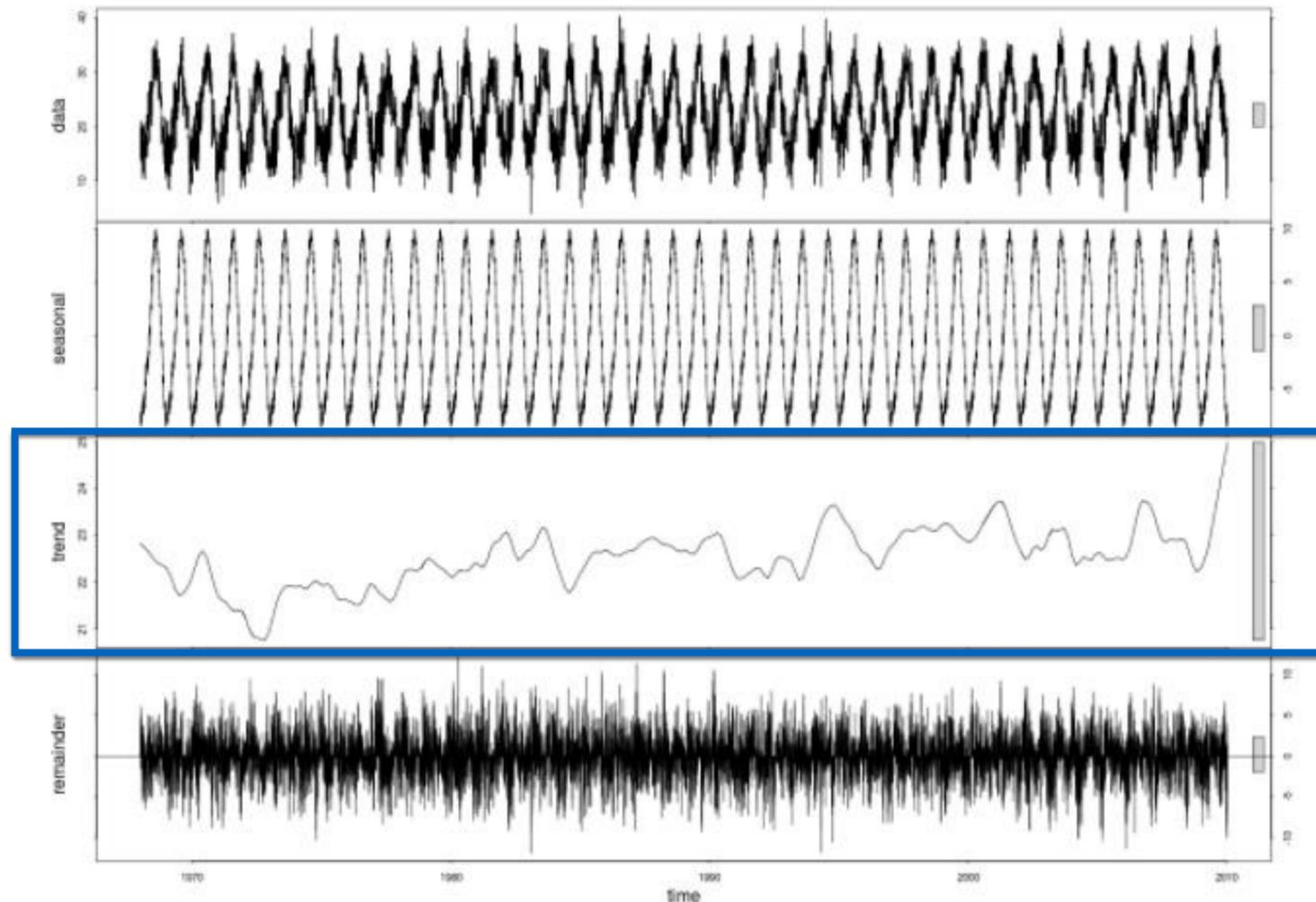
재화 증감량을 예측하는 회귀 모델 생성

- 일주일이나 월 단위의 주기성은 없는지, 재화 증감에 영향을 줄 수 있는 다른 요인은 없는지 다양하게 살펴본 후, 이것들을 회귀 모델의 설명 변수로 사용함
- 예) 서버 점검이 있는 때에는 증가량이 크게 하락하는 영향 → **점검 시간 변수**
시간대에 따른 주기성을 보임 → **시간 변수**
대규모 레이드가 진행될 때는 증가량이 별로 없음 → **레이드 변수**

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

재화 증감량을 예측하는 회귀 모델 생성

- 추가로 데이터의 추세를 반영하는 트렌드 변수를 사용



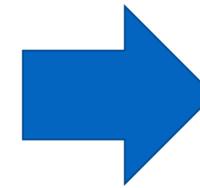
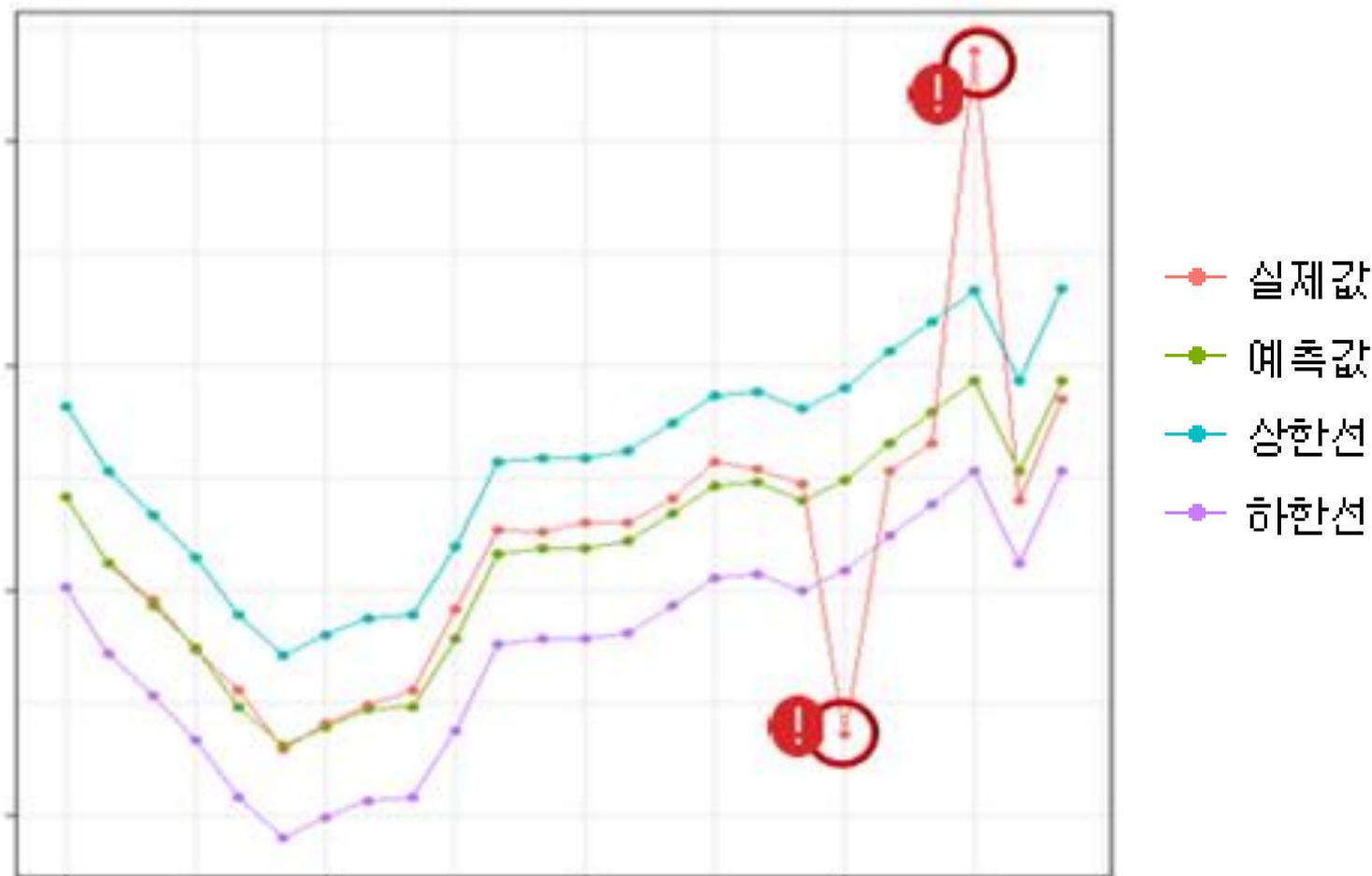
시계열 분석 방법 (STL 이용)

이전의 트렌드 값을 회귀 모델 변수로 사용

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

1차 이상 탐지

- 회귀 모델을 이용해 매시간별 재화 증감량을 예측
- 예측한 값과 예측 값에 대한 상/하한선을 구해, 실제 해당 시간의 증감량과 비교

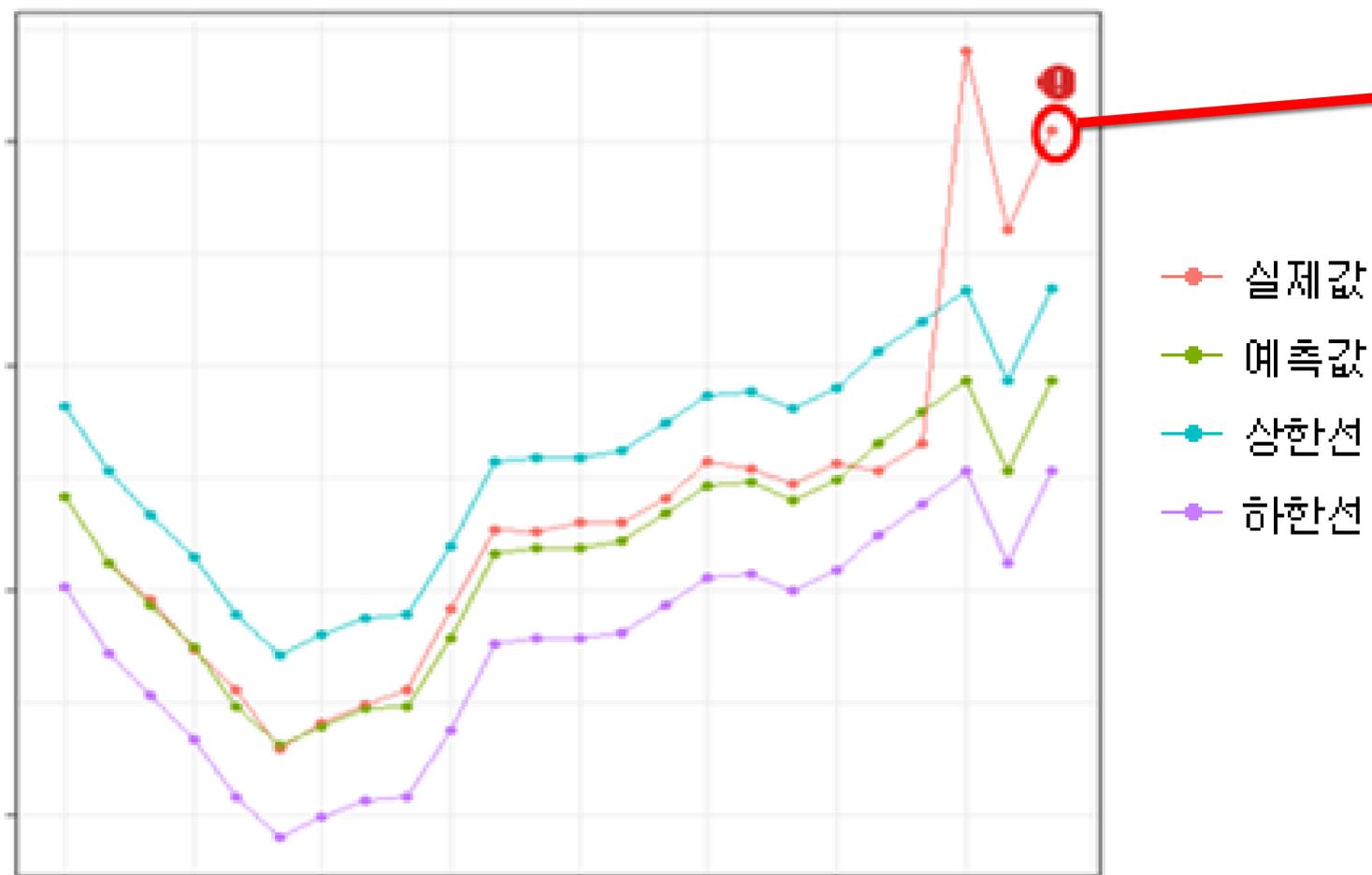


실제값이 예측 상/하한선을
초과하는 경우를
이상 데이터로 탐지

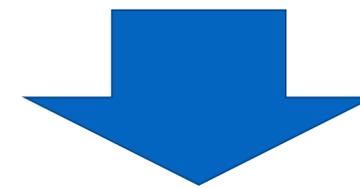
어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

1차 이상 탐지에 추가 Rule 적용

- 재화 증감량은 유저들의 행동에 따라 크게 영향을 받기 때문에, 이상값이 발생할 가능성이 높음



유저가 그동안 모아두었던
아이템을 상점에 한 번에 판매

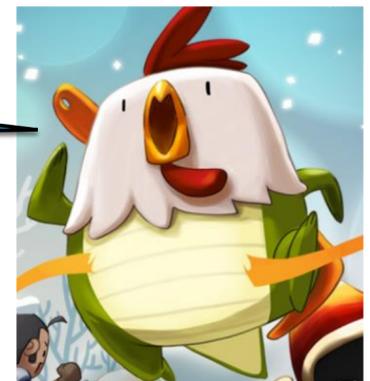


특정 시간 동안 N번 이상
이상값이 발생하는 경우 알람을 주는

추가 Rule 적용

일반적인 이상 데이터는 1차 방법에서 탐지됨 하지만 ...

만만한 일이 아니다



어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

2차 이상 탐지

- 급변하지는 않지만 게임 재화가 지속적으로 증가하는 현상이 발생할 경우
- 예) 소수의 어뷰저가 지속적으로 버그를 악용하는 경우



**상/하한선을 벗어나지 않기 때문에
1차 이상 탐지에서는 탐지되지 않음**

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

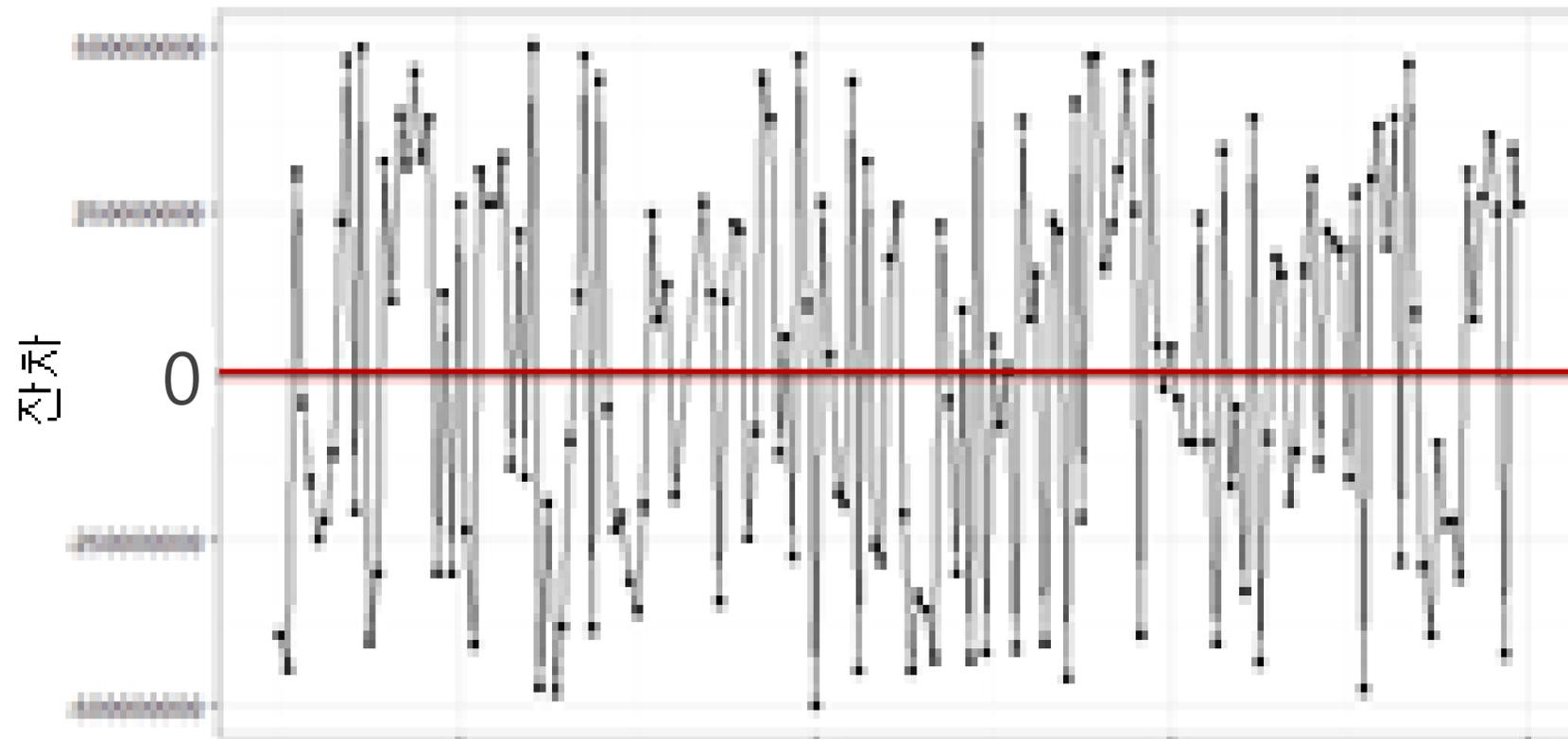
2차 이상 탐지

- 2차 이상 탐지는 잘못된 재화 증감량이 너무 작아 1차에서 발견되지 않은 값을 대상으로 함
- 즉, 예측 상/하한선을 넘지 않을 정도로 미미한 버그나 어뷰징이 발생하는 경우를 대비하는 절차

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

2차 이상 탐지

- 예측값과 실제값의 차이인 잔차를 이용
- 정상 상태라면, 잔차는 0을 평균으로 하는 정규 분포 형태로 나타남



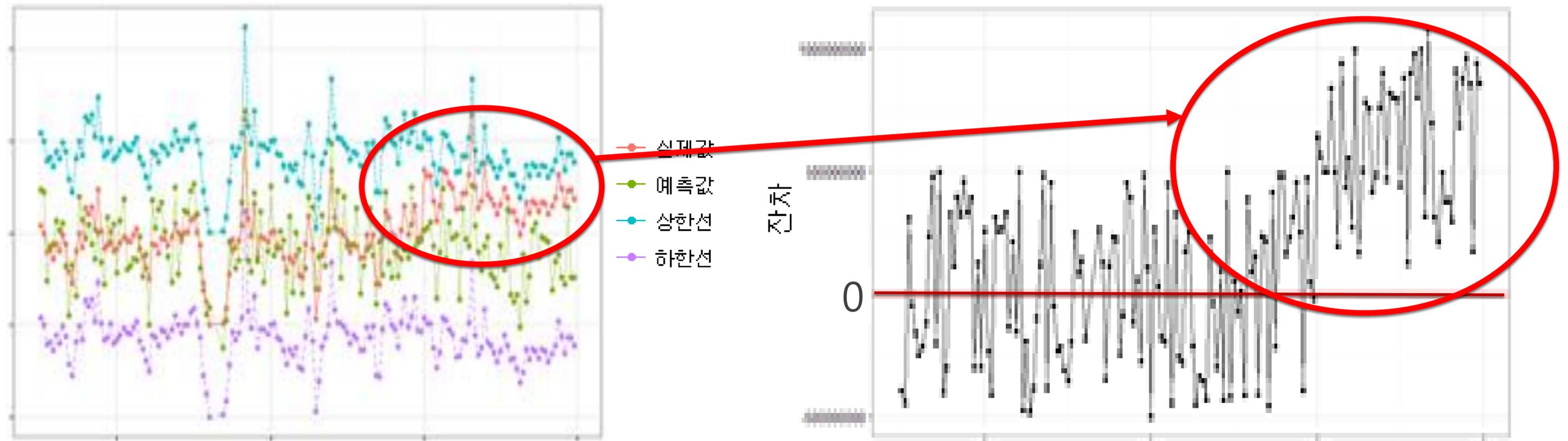
정상 데이터: 0을 기준으로 랜덤하게 분포

어떻게 하면 이상 데이터를 조기에 탐지할 수 있을 것인가?

2차 이상 탐지

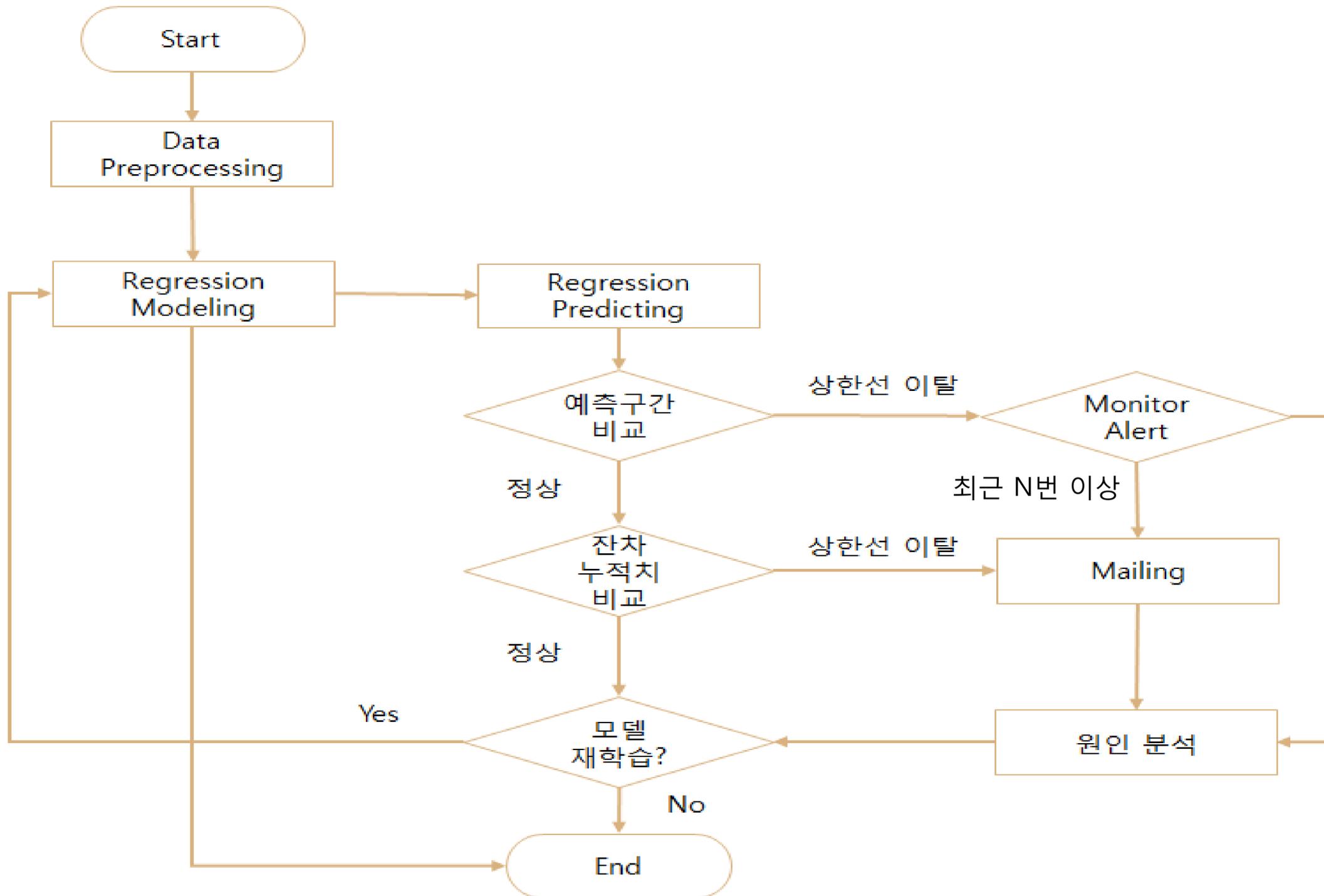
- 이상 현상 발생 시에는 잔차가 한 쪽으로 쏠린 분포를 나타냄

잔차의 누적치를 모니터링하여 일정 규모 이상이 되는 경우 이상으로 탐지



이상 탐지 프로세스 적용

이상 탐지 프로세스 적용



이상 탐지 프로세스 적용

이상 탐지 알람

- 1차, 2차 과정에서 탐지가 되면 운영자에게 상세한 증감 내용을 메일로 알람

최근 N번 이상치 발생

시간	증감량	상한값	탐지
03시	1000	200	○

상세 증감 항목

- 사냥 증가량: 800, 최근 N주 동일
요일/시간대 대비 XX% 증가



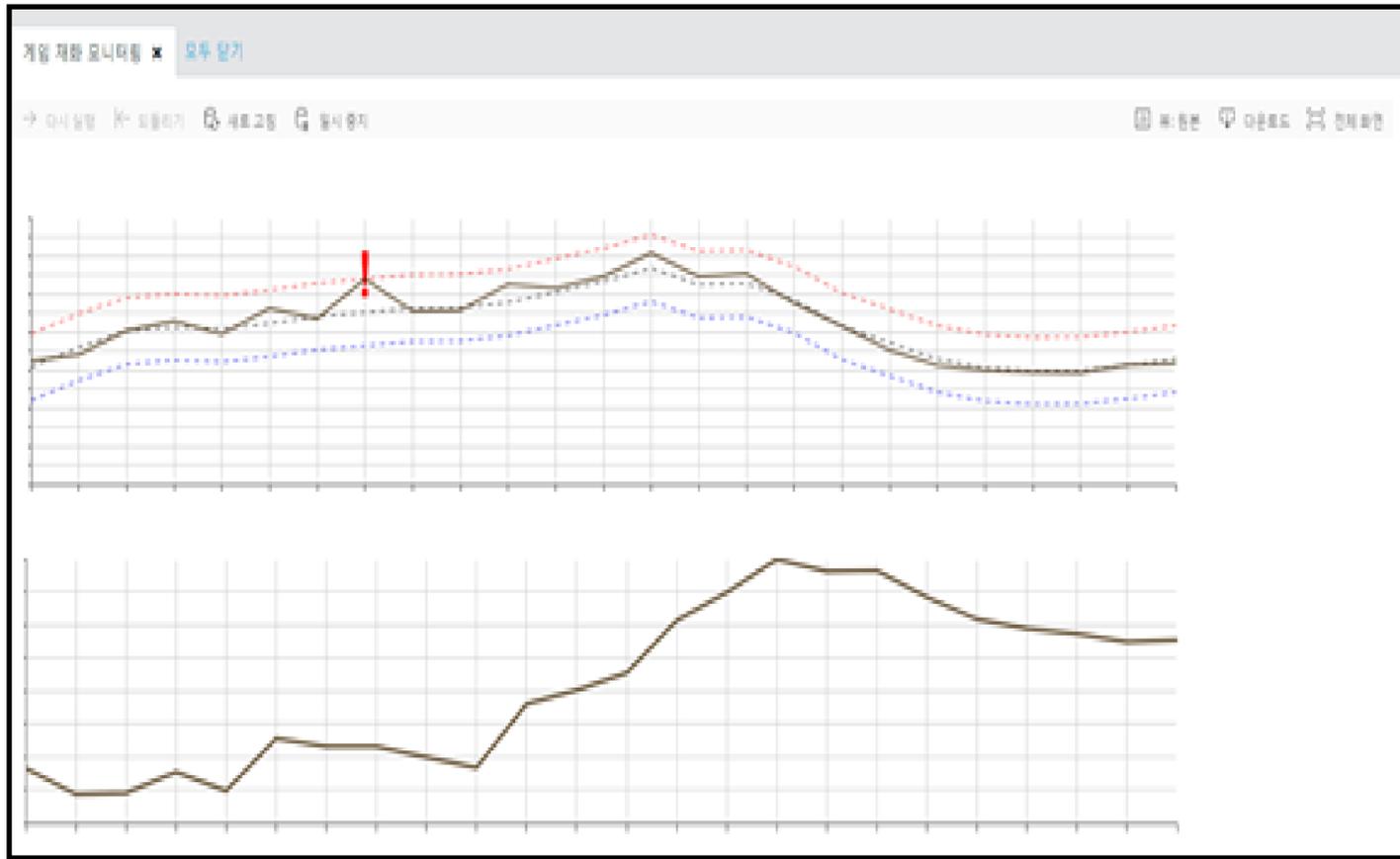
서비스 운영자

이상 원인
상세 분석

이상 탐지 프로세스 적용

모니터링 페이지 구축

- 재화 증감에 대한 모니터링 페이지를 통한 확인도 가능



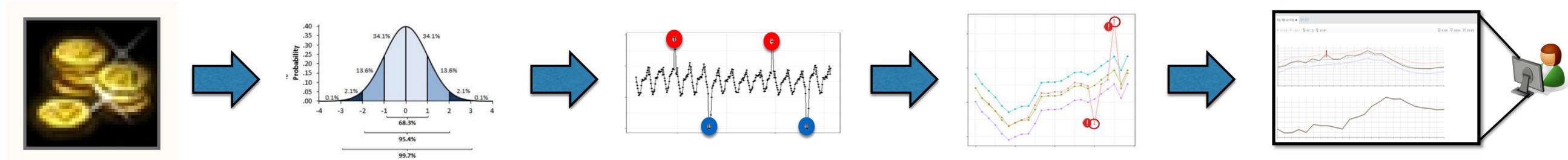
서비스 운영자

증감 현황 파악

마무리

마무리

1. 게임 재화 이상 탐지 왜 필요할까?
2. 일반적으로 이야기하는 이상 데이터란?
3. 게임 재화에서 나타날 수 있는 이상 데이터는?
4. 구체적인 탐지 방법
5. 이상 탐지 프로세스 적용



마무리

R timesuperin 패키지

- 시계열을 이용한 회귀 분석을 할 수 있도록 개발한 패키지
`devtools::install_github("danbi-ncsoft/timesuperin")`
R에서 위 경로로 설치 가능 (코드 수정은 ing...)

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.