



[weekly project] 게이트 라운드 변경에 따른 고객 유지율 변화

🕒 생성 일시	@2023년 12월 15일 오후 3:26
🏷 태그	AB Test EDA chi-square project t-test
🔗 코드 바로 가기	https://github.com/ujeongO/multicampus-2023-DataVisualization/tree/main/AB_test

프로젝트 소개 🍪

- 모바일 게임 Cookie Cats 유저가 첫번째 게이트에 도달하는 시점을 30 라운드에서 40 라운드로 옮겼을 때, 리텐션에 어떤 변화가 있는지를 알아보기 위한 AB Test를 가지고 EDA 및 t-test, 카이제곱 검정을 진행하였다.

Data

Mobile Games A/B Testing - Cookie Cats

<https://www.kaggle.com/datasets/mursideyarkin/mobile-games-ab-testing-cookie-cats>



90189 rows × 5 columns 로 구성

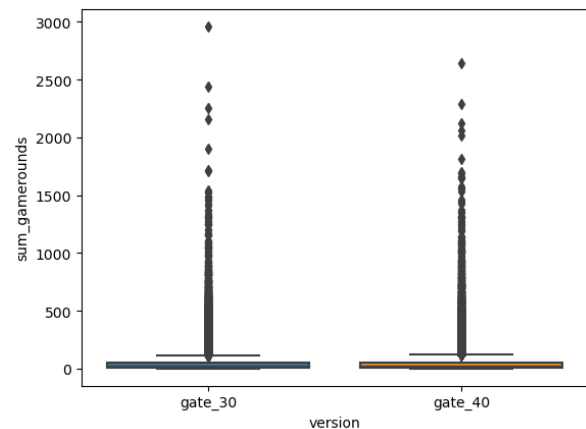
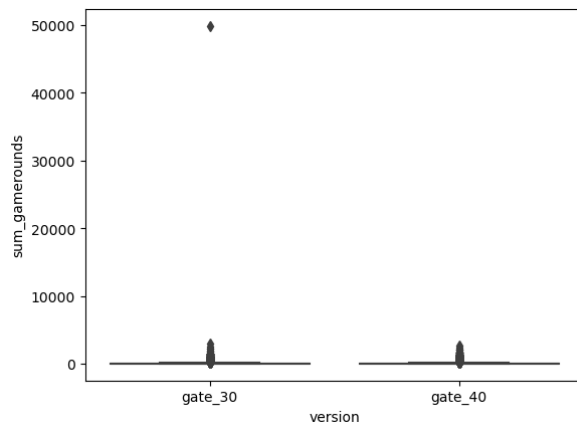
▼ 데이터 상세

- userid
- version: 첫 게이트에 도달하는 시점(gate_30, gate_40)
- sum_gamerounds: 설치 후 첫 14일 동안 플레이한 게임 라운드 수
- retention_1: 설치 후 1일 후에 다시 돌아와서 플레이 했는지 여부 (True, False)
- retention_7: 설치 후 7일 후에 다시 돌아와서 플레이 했는지 여부(True, False)

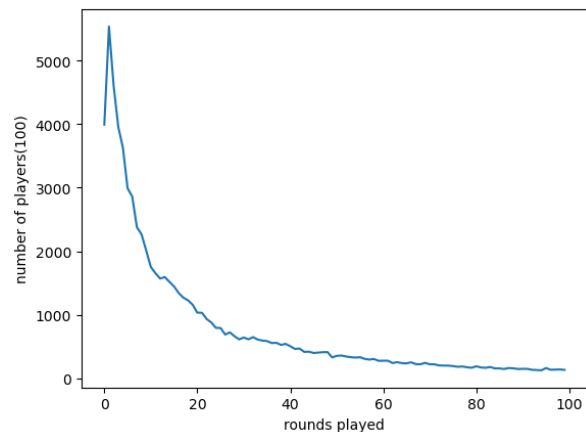
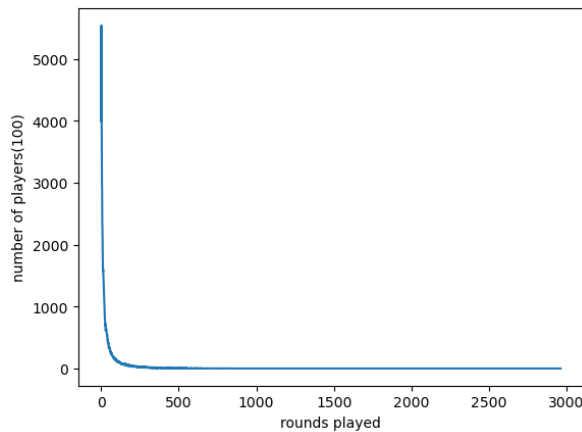
EDA

```
userid          0
version         0
sum_gamerounds  0
retention_1     0
retention_7     0
dtype: int64
```

- 결측치는 없었다.



- version에 대한 게임 플레이 횟수를 boxplot으로 그렸을 때, 1개의 이상치가 발견되었다. 다른 유저에 비해 비정상적으로 높은 플레이 횟수라고 판단하여 해당 데이터는 삭제하고 진행하였다. (버그나 매크로인 것 같다...)
- 이상치 삭제 후 boxplot을 다시 그렸을 때, gate_30과 40이 비슷한 결과를 띄는 것을 확인할 수 있었다.

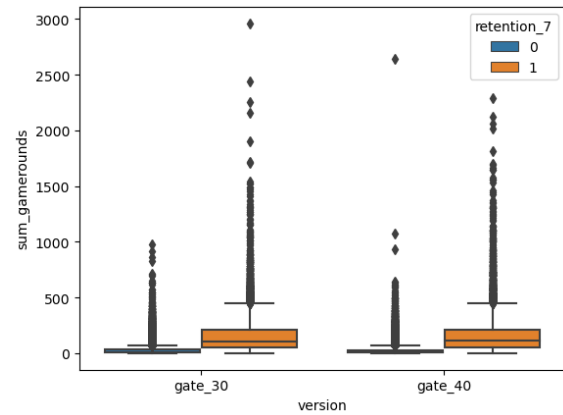
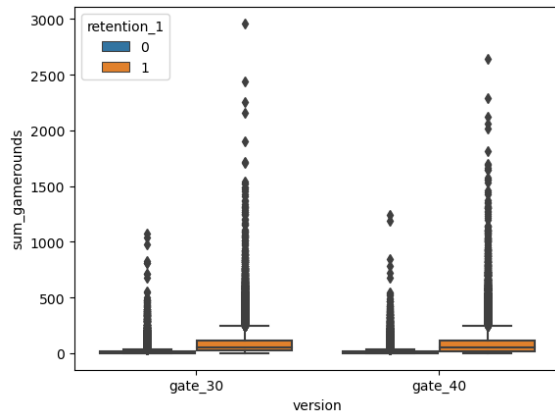


- 게임 플레이 횟수를 선그래프로 나타내면 왼쪽과 같다. 유저들이 몰려있는 부분만 확대해서 보면 오른쪽 그래프와 같이 유저의 게임 플레이 횟수가 0~20회에 몰려있는 것을 확인할 수 있다.

가설 설정 및 검정 결과

▼ 가설 1 - version & sum_gamerounds

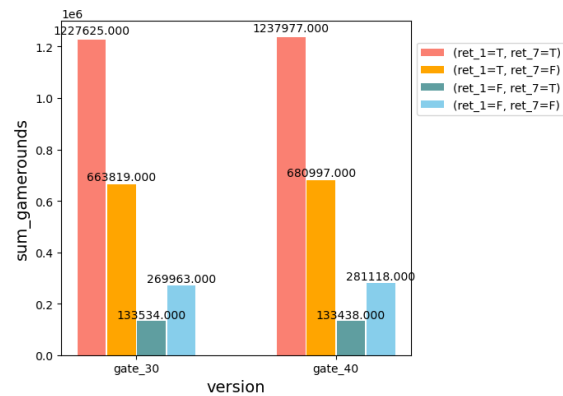
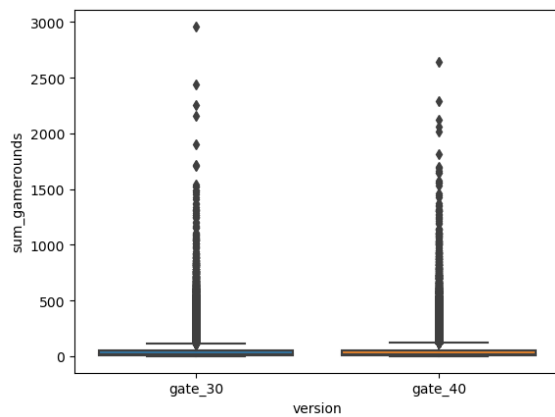
1 게임의 버전에 따라 유저들의 게임 플레이량에 차이가 있을 것이다.



- version-retention_1, version-retention_7 모두 크게 차이가 없어 보인다.
- 두 그룹의 평균 차이를 알아보기 위해 아래에서 t-test를 진행해보았다.

<t-test>

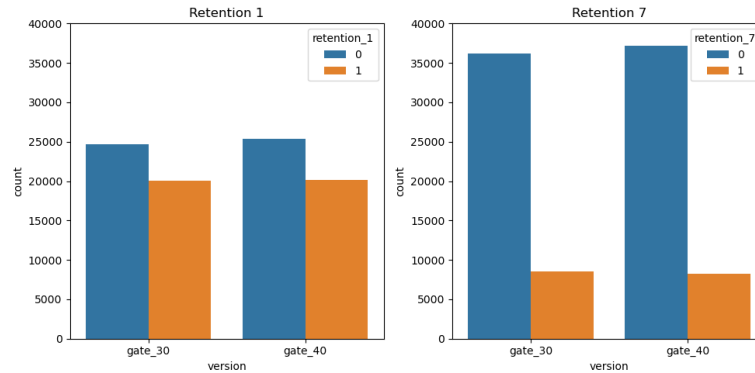
- levene test 결과 두 그룹이 등분산 가정을 만족했다.
- t-test 결과 유의확률이 0.9495로 유의수준 0.05보다 크기때문에 귀무가설을 기각할 수 없다. 즉, gate_30의 평균 플레이 횟수는 gate_40의 평균 플레이 횟수와 차이가 없다고 할 수 있다.



- 그래프를 봐도 version별 플레이한 라운드 수가 별 차이 없는 것을 확인할 수 있었다.

▼ 가설 2 - version & retention

2 게이트를 레벨 30에서 40으로 이동하면 사용자의 일주일 후의 유지율이 증가할 것이다.



- 1일 후 재접속 유저 수와 7일 후 재접속하는 유저 수의 그래프를 그려보았다. 그 결과 gate_30과 40의 재접속 유저의 수는 별 차이가 없어 보였다.
- 버전과 retention 사이에 관계를 자세히 알아보기 위해 아래에서 카이 제곱 검정을 진행해 보았다.

<chi-square 검정>

retention_1	False	True
version		
gate_30	24665	20034
gate_40	25370	20119

- p-value가 0.075로 유의 수준인 0.05보다 커서 귀무가설을 기각할 수 없다. 즉, version과 retention_1 두 변수는 독립이며 서로 강한 관련성이 없다.

→ 🍪🐱 게임의 첫 게이트가 30에 있든지 40에 있든지 첫 날 접속률에는 큰 영향을 미치지 않는다.

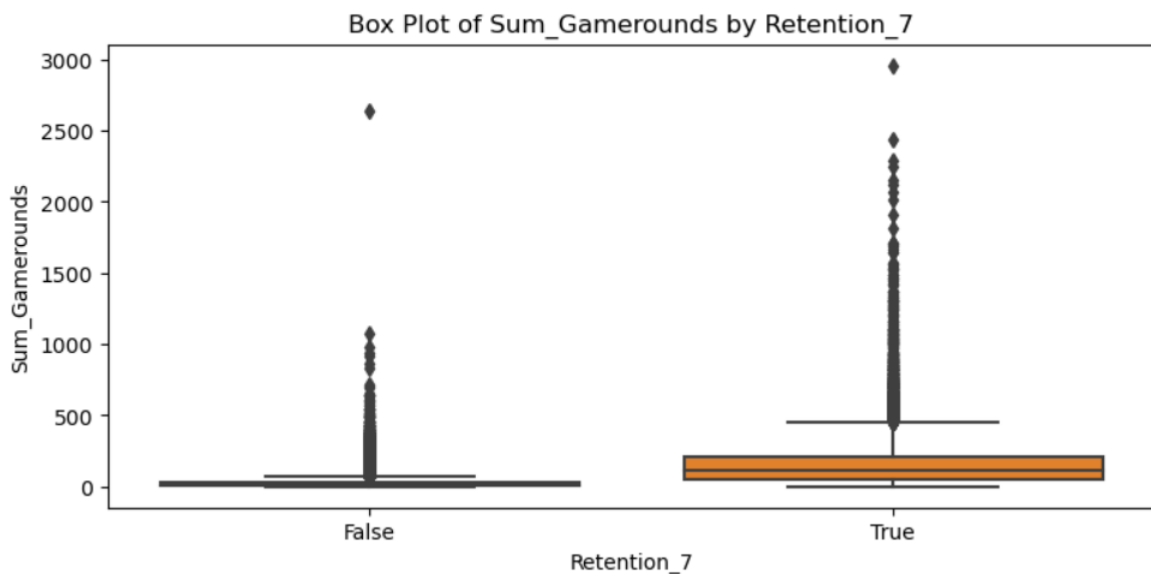
retention_7	False	True
version		
gate_30	36198	8501
gate_40	37210	8279

- p-value가 0.0016으로 유의 수준인 0.05보다 작아서 귀무가설을 기각 및 대립가설 채택, 즉 version과 retention_7 두 변수는 독립이 아니며 서로 강한 관련성이 있다.

→ 🍪🐱 게임의 첫 게이트 위치에 따라 일주일 뒤 유지율에는 큰 영향을 미친다.

▼ 가설 3 - sum_gamerounds & retention7

3 게임에 더 많은 시간을 투자한 유저들은 일주일 후의 유지율이 더 높을 것이다.



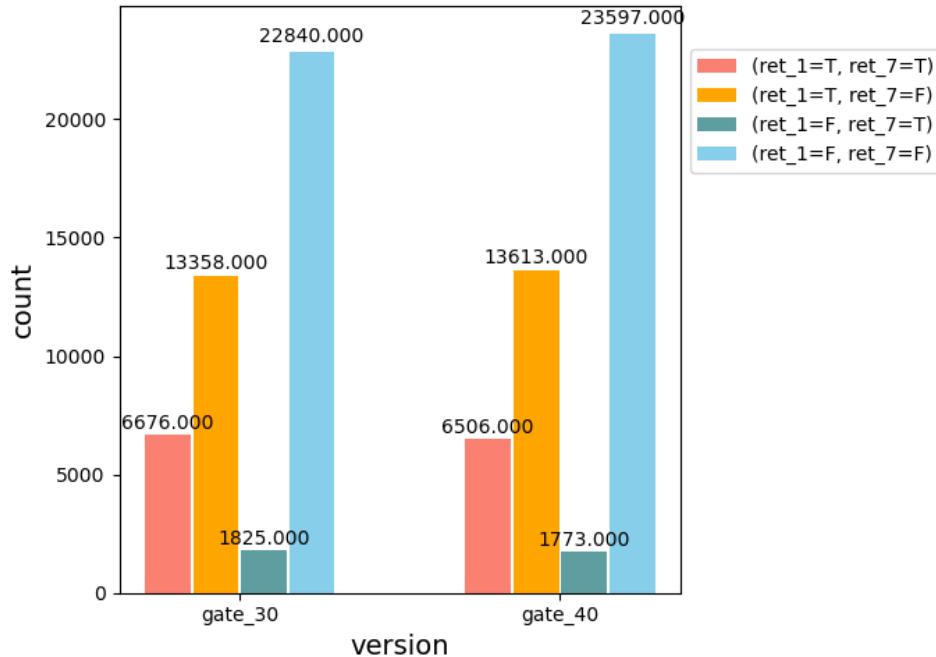
- sum_gamerounds와 retention_7의 boxplot을 찍어본 결과, 1일보다 7일 후 유지된 그룹의 플레이 수의 중앙값과 사분위수가 높았다.
- retention_7이 True일 때 평균이 정말 높은지 확인하기 위해 t-test를 진행해보았다. 그 결과 p-value가 1.0으로 유의수준 0.05보다 크기 때문에 귀무가설을 기각할 수 없었다.

즉, retention_7이 False일때 평균은 True일때 평균보다 크지 않다고 할 수 있다.

- 따라서 7일 후에 접속한 유저들이 게임을 더 많이 플레이 했을 것이라고 볼 수 있다.

Insight

- 게임의 첫 번째 게이트를 더 높이 설정해도 전반적인 평균 게임 플레이 횟수에 영향을 미치지 않았다.
 - 옮긴 게이트 위치가 사용자 이탈을 촉진하지 않았거나 사용자들이 계속해서 게임에 참여했을 가능성이 있다고 예상해볼 수 있다.
- **version과 retention_1**
 - retention_1은 유저가 게임을 시작한 1일 후에 다시 돌아와서 플레이 했는지 여부를 나타내는데, 두 변수가 독립이기 때문에 version이 첫 날 접속에는 큰 영향을 미치지 않는다는 것을 알 수 있었다.
- **version과 retention_7**
 - 게임의 변경 사항이 사용자가 더 오랜 기간 동안 게임을 플레이하게 하는 데 영향을 미칠 수 있다는 것을 시사할 수 있다.



- 알림 시스템 및 보상 시스템이 있으면 7일 후에 들어오는 전체 접속자 수를 늘릴 수 있을 것 같다. 예를 들어, gate를 직면했을 때 기다리는 시간을 줄여주거나 결제 비용을 할인해주는 등의 혜택을 제공해주면 1일 후와 7일 후에 접속하는 유저가 많아질 것이라고 기대할 수 있다.
- 게임 내 커뮤니티를 활성화한다면 유저 접속률이 오를 것이라고 기대한다. (ex. 메이플스토리, 테일즈런너 등)

<추가 분석이 필요할 듯>

- 사용자 행동과 관련한 다른 변수들에 대한 분석이나 사용자 특성을 세분화하는 작업을 진행해 미묘한 변화를 찾아 게이트 이동이 어떻게 영향을 미치는지 확인하는 것이 필요할 것 같음. (인 앱 결제, 사용자 연령, 매크로)
- 설문조사를 통해 사용자 피드백을 받아 게임 게이트 위치에 대한 선호도를 파악하는 추가 작업을 진행하면 사용자 접속률에 영향을 미치는 요소들을 더 자세히 이해하고 개선할 수 있지 않을까...