



숙명여자대학교 경제학부 16학번 조유민

# 게임 ESD 서비스 크롤링

-Steam 기반 게임 및 유저 분석-

# 01 중간발표 리마인드

## 분석 방향 수정

### 04 지난 애로사항 및 한계점 앞으로 보완해야 할 부분은?

✓ "새로운 환경에서 로그인 되었습니다." (로그인 실패)  
→ 별도의 selenium setting 필요

✓ 신규 구매 여부를 파악할 수는 없음  
→ 현재 플레이 중인 유저 수나 최근 리뷰 수  
할인 행사가 해당 게임의 실질 구매로 이어

✓ 부정적인 평가를 받는 게임에 대한 분석이 미흡할 것으로  
→ 최고 매출/실시간 인기 게임은 물론이고  
할인 게임도 대체적으로 긍정적인 평가를

### 03 크롤링 대상 및 목표 할인 게임 분석

insight 하루 최고 유저 플레이 수 대비 실시간 유저 플레이 수가 적은 게임은 몰입도가 약한 게임일 것이다.

검증 방법 하루 최고 유저 플레이 수 대비 실시간 유저 플레이 수가 적은 게임들(set B)의 최근 긍정적

insight 최근 리뷰 수가 적은 게임은 할인 행사로 리뷰

검증 방법 할인 당시 최근 리뷰 수가 적었던

1 할인 게임들의 전반적인 특징 분석  
→ 호감도와 장르, 태그 및 가격, 발매일 등 그래프로 시각화

### 02 크롤링 대상 및 목표 & DB 구조 및 저장 전략 최고 매출 게임 분석

Long-run 게임의 특징은?

가격대 분포는? 무료 게임 비율이 높을까?  
무조건 호감도가 높을까?  
어떤 장르가 대세일까?  
많이 사용하는 검색 태그는 무엇일까?

STEP 1

Top selling에서 30개 게임 크롤링 후  
table 1에 저장  
※ 일별 변동이 크지 않기 때문에 일회성 크롤링  
단, 즉시 모를 가능성에 대비해서  
첫 날과 마지막 날 총 2번 크롤링  
[직접 생성한 column]  
- ID: 게임 고유 번호 임의로 생성(Primary key)

<Table 1>

ID	Title	Original Price	Final Price	Genre	Popular Tags	Review	link
1	Grand Theft Auto V	₩33,000	₩16,500	Action, Adventure	Open World, Action, Multiplayer, Crime, ...	Mostly Positive	271390/Grand_Theft_Auto_V/
2	Destiny 2	NA	Free to Play	Action, Adventure, Free to Play	FPS, Multiplayer, Sci-fi, Online Co-op, ...	Mostly Positive	1085660/Destiny_2/

# 01 중간발표 리마인드

## 분석 방향 수정

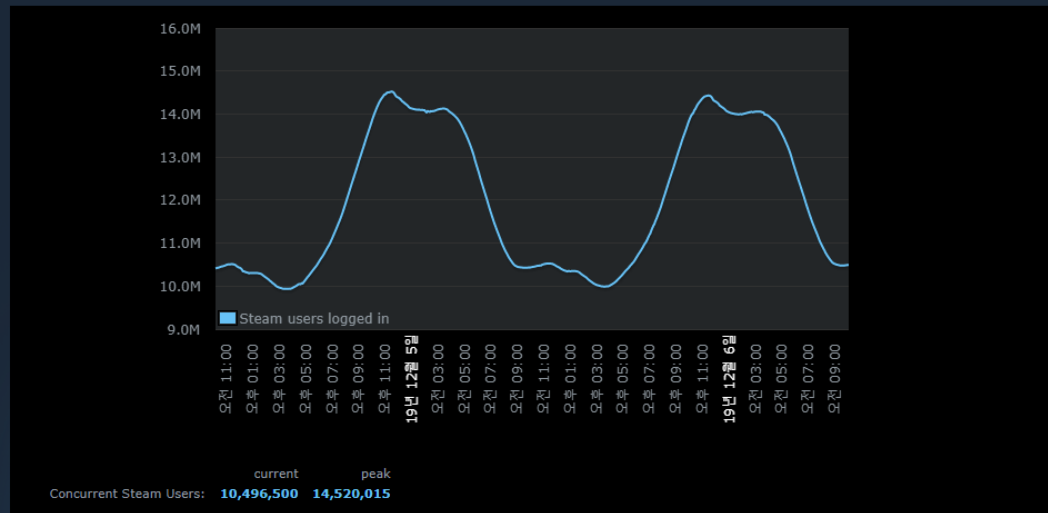
Steam Stats

### STEAM & GAME STATS

UPDATED: 6 DECEMBER, 2019 @ 10:25AM

Our on-going goal with Steam is to improve the service we offer customers. We believe that by sharing this data, we'll be able to spot problems earlier, improve the Steam service more efficiently, and ultimately build better products and experiences. Let us know what you think.

#### Concurrent Steam Users (most recent 48 hours)



#### Top games by current player count

CURRENT PLAYERS	PEAK TODAY	GAME
268,248	542,145	Dota 2
265,168	614,239	Counter-Strike: Global Offensive
106,777	622,679	PLAYERUNKNOWN'S BATTLEGROUNDS
88,426	123,283	Halo: The Master Chief Collection
69,367	104,987	Destiny 2
61,406	110,504	Tom Clancy's Rainbow Six Siege

#### Steam Hardware & Software Survey

Steam Hardware & Software Survey

#### Steam Stats

Steam Download Stats

Steam Support Stats

“할인 행사에 따른 유저들의 구매력의 변화가 나타나는가?”

“가격대, 할인율 등의 범주에 따른 유의미한 차이가 있는가?”





# 02 크롤링 및 DB 저장

## price/details/current player

### STEP 1 가격과 실시간 접속자 수 크롤링

Results exclude some products based on [your preferences](#)

Game	Price
Red Dead Redemption 2	₩ 66,000
Halo: The Master Chief Collection	₩ 39,900
Darksiders Genesis	₩ 35,400

Top games by current player count

CURRENT PLAYERS	PEAK TODAY	GAME
292,795	542,145	Dota 2
267,935	614,239	Counter-Strike: Global Offensive
247,799	622,679	PLAYERUNKNOWN'S BATTLEGROUNDS
85,609	102,916	Halo: The Master Chief Collection

topsell_price	
PK	ID
	TITLE
	ORIGINAL_PRICE
	FINAL_PRICE
	DISCOUNT
	LINK

concur_price	
PK	ID
	TITLE
	ORIGINAL_PRICE
	FINAL_PRICE
	DISCOUNT
	LINK

cpt_top100	
PK	ID
	TIME_OF_CRAWL
	TITLE
	CUR_USER
	PEAK_TODAY

### STEP 2 2개 테이블 소속 여부에 따라 3가지 type으로 분류하여 세부 정보 크롤링

longrun_game	
PK	ID
FK	TITLE
	RELEASE_DATE
	GENRE
	POPULAT_TAGS
	RATINGS

rising_game	
PK	ID
FK	TITLE
	RELEASE_DATE
	GENRE
	POPULAT_TAGS
	RATINGS

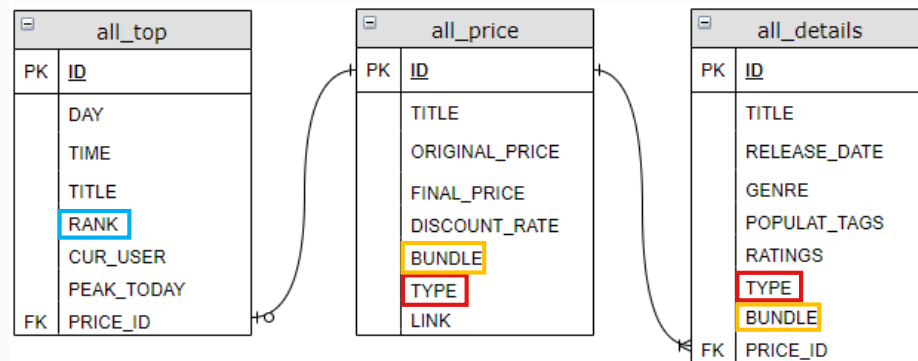
topseller_game	
PK	ID
FK	TITLE
	RELEASE_DATE
	GENRE
	POPULAT_TAGS
	RATINGS
	BUNDLE

- DAY = 일
- TIME = 시
- RANK = 순위
- BUNDLE = 묶음 상품 여부
- TYPE = 게임 타입
- PRICE\_ID = all\_price의 ID

### STEP 3 최종 테이블 - 테이블 병합 및 컬럼 추가

#### E-R Diagram

[2019-11-17~2019-11-29]



※ all\_price 테이블 상의 게임만 추출

# 02 크롤링 및 DB 저장

## price/details/current player

STEP 1 가격과 실시간 접속자 수 크롤링

STEP 2 2개 테이블 소속 여부에 따라 3가지 type으로 분류하여 세부 정보 크롤링

Results exclude some products based on [your preferences](#)

NEW AND TRENDING	TOP SELLING	WHAT'S BEING PLAYED	UPCOMING
Red Dead Redemption 2			
Halo: The Master Chief Collection			
Darksiders Genesis			

Top games by current player count

CURRENT PLAYERS	PEAK TODAY	GAME
292,795	542,145	Dota 2
267,935	614,239	Counter-Strike: Global Offensive
247,799	622,679	PLAYERUNKNOWN'S BATTLEGROUNDS
85,609	102,916	Halo: The Master Chief Collection

"Longrun game은 판매량이나 접속량 모두 높은 게임이므로 가격대가 높지 않은 편이고 평가도 긍정적이지 않을까?"

longrun_game	rising_game	topseller_game
FK TITLE	FK TITLE	FK TITLE
RELEASE_DATE	RELEASE_DATE	RELEASE_DATE
GENRE	GENRE	GENRE
POPULAT_TAGS	POPULAT_TAGS	POPULAT_TAGS
RATINGS	RATINGS	RATINGS
		BUNDLE

"Rising game은 많이 팔리지는 않지만 사람들이 요즘 많이 하고 있는 게임이므로 가격대가 높거나 할인율이 낮지 않을까?"

STEP 3 최종 테이블 - 테이블 병합 및 컬럼 추가

E-R Diagram

[2019-11-17~2019-11-29]

PK	ID
	TITLE
	ORIGINAL_PRICE
	FINAL_PRICE
	DISCOUNT
	LINK

PK	ID
	TITLE
	ORIGINAL_PRICE
	FINAL_PRICE
	DISCOUNT
	LINK

"Topseller game은 사람들이 구매는 했지만 지금 당장 하는 게임은 아니므로 접속자 수가 적고 시간대와 큰 상관관계가 없지 않을까?"

TITLE
TIME_OF_CRAWL
TITLE
CUR_USER
PEAK_TODAY

FK	PRICE_ID
	TITLE
	TIME
	TITLE
	RANK
	CUR_USER
	PEAK_TODAY

FK	PRICE_ID
	TITLE
	ORIGINAL_PRICE
	FINAL_PRICE
	DISCOUNT_RATE
	BUNDLE
	TYPE
	LINK

FK	PRICE_ID
	ID
	TITLE
	RELEASE_DATE
	GENRE
	POPULAT_TAGS
	RATINGS
	TYPE
	BUNDLE

- DAY = 일
- TIME = 시
- RANK = 순위
- BUNDLE = 묶음 상품 여부
- TYPE = 게임 타입
- PRICE\_ID = all\_price의 ID

※ all\_price 테이블 상의 게임만 추출

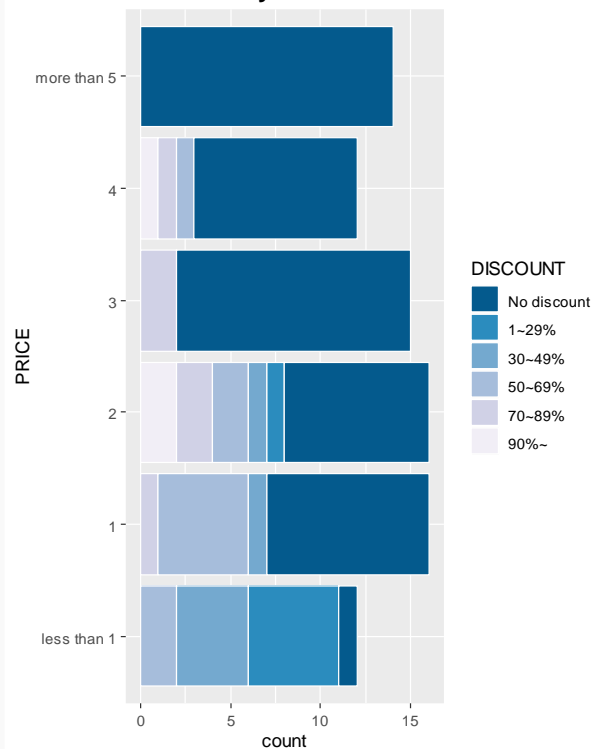
# 03 분석 결과 시각화

STEP 4

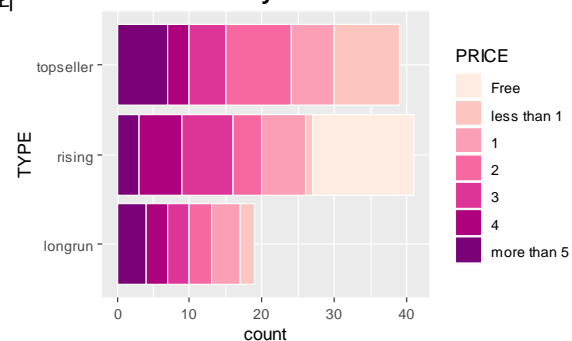
(분석을 위한)테이블 병합 및 컬럼 추가

- FINAL\_PRICE -> **PRICE**(7개 범주)
- DISCOUNT\_RATE -> **DISCOUNT**(6개 범주)
- GENRE -> 게임별 대표 장르 1개만 사용
- RATINGS, RELEASE\_DATE -> 범주 개수 ↓

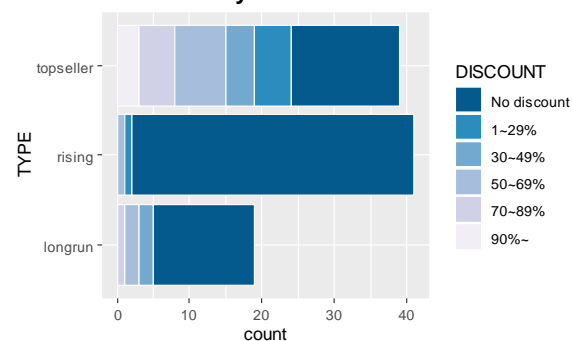
PRICE by DISCOUNT ※ Free 게임 제외



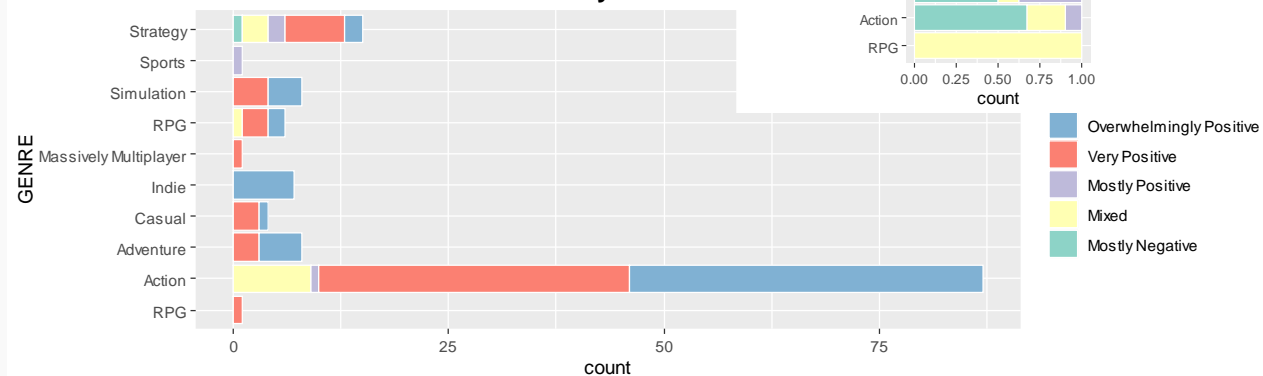
TYPE by PRICE



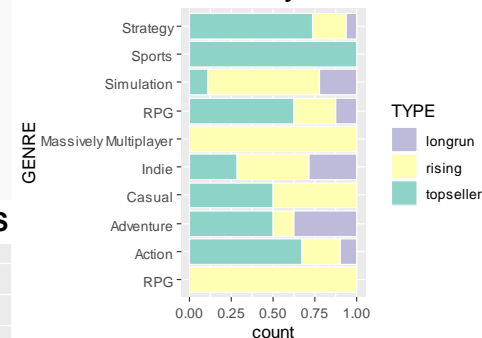
TYPE by DISCOUNT ※ Free 게임 제외



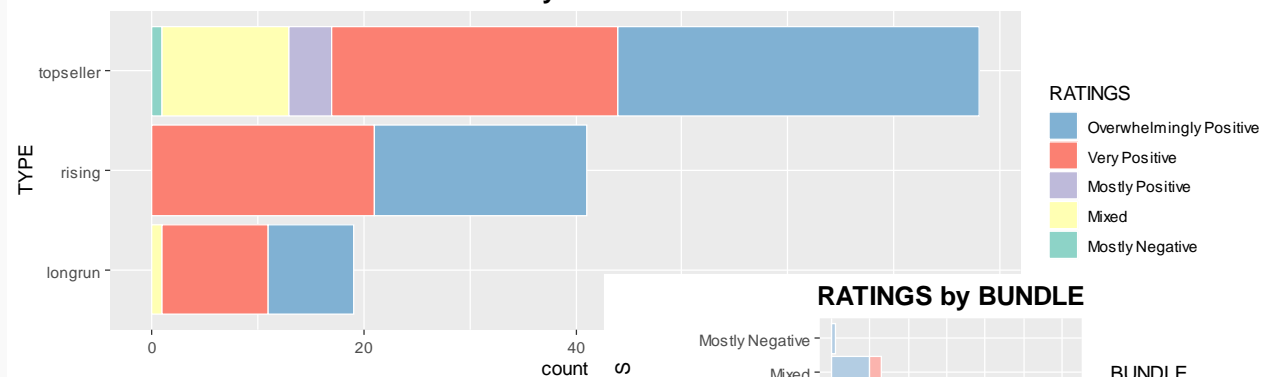
GENRE by RATINGS



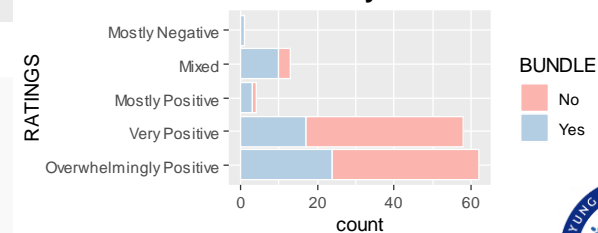
GENRE by TYPE



TYPE by RATINGS



RATINGS by BUNDLE

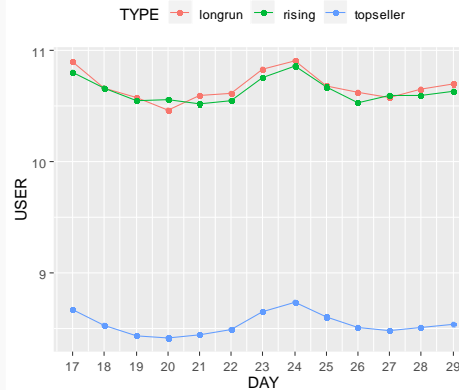


# 03 분석 결과 시각화

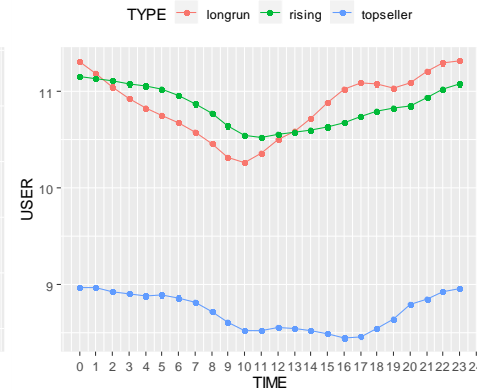
## STEP 4 (분석을 위한)테이블 병합

- all\_top과 all\_price, all\_detail을 각각 병합  
-> all\_top\_p, all\_top\_d 생성
- DAY 기준으로 CUR\_USER의 평균을 시각화
- 특정 날짜의 TIME 기준으로 CUR\_USER 시각화

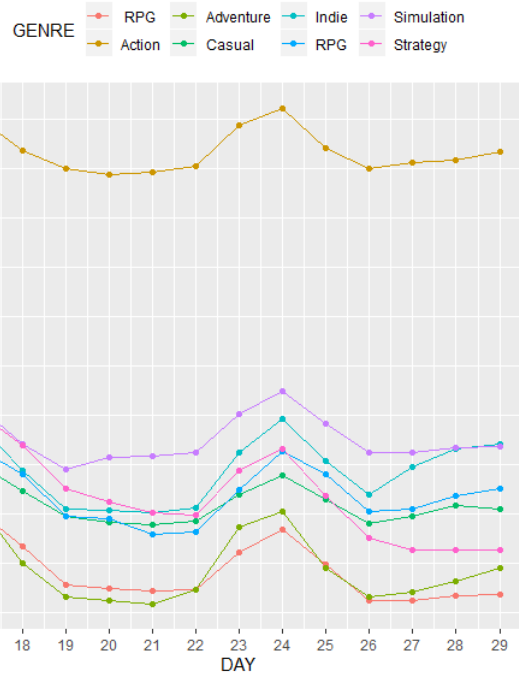
1 2019-11-17 ~ 2019-11-29  
Current user by TYPE



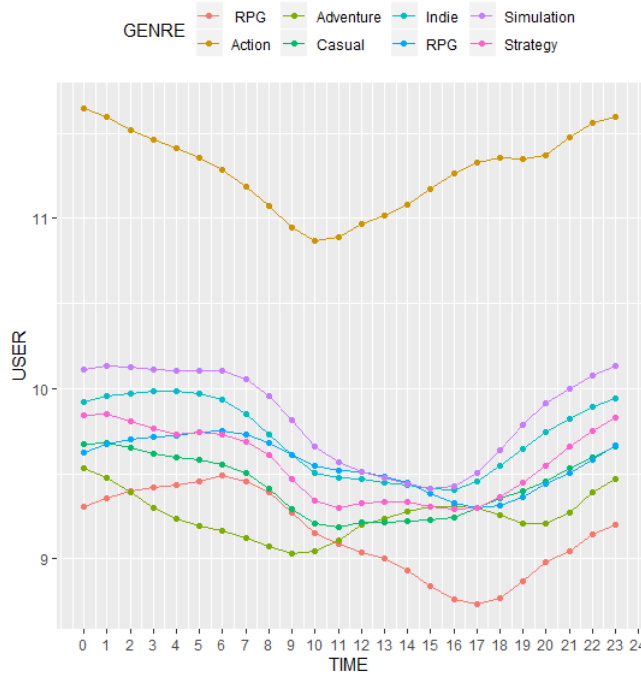
2019-11-24  
Current user by TYPE



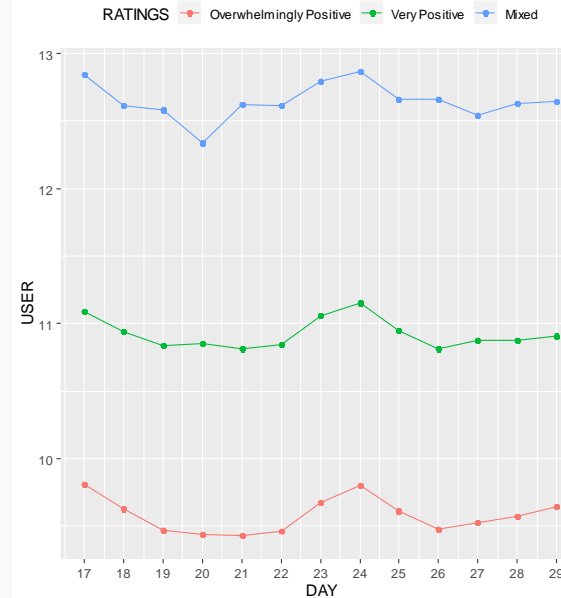
2 2019-11-17 ~ 2019-11-29  
Current user by GENRE



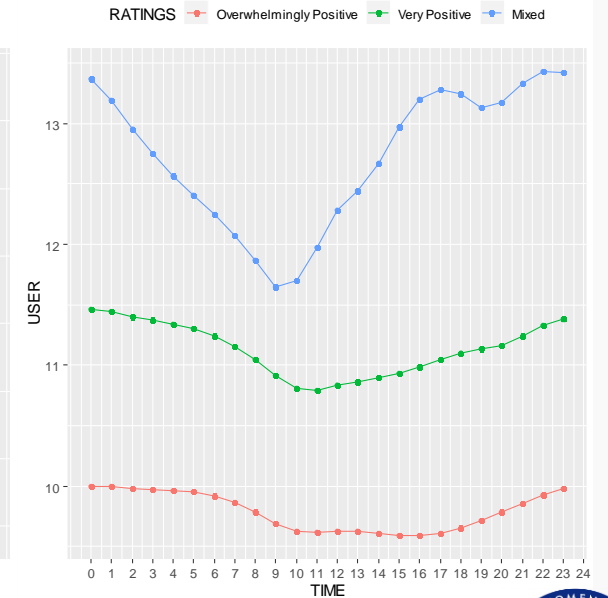
2019-11-24  
Current user by GENRE



3 2019-11-17 ~ 2019-11-29  
Current user by RATINGS



2019-11-24  
Current user by RATINGS

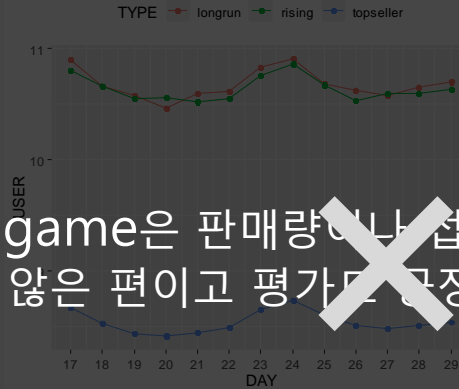


# 03 분석 결과 시각화

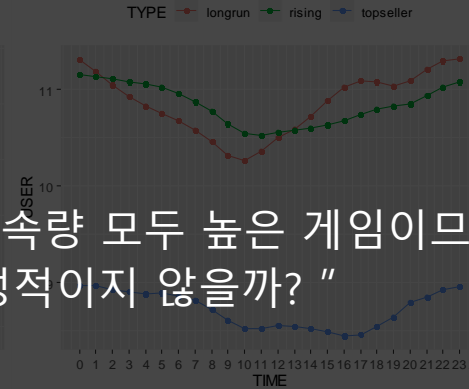
## STEP 4 (분석을 위한)테이블 병합

- all\_top과 all\_price, all\_detail을 각각 병합  
-> all\_top\_p, all\_top\_d 생성
- DAY 기준으로 CUR\_USER의 평균을 시각화
- 특정 날짜의 TIME 기준으로 CUR\_USER 시각화

1 2019-11-17 ~ 2019-11-29  
Current user by TYPE

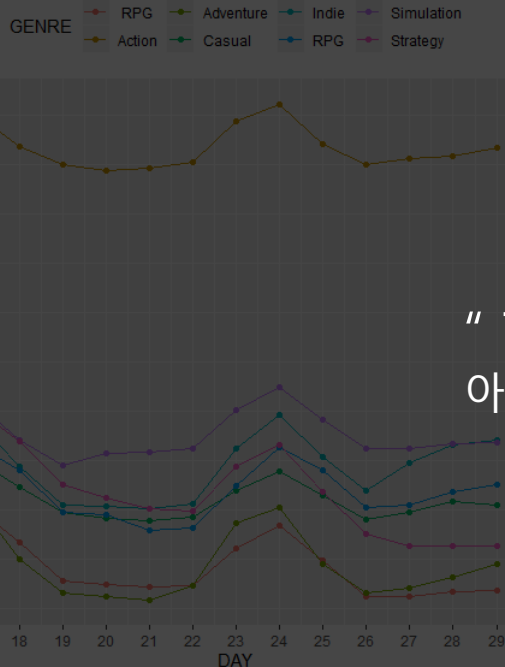


2019-11-24  
Current user by TYPE

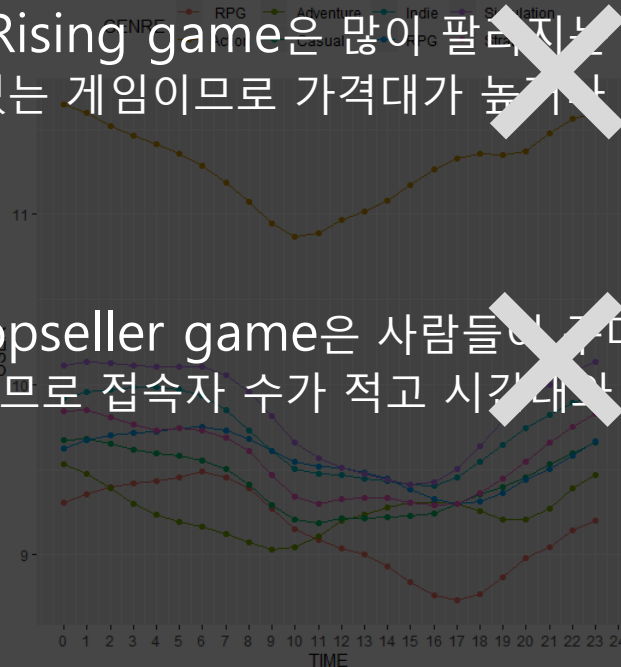


"Longrun game은 판매량이나 접속량 모두 높은 게임이므로 가격대가 높지 않은 편이고 평가도 긍정적이지 않을까?"

2 2019-11-17 ~ 2019-11-29  
Current user by GENRE

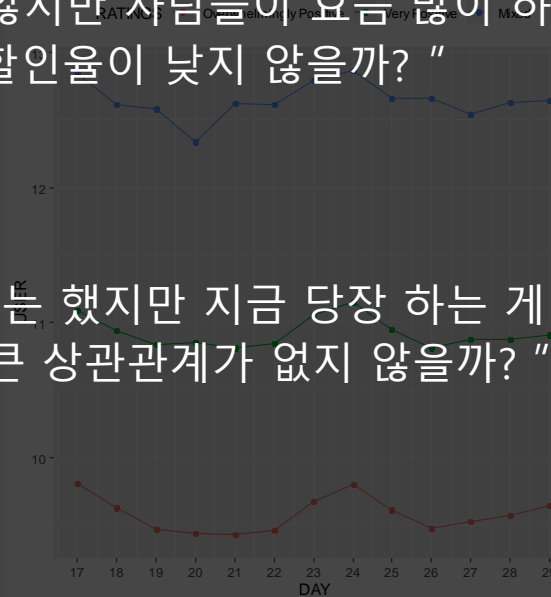


2019-11-24  
Current user by GENRE

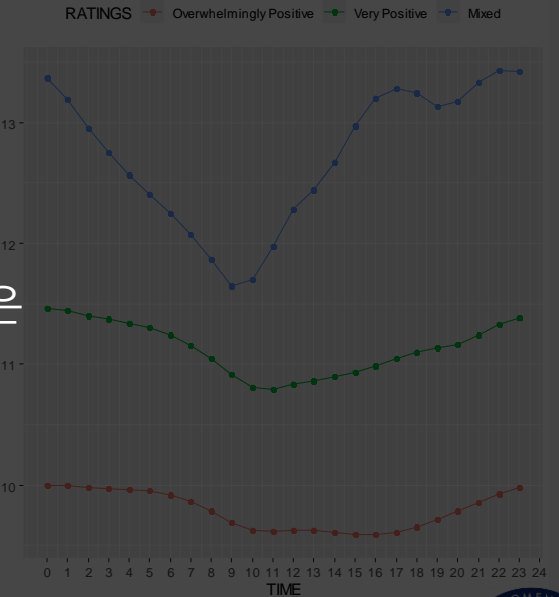


"Rising game은 많이 팔리지는 않지만 사람들이 요즘 많이 하고 있는 게임이므로 가격대가 높지 않고 할인율이 낮지 않을까?"

3 2019-11-17 ~ 2019-11-29  
Current user by RATINGS



2019-11-24  
Current user by RATINGS



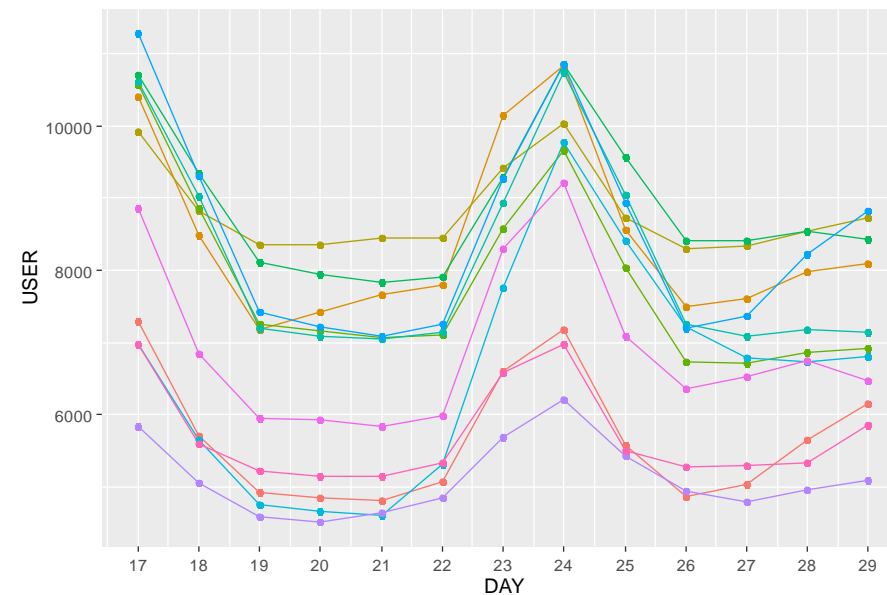
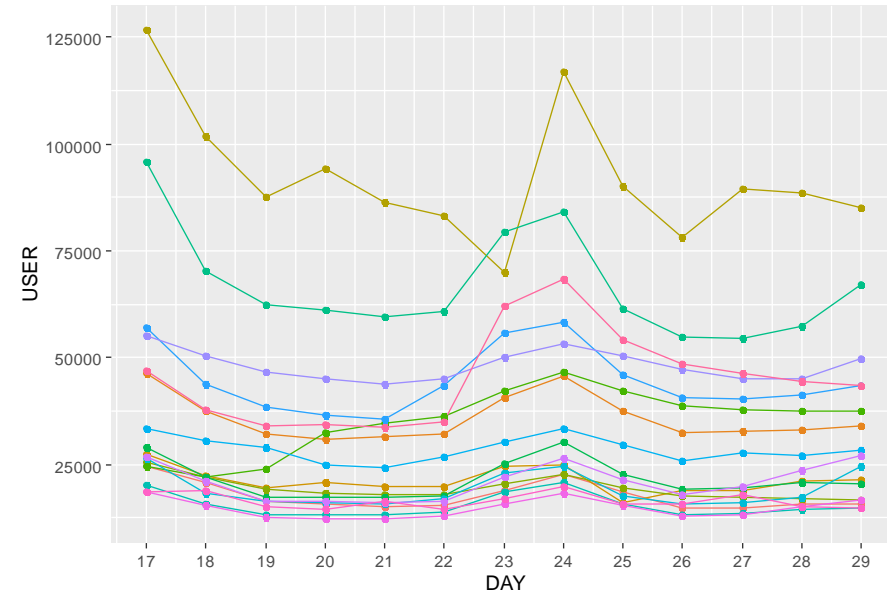
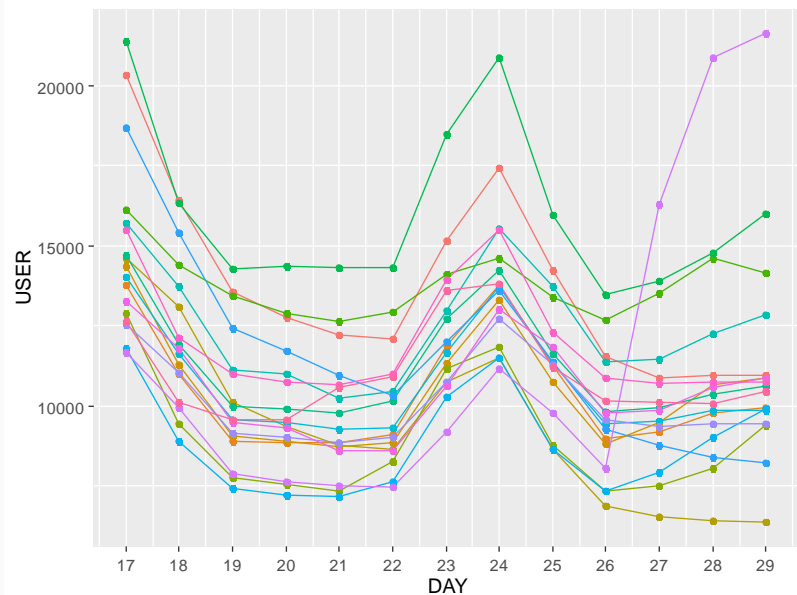
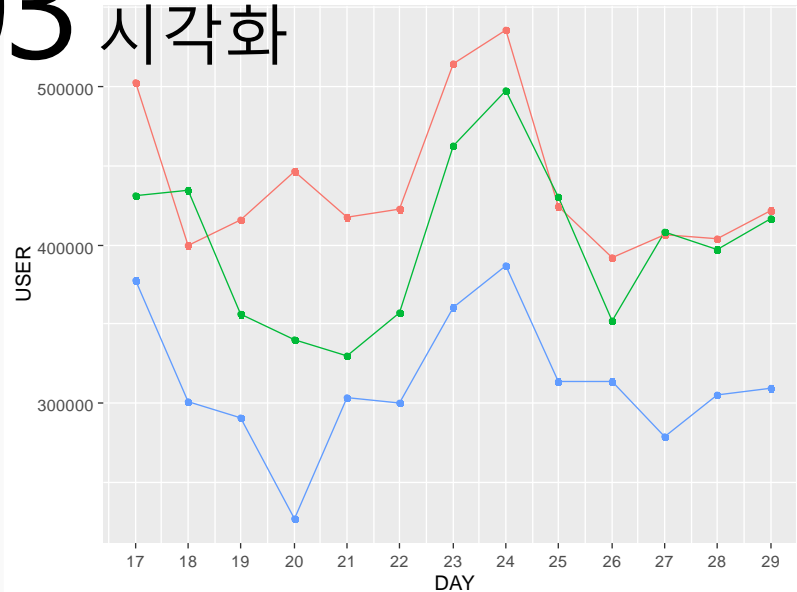
"Topseller game은 사람들이 구매는 했지만 지금 당장 하는 게임은 아니므로 접속자 수가 적고 시간대별 큰 상관관계가 없지 않을까?"



# 03 분석 결과

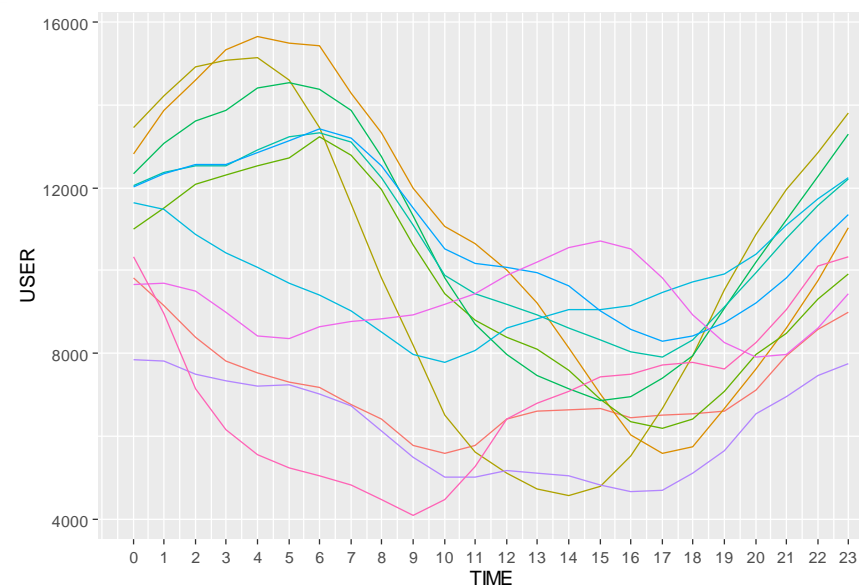
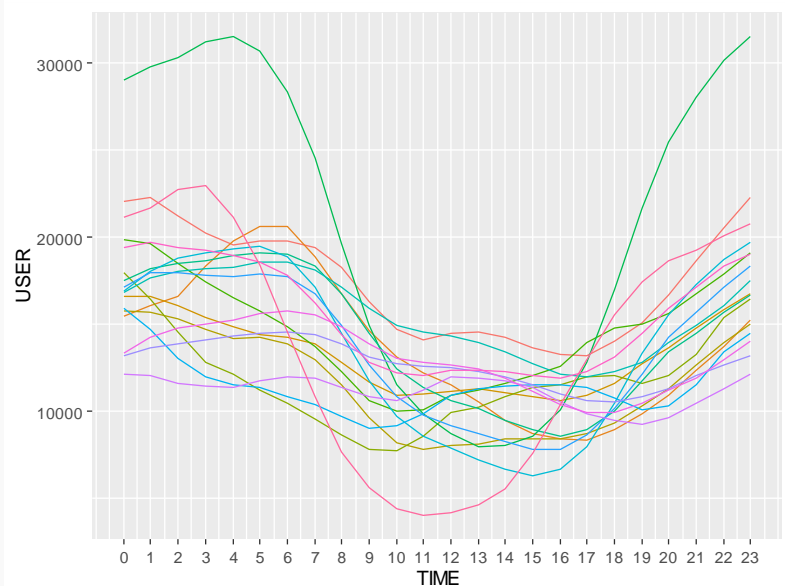
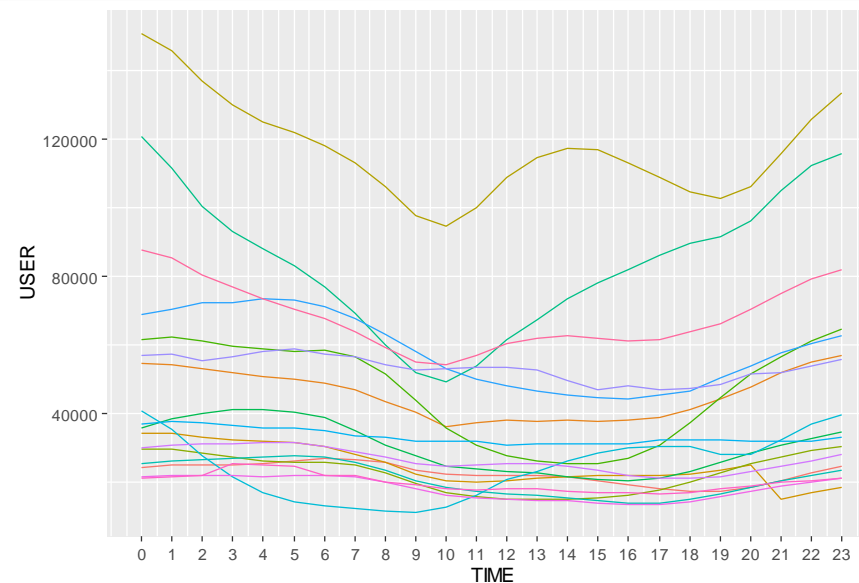
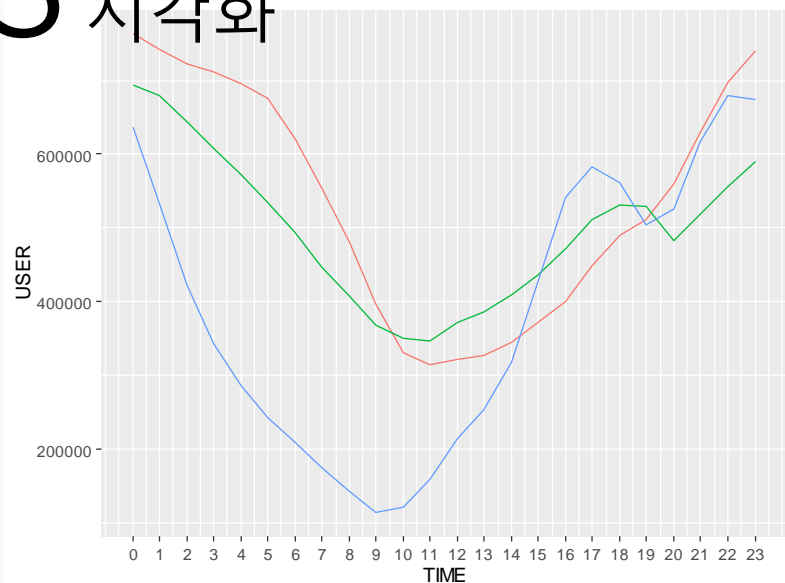
## 일별 평균 접속자 수

### 시각화



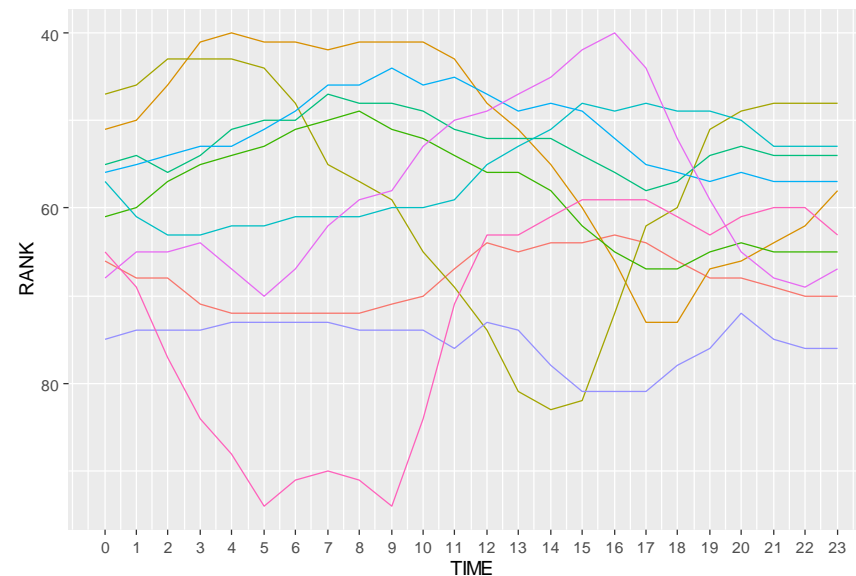
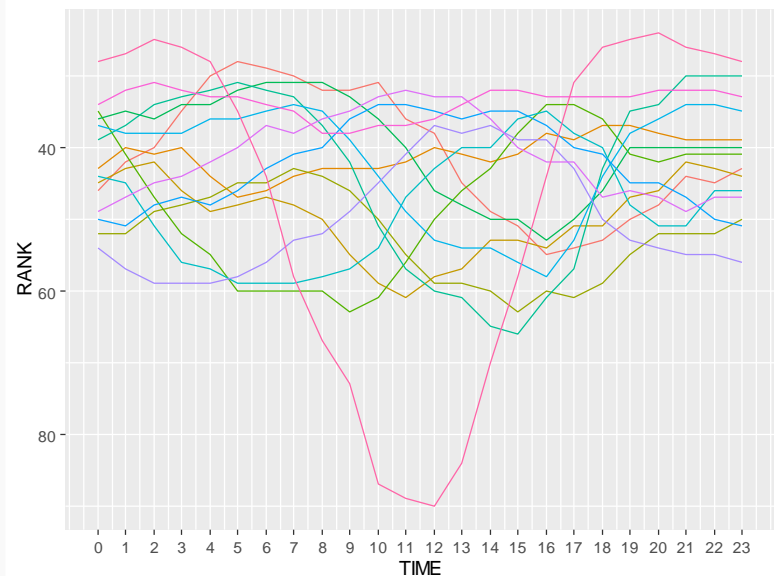
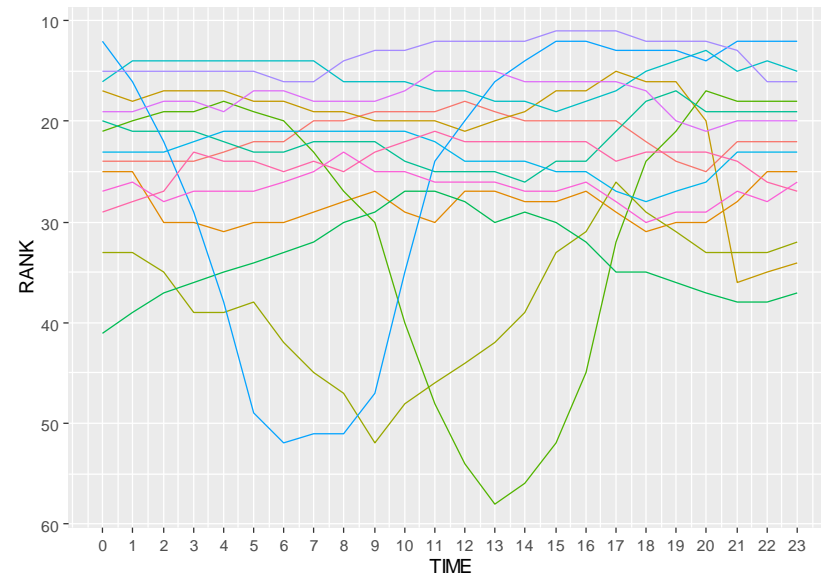
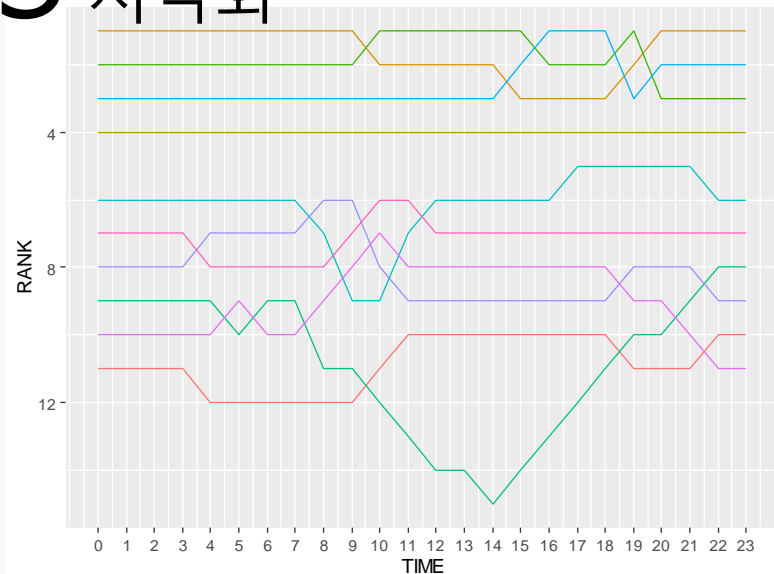
# 03 분석 결과 시각화

2019-11-24 시간별 접속자 수



# 03 분석 결과 시각화

2019-11-24 시간별 순위

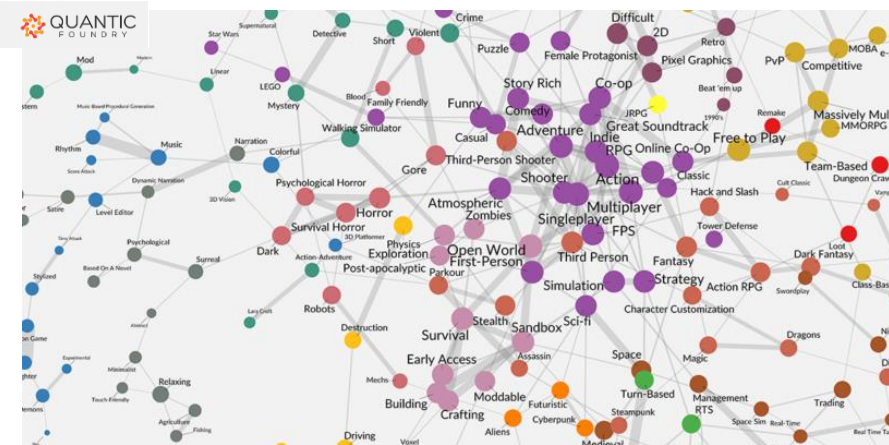
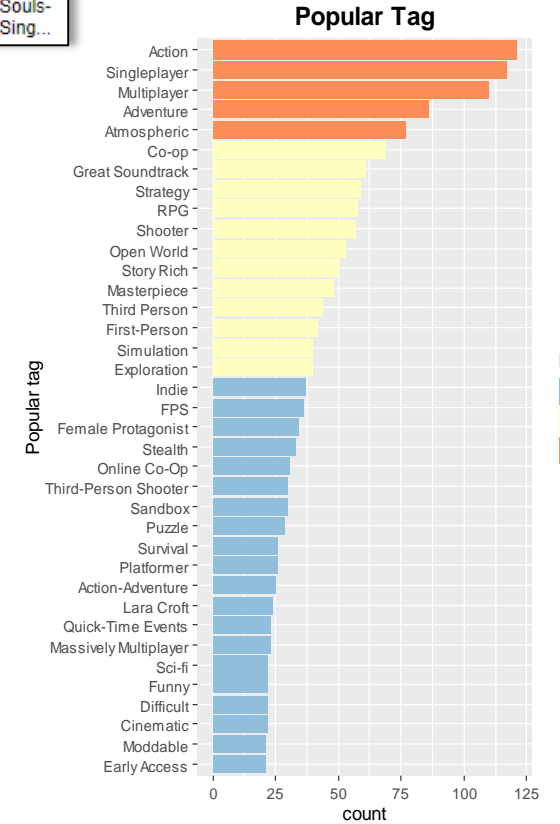


# 04 마무리 애로사항 및 한계점

레벨이 많은 범주형 변수를 제대로 다루지 못하였음

→ 장르와 검색 태그의 경우 다른 시각화나 범주화를 시도해야 함

GENRE	POPULAR_TAGS
Simulation, Sports	Simulation, Sports, Management, Soccer, Footba...
Action, Adventure	Star Wars, Action, Adventure, Souls-like, Sing...



검색 태그 간 연관성을 바탕으로 범주화  
→ 장르를 새롭게 정의

수집한 데이터의 양이 너무 적고 도메인 지식이 부족하였음  
→ 시각화에서 결과 해석으로 이어지는 부분이 미흡함