

SQL을 활용한 EDA 프로젝트

# App Store Game Data EDA

앱 스토어 게임 데이터를 활용한 EDA 분석

TABLE OF CONTENTS

# 목차 소개

01

프로젝트 소개 및 주제 선정 이유

02

프로젝트 세부 분석

03

결론 및 마케팅 전략

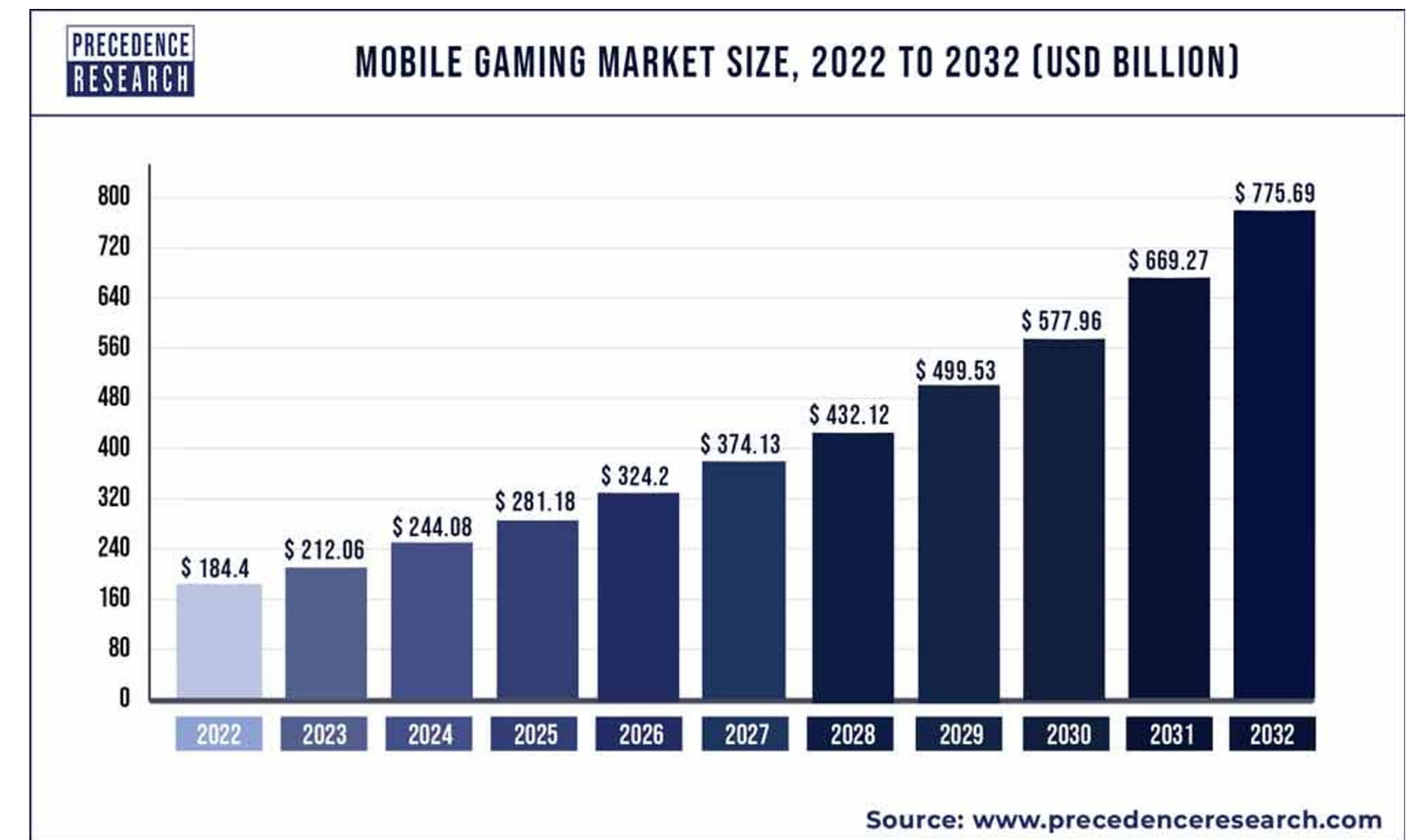
# 01 | 프로젝트 소개 및 주제 선정 이유

## 배경 | 모바일 게임 시장의 성장

- \$184.4 Billion → \$775.69 Billion 가파른 성장 예측
- 전체 게임 산업 매출의 58.9% 차지

## 기대효과

- 사용자 참여 증진
- 시장 트렌드 파악
- 마케팅 전략 최적화



# 01 | 프로젝트 소개 및 주제 선정 이유

“ 핵심 성과 지표(Key Performance Indicator)  
사용자 평점 & 사용자 평점 수 ”



## 01 | 사용자 평점(averageUserRating)

- 게임의 품질, 사용자 만족도와 게임 경험을 반영
- 사용자 평균 평점이 높을수록 긍정적인 반응
- 기대효과: 사용자 유입 ↑ & 장기적인 성공

## 02 | 사용자 평점 수(userRatingCount)

- 게임의 인기와 시장 내 활동성을 반영
- 평점 수가 많을 수록 유저 게임 경험 ↑
- 기대효과: 게임의 인지도 & 시장 점유율 증가

# 01 | 프로젝트 소개 및 주제 선정 이유

## 팀원 소개

**윤장원** : 게임 카테고리화 및 출시 패턴,  
게임 설명 텍스트와 평점 연관성 분석

**박서연** : 사용자 평점 수와 평균 평점의  
상관관계, 파일 크기와 평균 평점 관계

**윤정희** : 연령 적합성, 최소 요구 OS 버전,  
출시 날짜가 평점 개수에 미치는 영향 분석

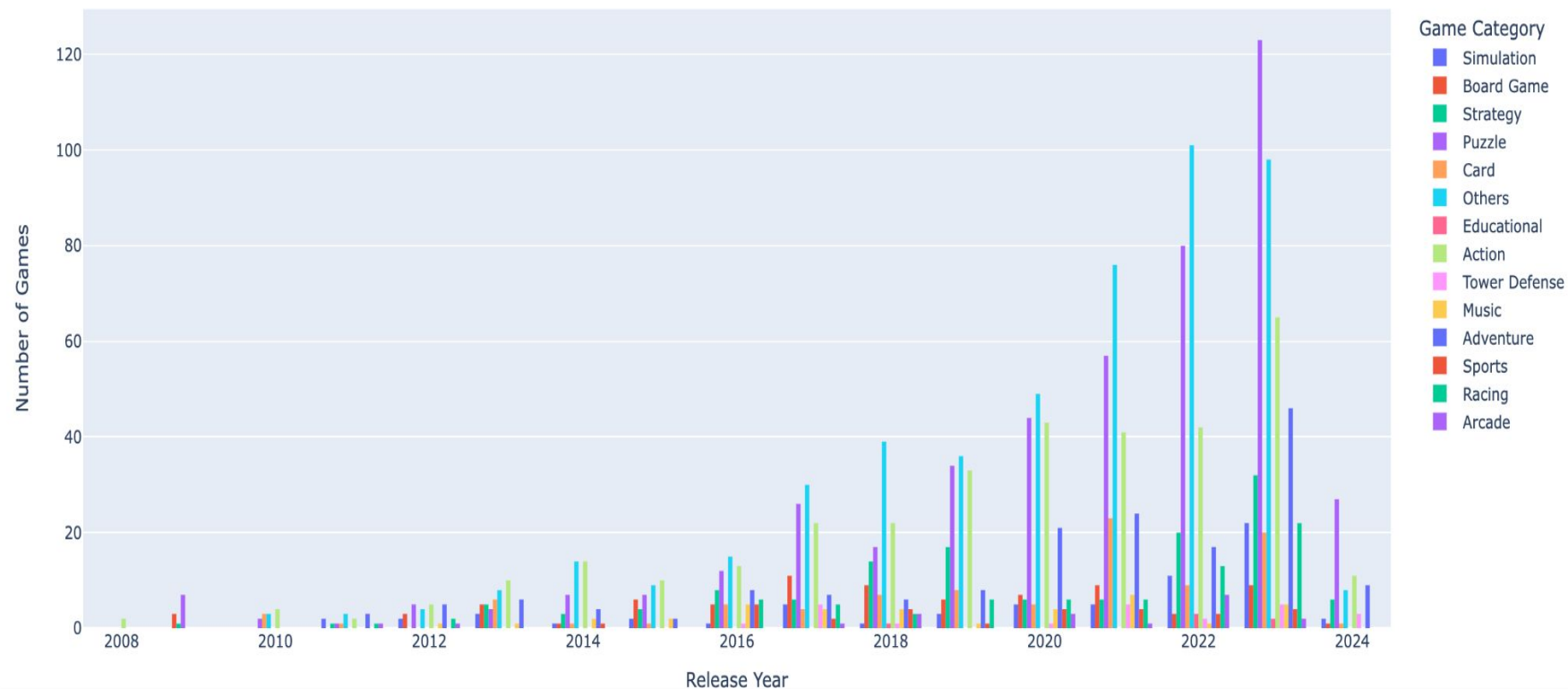
**백승훈** : 애플 로그인 Game Center 연동  
여부와 사용자 평가 수 관계 분석

## 활용 기술

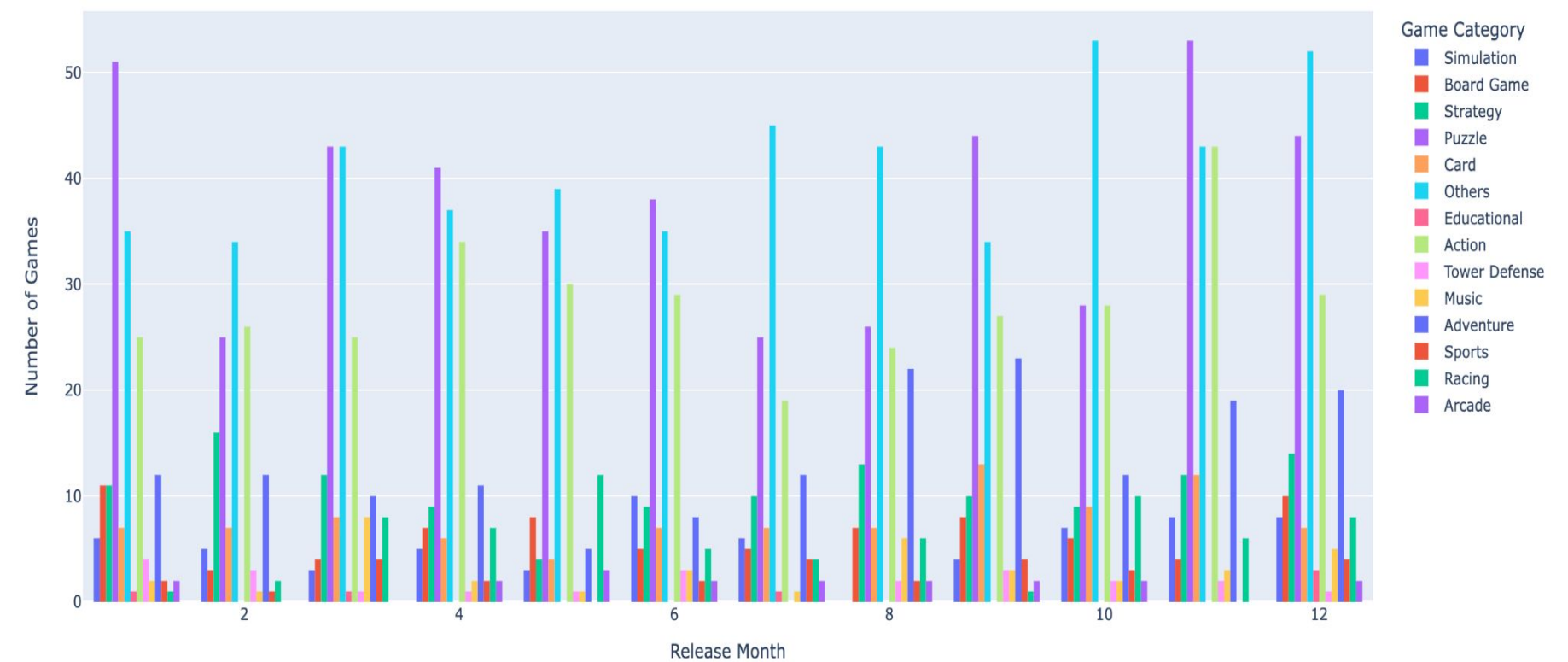


# 02 | 게임 카테고리별 출시 패턴

Number of Games Released per Year by Category



Count of Games Released per Month by Category



## 01 | 연도별 게임 출시 패턴

- 2017년부터 증가, 2020년 후로는 급증
- Puzzle, Action, Simulation의 출시 수 높음
- Puzzle 카테고리 180%의 성장률을 기록

## 02 | 월별 게임 출시 패턴

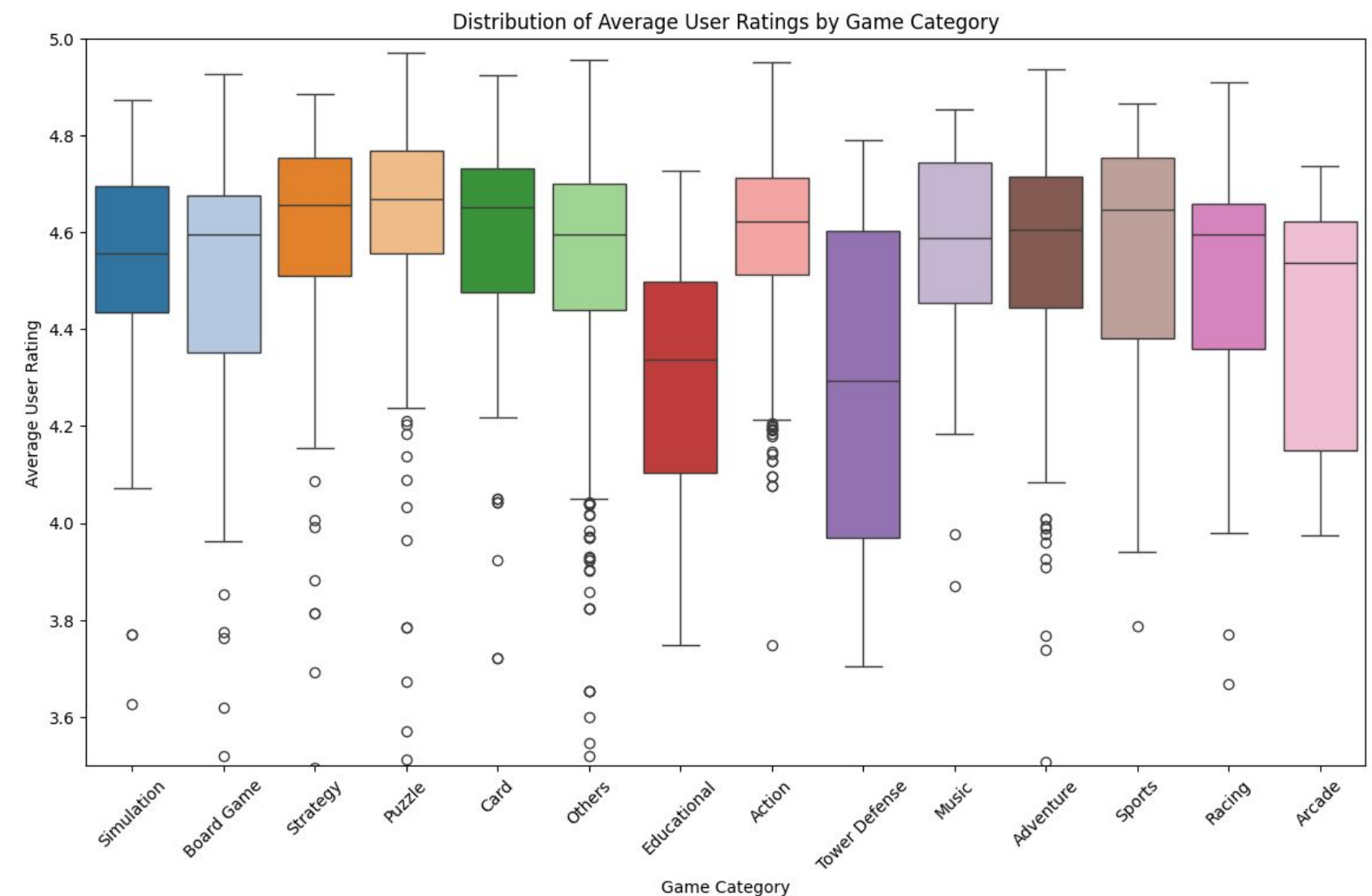
- 주로 10월 부터 1월사이 출시 수 ↑,
- Puzzle, Action 출시 수 높음
- 특히, 11월에 가장 많이 출시

## 02 | 카테고리별 평균 사용자 평점

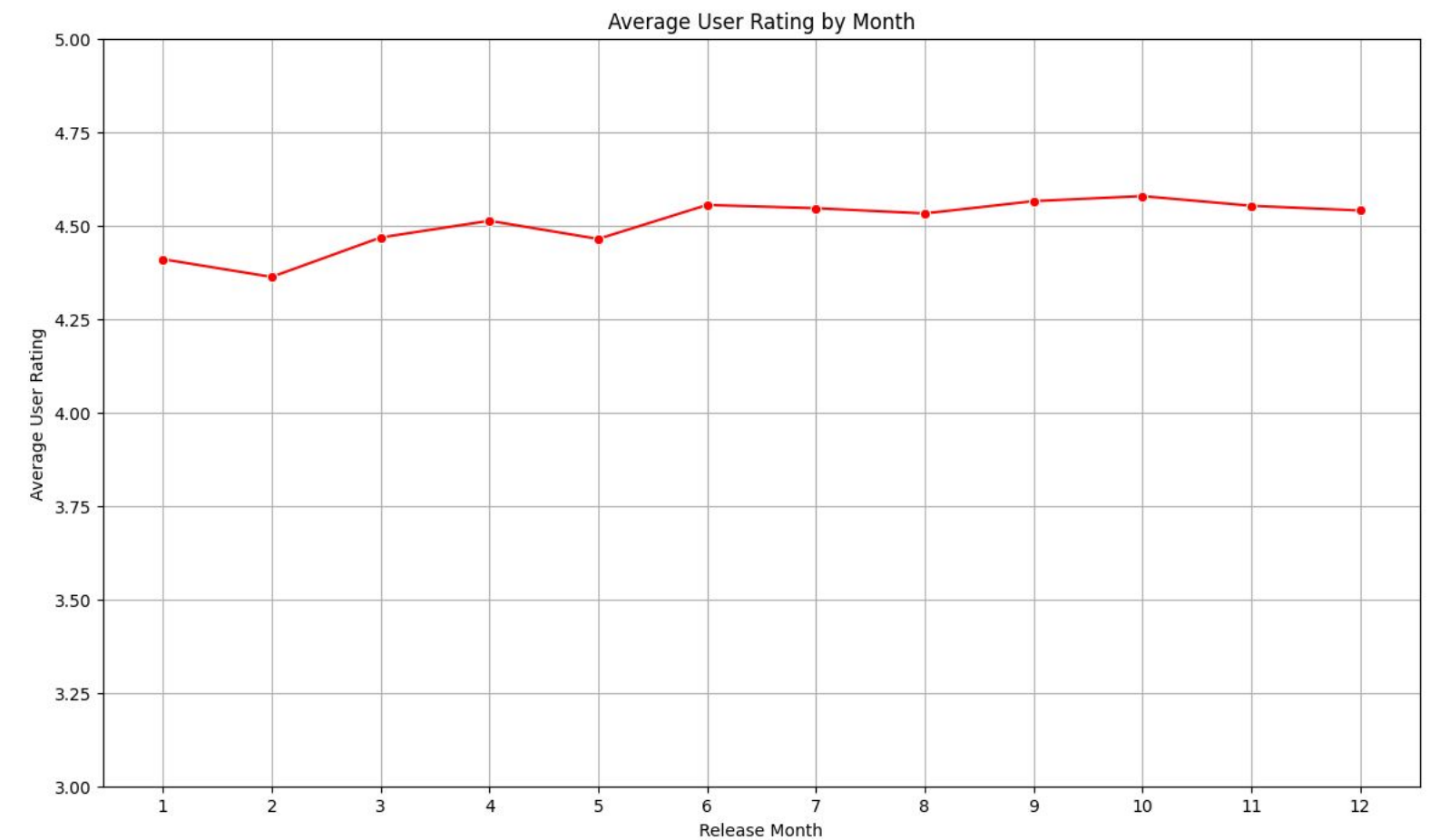
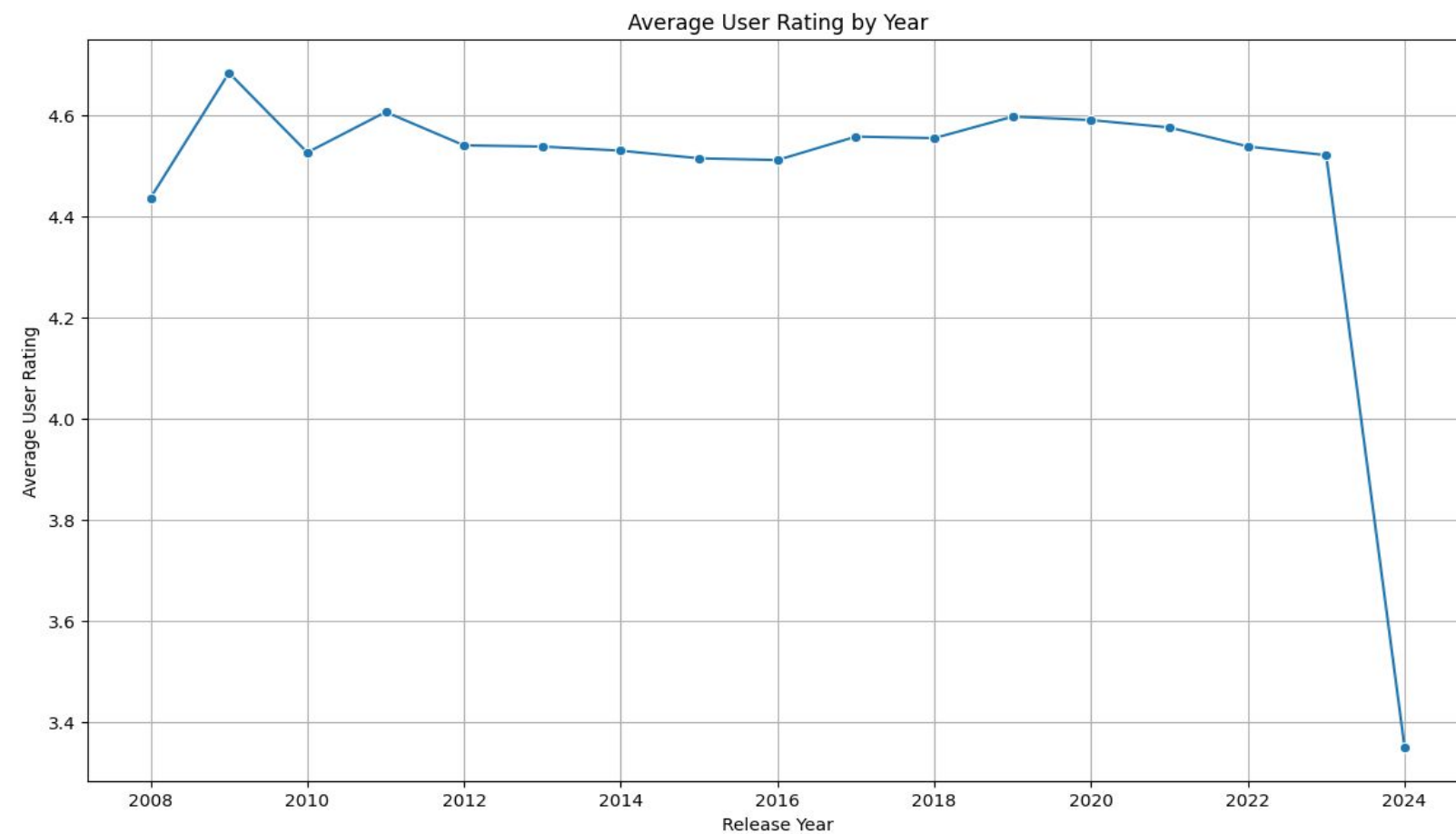
평균 평점은 주로 4~4.8 사이에 위치

Strategy, Puzzle, Action:  
변동성 ↓, 일관된 긍정적인 평가

Educational, Tower Defense, Arcade:  
변동성 ↑, 다양한 평가



# 02 | 연도별 & 월별 사용자 평점 변화



## 01 | 연도별 게임 사용자 평점 패턴

- 연도별 큰 변화 없이 일정하게 유지 (4.4~4.6 사이)

## 02 | 월별 게임 사용자 평점 패턴

- 월별 큰 변화 없이 일정하게 유지 (4.4~4.6 사이)

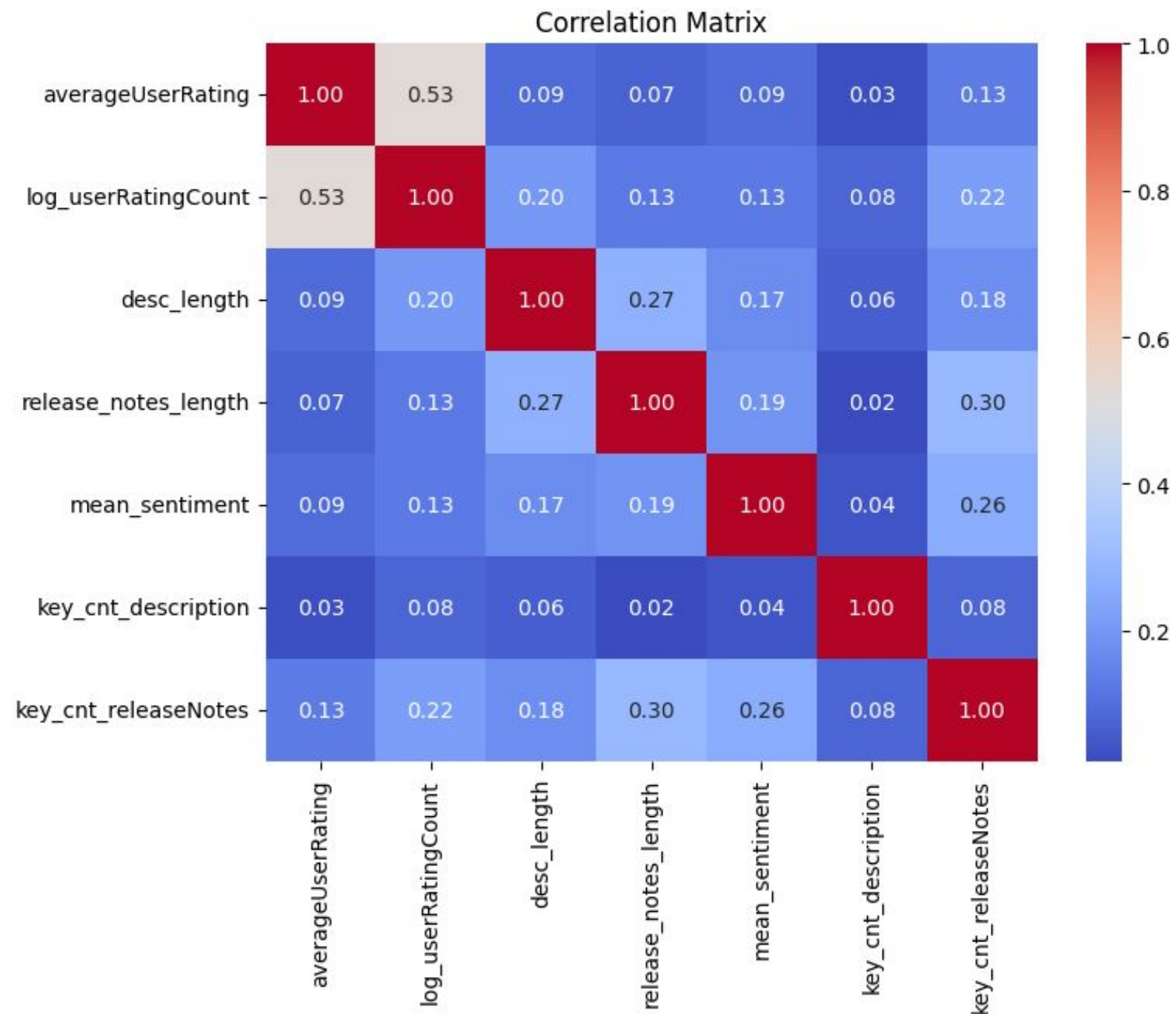
“

사용자 평균 평점은 시간과 관계 없다.

”

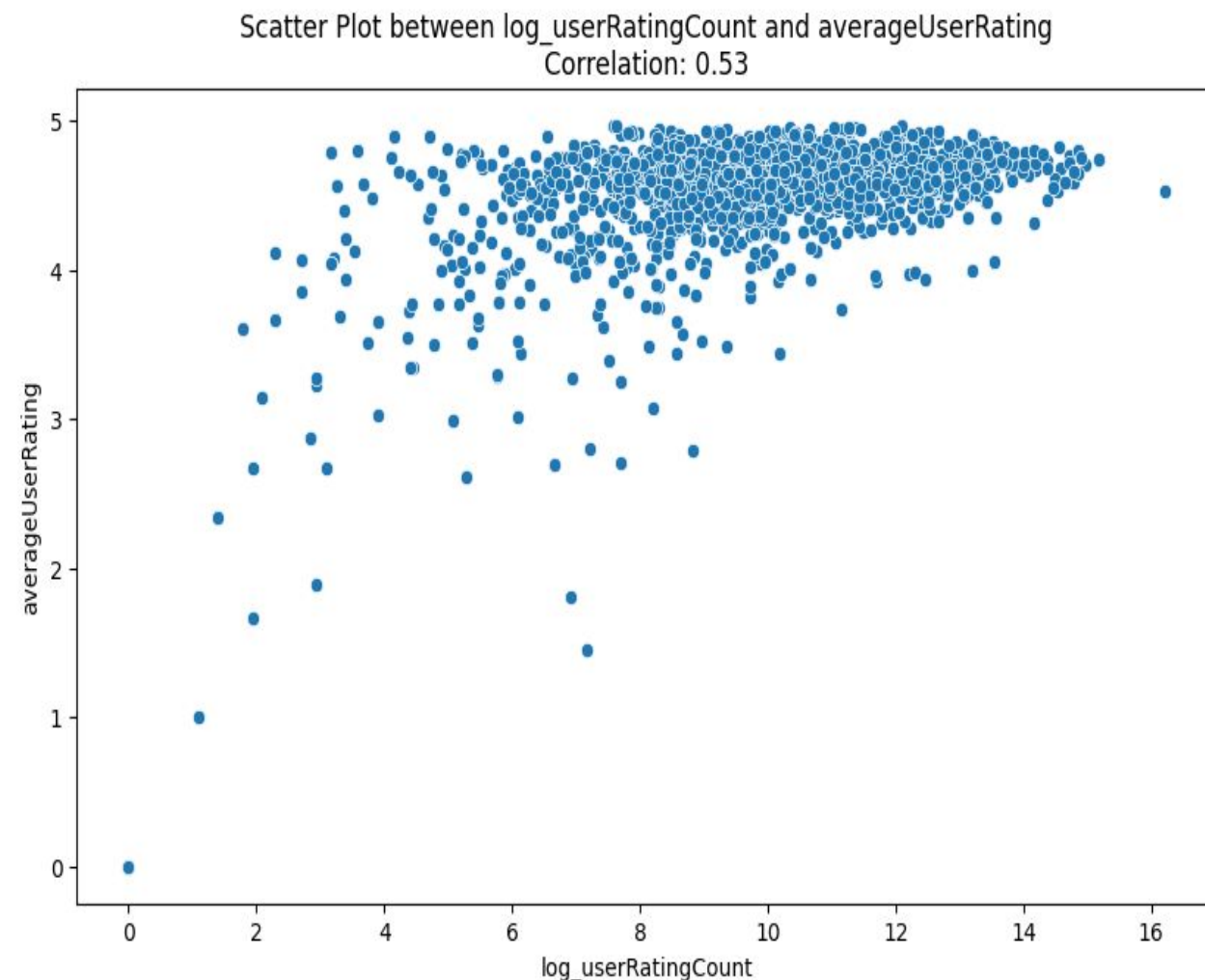


## 02 | 게임 텍스트 분석

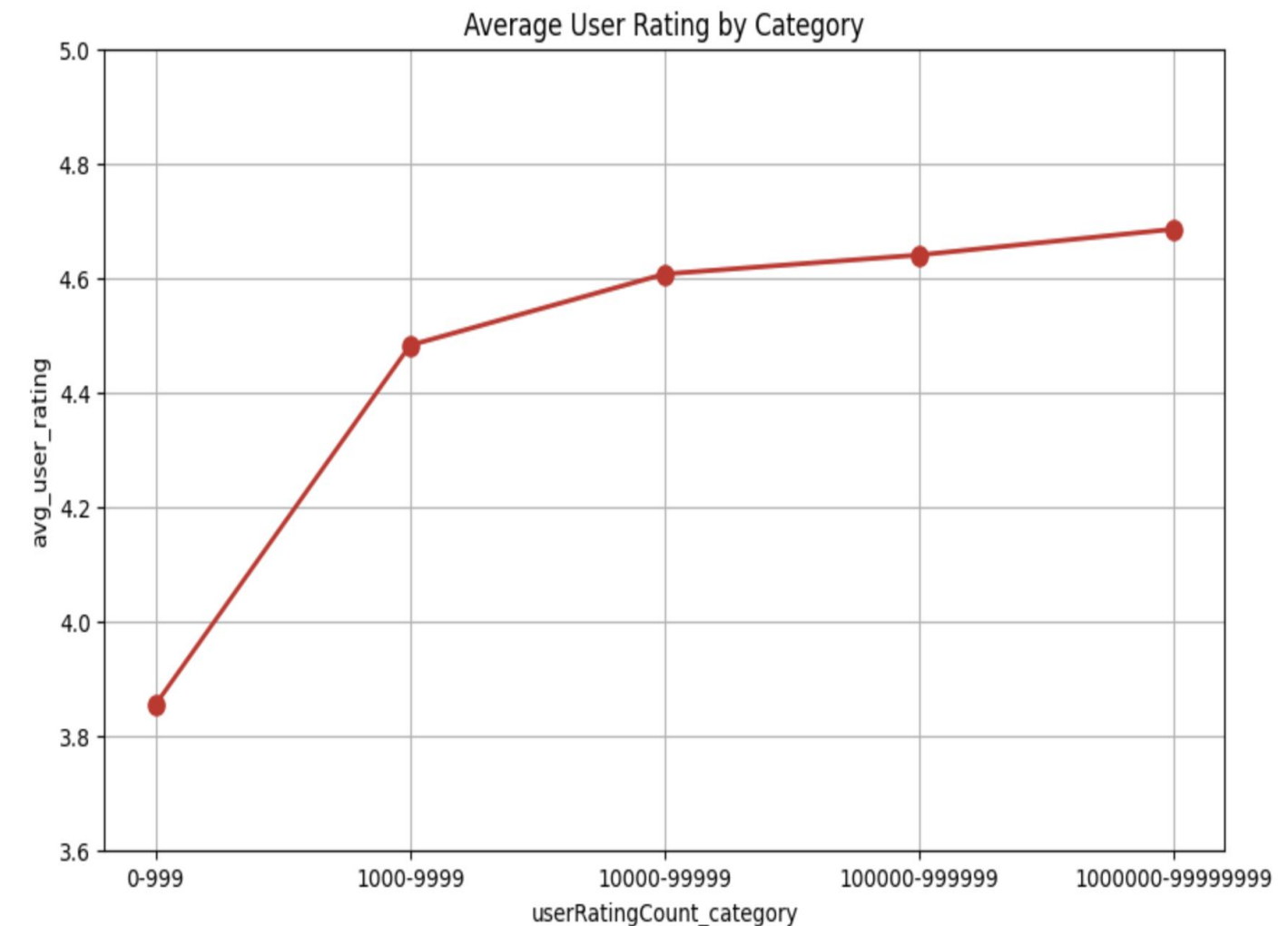


- **유저 평균 점수:**
  - 릴리스 노트 키워드 수와 약간의 상관관계(0.13)
- **유저 평점 개수:**
  - 설명 및 릴리스 노트 길이와 약간의 상관관계(0.2)
  - 평균 감정 점수와 약간의 상관관계(0.13)
  - 릴리스 노트 키워드 수와 약간의 상관관계(0.22)
- **높은 상관관계 부재 이유:**
  - 부족한 데이터, 비선형의 관계

## 02 | 사용자 평가 개수와 평점



사용자 평가 개수와 평점 간의 상관 관계

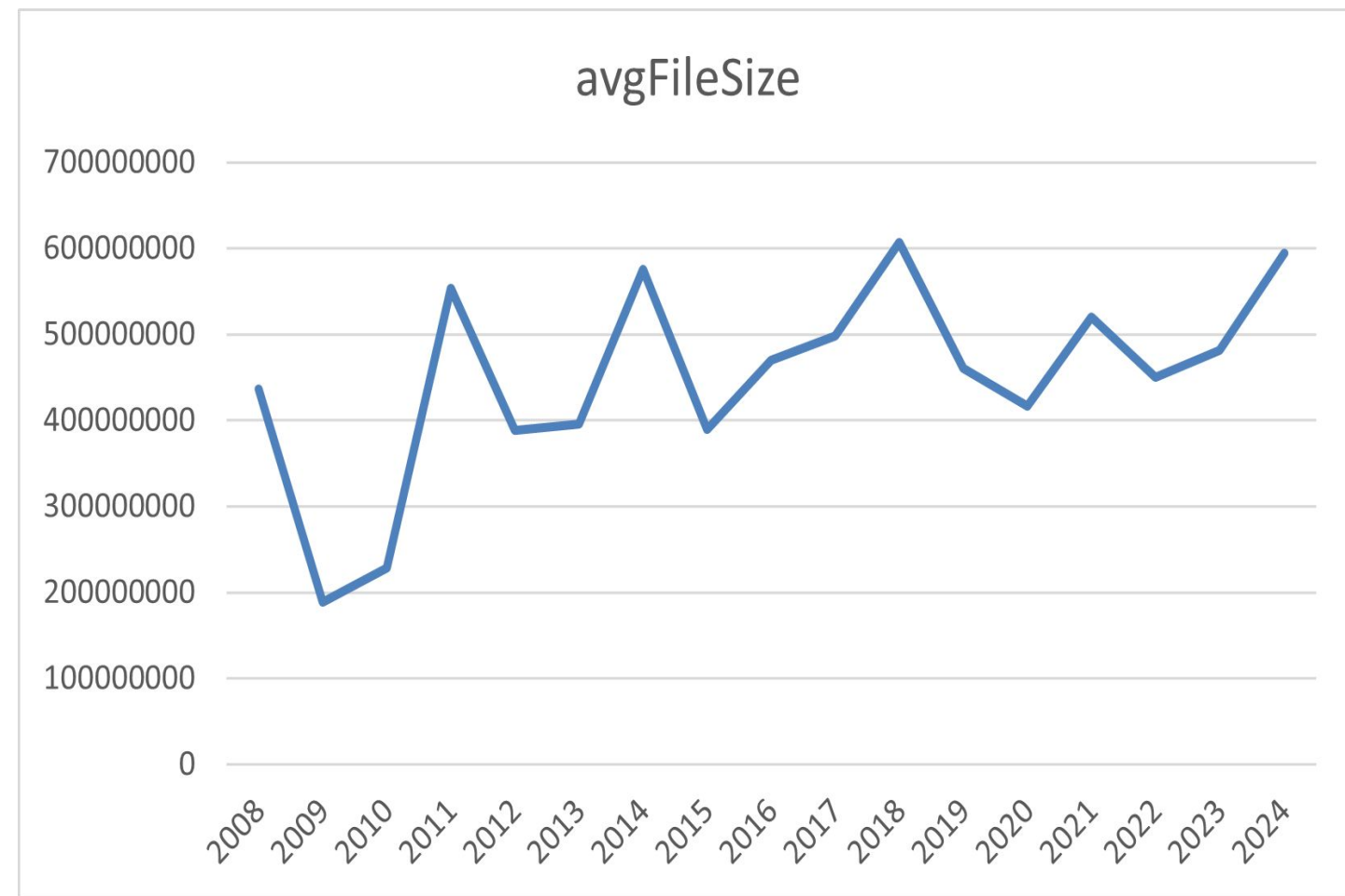


평가 개수 그룹별 평균 평점 추이

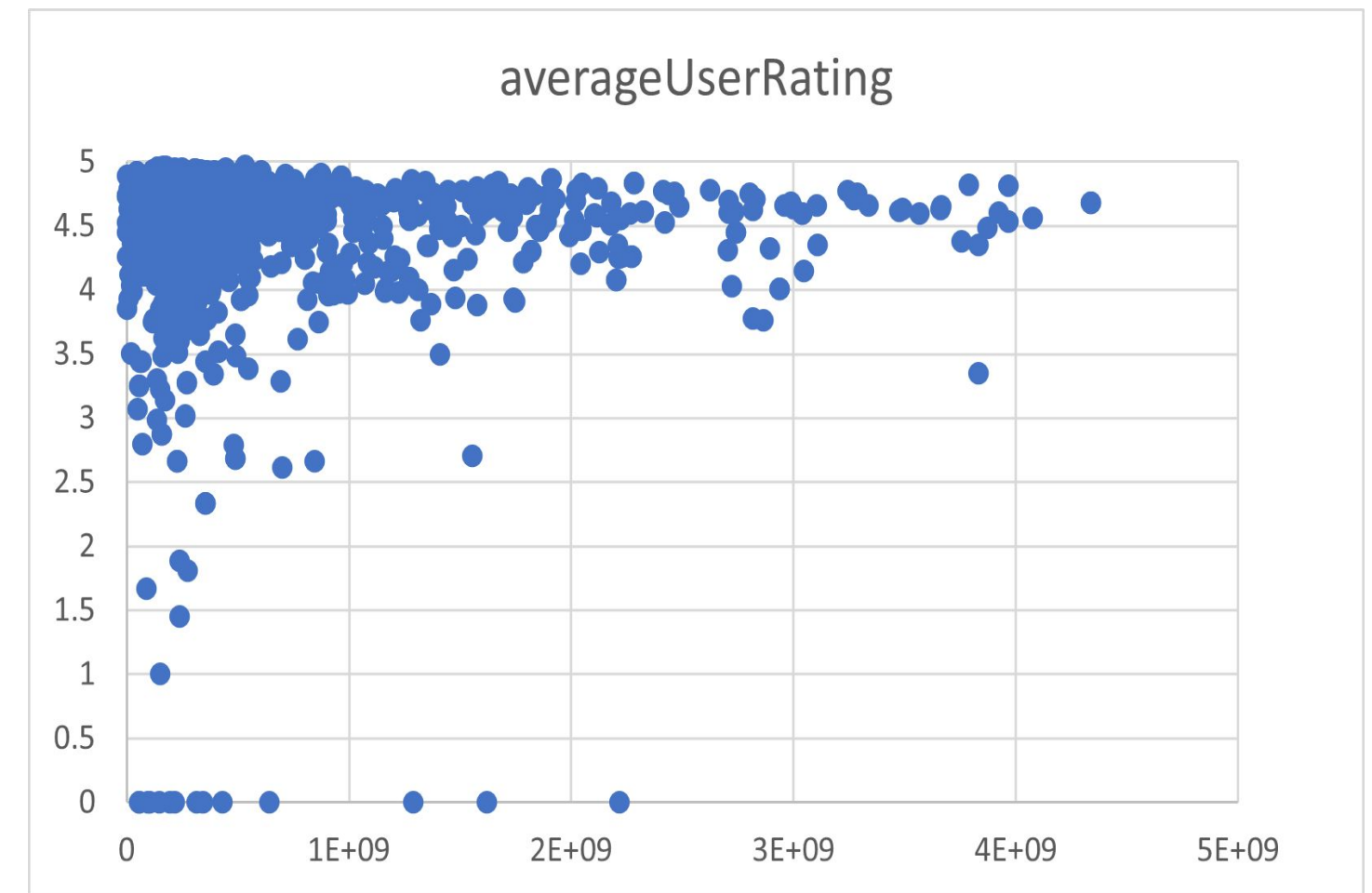
“ 사용자 평가 개수와 평균 평점은 양의 상관 관계이다. ”

- 많은 평가 개수의 게임이 높은 평점을 얻는 경향
- 어플 평가 독려를 통한 평균 평점 증대 기대

## 02 | 파일 사이즈와 평점



시간에 따른 파일 사이즈 변화 추이



파일 사이즈와 평점 간의 상관 관계

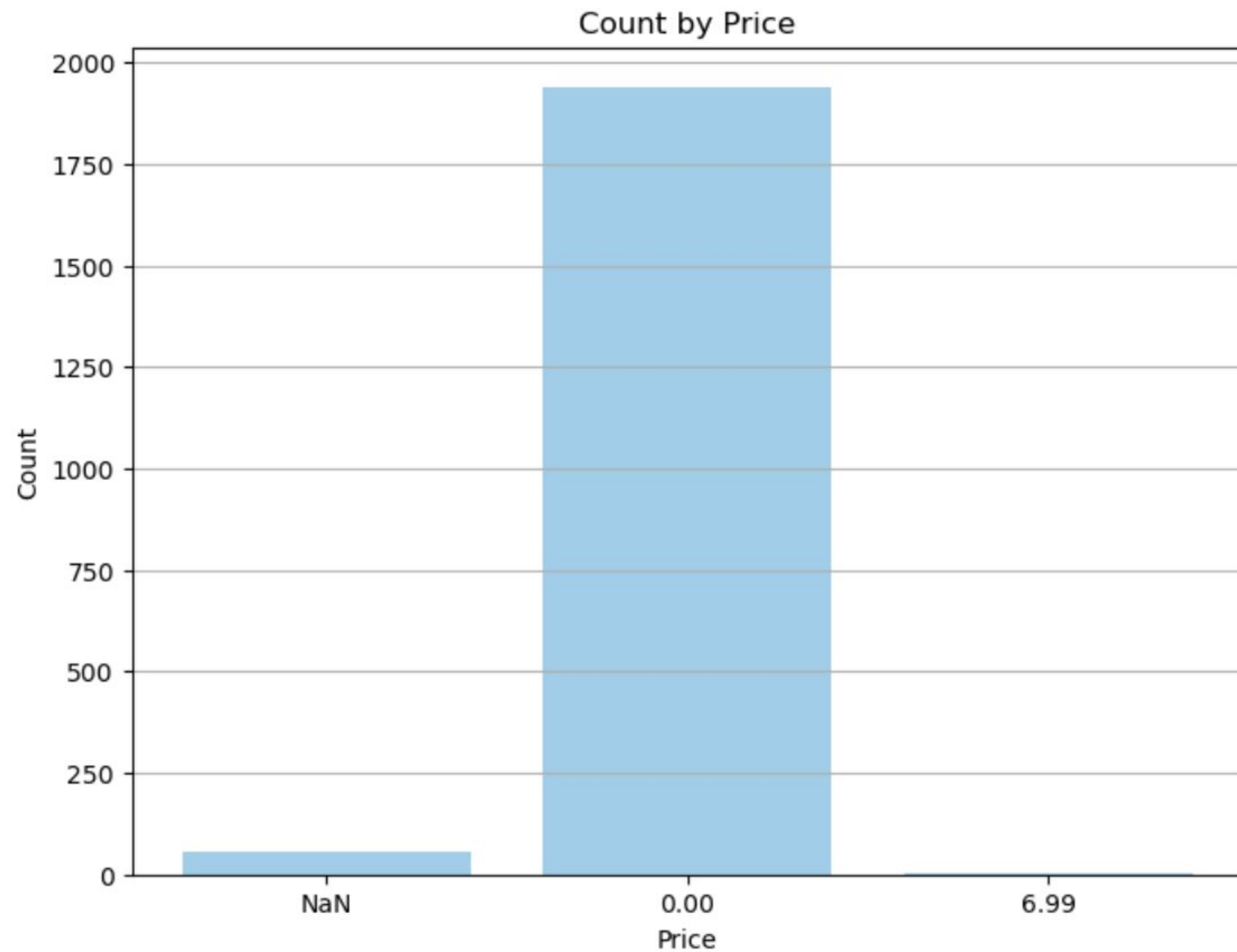
“

파일 사이즈와 평균 평점은 무관하다.

”

- 시간에 따른 파일 사이즈의 뚜렷한 변화 없음
- 파일 사이즈와 평균 평점 간 매우 낮은 상관관계 (-0.026)

## 02 | 가격과 평점



가격별 어플 개수

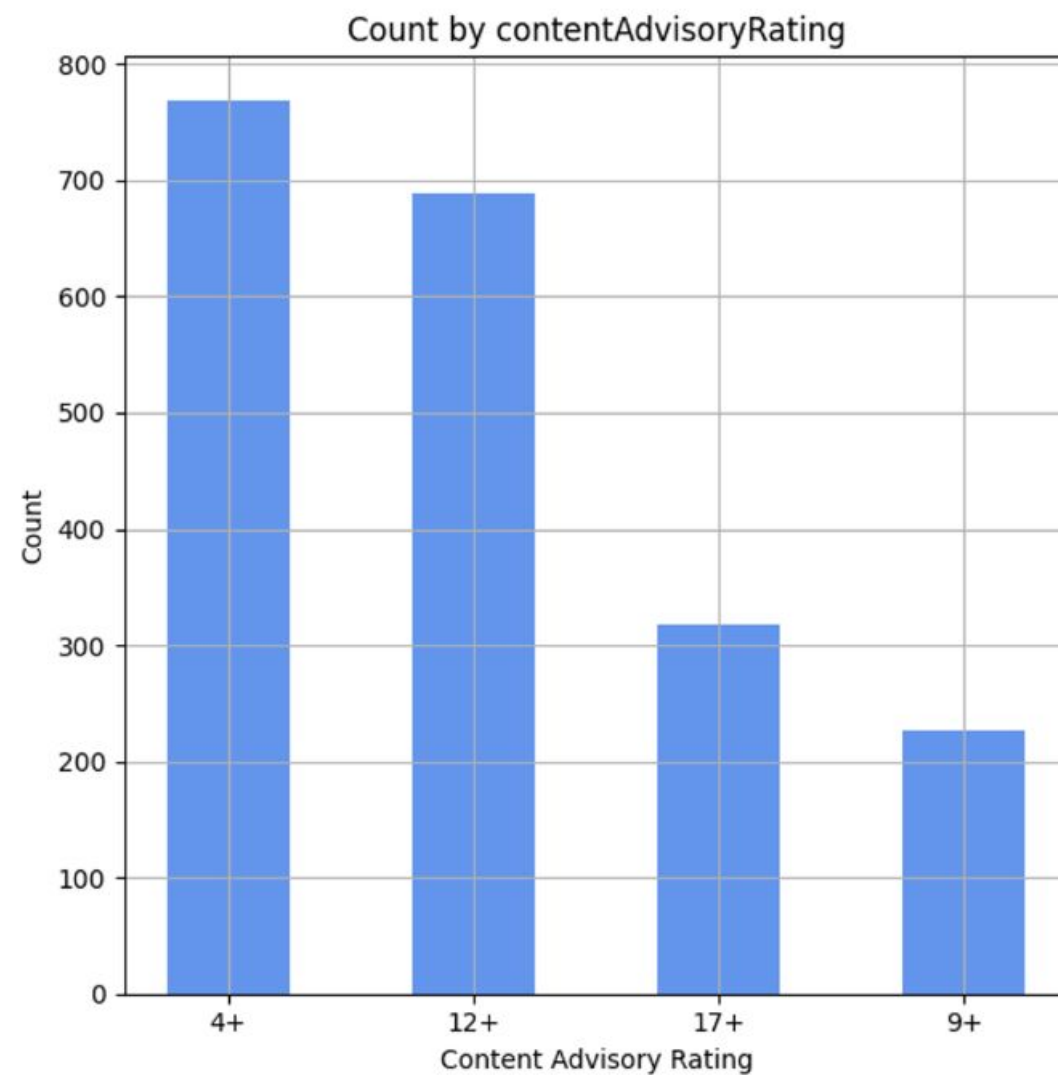
“

가격과 평점 간의 관계는 확인이 불가하다.

”

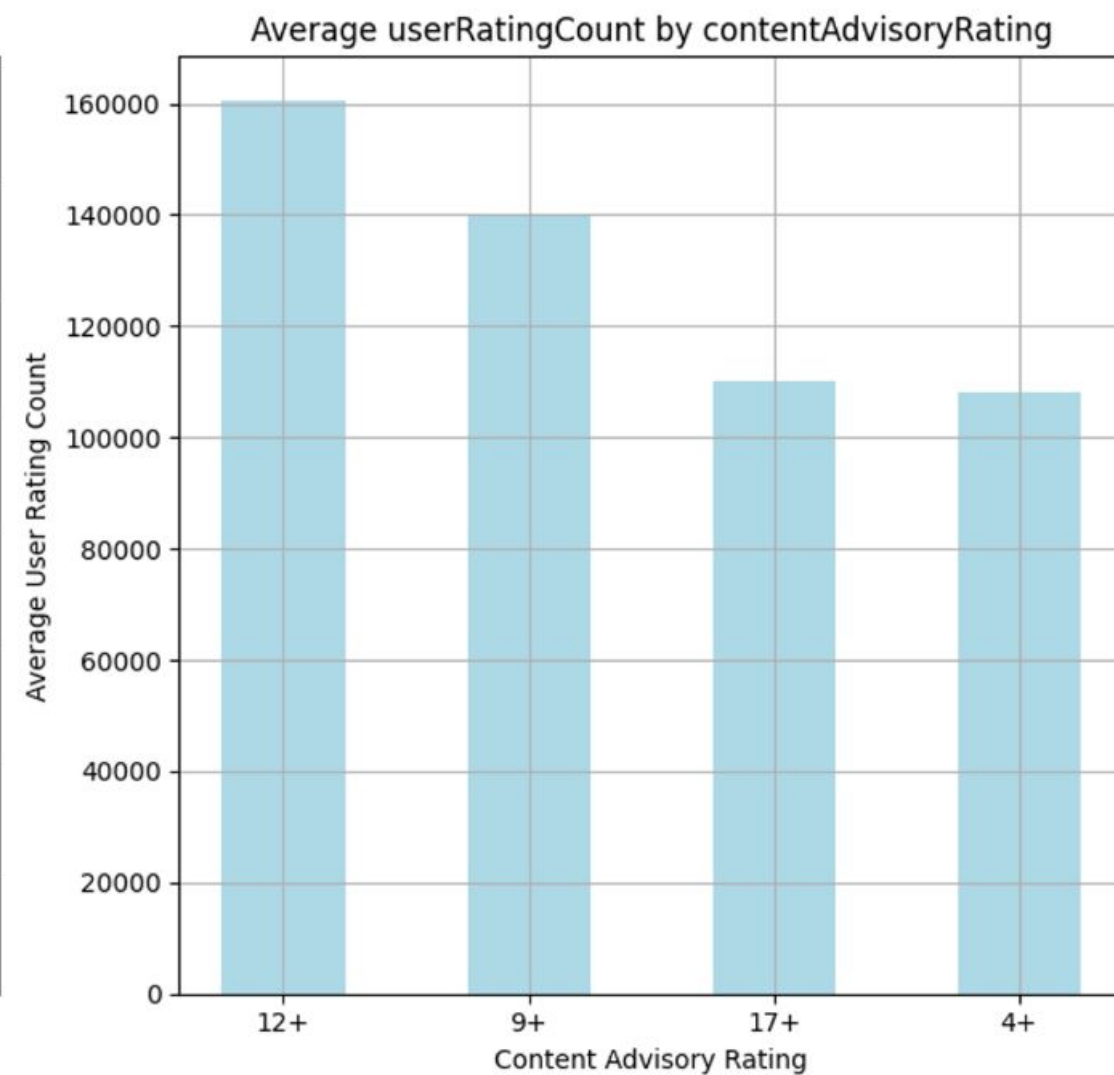
- 가격 (price) 칼럼에 0 값이 대부분 (1941/2000)으로 해당 칼럼을 이용한 EDA는 어려움
- 많은 게임들은 다운로드 유도를 위해 무료로 앱스토어에 출시하고 인앱 결제를 유도하는 것으로 보임

## 02 | 연령 적합성과 평점 개수



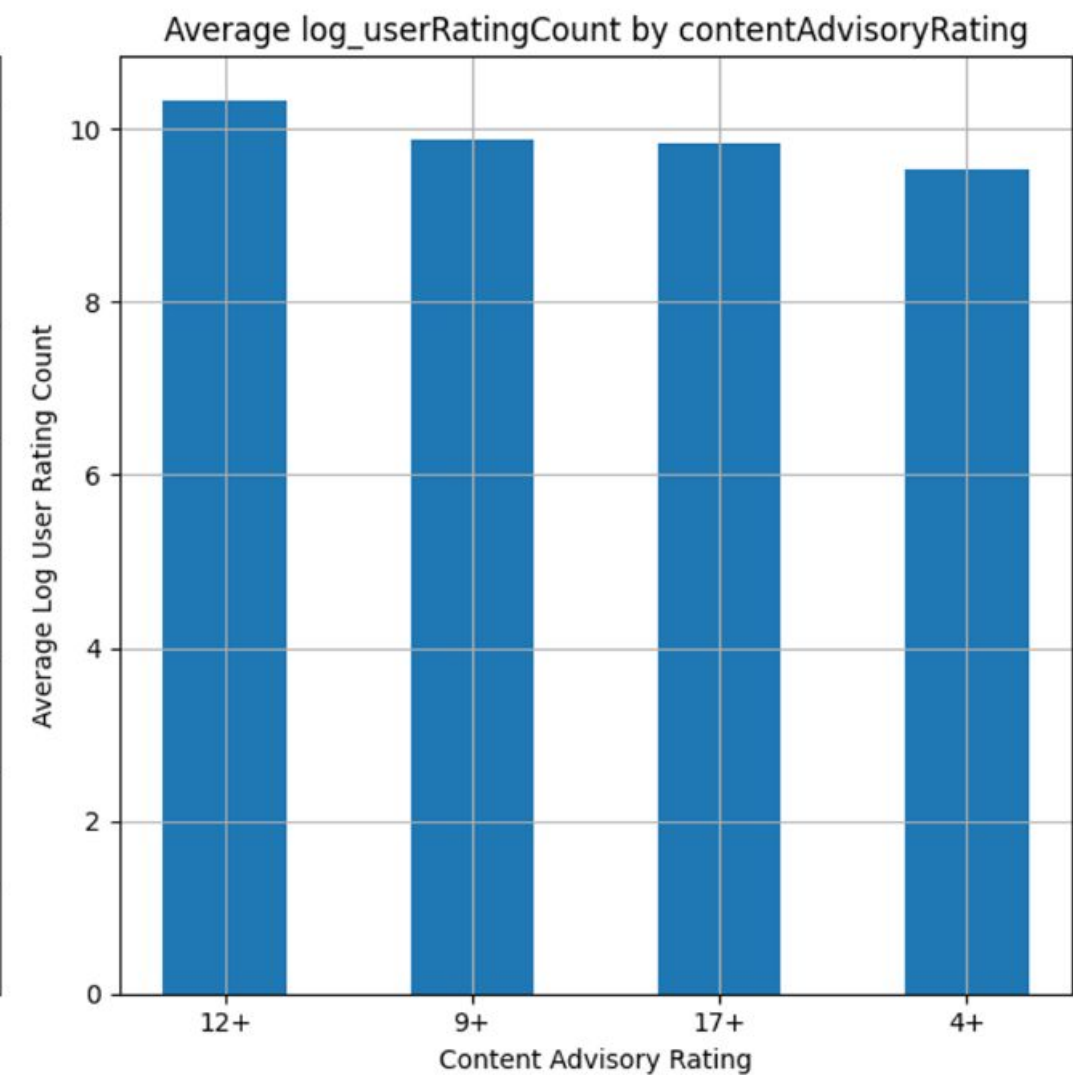
각 연령 적합성 분포

4+ > 12+ > 17+ > 9+



각 연령 적합성별 평균 평점 개수

12+ > 9+ > 17+ > 4+



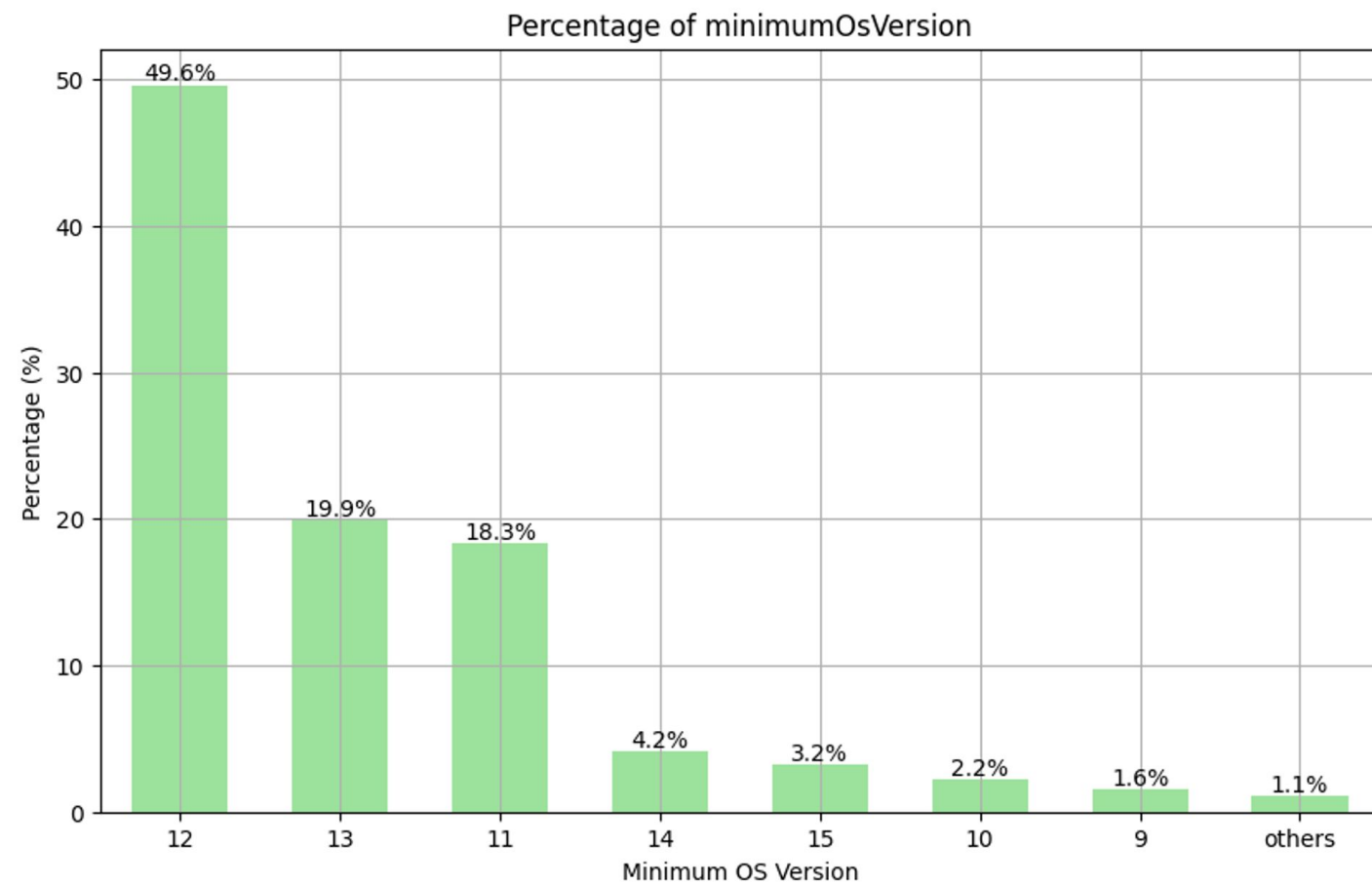
로그 변환 적용 이후의  
각 연령 적합성별 평균 평점 개수

로그 변환 -> 이상치 완화 및 분산 안정화

“ 연령 적합성과 평점 개수는 무관하다. ”  
두 속성간 상관계수 = 0.091

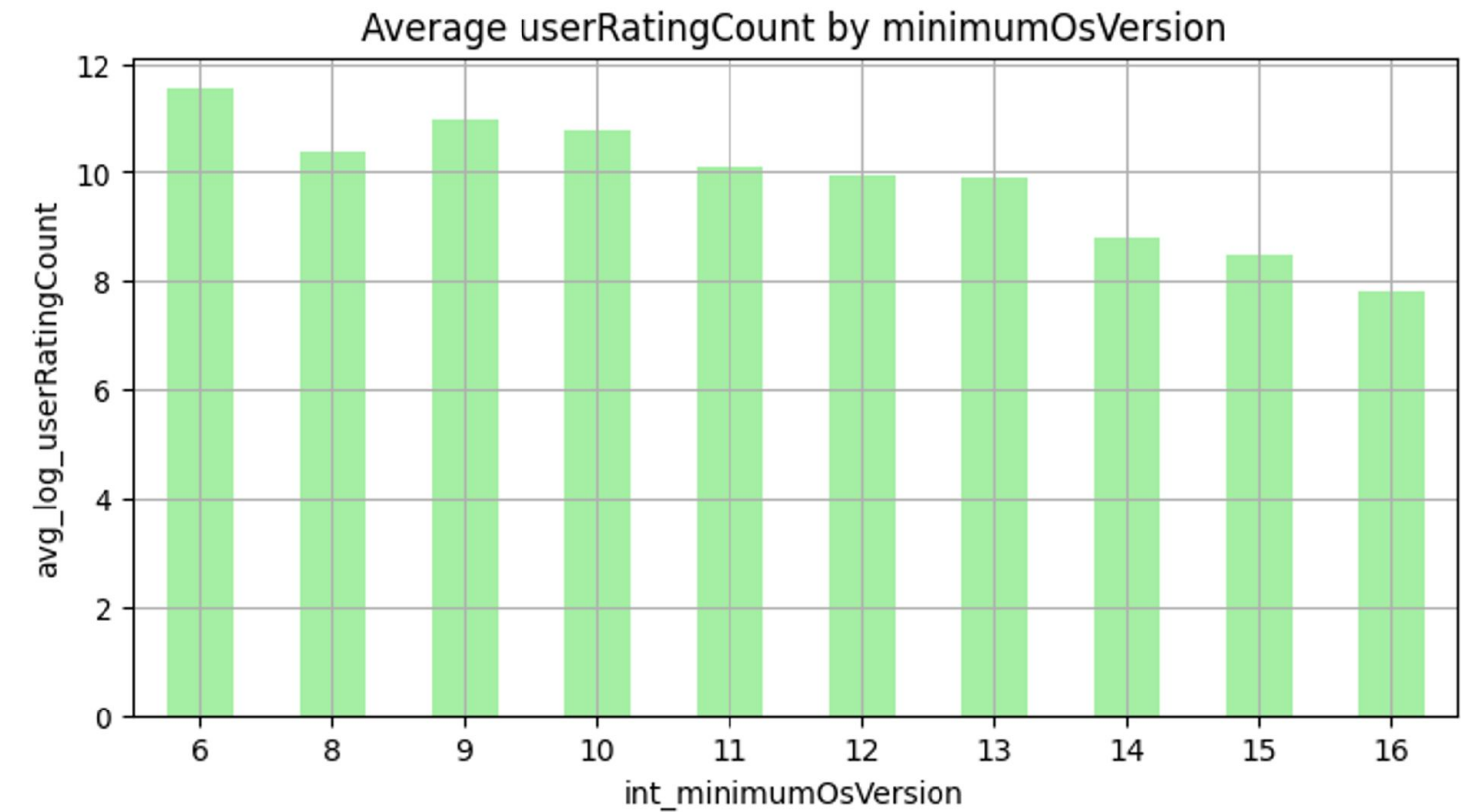


## 02 | 최소 요구 OS 버전과 평점 개수



### 각 최소 요구 OS 버전 분포

- iOS 12, 13, 11이 대부분을 차지하며, 특히 iOS 12는 거의 절반의 비율을 차지함



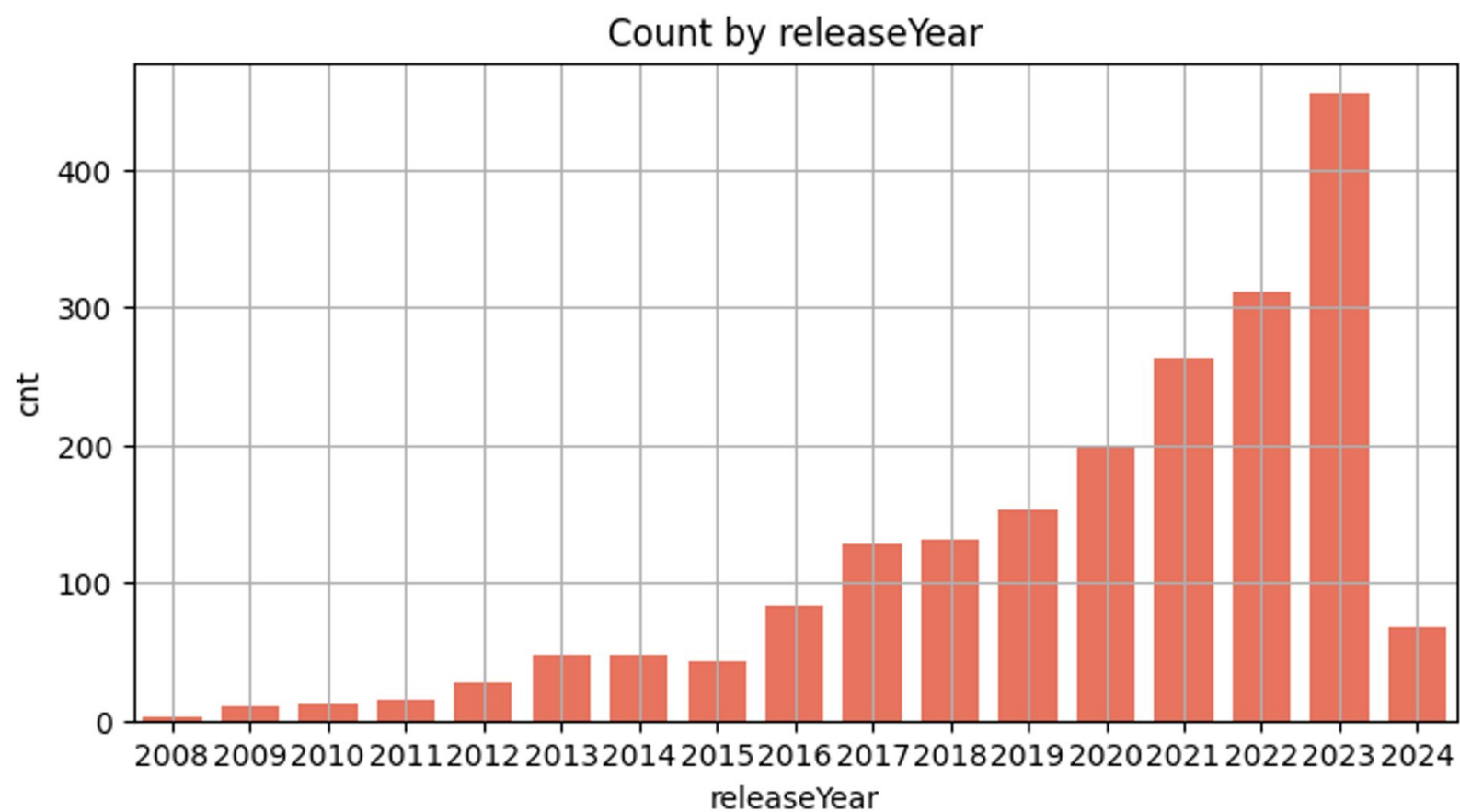
### 로그 변환 적용 이후의 각 최소 요구 OS 버전별 평균 평점 개수

- 최소 요구 OS 버전이 낮을수록 평점 개수가 높은 경향을 보임

“ 최소 요구 운영체제 버전과 평점 개수는 무관하다. ”

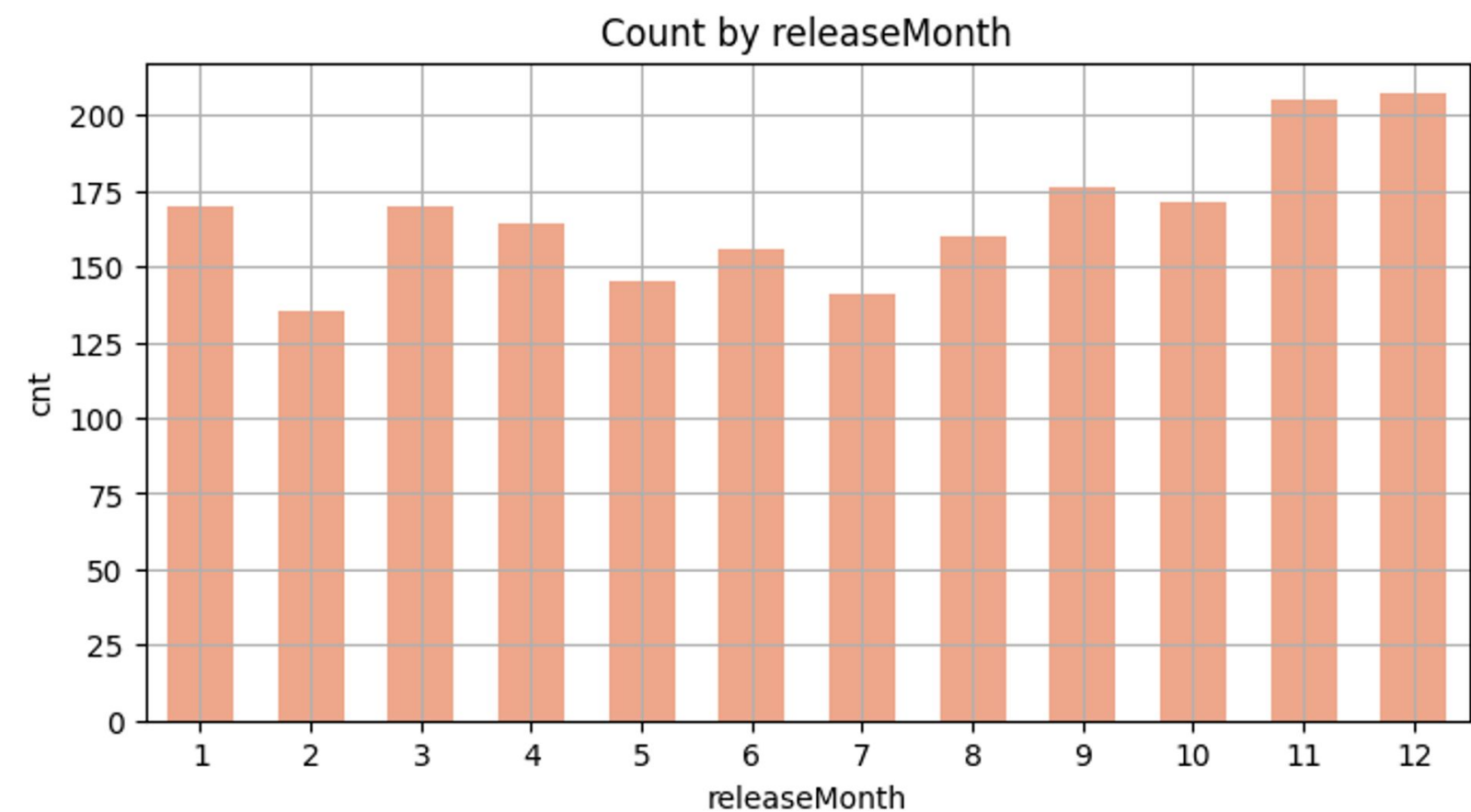
두 속성간 상관계수 = - 0.148  
(예상대로 음수(-)로 나왔지만 절대값이 작아 무관하다고 판단함)

## 02 | 출시 날짜와 평점 개수



### 출시일의 연도별 분포

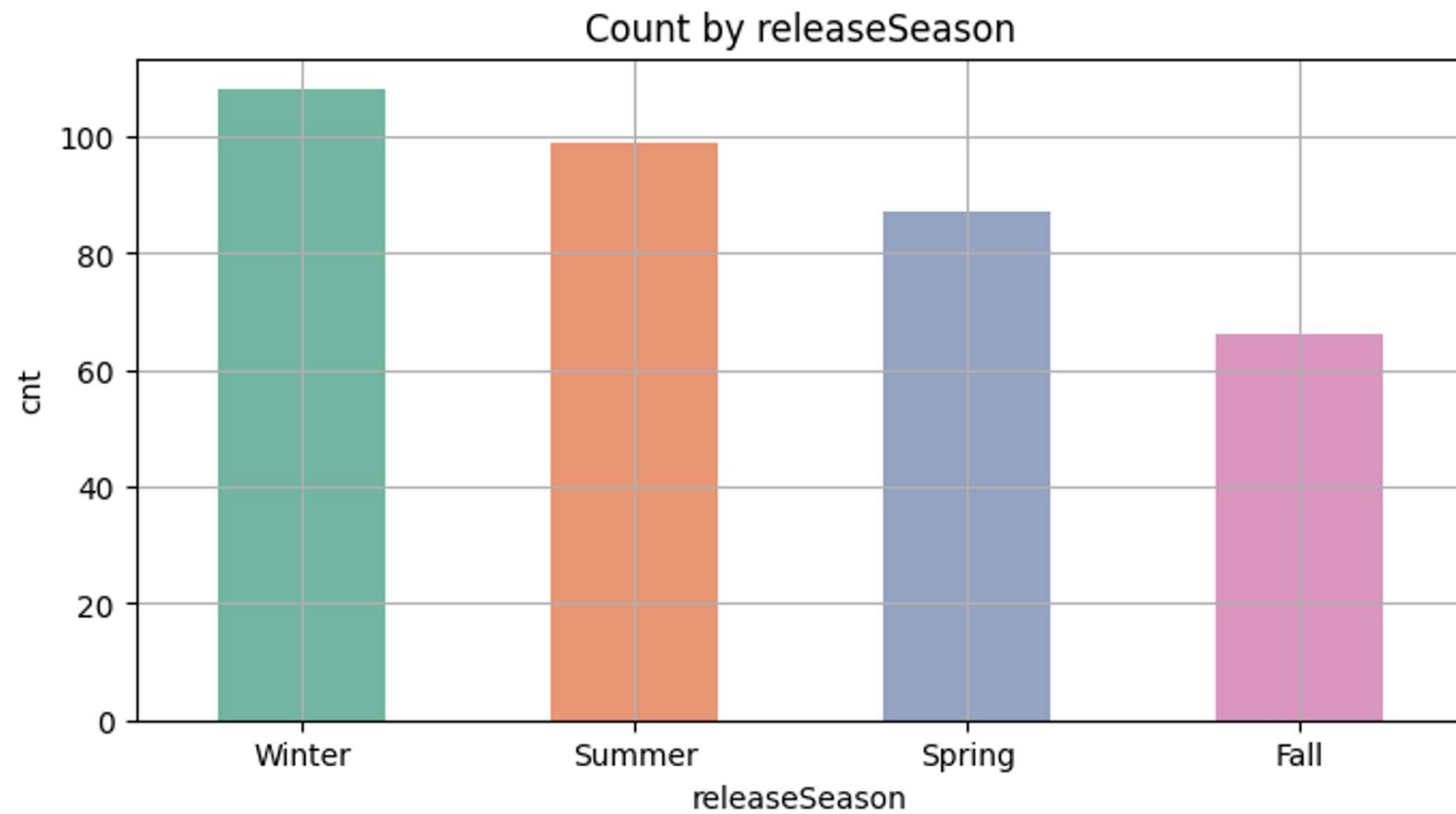
- 해가 갈수록 점점 더 많은 앱이 출시되는 경향
- 데이터의 마지막 업데이트 = 2024년 2월
  - 2024년도엔 2023년도보다 더 많은 앱들이 출시될 것으로 예상



### 출시일의 월별 분포

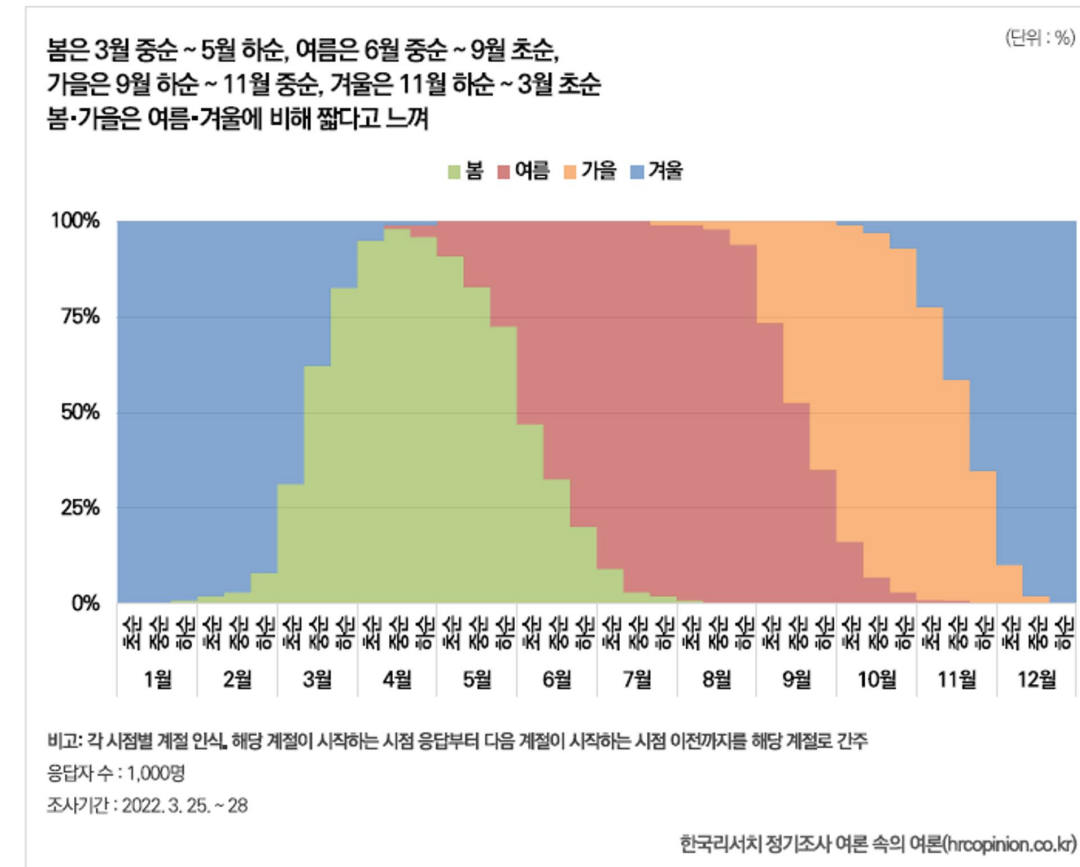
- 11월, 12월에 출시된 앱들이 가장 많고, 2월에 출시된 앱들이 가장 적음
  - 실제로 2월은 열두 달 중 가장 짧은 달

## 02 | 출시 날짜와 평점 개수



### 출시일의 계절별 분포

- 겨울 > 여름 > 봄 > 가을 순서로 출시된 앱이 많음

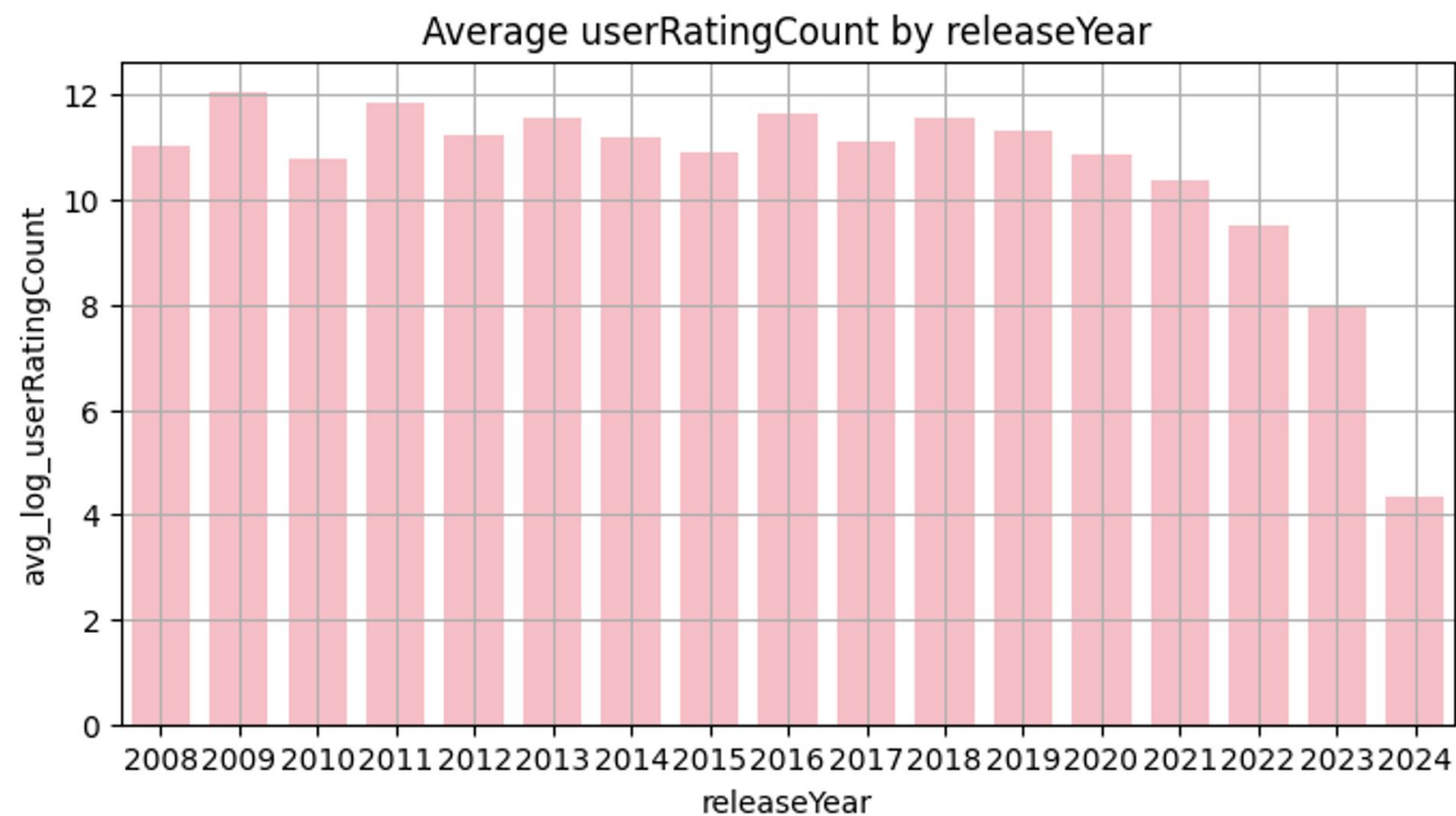


### 참고 자료

- 계절별 체감 길이를 반영하여 계절 구분
- <https://hrcopinion.co.kr/archives/21702>



## 02 | 출시 날짜와 평점 개수



### 로그 변환 적용 이후의 각 출시 연도별 평균 평점 개수

- 2008년도~2017년도까지는 약간의 변동 있음
- 2018년도부터는 계속해서 평균 평점 개수 감소 추세를 보이고 있음

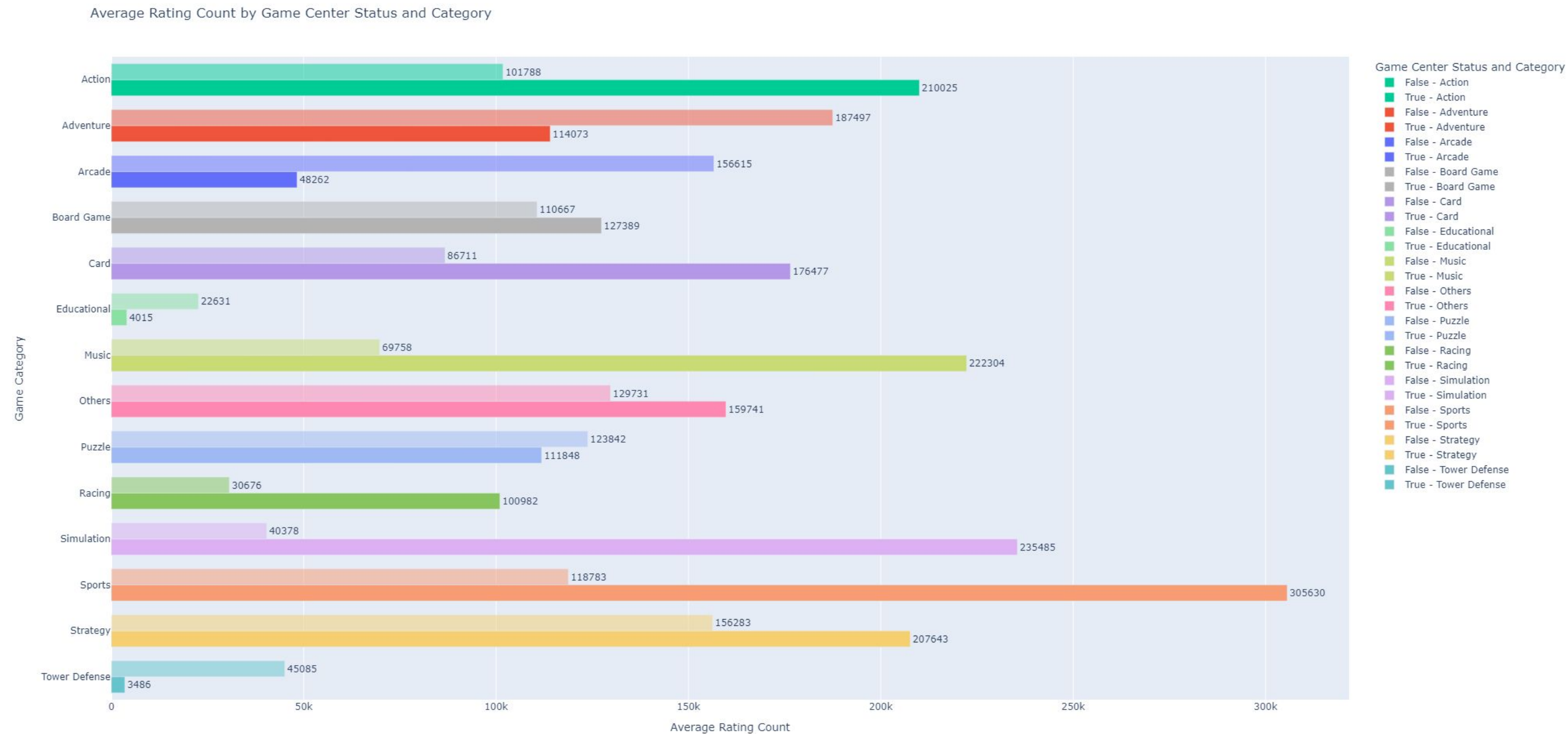
“ 출시 날짜와 평점 개수는 음의 상관관계를 가진다. ”

두 속성간 상관계수 = - 0.509

(출시 날짜가 빠를수록 평점 개수가 높다. ↔ 출시 날짜로부터 오래됐을수록 평점 개수가 높다.)

# 02 | Game Center 활성화 여부와 평가 수

Game Center 활성화 여부에 대한 평균 평가 수

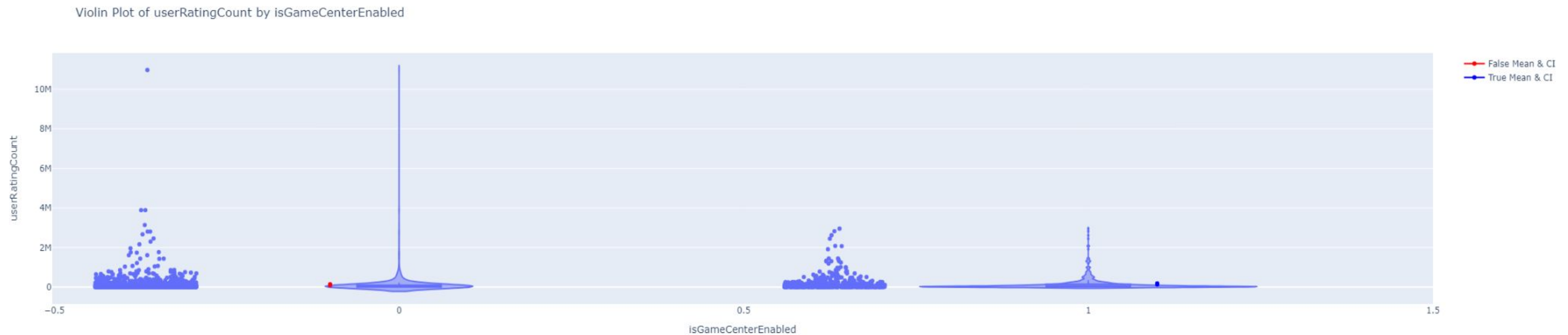


- Game Center **활성** 게임 평균 평가 수 > Game Center **비활성** 게임 평균 평가 수

# 02 | Game Center 활성화 여부와 평가 수

## Game Center 활성화 여부에 대한 평균 평가 수 상관관계

포인트 바이스리얼 상관계수(Point-Biserial Correlation)  
검증 방법 Mann-Whitney U 검정



- U 통계량: 473,409.0  
U 값이 높을수록 두 그룹 간의 차이가 큼 U 값의 범위는 두 그룹의 표본 크기에 따라 달라지며, 주어진 p-value와 함께 해석
- P-value: 6.99e-09  
p-value는 귀무가설이 참일 확률을 나타내며, 매우 작은 값(0.05보다 작음)은 귀무가설을 기각할 수 있음
- 귀무가설 기각: Game Center 연동 여부가 사용자 평가 수에 유의미한 영향을 미침  
Game Center 연동된 앱이 평균적으로 더 많은 사용자 평가 수를 가짐

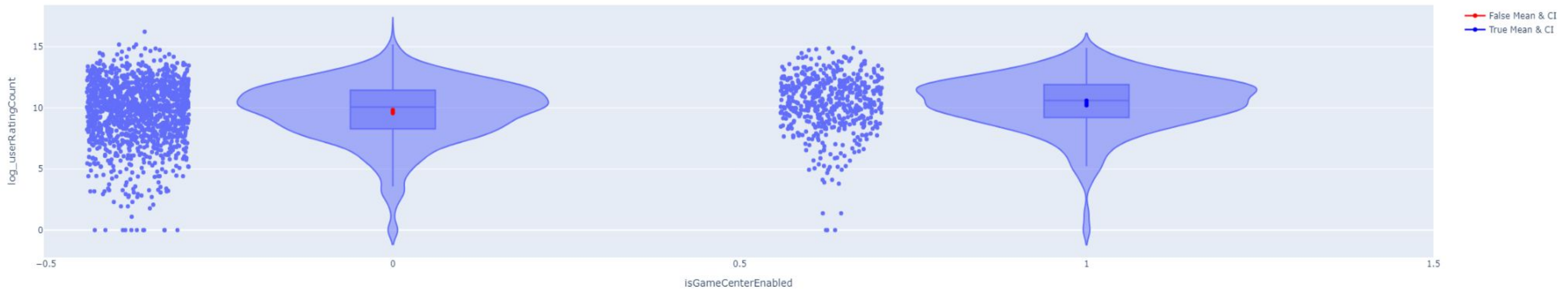
# 02

## Game Center 활성화 여부와 평가 수

Game Center 활성화 여부에 대한 평균 평가 수 상관관계-로그 변환

포인트 바이스리얼 상관계수(Point-Biserial Correlation)  
검증 방법 T-검정

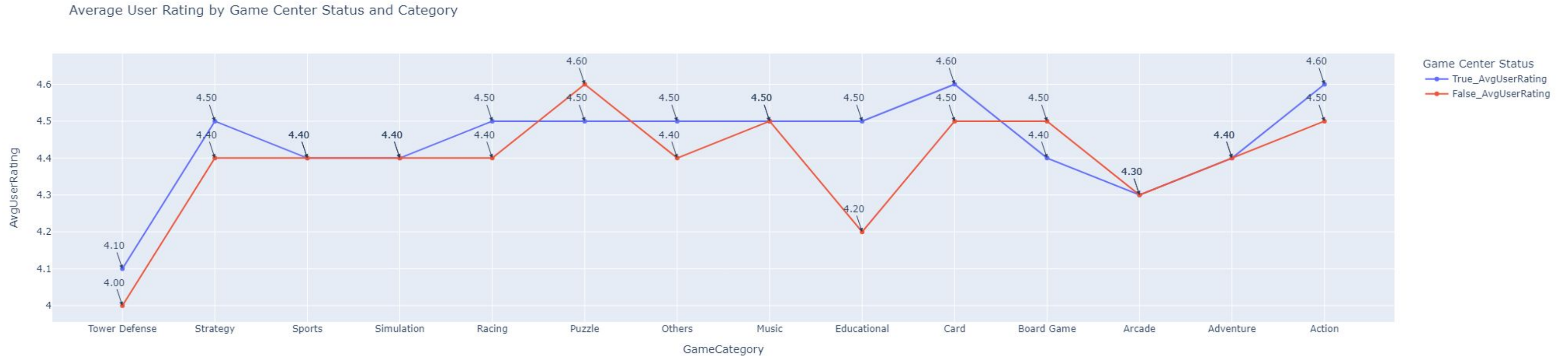
Violin Plot of log\_userRatingCount by isGameCenterEnabled



- T-검정 통계량: 6.09  
일반적으로 2보다 크면 두 그룹 간의 차이가 유의미함, 6.09는 매우 큰 차이를 의미
- P-value: 6.99e-09  
p-value는 귀무가설이 참일 확률을 나타내며, 매우 작은 값(0.05보다 작음)은 귀무가설을 기각할 수 있음
- 귀무가설 기각: Game Center 연동 여부가 사용자 평가 수에 유의미한 영향을 미침  
Game Center 연동된 앱이 평균적으로 더 많은 사용자 평가 수를 가짐

# 02 | Game Center 활성화 여부와 평점 평균

Game Center 활성화 여부에 대한 카테고리별 평균 평점 & 상관관계



- Game Center 활성화 게임의 카테고리별 평균 평점의 변동성이 낮고 유의미한 차이를 보이지 않음
- T-검정 통계량: -0.665  
일반적으로 2보다 크면 두 그룹 간의 차이가 유의미함, -0.665는 차이가 없음 의미
- P-value: 0.5061287730249664  
p-value는 귀무가설이 참일 확률을 나타내며, 매우 작은 값(0.05보다 작음)은 귀무가설을 채택
- 귀무가설 채택: Game Center 연동 여부가 사용자 평점에 유의미한 영향이 없음



# 02 | Game Center 활성화 여부와 게임 분포

## Game Center 활성화 여부에 대한 게임 분포

Game Center 활성화 게임의 평가 수가 많다는 것은 Game Center 활성화 게임의 수가 많다는 것을 의미한다고 판단

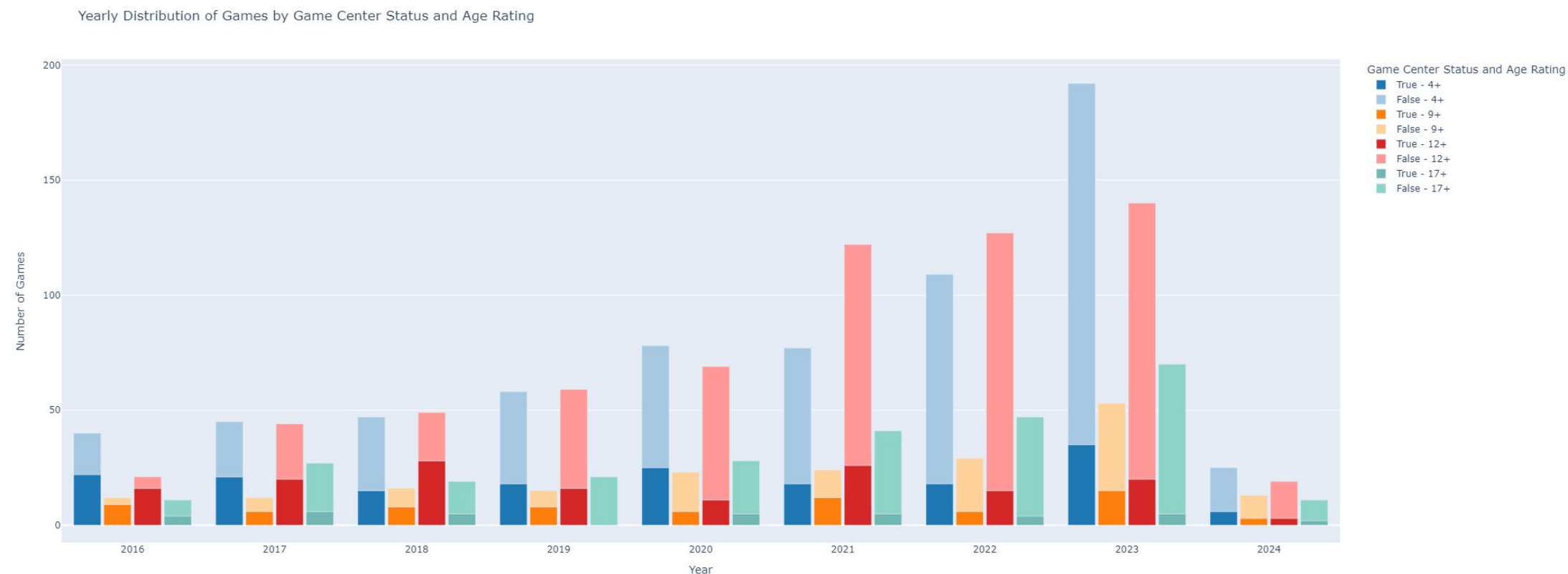


- 2017년부터 Game Center 활성화 여부의 역전 현상이 시작되어 큰 차이를 보여주고 있으며 Game Center 활성화 게임이 증가 추세를 보이긴하지만 일정한 변동성을 보이는 반면, Game Center 비활성 게임은 변동성이 급격하게 커지는 현상 확인
- 이를 통해 Game Center 활성화 여부는 평점과 평점 개수에 영향을 미치는 필수 조건이 아님을 알 수 있음
- 따라서 Game Center 비활성화 게임이 압도적으로 많은 점에 대해 의문을 가지고 이어지는 다음 가설 제시

# 02 | Game Center 활성화 여부와 특정 유형 & 특정 유저 기반 게임

Game Center 활성화 여부에 대한 특정 유저 기반 게임 & 상관관계

연도별 이용 연령 등급 Game Center 활성화 여부(2016\_2024)



## 데이터 범위 제한

### 데이터의 완전성 및 신뢰성

- 2008년부터 2015년까지의 데이터는 불완전하거나 카테고리별로 게임이 고르게 분포하지 않음.
- 분석 결과의 신뢰성을 떨어뜨릴 수 있음.
- 2016년 이후의 데이터는 보다 완전하고 신뢰할 수 있는 데이터일 가능성이 높음.

### 데이터의 최신성

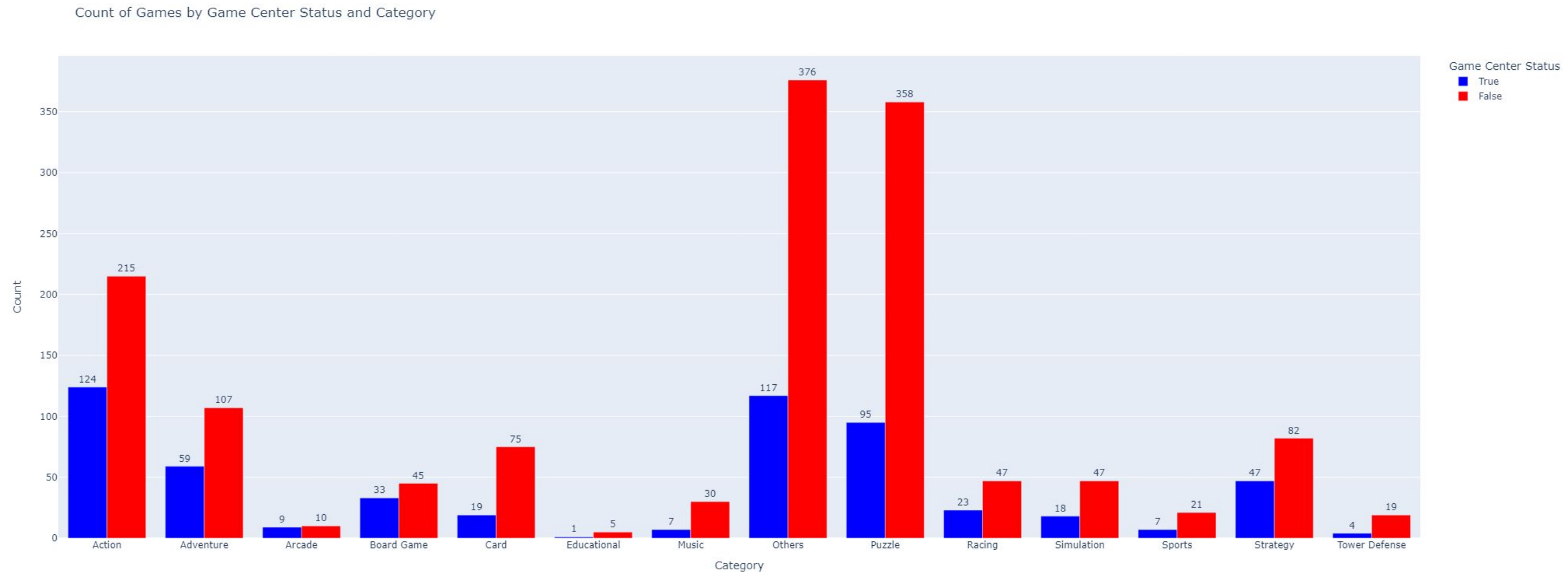
- 최신 데이터는 최근의 사용자 행동 및 게임 출시 패턴을 더 잘 반영
- 2016년 이후의 데이터를 분석하면 현재 시장 상황을 더 잘 이해하고 예측할 수 있음

- 저연령층 이용 등급 게임에서 Game Center의 비활성 게임에 대한 값이 상대적으로 높은 값으로 나타남
- Chi-square 통계량: 28.14  
클수록 두 변수 간의 차이가 크다는 것을 의미. 특정 값이 클지 여부는 자유도와 데이터의 크기에 따라 다름  
일반적으로, Chi-square 값이 높으면 두 변수 간의 관계가 유의미한 영향이 있음
- P-value: 1.13e-07  
p-value는 귀무가설이 참일 확률을 나타내며, 매우 작은 값(0.05보다 작음)은 귀무가설을 기각
- 귀무가설 기각: 저연령층 게임과 Game Center 비활성화 여부 간에 유의미한 상관관계가 있음

# 02 | Game Center 활성화 여부와 특정 유형 & 특정 유저 기반 게임

Game Center 활성화 여부에 대한 특정 유형 게임

카테고리별 Game center 활성화 여부



- 전반적으로 Game Center 비활성 게임 수가 더 많았으며, 특정 유형의 게임에서 Game Center 활성화 여부에 따른 차이가 두드러지게 나타남



# 02 | Game Center 활성화 여부와 특정 유형 & 특정 유저 기반 게임

## Game Center 활성화 여부에 대한 특정 유형 게임 & 상관관계

### 카테고리별 Game center 활성화 여부

카테고리: Action

Chi-square test statistic: 65.84907927034259

P-value: 2.4244638713748544e-08

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Board Game

Chi-square test statistic: 21.21706414797324

P-value: 0.06874481892831955

귀무가설 채택: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 없습니다.

카테고리: Puzzle

Chi-square test statistic: 40.56994715162068

P-value: 0.00037178604394093876

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Strategy

Chi-square test statistic: 35.232470684174025

P-value: 0.0007797705780564212

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Card

Chi-square test statistic: 16.511019820081604

P-value: 0.2226318845492597

귀무가설 채택: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 없습니다.

카테고리: Others

Chi-square test statistic: 96.00873010606932

P-value: 2.747304075467082e-14

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Adventure

Chi-square test statistic: 32.01448789278405

P-value: 0.002390445811291058

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Arcade

Chi-square test statistic: 7.444179894179895

P-value: 0.38414034180280593

귀무가설 채택: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 없습니다.

카테고리: Racing

Chi-square test statistic: 20.97030936993934

P-value: 0.012782906132182815

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Simulation

Chi-square test statistic: 39.87821477183179

P-value: 0.0001446142779179281

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Music

Chi-square test statistic: 25.26571428571428

P-value: 0.008334020559388635

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Sports

Chi-square test statistic: 17.599999999999998

P-value: 0.024433630495543302

귀무가설 기각: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 있습니다.

카테고리: Tower Defense

Chi-square test statistic: 9.078947368421051

P-value: 0.24703755488983178

귀무가설 채택: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 없습니다.

카테고리: Educational

Chi-square test statistic: 2.4000000000000004

P-value: 0.3011942119122021

귀무가설 채택: 이 카테고리에서 Game Center 비활성화 여부와 유의미한 상관관계가 없습니다.

- Game Center 비활성화 여부에 대한 카테고리별 게임의 유의미한 상관관계를 확인 할 수 있음

# 03 | 마케팅 전략

## 카테고리 별 출시 전략

- 시즌별 전략:
  - 겨울 시즌에 맞추어 게임 출시 (11월)
- 인기 카테고리 활용:
  - Puzzle & Action 카테고리 출시
  - 인기가 높으며 안정적인 수요 O
- 평점 관리:
  - 개발 초기 단계 사용자 경험 최적화 필요

## 게임 설명 관련 전략

- Release Notes 전략:
  - 디테일한 업데이트 노트 작성 필요
  - 업데이트 키워드를 넣어 작성
  - 긍정적인 텍스트로 작성
- 게임 설명 전략:
  - 자세하고 긴 설명 작성 필요
  - 긍정적인 텍스트로 작성

## 03 | 마케팅 전략

### 평점 개수와 평균 평점 간의 관계

- 사용자에게 팝업, 푸시 등을 통한 게임 평가 독려하여 높은 평점 기대
- 사용자들이 게임에 대해 충분히 경험했을 때 만족도 조사
  - 만족 응답 시에만 어플 평가 요청
  - 불만족 응답 시 피드백 요청

### 가격과 평균 평점 간의 관계

- 더 많은 다운로드 유도를 위해 앱스토어에는 0원으로 출시
- 인앱 결제를 통해 돈 지출은 꺼리는 가벼운 사용자들이 게임을 떠나지 않게 하고 동시에 유료 사용자들에게서는 매출 기대 가능

# 03 | 마케팅 전략

## 연령 적합성과 평점 개수

- 12+ 등급의 타겟팅 강화
  - SNS 광고, 인플루언서 마케팅 등
- 4+, 9+ 등급을 타겟팅하고자 한다면, 가족 친화적 마케팅 및 부모 참여 유도
  - 교육적 요소, 안전성 강조
- 성인 사용자 대상 고급 기능 제공 및 소셜 미디어 활용

## 최소 요구 OS 버전과 평점 개수

- iOS 11 또는 iOS 12를 최소 요구 운영체제로 설정
- iOS 호환성을 위한 정기적 업데이트 및 사용자 피드백 반영

# 03 | 마케팅 전략

## 출시 날짜와 평점 개수

- 봄과 가을 출시 고려, 겨울과 여름에는 특별 프로모션
  - 겨울과 여름에는 게임 앱 출시가 가장 많으므로 시장 내 경쟁 치열
- 사용자 리뷰 유도 및 인센티브 제공
  - ex: 게임 내 보상

# 03 | 마케팅 전략

## Game Center 연동과 평가 수, 평균 평점의 마케팅 전략

- 마케팅 전략
  - 커뮤니티 전체가 일정 목표 평점 수를 달성하면, 모든 참여자에게 보상을 제공하는 이벤트 개최
  - 매달 가장 창의적이고 도움이 되는 리뷰를 남긴 사용자를 선정하여 '리뷰왕' 칭호와 보상을 제공
  - 간단한 설문조사를 통해 사용자 경험을 평가하고, 긍정적인 평가를 유도하는 질문 포함

## Game Center 연동과 특정 유저, 특정 유형 게임의 마케팅 전략

- 마케팅 전략
  - 교육적 요소를 게임 내 통합하여 부모들이 자녀에게 게임을 허용하는 데 긍정적 영향을 미치도록 함
  - 학습과 재미를 동시에 제공하는 콘텐츠 개발
  - 인기 카테고리 게임 출시

# 04 | 기대효과와 한계점

## 기대 효과

- 고도화된 게임 시장 분석
  - 사용자 선호도, 행동 패턴, 게임 평가에 대한 인사이트 제공
  - 출시와 마케팅 전략 계획 및 실행에 도움
- 사용자 만족도 향상
  - 키워드 및 감정 분석을 통한 게임 이해도 및 기대치 명확화
  - 평점과 사용자 만족도 증대
- 경쟁 우위 확보
  - 기술적 요소 분석을 통한 독특한 가치 제안 개발
  - 게임의 독창성과 매력 향상
  - 시장 점유율 확대 및 브랜드 인지도 상승
- 리스크 관리 및 예측 강화
  - 소셜 통합 기능의 사용자 참여 및 만족도 영향 분석
  - 개발 및 마케팅 비용 효율적 관리
  - 프로모션 전략 최적화
- 결론
  - 데이터 분석은 게임 개발 및 마케팅 전략 성공에 중요한 정보 제공
  - 사용자 경험과 시장 수용성에 대한 이해 심화
  - 전략적 의사결정을 통한 경쟁력 강화 및 시장 점유율 확대
  - 혁신적인 제품과 서비스 제공에 필요한 전략적 기반 마련

## 한계점(Limitation)

- 선택 편향 및 측정 편향
  - 데이터 샘플의 편향 가능성
  - 데이터 전처리 및 측정 과정에서의 오류 가능성
- 상관관계와 인과관계 혼동
  - 상관관계가 인과관계를 의미하지 않음
  - 숨겨진 변수의 영향 가능성
- 다양한 데이터의 필요성
  - 한정된 지표만을 분석
  - 더욱 심층적이고 다각적인 분석의 제한
  - 그래픽 데이터, 광고 효과 데이터, 수익 데이터 등의 필요성

## 출처

- Gajamuthu, K. (n.d.). *GameSphere* 📍: 2000 App Store Insights & Ratings. Kaggle. <https://www.kaggle.com/datasets/kanchana1990/gamesphere-2000-app-store-insights-and-ratings>
- Precedence research. (2023, July). *Mobile Gaming Market (By Platform: IOS, Android; By Age Group: Below 24 Years, 24-44 Years, Above 44 Years; By Business Model: Freemium, Paid, Free, Paymium) - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023-2032*. Precedence. <https://www.precedenceresearch.com/mobile-gaming-market>
- 한국콘텐츠진흥원. (n.d.). 글로벌 게임산업 성장둔화 속 국내 게임산업 성장 유지' 콘진원, <2023 대한민국 게임백서> 발간. 한국콘텐츠진흥원. <https://www.kocca.kr/kocca/koccanews/reportview.do?menuNo=204767&nttNo=659>
- Matt Galligan, The right way to ask users to review your app, Medium, 2014.12.05 <https://medium.com/circa/the-right-way-to-ask-users-to-review-your-app-9a32fd604fca>
- 엘리스테어 크롤, 벤저민 요스코비츠, "린 분석", 한빛미디어, (2014)



**Thank You.**