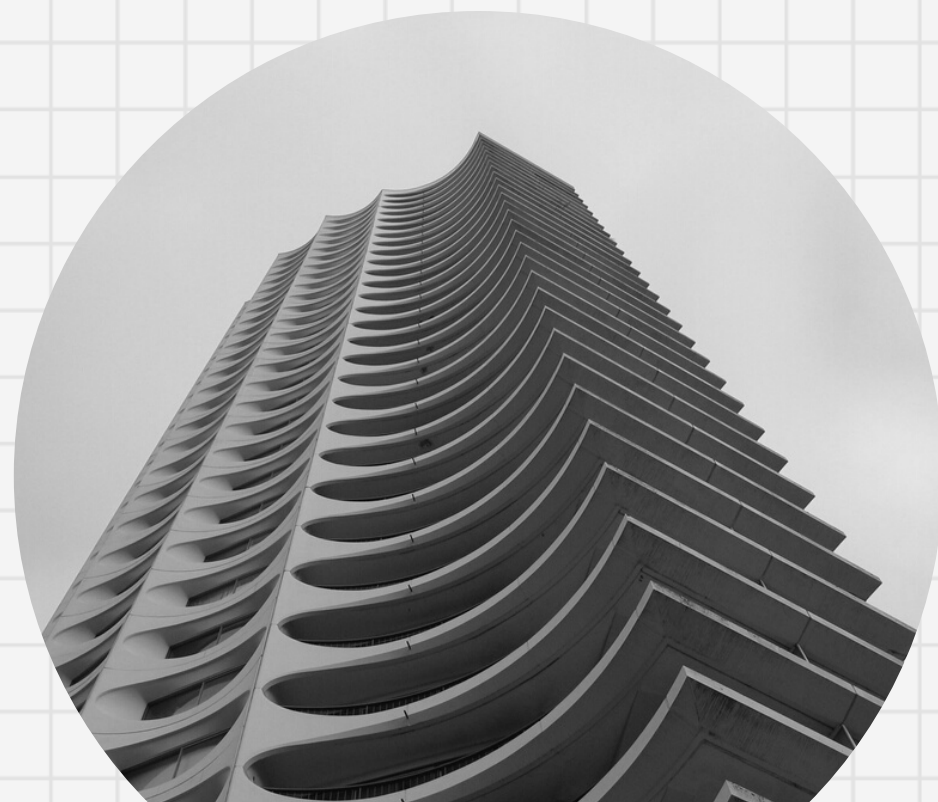


CODE STATES

GAME DATA ANALYSIS

조영재



목차

1. 간단 DATA 처리과정
2. 지역에 따른 게임 선호
3. 연도별 게임의 트렌드
4. 인기게임 분석
5. Insight

간단 Data 처리

결측치 처리

같은 이름의 게임의 다른 Data로 대체

같은 이름의 게임에 모두 결측치가 있을 시 처리하지 못함

Year 칼럼

두 자릿수 데이터 처리

더 정확한 Year 데이터를 얻을 수 있음

Global Sales

EU, JP, NA 칼럼의 합계

전세계 세일즈의 동향 파악에 용이함

Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
Madden NFL 07	XB	2006.0	Sports	Electronic Arts	0.97	0.03	0.00	0.03
Madden NFL 07	GC	2006.0	Sports	Electronic Arts	0.48	0.13	0.00	0.02
Madden NFL 07	X360	2006.0	Sports	Electronic Arts	1.66	0.00	0.01	0.13
Madden NFL 07	PS3	2006.0	Sports	Electronic Arts	0.47	0.00	0.01	0.04
Madden NFL 07	DS	2006.0	Sports	Electronic Arts	0.20	0.00	0.00	0.02
Madden NFL 07		NaN	Sports	Unknown	0.77	0.03	0.00	0.04
Madden NFL 07	GBA	2006.0	Sports	Electronic Arts	0.02	0.01	0.00	0.00
Madden NFL 07	PS2	2006.0	Sports	Electronic Arts	3.63	0.24	0.01	0.61
Madden NFL 07	Wii	2006.0	Sports	Electronic Arts	0.46	0.00	0.00	0.04

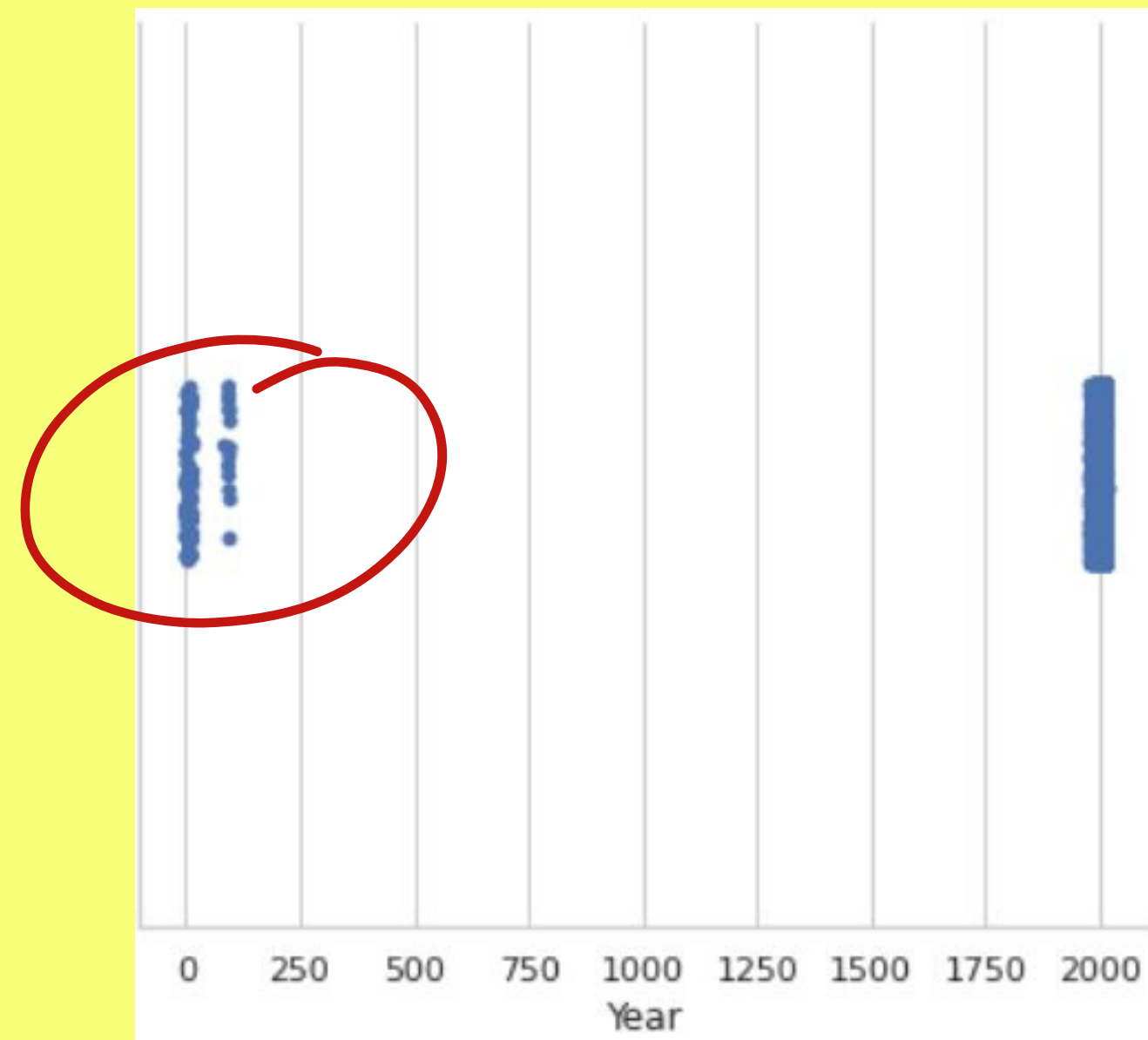
이름이 같은 Game의 경우

다른 열에서 Data 대체

(모든 열에서 빈 값이면 대체 불가능)

Year 칼럼 이상치 발견

11.0	10
9.0	10
8.0	10
1980.0	9
10.0	7
3.0	7
6.0	7
16.0	5
15.0	5
5.0	4
4.0	4
95.0	4
98.0	3
7.0	3
96.0	3
2017.0	3
2.0	3
0.0	3
14.0	2
1.0	2
97.0	2
13.0	2




대체 방식

확인 결과 앞의 두자리가 빠진 것을 확인

코딩을 통해 데이터를 조금 더 정제

```
[316] def update_year(year):  
    if year < 24:  
        return 2000 + year  
    elif year >= 24 and year < 100:  
        return 1900 + year  
    else:  
        return year  
  
df['Year'] = df['Year'].apply(update_year)
```



2016.0	344
1999.0	338
1997.0	289
1996.0	263
1995.0	219
1994.0	121
1993.0	60
1981.0	46
1992.0	43
1991.0	41
1982.0	36
1990.0	34

Sale 칼럼 Data 정제



칼럼별 단위 통일

M, K 등의 단위가 혼용

단순 숫자형으로 바꾸어 알아보기
쉽게 데이터를 정제



Global_Sale 생성

각 지역별 Sale를 통합

전세계 판매량을 한눈에 알아볼 수 있음

칼럼별 단위 통일

Microsoft Game Studios	0.36	0.08	0M	0.02
Warner Bros. Interactive Entertainment	0.3	0.11	0.01	0.04
Square Enix	0.63	0.42	0.75	0.26
Konami Digital Entertainment	0.08	0	0	0.01
Bethesda Softworks	0.59	0.35	0.01	0.09
Inti Creates	0	0	0.01	0
Namco Bandai Games	0	0K	0.38	0

확인 결과 Dataset의 단위가 Million

코딩을 통해 모든 단위를 없애고 단순 순자형으로 치환
하여 더 알아보기 쉽게 함

Global_Sales

생성

NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0.52	0.01	0.12	0.05	0.70
0.00	0.04	0.00	0.57	0.61
5.17	4.05	4.34	0.79	14.35
0.23	0.16	0.00	0.03	0.42
0.13	0.05	0.00	0.00	0.18
0.25	0.02	0.00	0.03	0.30
0.00	0.00	0.02	0.00	0.02

NA, EU, JP, Other을 합한 Global_Sales.

각 지역별로 파악하기는 용이한 데이터지만 전세계 적으로 판매량을 집계하기는 어려워 새로운 칼럼 생성

지역별 게임 선호도



지역별 게임 선호도 구분

지역별 장르 선호도를 통한 지역구분

지역별 게임 선호도 차이 존재유무

지역별로 선호하는 장르에 차이가 있는지를 설명

지역별 선호 장르

각 지역별 선호도 차이 유무

각 지역별 선호 장르는 대개 비슷하나 일본은 다른 특성을 지님

North America: 'Action', 'Sports', 'Shooter', 'Platform', 'Misc'

Europe: 'Action', 'Sports', 'Shooter', 'Racing', 'Misc'

Japan: 'Role-Playing', 'Action', 'Sports', 'Platform', 'Misc'

Other: 'Action', 'Sports', 'Shooter', 'Racing', 'Misc'



Japan의 경우 nes,snes 시절부터 내려오던
JRPG 장르의 역사가 깊어 코어 팬층이 많아
Role-Playing 장르가 선호가 높음

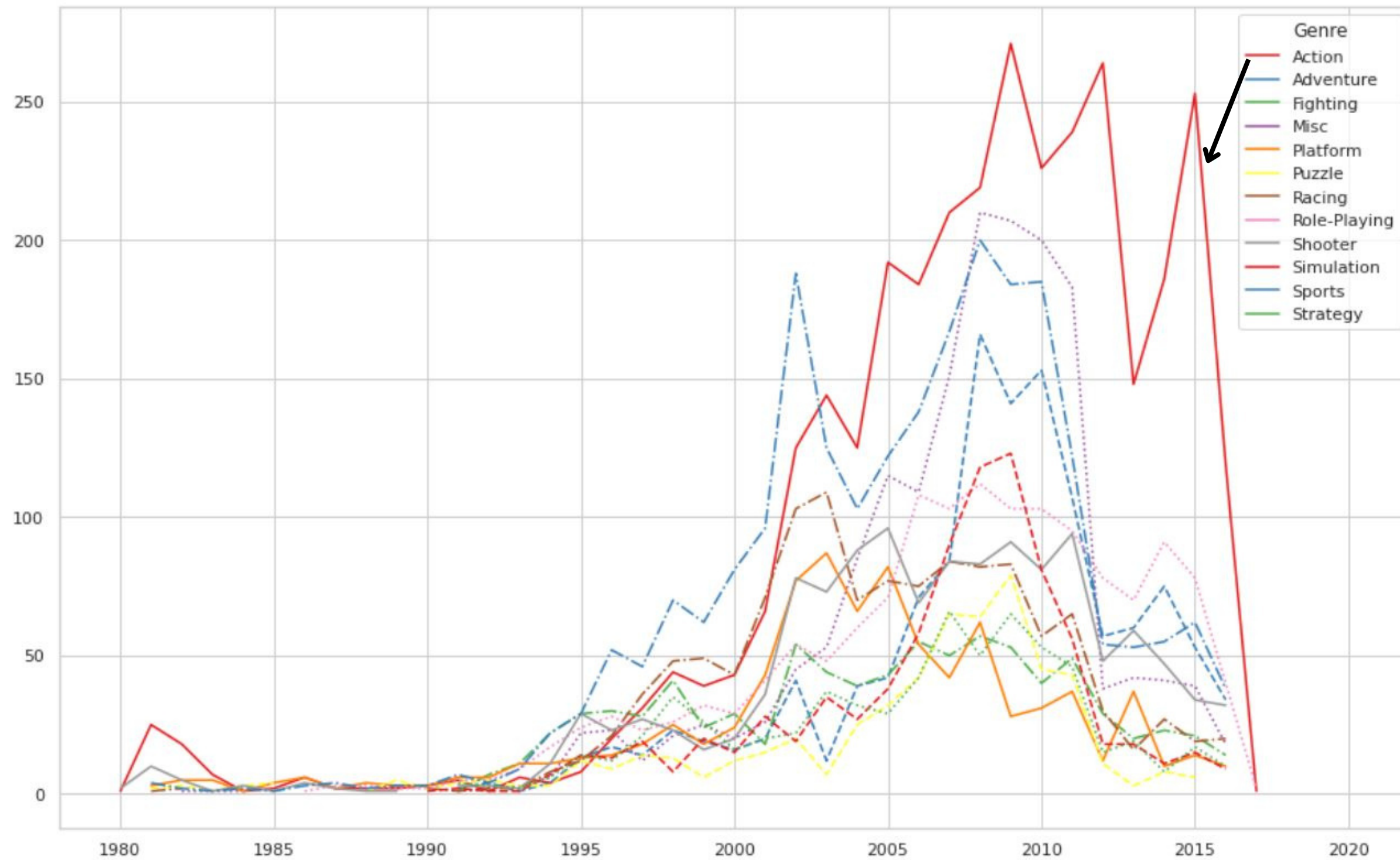
연도별 게임의 트렌드

각 연도별 장르 출고량 변화 시각화

장르별 점유율 변화 추이

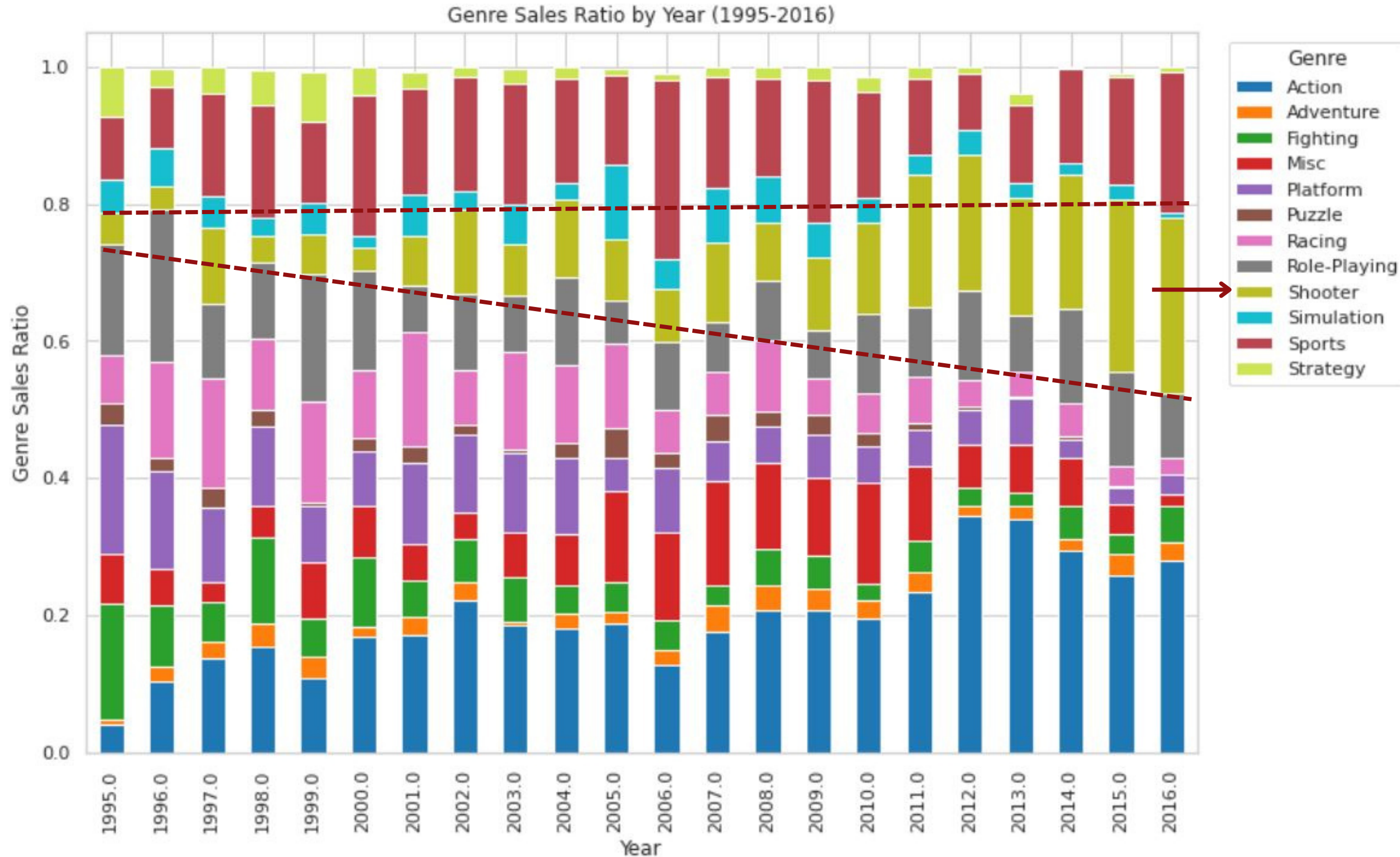
플랫폼 별 게임 출시 추이

각 연도별 장르 출고량 변화 시각화

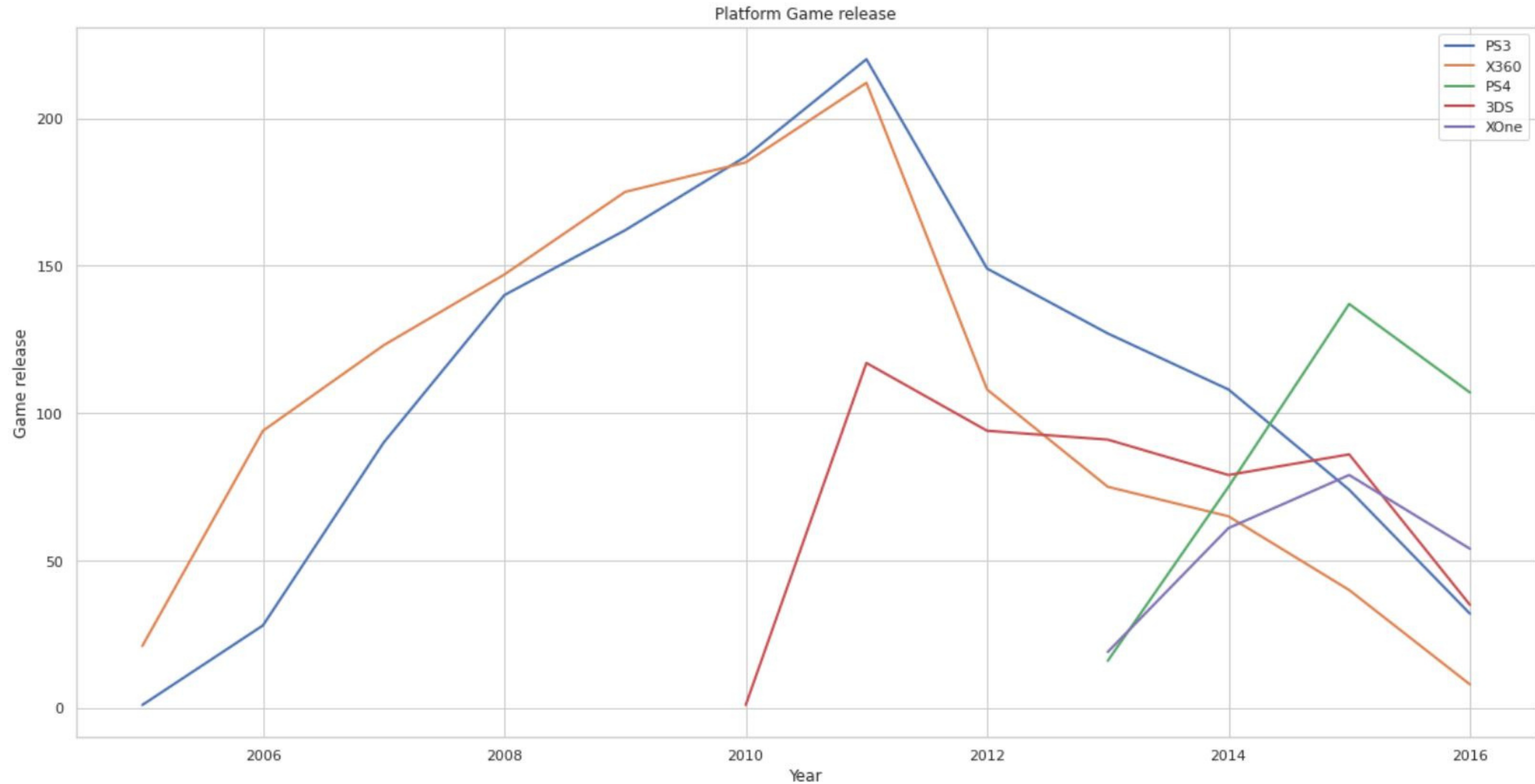


각 연도별 장르 출고량 변화 시각화

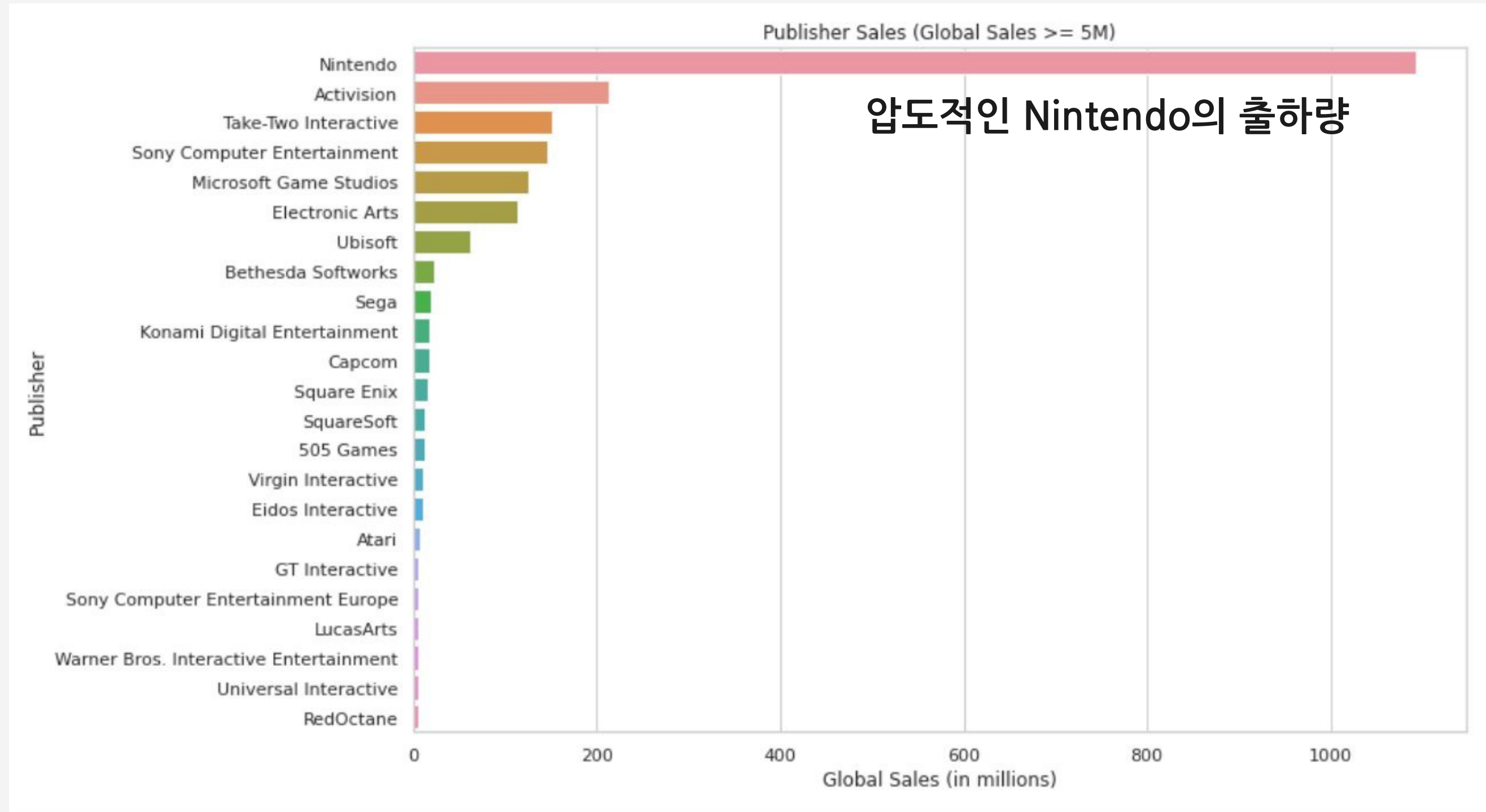
장르별 점유율 변화 추이



각 연도별 플랫폼 출고량 변화 시각화



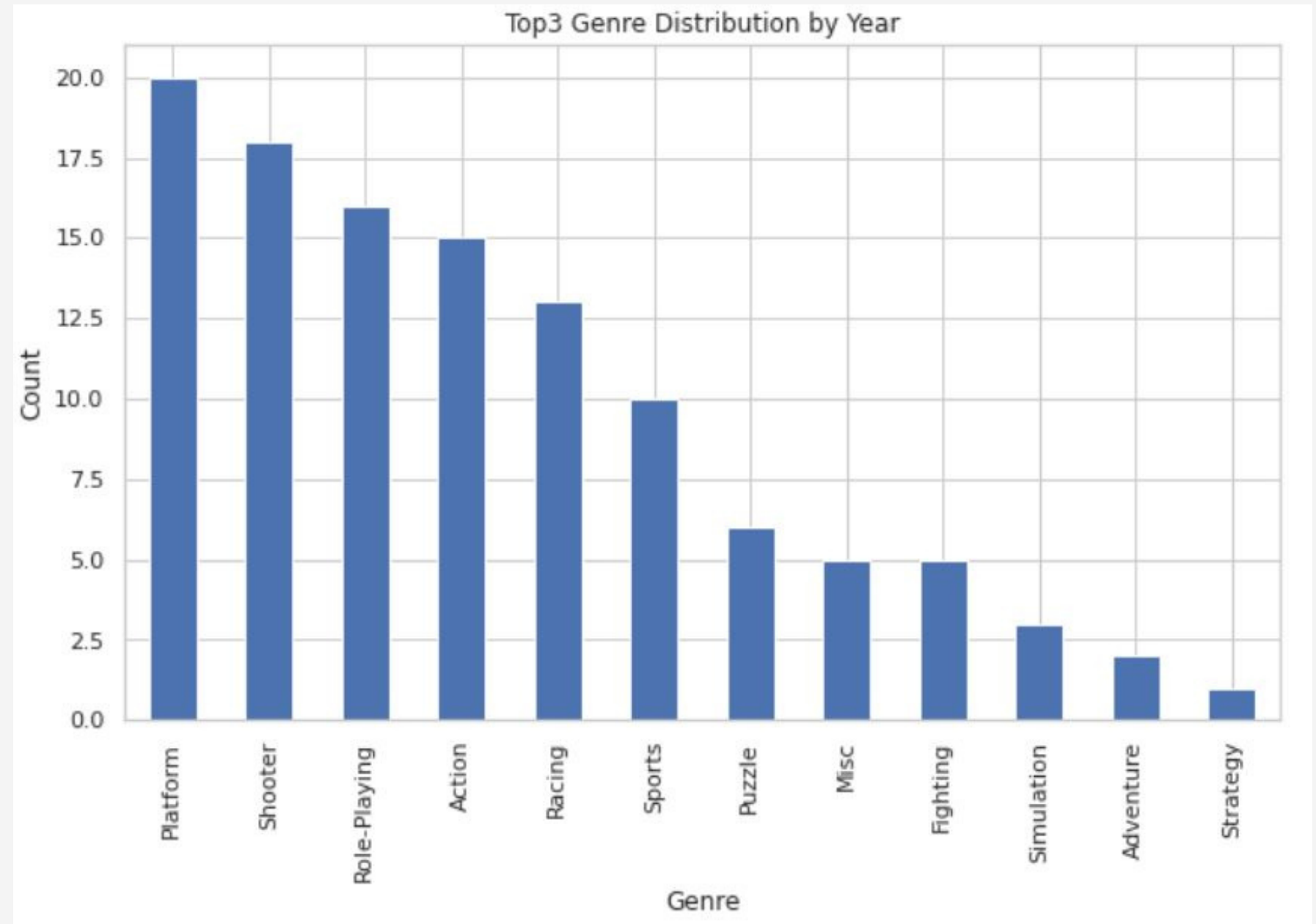
인기게임 분석



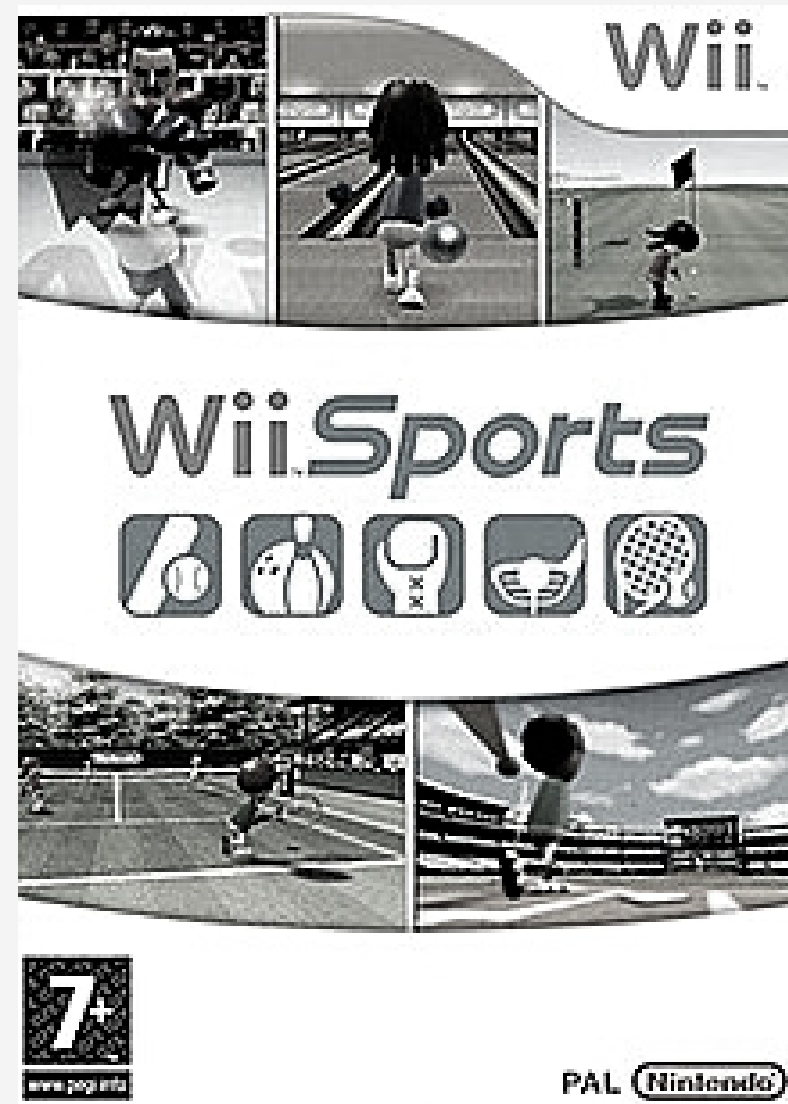
인기게임 분석

연도별 TOP 3 출하량 게임 장르

Platform	20
Shooter	18
Role-Playing	16
Action	15
Racing	13
Sports	10
Puzzle	6
Misc	5
Fighting	5
Simulation	3
Adventure	2
Strategy	1



닌텐도는 왜?



닌텐도는 왜?



압도적 IP 파워

Business Insight

상황 설정

데이터량이 풍부한 2016년 까지를 기준으로 삼아

다음 분기부터 개발에 착수

2018~9년 발매를 목표로 삼는다



장르 선정

가장 선호하는 장르인 Action과 성장세인 Shooter 장르로 선정



플랫폼 선정

성장세를 보이고 있는 PS4 및 PC 멀티 플랫폼



발매 지역 및 발매 시기 선정

전세계 발매를 통해 출고량을 극대화

감사합니다!

