



게임 트렌드 분석

데이터를 통한 분석 비교, 및 게임 개발 방향 조사



목차

1. 분석목표
2. 지역 별 게임 선호도 분석
3. 연도 별 게임 트렌드 분석
4. 높은 출고량의 게임 에 대한 분석
5. 개발 방향



1. 분석 목표

- 외부 데이터를 EDA를 통해 우리에게 유의미한 데이터로 처리한다.
- 처리한 데이터를 지역, 연도 등 여러 방면으로 분석을 통해 정보를 이해한다.
- 분석을 통한 이해를 바탕으로 우리가 개발해야 할 게임의 방향성을 정한다.

데이터 출처

<https://ds-lecture-data.s3.ap-northeast-2.amazonaws.com/datasets/vgames2.csv>

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	total
1	Candace Kane's Candy Factory	DS	2008.0	Action	Destineer	40000.0	0.0	0.0	0.0	40000.0
2	The Munchables	Wii	2009.0	Action	Namco Bandai Games	170000.0	0.0	0.0	10000.0	180000.0
3	Otome wa Oanesama Boku ni Koi Shiteru Portable	PSP	2010.0	Adventure	Alchemist	0.0	0.0	20000.0	0.0	20000.0
4	Deal or No Deal: Special Edition	DS	2010.0	Misc	Zoo Games	40000.0	0.0	0.0	0.0	40000.0
5	Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010.0	Platform	D3Publisher	120000.0	90000.0	0.0	40000.0	250000.0
...
16594	Ice Age 2: The Meltdown	GC	2006.0	Platform	Vivendi Games	150000.0	40000.0	0.0	10000.0	200000.0
16595	Rainbow Islands: Revolution	PSP	2005.0	Action	Rising Star Games	10000.0	0.0	0.0	0.0	10000.0
16596	NBA 2K16	PS3	2015.0	Sports	Take-Two Interactive	440000.0	190000.0	30000.0	130000.0	790000.0
16597	Toukiden: The Age of Demons	PSV	2013.0	Action	Tecmo Koei	50000.0	50000.0	250000.0	30000.0	380000.0
16598	The King of Fighters '95	PS	1996.0	Fighting	Sony Computer Entertainment	0.0	0.0	160000.0	10000.0	170000.0

16548 rows × 10 columns



2. 지역 별 게임 선호도 분석

- 데이터셋에 존재하는 북미, 유럽, 일본, 기타 등 각나라별로 비교하여 분석한다
- 나라에 따라 선호하는 게임이 다른가에 대해 비교 분석 한다



수학적 검정

χ^2 검정(카이제곱 검정)이란?

수학식을 계산하여 얻은 값을 통해 데이터들간에 관계가 있는지 아님 그냥 우연히 만들어진건지 비교하기 위한 가설검정방법

- 귀무가설(H_0): 나라와 출고량과는 관계가 없다 $p\text{-value} > 0.1(10\%)$
- 대안가설(H_1): 나라와 출고량과는 관계가 있다 $p\text{-value} < 0.05(5\%)$

```

from scipy.stats import chi2_contingency

df1 = df.loc[:, 'Genre']
df2 = df.loc[:, 'NA_Sales': 'Other_Sales']

GenbyCon = pd.concat([df1, df2], axis=1)
GenbyCon.set_index('Genre', inplace=True)
GenbyCon = GenbyCon[~((GenbyCon['NA_Sales'] == 0) & (GenbyCon['EU_Sales'] == 0) & (GenbyCon['JP_Sales'] == 0) & (GenbyCon['Other_Sales'] == 0))] # 모든 값이 0인 row가 존재, 삭제 필요
GenbyCon_total = GenbyCon.groupby('Genre').mean() # 그룹화 시킴

print('p-value:', chi2_contingency(GenbyCon_total, correction = False)[1])
print('\n')
print('p-value:', chi2_contingency(GenbyCon_total, correction = True)[1])

#p-value가 0이므로 귀무가설을 기각, 장르와 나라간의 관계는 독립적이지 않다.(지역에 따라 선호하는 게임이 있다.)

```

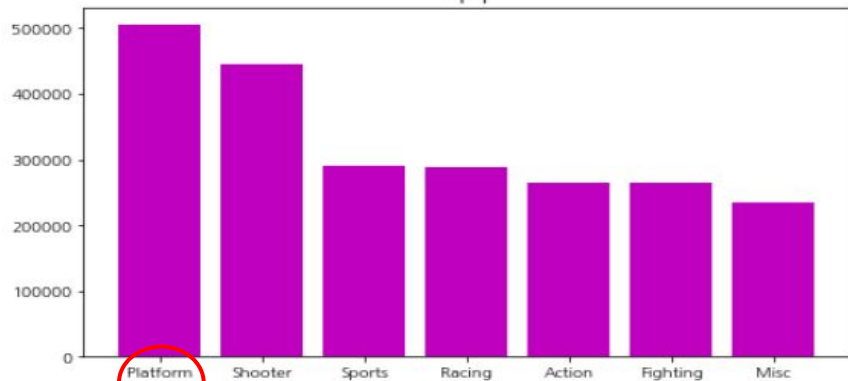
p-value : 0.0

p-value : 0.0

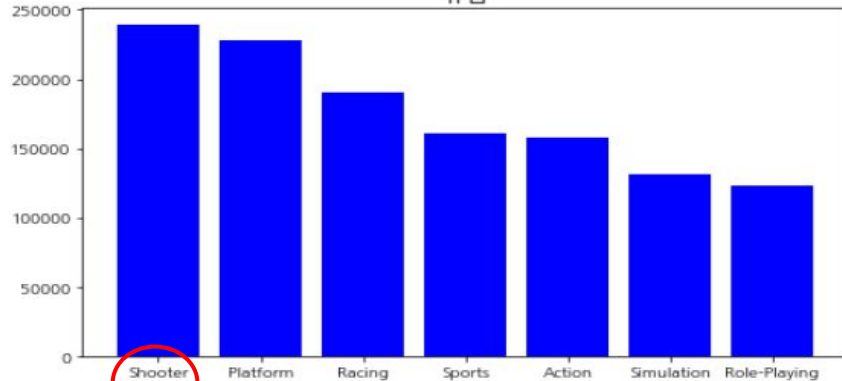
결론 : $p\text{-value} < 0.05(5\%)$ 이므로 서로 나라와 출고량은 서로 관계가 있다

나라별 인기 장르

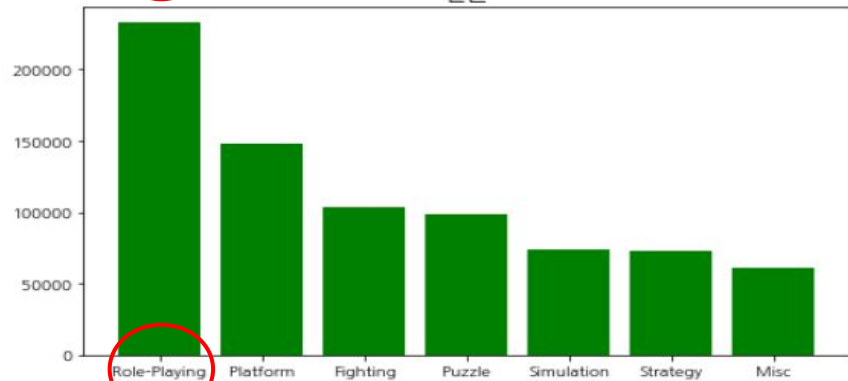
북미



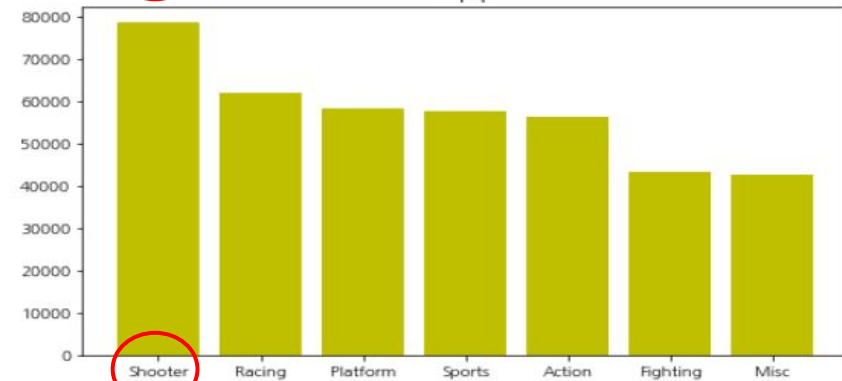
유럽



일본



기타





3. 연도별 게임 트렌드 분석

- 연도에 따라 어떤 게임들이 출시 했는지 분석한다
- 연도 별로 유행하는 게임 장르가 있었는지 분석한다

- 너무 오래된 정보는 현시장과는 의미가 없다고 판단하여 2006년 부터 게임을 분석
- 2017 이후의 정보는 현저히 부족하여 무의하다고 판단하여 제외

```
print('p-value :',chi2_contingency(Year_Gen, correction = False)[1])  
print('\n')  
print('p-value :',chi2_contingency(Year_Gen, correction = True)[1])
```

p-value : 6.024932453033478e-121

p-value : 6.024932453033478e-121

e-121는 소숫점 121번째부터라는 의미

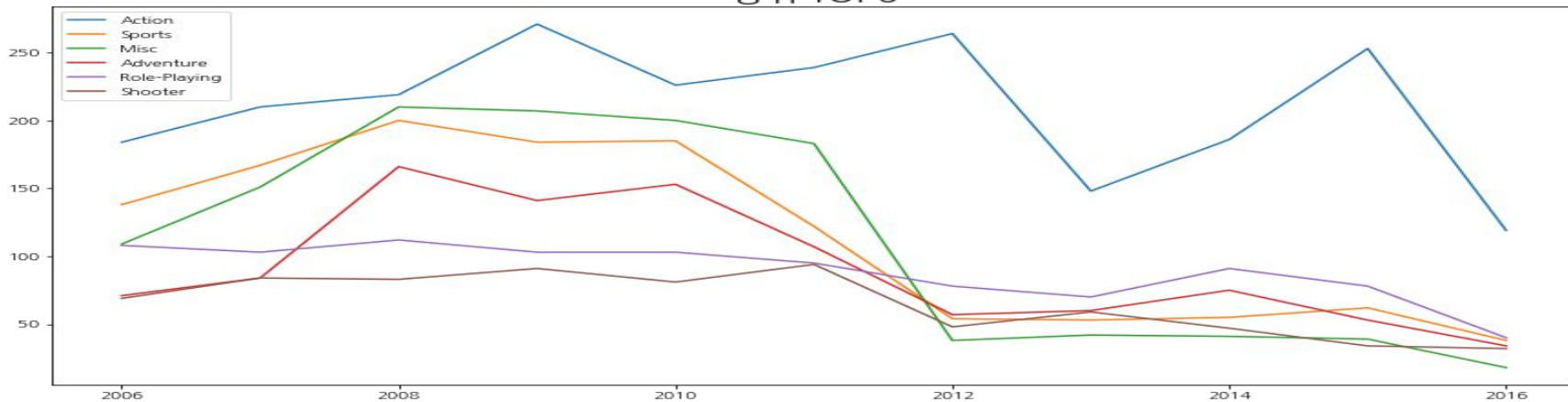
0.000000.....6024.....

0에 거의 가깝다는 뜻

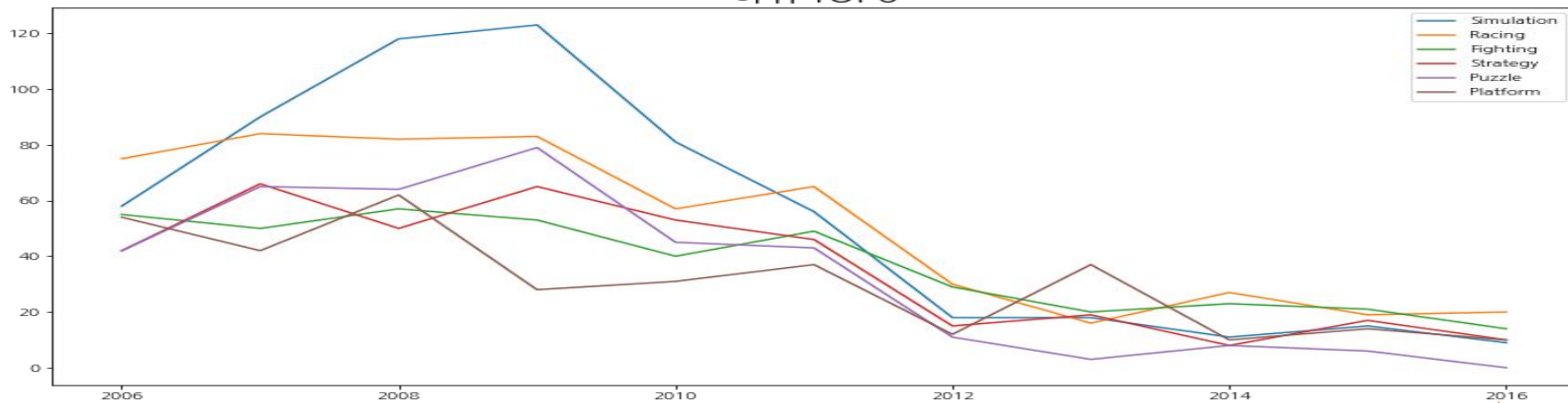
결론 : $p\text{-value} < 0.05(5\%)$ 이므로 서로 나라와 출고량은 서로 관계가 있다

출시게임수

상위 TOP6



하위 TOP6



연도

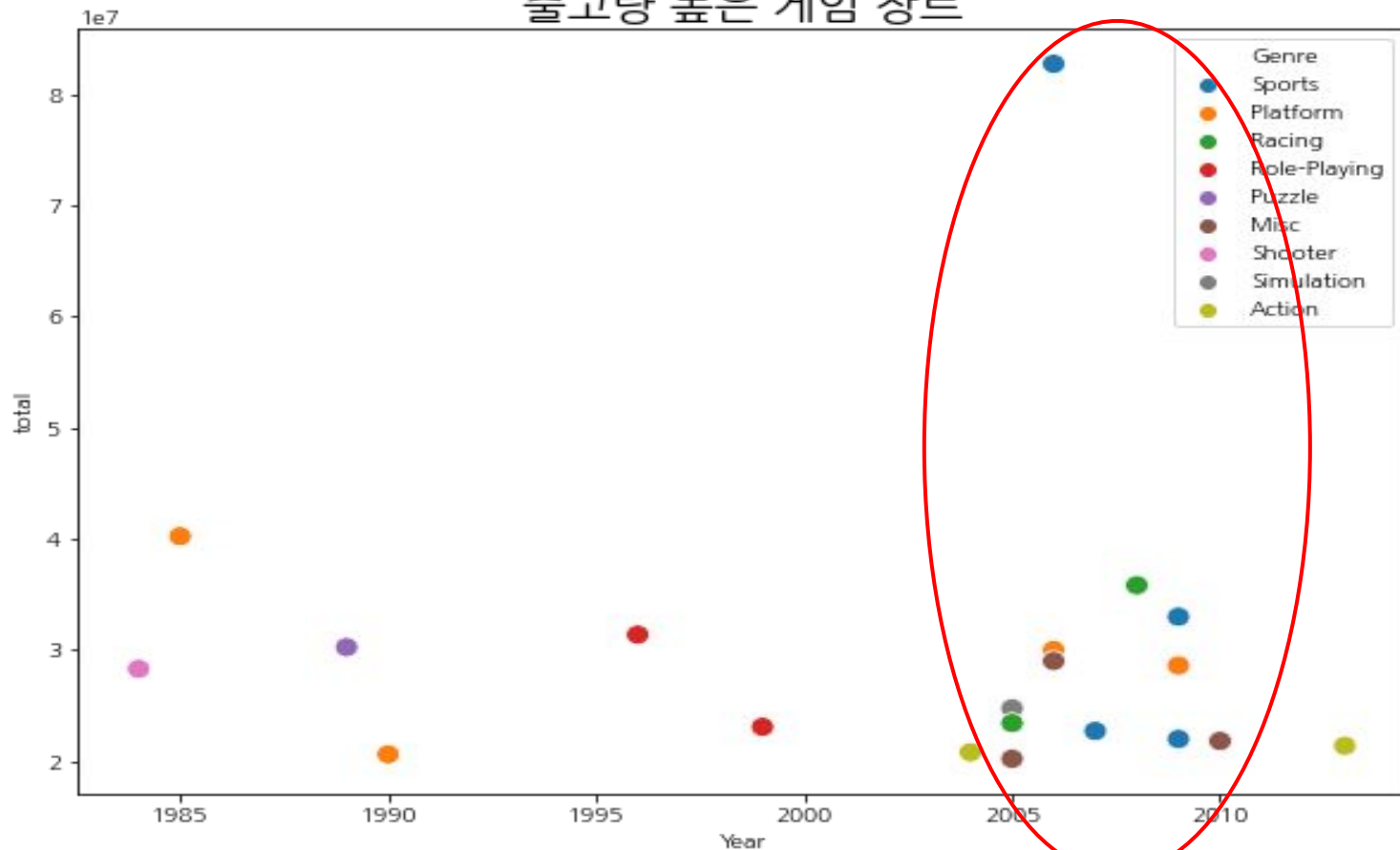


4. 높은 출고량의 게임에 대한 분석

- 높은 출고량의 게임은 무엇이 있는지 분석한다
- 높은 출고량의 게임에 대해 특성을 분석한다

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	total
2909	Wii Sports	Wii	2006.0	Sports	Nintendo	41490000.0	29020000.0	3770000.0	8460000.0	82740000.0
4200	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29080000.0	3580000.0	6810000.0	770000.0	40240000.0
12580	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15350000.0	12880000.0	3790000.0	3310000.0	35830000.0
3353	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15150000.0	11010000.0	3280000.0	2960000.0	33000000.0
7830	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11270000.0	8890000.0	10220000.0	1000000.0	31380000.0
1204	Tetris	GB	1989.0	Puzzle	Nintendo	23200000.0	2260000.0	4220000.0	580000.0	30260000.0
6938	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11380000.0	9230000.0	6500000.0	2900000.0	30010000.0
3015	Wii Play	Wii	2006.0	Misc	Nintendo	14030000.0	9200000.0	2930000.0	2850000.0	29010000.0
16532	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14590000.0	7060000.0	4700000.0	2260000.0	28610000.0
10489	Duck Hunt	NES	1984.0	Shooter	Nintendo	26930000.0	630000.0	280000.0	470000.0	28310000.0
11138	Nintendogs	DS	2005.0	Simulation	Nintendo	9070000.0	11000000.0	1930000.0	2750000.0	24750000.0
10651	Mario Kart DS	DS	2005.0	Racing	Nintendo	9810000.0	7570000.0	4130000.0	1920000.0	23430000.0
9417	Pokemon Gold/Pokemon Silver	GB	1999.0	Role-Playing	Nintendo	9000000.0	6180000.0	7200000.0	710000.0	23090000.0
3436	Wii Fit	Wii	2007.0	Sports	Nintendo	8940000.0	8030000.0	3600000.0	2150000.0	22720000.0
10900	Wii Fit Plus	Wii	2009.0	Sports	Nintendo	9090000.0	8590000.0	2530000.0	1790000.0	22000000.0
4344	Kinect Adventures!	X360	2010.0	Misc	Microsoft Game Studios	14970000.0	4940000.0	240000.0	1670000.0	21820000.0
3484	Grand Theft Auto V	PS3	2013.0	Action	Take-Two Interactive	7010000.0	9270000.0	970000.0	4140000.0	21390000.0
14670	Grand Theft Auto: San Andreas	PS2	2004.0	Action	Take-Two Interactive	9430000.0	400000.0	410000.0	10570000.0	20810000.0
13112	Super Mario World	SNES	1990.0	Platform	Nintendo	12780000.0	3750000.0	3540000.0	550000.0	20620000.0
13230	Brain Age: Train Your Brain in Minutes a Day	DS	2005.0	Misc	Nintendo	4750000.0	9260000.0	4160000.0	2050000.0	20220000.0

출고량 높은 게임 장르

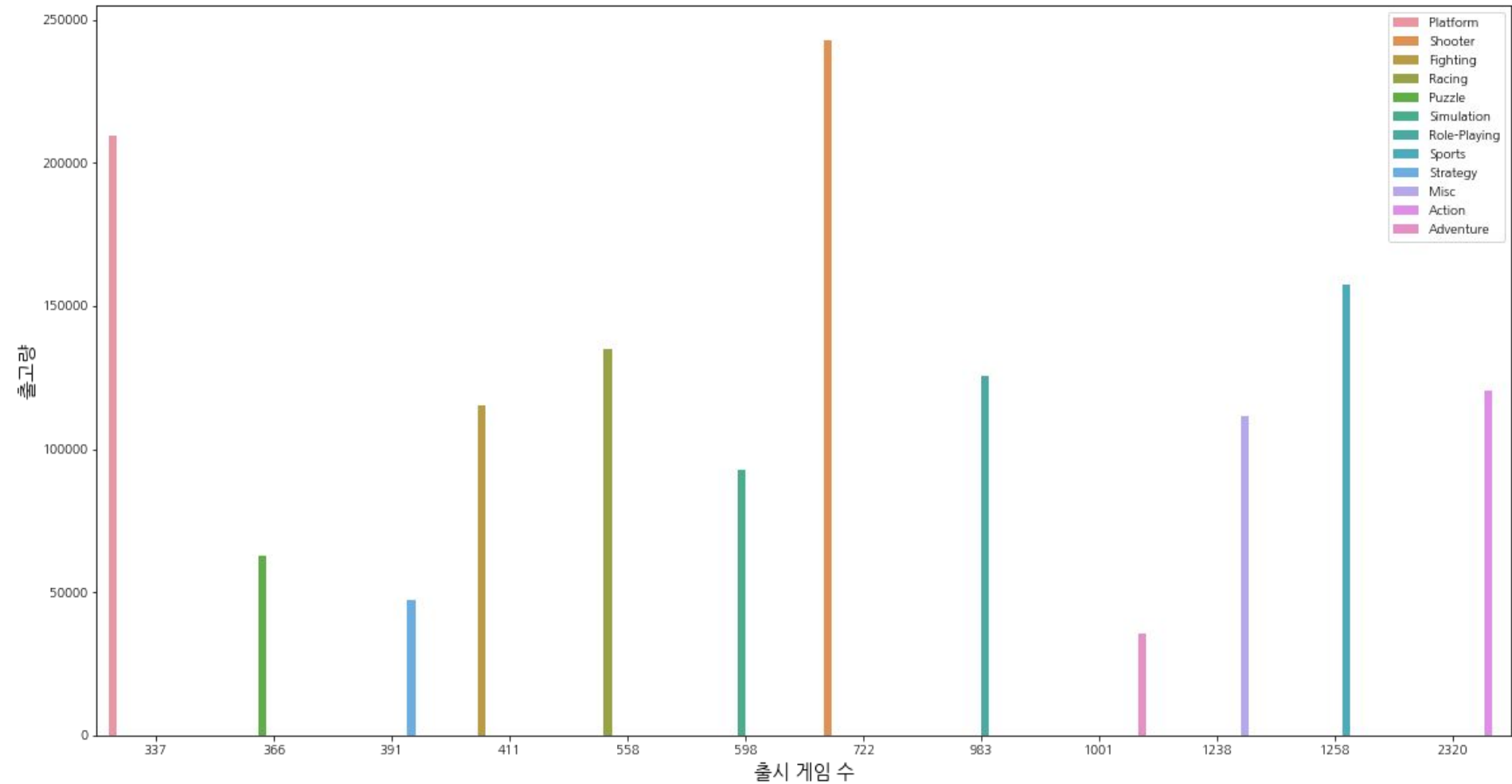




5. 개발 방향

- 수요가 많으면서 공급이 적은 장르를 분석한다
- 데이터를 통해 우리가 우리가 설계해야할 게임을 분석한다

- 오래된 데이터는 우리의 개발방향에 대해 의미가 없다고 판단, 2006년부터 분석
- 2017이후의 데이터는 현저히 적어 무의미하다고 판단, 제외



게임당 평균 출고량

1. Platform(26.4%)

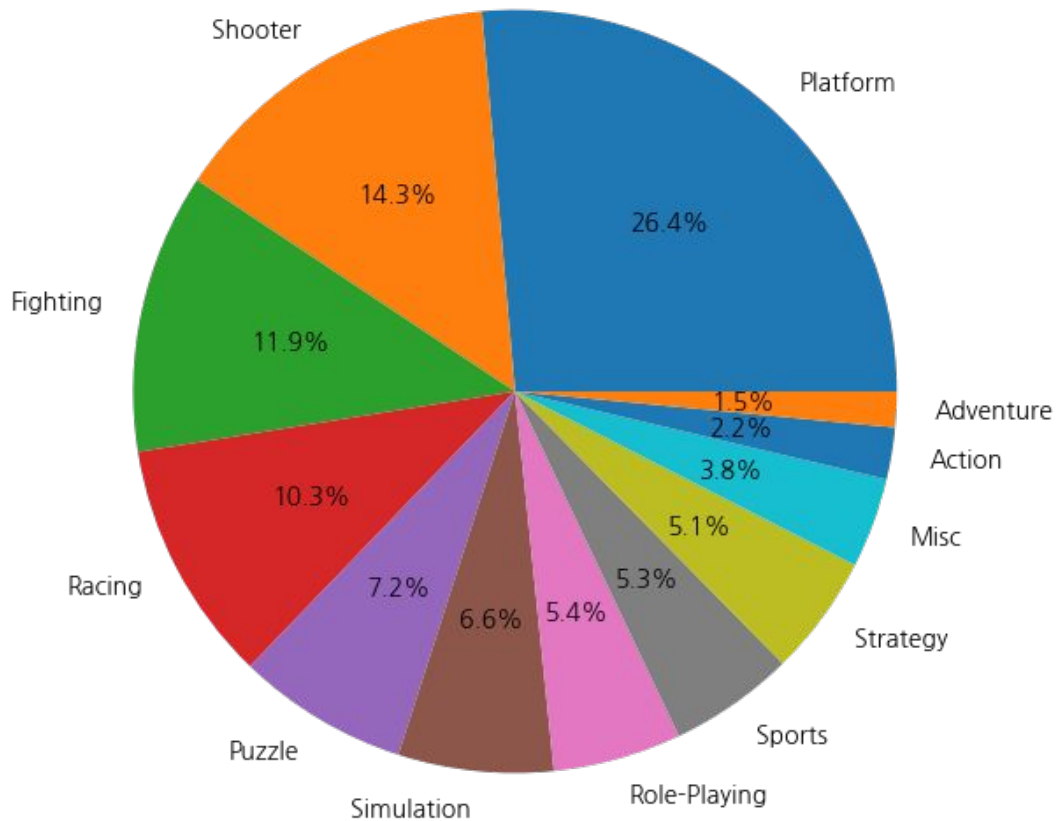
2. Shooter(14.3%)

3. Fighting(11.9%)



52.6%

장르별 출고량 비율





END