

· SIMPLE PRESENTATION ·

# Data Project

AI\_18\_박찬선

# CONTENTS

## 001

### 프로젝트의 목표 및 설정

- 프로젝트의 목표 설명
- 프로젝트 진행을 위한 상황 설정

## 002

### 데이터셋 설명

- 분석에 사용한 데이터 설명

## 003

### 데이터 분석을 위한 EDA

- EDA 적용이 필요한 사례
- EDA를 적용 후 데이터셋

## 004

### 데이터 분석 시각화

- 지역에 따른 게임 장르 선호도
- 연도별 게임 트렌드 변화
- 인기가 많은 게임 분석

## 005

### 결론 도출

## 프로젝트의 목표 및 설정



"2017년에 새롭게 콘솔형 게임 시장에 진출하려는 국내 기업"

# 데이터셋 설명

10개의 컬럼과 약 16400개의 행으로 이루어진 데이터셋

## 열 종류

Name : 게임의 이름

Platform : 게임 플랫폼의 이름

Year : 게임이 출시된 연도

Genre : 게임의 장르

Publisher : 게임을 배급한 회사

NA\_Sales : 북미지역에서의 출고량

EU\_Sales : 유럽지역에서의 출고량

JP\_Sales : 일본지역에서의 출고량

Other\_Sales : 기타지역에서의 출고량

(단위: 100만)

Unnamed: 0		Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
0	1	Candace Kane's Candy Factory	DS	2008.0	Action	Destineer	0.04	0	0	0
1	2	The Munchables	Wii	2009.0	Action	Namco Bandai Games	0.17	0	0	0.01
2	3	Otome wa Oanesama Boku ni Koi Shiteru Portable	PSP	2010.0	Adventure	Alchemist	0	0	0.02	0
3	4	Deal or No Deal: Special Edition	DS	2010.0	Misc	Zoo Games	0.04	0	0	0
4	5	Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010.0	Platform	D3Publisher	0.12	0.09	0	0.04

# 데이터 분석을 위한 EDA

	Unnamed: 0	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
4	5	Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010.0	Platform	D3Publisher	0.12	0.09	0	0.04
10	11	Ford Racing	PS	2001.0	Racing	Empire Interactive	480K	0.33M	0K	0.06
31	32	Rock Band	X360	NaN	Misc	Electronic Arts	1.93	0.34	0	0.21
74	75	Age of Empires III: Complete Collection	PC	9.0	Strategy	Microsoft Game Studios	0.01	0.12	0	0.03
7346	7347	Madden NFL 13	PS3	2012.0	Sports	Electronic Arts	2.11	0.23	0	0.22
8717	8718	Madden NFL 13	PS3	2012.0	Sports	Electronic Arts	0	0.01	0	0

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0      16598 non-null  int64
1   Name            16598 non-null  object
2   Platform        16598 non-null  object
3   Year            16327 non-null  float64
4   Genre           16548 non-null  object
5   Publisher       16540 non-null  object
6   NA_Sales        16598 non-null  object
7   EU_Sales        16598 non-null  object
8   JP_Sales        16598 non-null  object
9   Other_Sales     16598 non-null  object
```

본격적인 데이터 분석에 앞서 데이터셋의  
결측치, 이상치를 파악하고 조정하는 작업이 필요함.



# 데이터 분석을 위한 EDA

Unnamed: 0		Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
4	5	Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010.0	Platform	D3Publisher	0.12	0.09	0	0.04
10	11	Ford Racing	PS	2001.0	Racing	Empire Interactive	480K	0.33M	0K	0.06
31	32	Rock Band	X360	NaN	Misc	Electronic Arts	1.93	0.34	0	0.21
74	75	Age of Empires III: Complete Collection	PC	9.0	Strategy	Microsoft Game Studios	0.01	0.12	0	0.03
7346	7347	Madden NFL 13	PS3	2012.0	Sports	Electronic Arts	2.11	0.23	0	0.22
8717	8718	Madden NFL 13	PS3	2012.0	Sports	Electronic Arts	0	0.01	0	0



		Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	All_Sales
1113		Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010	Platform	D3Publisher	0.12	0.09	0.0	0.04	0.25
257		Age of Empires III: Complete Collection	PC	2009	Strategy	Microsoft Game Studios	0.01	0.12	0.0	0.03	0.16
4614		Ford Racing	PS	2001	Racing	Empire Interactive	0.48	0.33	0.0	0.06	0.87
7589		Madden NFL 13	PS3	2012	Sports	Electronic Arts	2.11	0.24	0.0	0.22	2.57

# 데이터 분석을 위한 EDA

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0      16598 non-null  int64
1   Name            16598 non-null  object
2   Platform        16598 non-null  object
3   Year            16327 non-null  float64
4   Genre           16548 non-null  object
5   Publisher       16540 non-null  object
6   NA_Sales        16598 non-null  object
7   EU_Sales        16598 non-null  object
8   JP_Sales        16598 non-null  object
9   Other_Sales     16598 non-null  object
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16326 entries, 0 to 16325
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Name            16326 non-null  object
1   Platform        16326 non-null  object
2   Year            16326 non-null  int64
3   Genre           16296 non-null  object
4   Publisher       16291 non-null  object
5   NA_Sales        16326 non-null  float64
6   EU_Sales        16326 non-null  float64
7   JP_Sales        16326 non-null  float64
8   Other_Sales     16326 non-null  float64
9   All_Sales       16326 non-null  float64
```

해당 조정된 새로운 데이터셋을 사용하여  
여러 시각화 도구를 통해  
어떤 게임을 기획해야 하는지 인사이트 도출

# 조정된 데이터셋

10개의 컬럼과 약 16400개의 행으로 이루어진 데이터셋

## 열 종류

Name : 게임의 이름

Platform : 게임 플랫폼의 이름

Year : 게임이 출시된 연도

Genre : 게임의 장르

Publisher : 게임을 배급한 회사

NA\_Sales : 북미지역에서의 출고량

EU\_Sales : 유럽지역에서의 출고량

JP\_Sales : 일본지역에서의 출고량

Other\_Sales : 기타지역에서의 출고량

All\_Sales : 전체 지역에서의 출고량

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	All_Sales
0	'98 Koshien	PS	1998	Sports	Magical Company	0.15	0.10	0.12	0.03	0.40
1	.hack//G.U. Vol.1//Rebirth	PS2	2006	Role-Playing	Namco Bandai Games	0.00	0.00	0.17	0.00	0.17
2	.hack//G.U. Vol.2//Reminisce	PS2	2006	Role-Playing	Namco Bandai Games	0.11	0.09	0.00	0.03	0.23
3	.hack//G.U. Vol.2//Reminisce (jp sales)	PS2	2006	Role-Playing	Namco Bandai Games	0.00	0.00	0.16	0.00	0.16
4	.hack//G.U. Vol.3//Redemption	PS2	2007	Role-Playing	Namco Bandai Games	0.00	0.00	0.17	0.00	0.17



**1 지역에 따른  
선호하는 게임 장르의  
차이 여부**

**2 연도별 게임의 트렌드**

연도별 판매량 변화  
연도별 배출된 게임 장르 비율  
연도별 배출된 플랫폼 비율

**3 인기가 많은 게임 분석**

연도별 게임 장르 판매량  
연도별 플랫폼 판매량

**4 결론 도출**



KEY POINT

## 데이터 분석 시각화

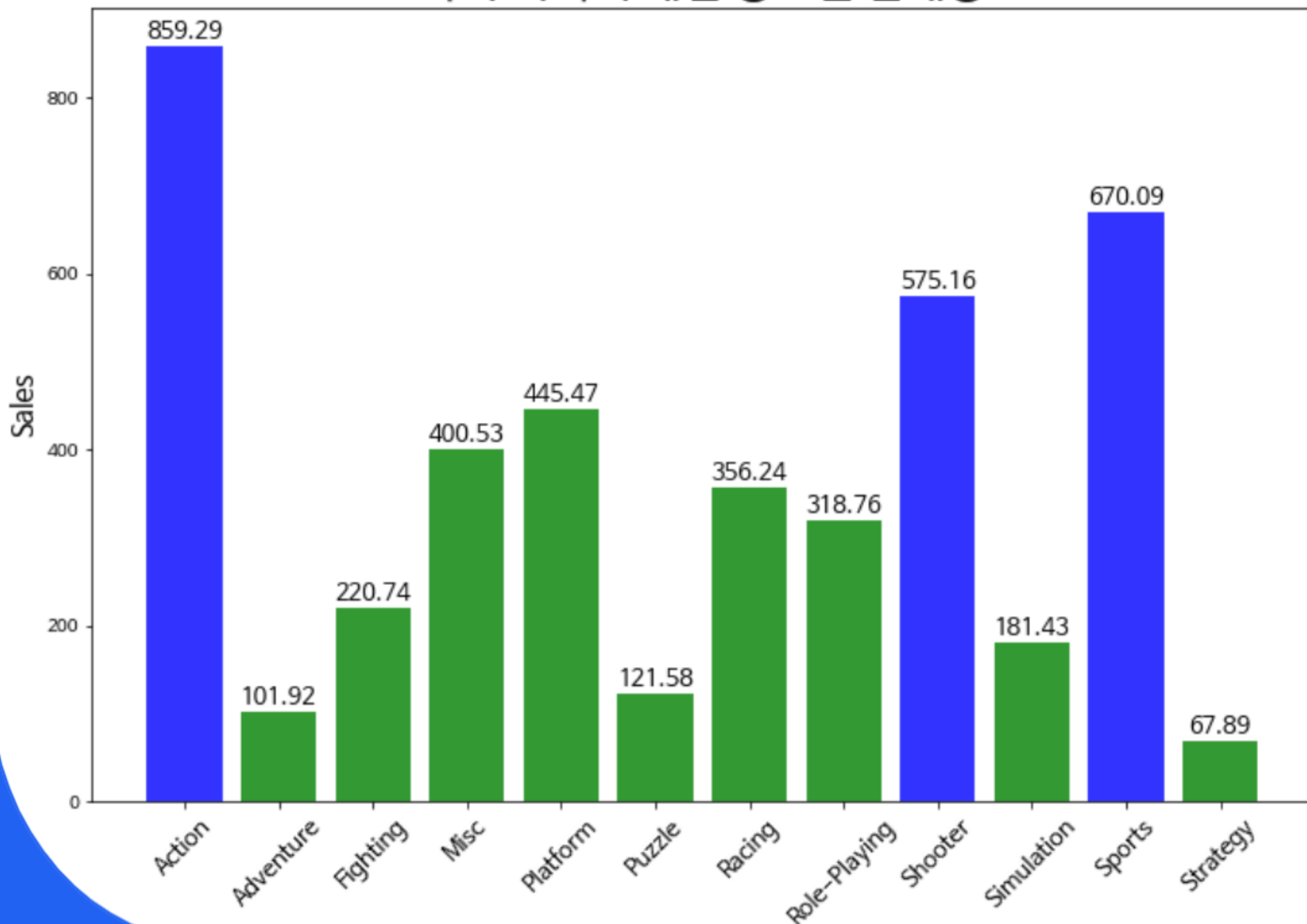
크게 3가지 주제로  
데이터 시각화를 통해 인사이트를 도출합니다.

**가정 상황**

"2017년에 새롭게 콘솔형 게임 시장에 진출하려는 국내 기업"

# 지역에 따른 선호하는 게임 장르의 차이 - 북미

북미 지역의 게임 장르별 판매량



---

북미 지역의  
게임 장르별 판매량 상위 3개 :  
Action > Sports > Shooter

---

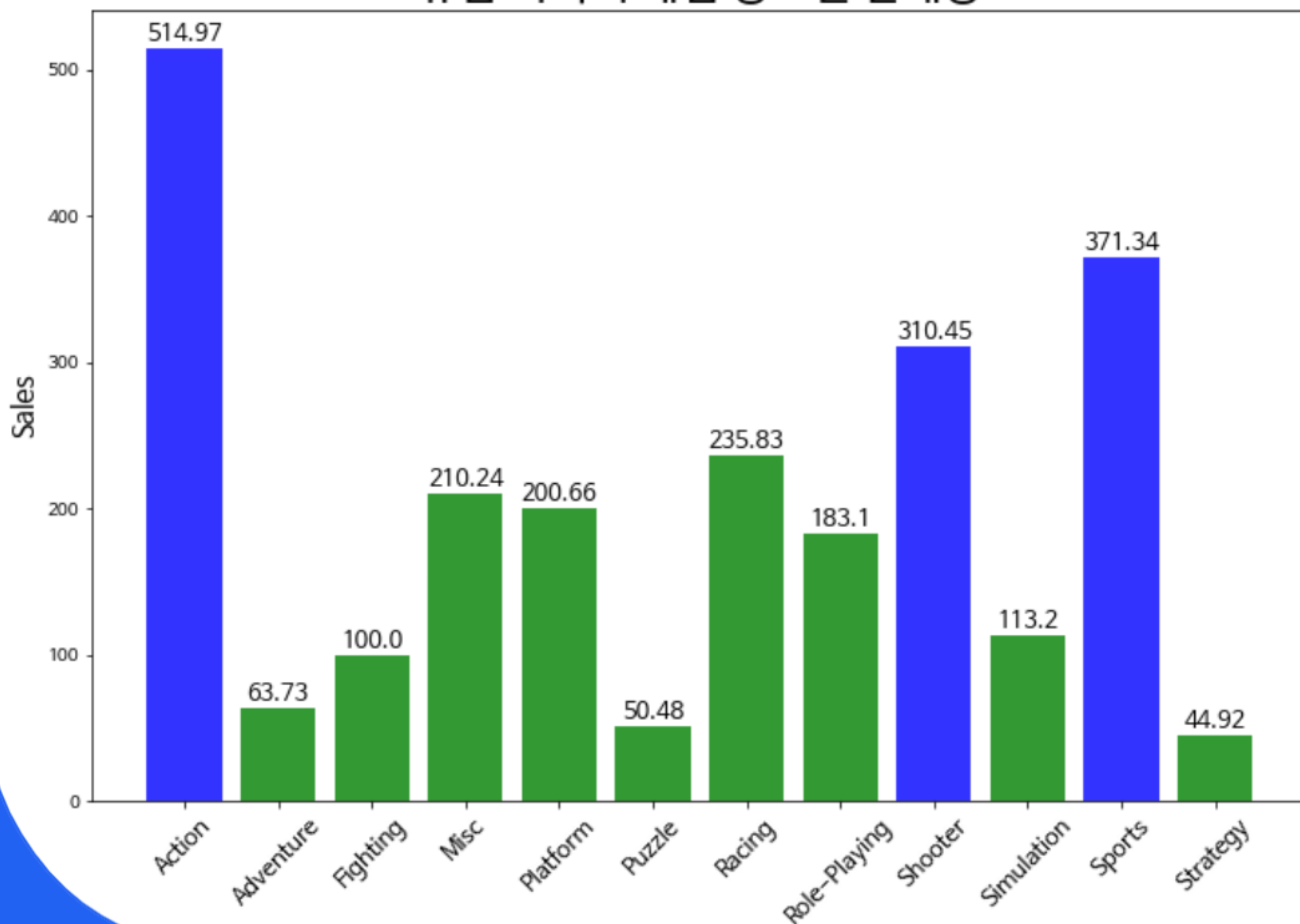
---

북미 지역의  
게임 장르별 판매량 하위 3개 :  
Strategy < Adventure < Puzzle

---

# 지역에 따른 선호하는 게임 장르의 차이 - 유럽

유럽 지역의 게임 장르별 판매량



---

유럽 지역의  
게임 장르별 판매량 상위 3개 :  
Action > Sports > Shooter

---

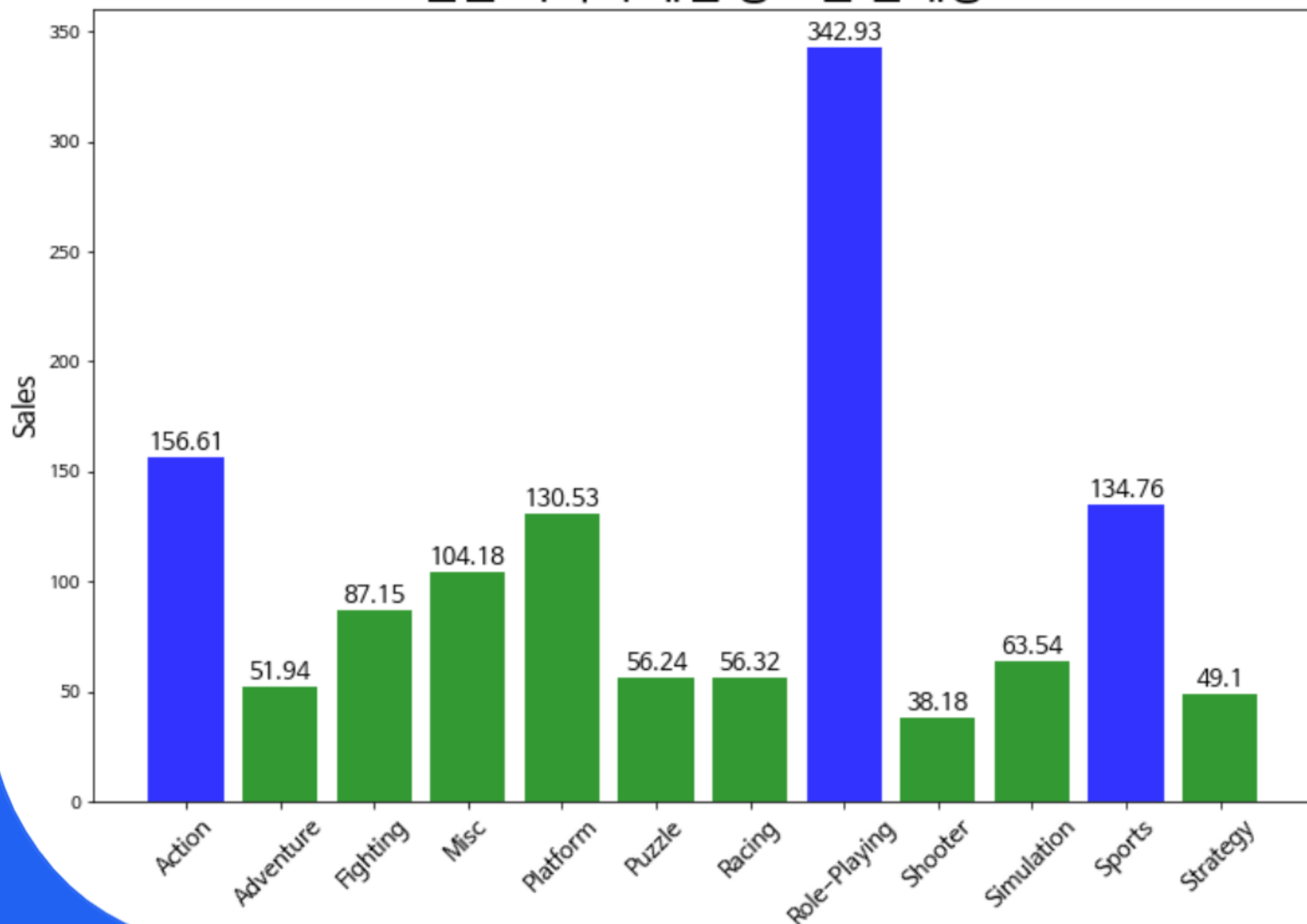
---

유럽 지역의  
게임 장르별 판매량 하위 3개 :  
Strategy < Puzzle < Adventure

---

# 지역에 따른 선호하는 게임 장르의 차이 - 일본

일본 지역의 게임 장르별 판매량



---

일본 지역의  
게임 장르별 판매량 상위 3개 :  
Role-Playing > Action > Sports

---

---

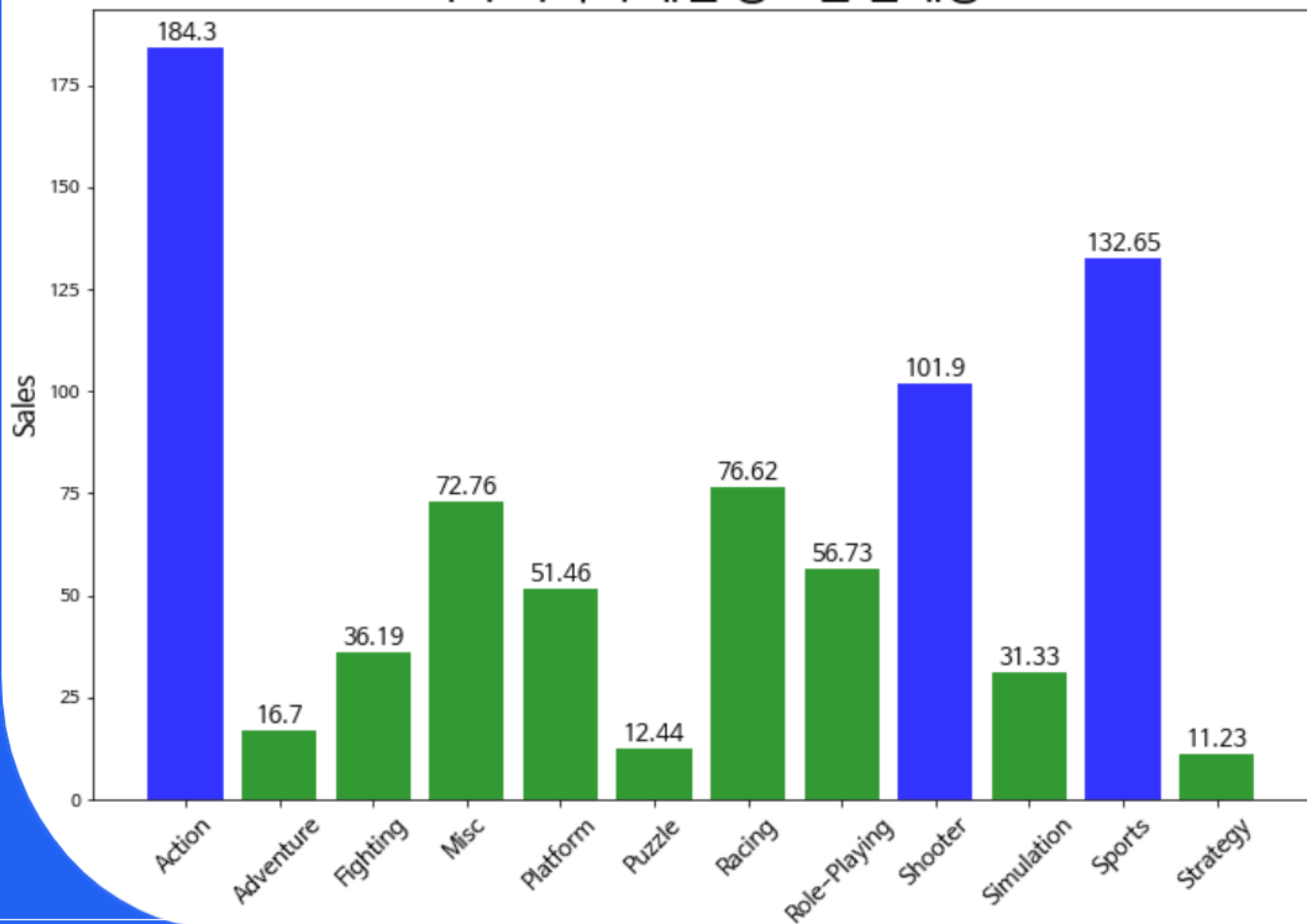
일본 지역의  
게임 장르별 판매량 하위 3개 :  
Shooter < Strategy < Adventure

---



# 지역에 따른 선호하는 게임 장르의 차이 - 기타

기타 지역의 게임 장르별 판매량



---

기타 지역의  
게임 장르별 판매량 상위 3개 :  
Action > Sports > Shooter

---

---

기타 지역의  
게임 장르별 판매량 하위 3개 :  
Strategy < Puzzle < Adventure

---

# 지역에 따른 선호하는 게임 장르의 차이

북미 지역의

게임 장르별 판매량 상위 3개 :  
**Action** > Sports > Shooter

유럽 지역의

게임 장르별 판매량 상위 3개 :  
**Action** > Sports > Shooter

일본 지역의

게임 장르별 판매량 상위 3개 :  
**Role-Playing** > Action > Sports

기타 지역의

게임 장르별 판매량 상위 3개 :  
**Action** > Sports > Shooter

북미 지역의

게임 장르별 판매량 하위 3개 :  
Strategy > Adventure > Puzzle

유럽 지역의

게임 장르별 판매량 하위 3개 :  
Strategy < Puzzle < Adventure

일본 지역의

게임 장르별 판매량 하위 3개 :  
Shooter < Strategy < Adventure

기타 지역의

게임 장르별 판매량 하위 3개 :  
Strategy < Puzzle < Adventure

## 지역에 따른 선호하는 게임 장르의 차이는

 있다

 없다

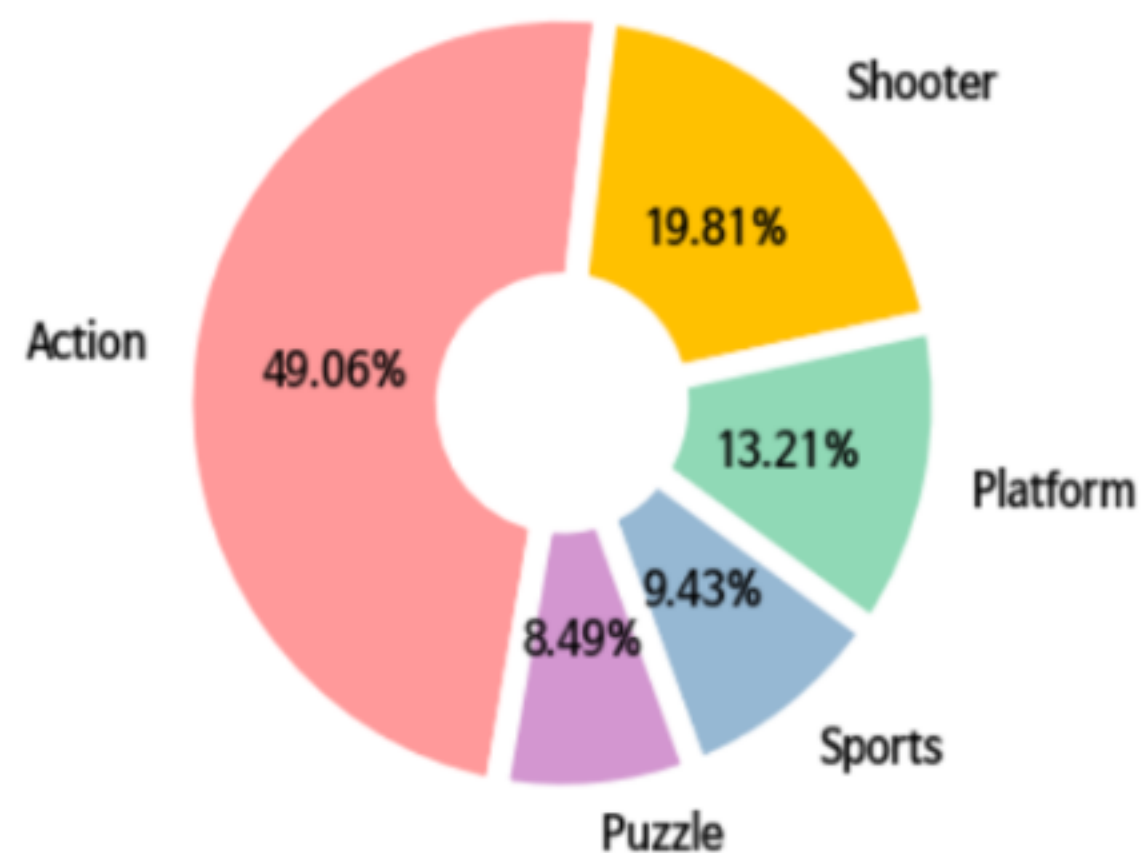
선호하는 게임 장르는 공통적으로  
Action, Sports, Shooter 가 상위권이다.  
반면,  
비선호하는 게임장르는 공통적으로  
Strategy, Adventure, Puzzle 가 하위권이다.

다만,  
일본은 Role-Playing을 선호하고  
Shooter는 비선호한다는 점이 특이점이다.

# 연도별 게임의 트렌드

## 연도별 배출된 게임 장르

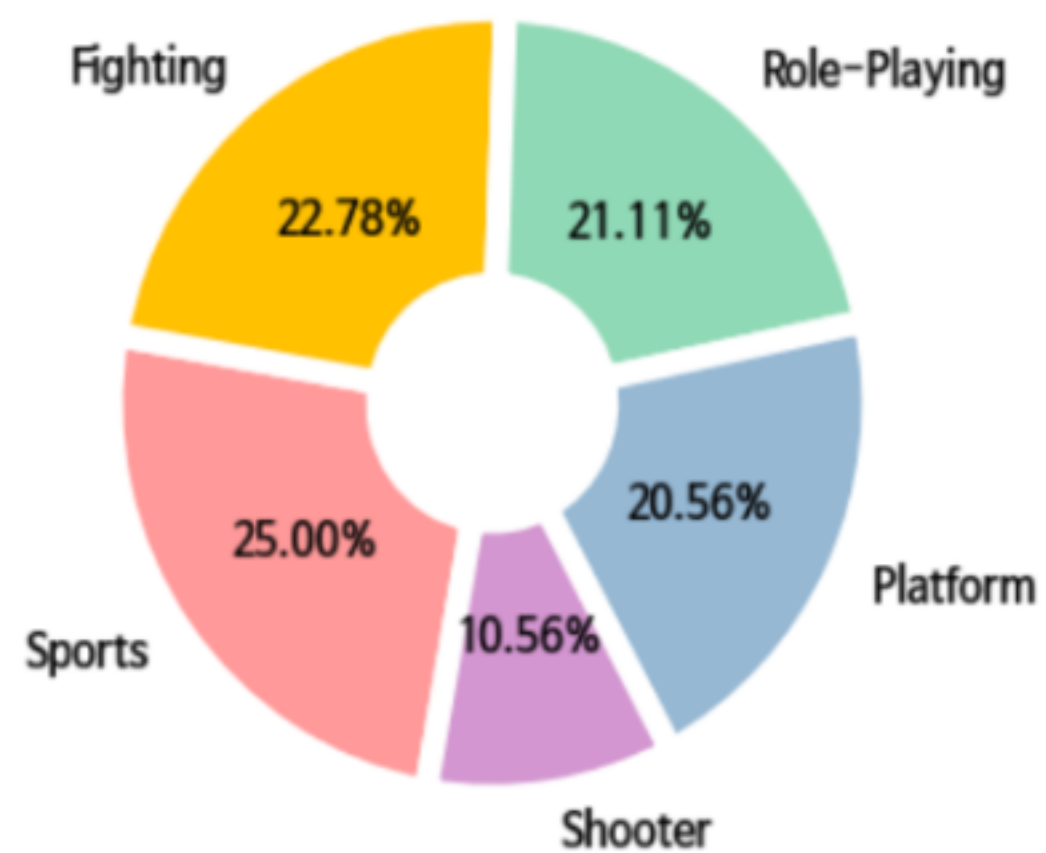
1980-1984년 판매된 장르 분포



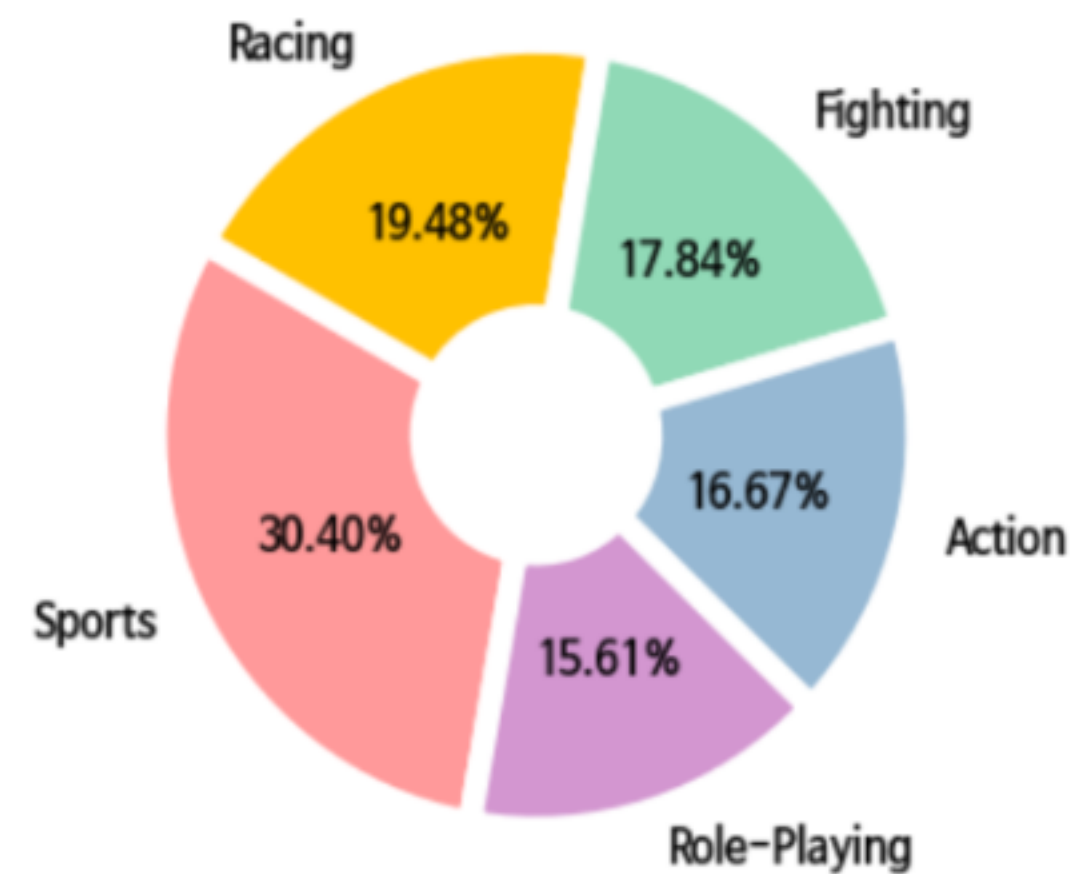
1985-1989년 판매된 장르 분포



1990-1994년 판매된 장르 분포



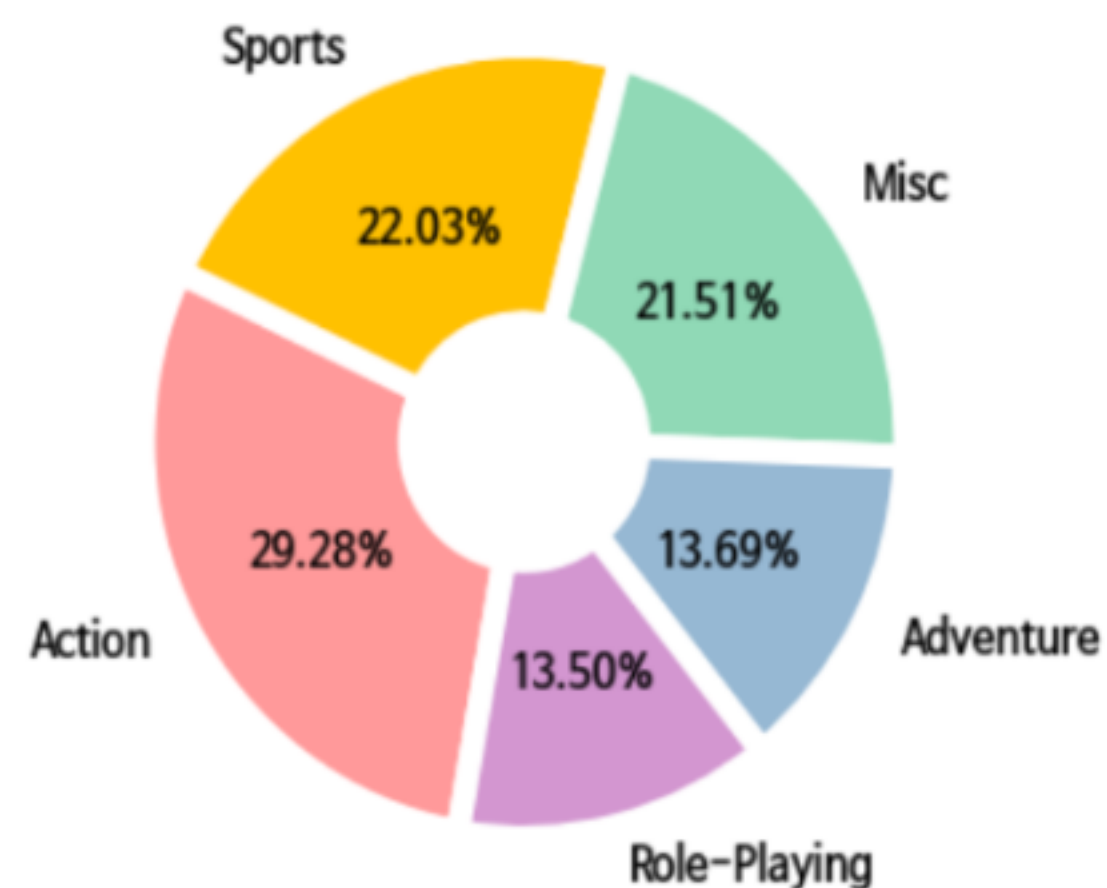
1995-1999년 판매된 장르 분포



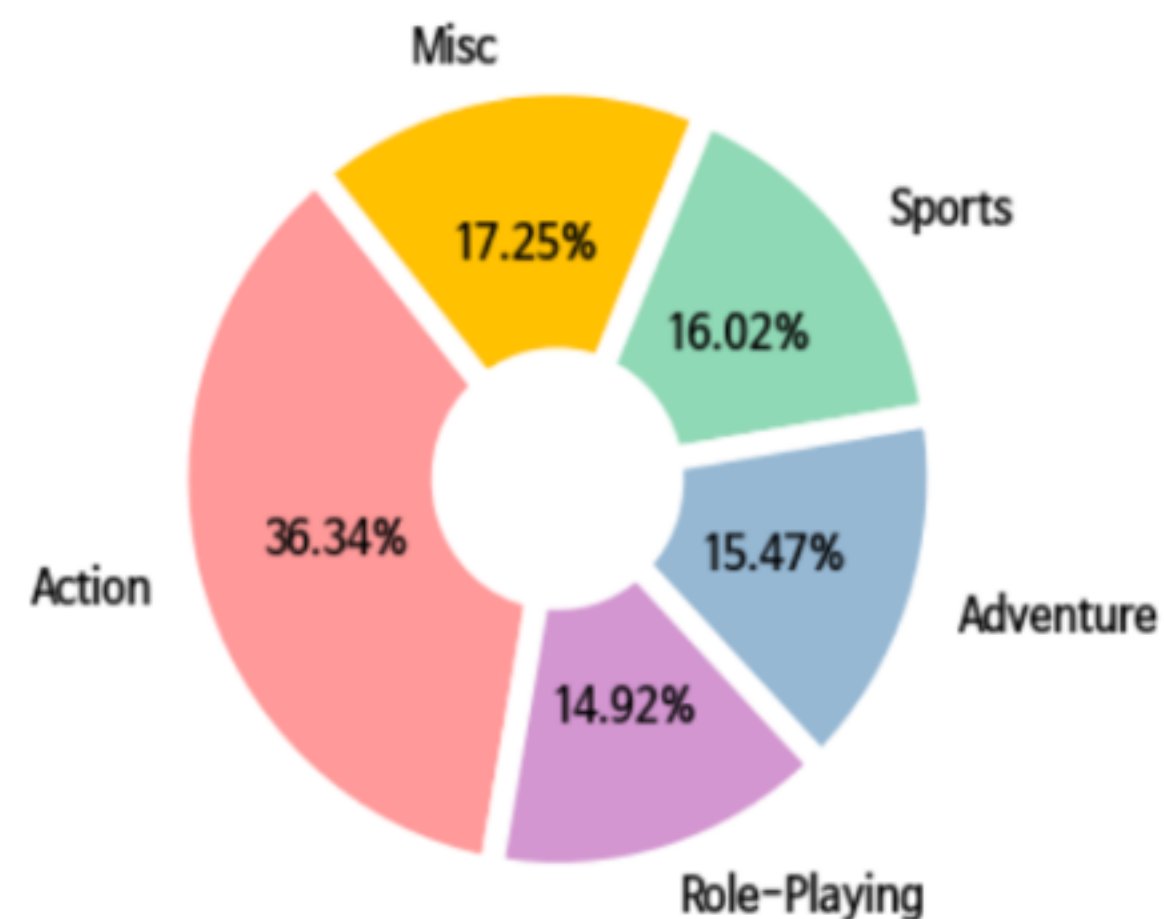
2000-2004년 판매된 장르 분포



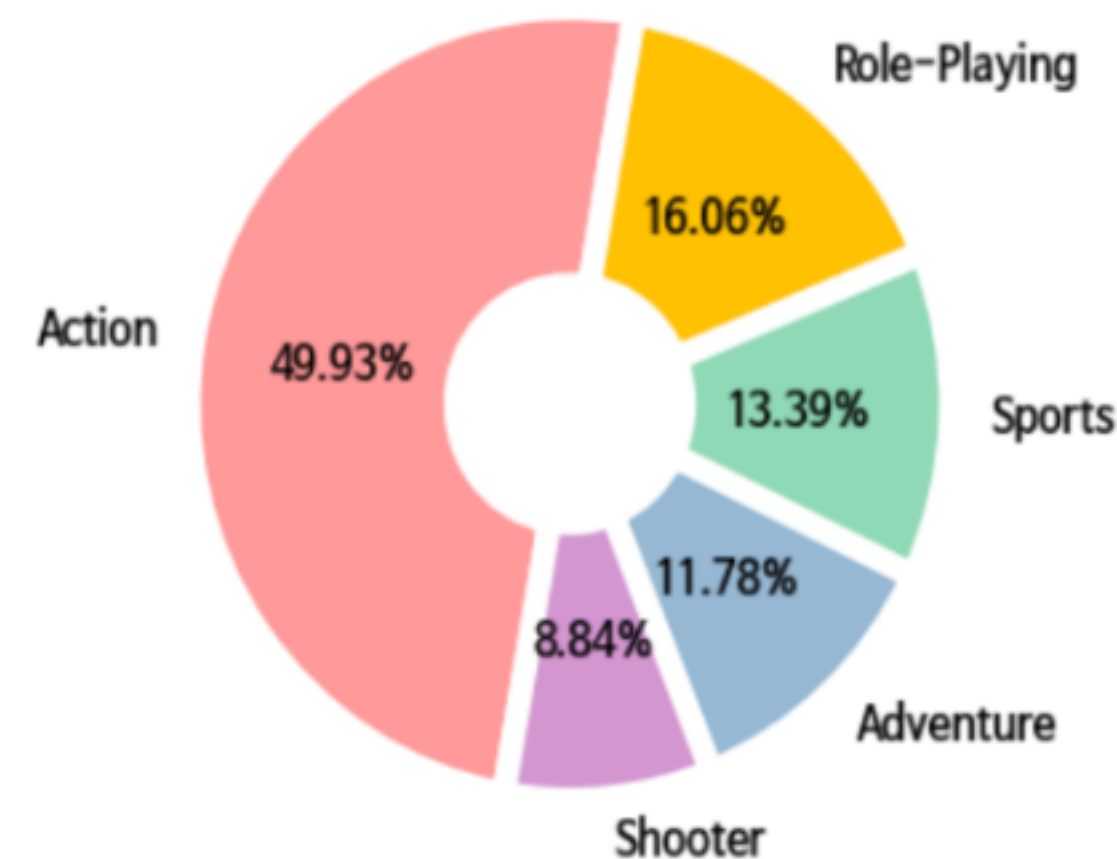
2005-2009년 판매된 장르 분포



2010-2014년 판매된 장르 분포



2015-2020년 판매된 장르 분포

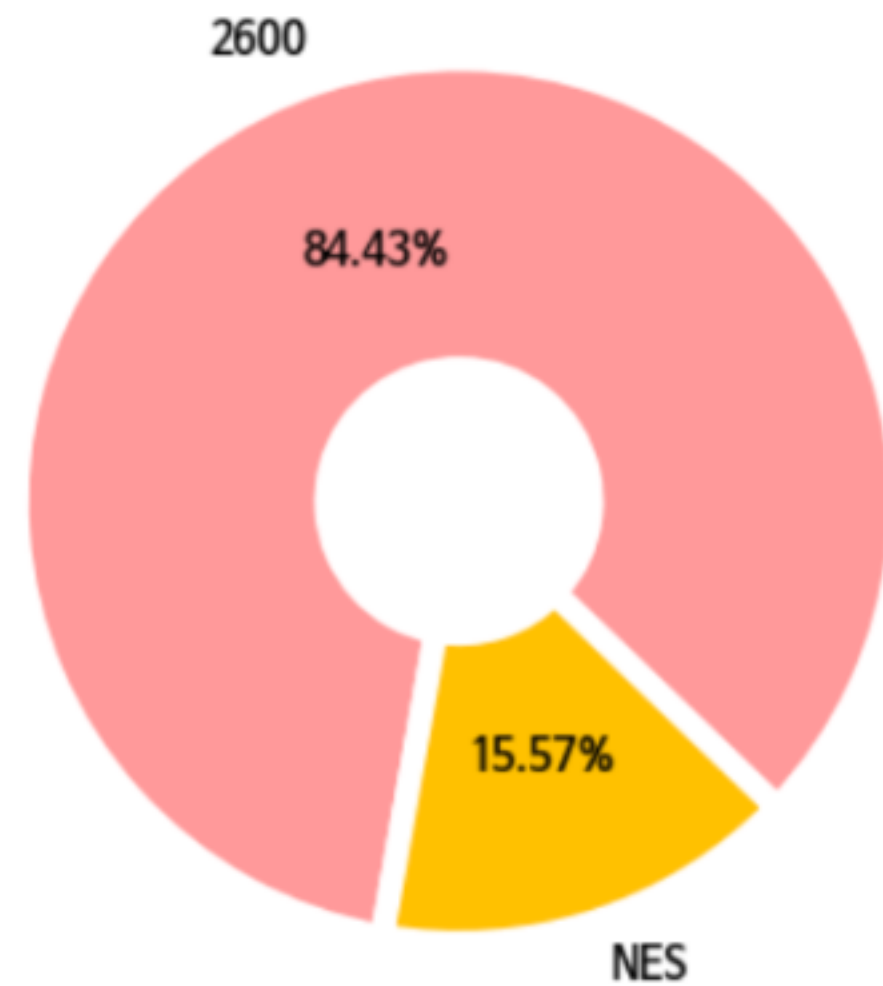




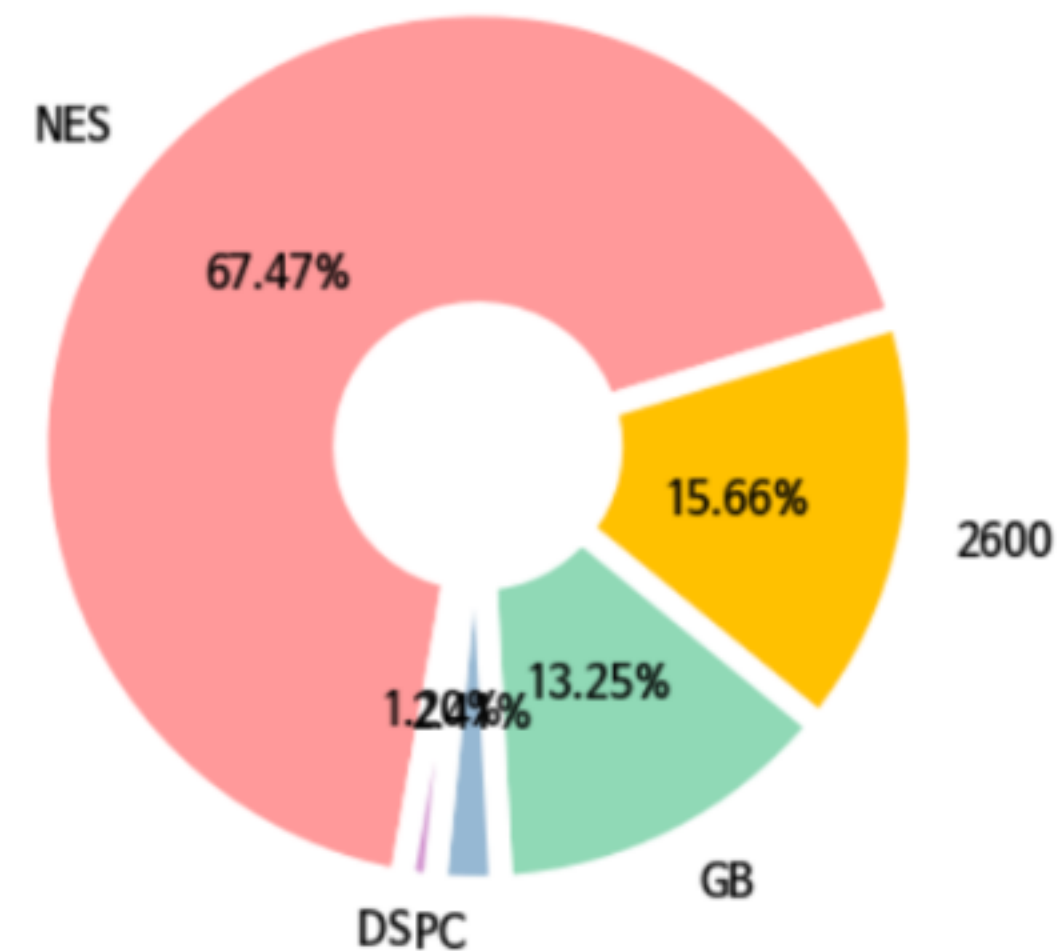
# 연도별 게임의 트렌드

## 연도별 배출된 플랫폼

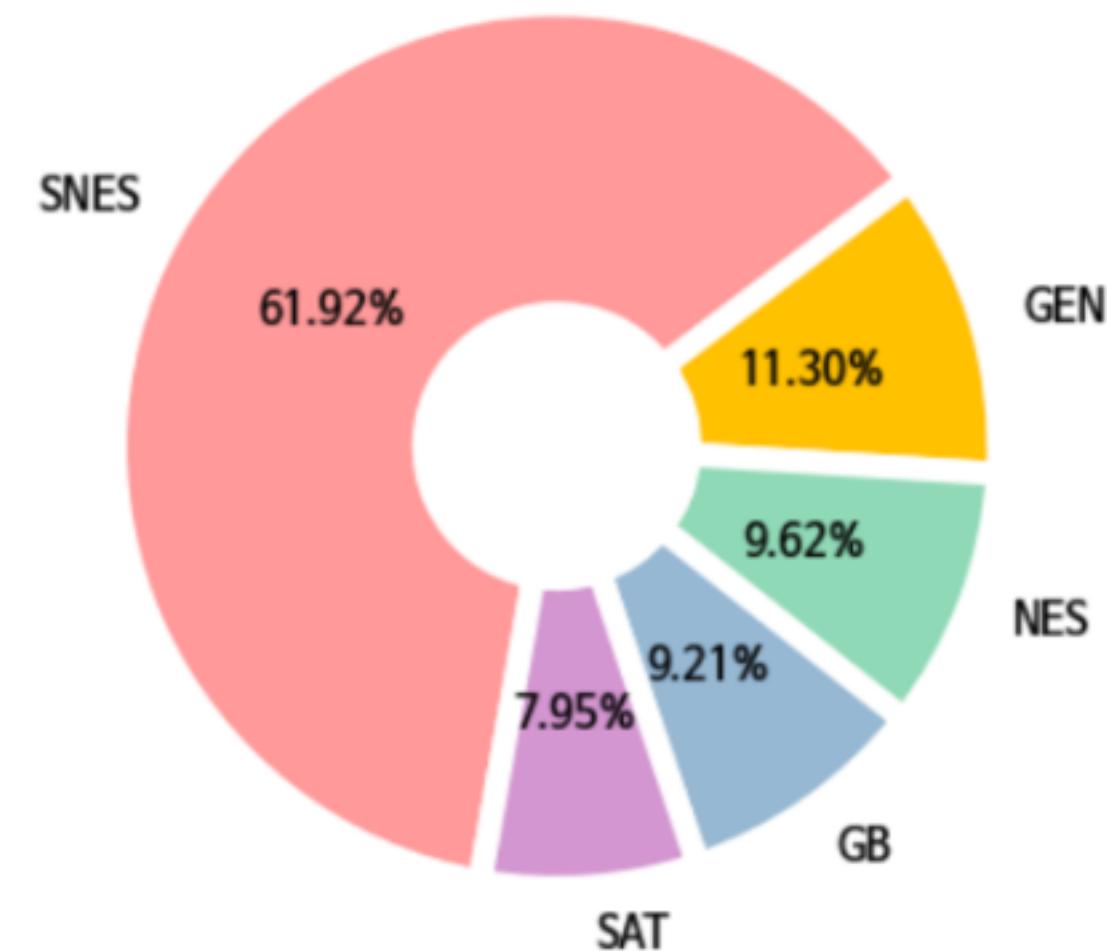
1980-1984년 판매된 플랫폼 분포



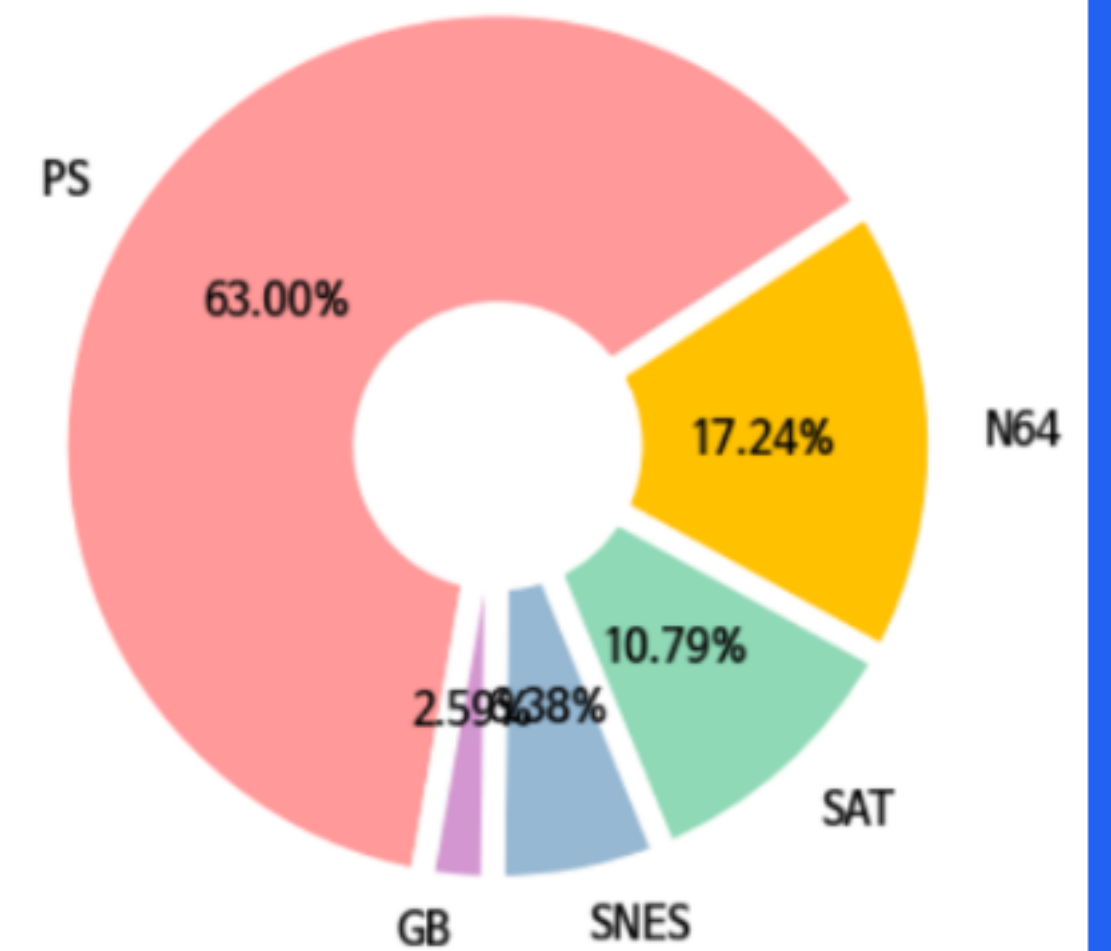
1985-1989년 판매된 플랫폼 분포



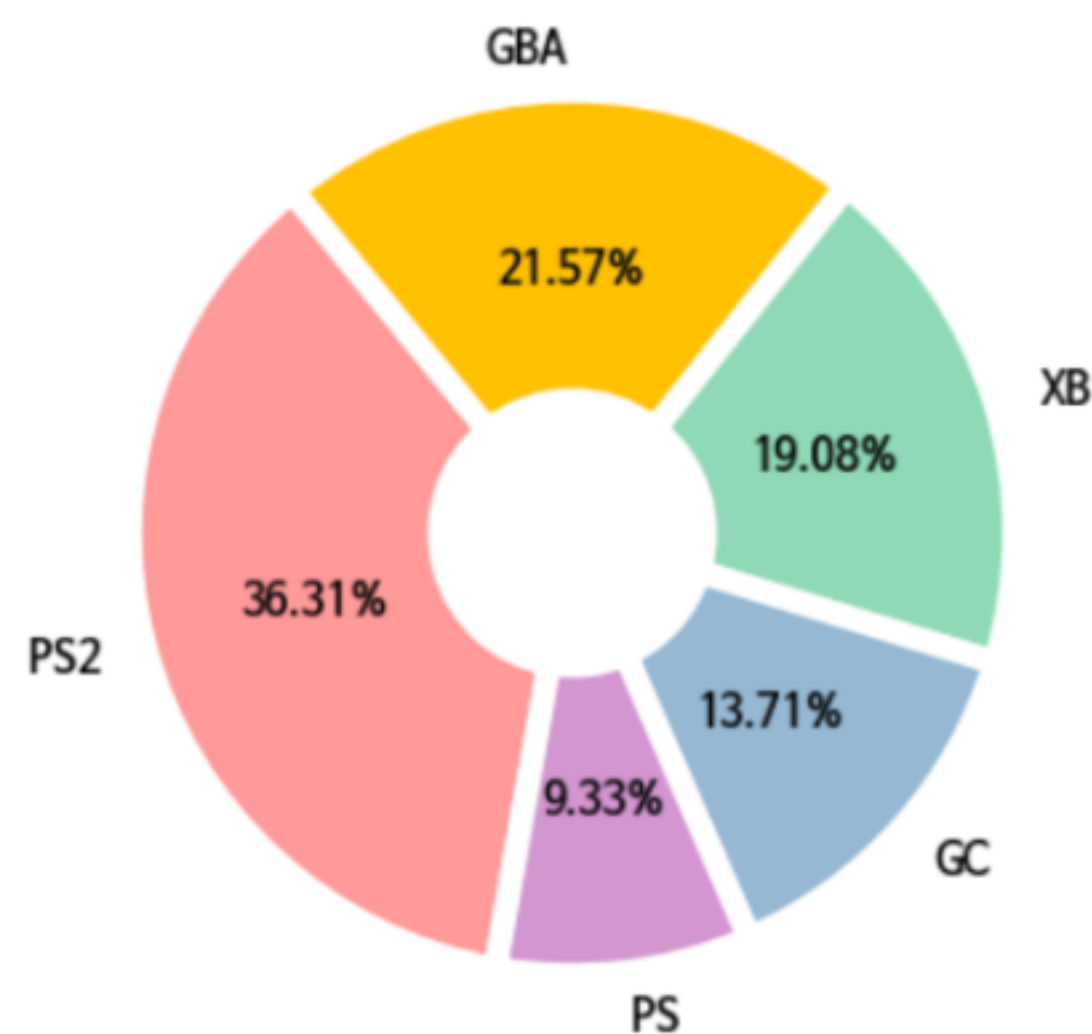
1990-1994년 판매된 플랫폼 분포



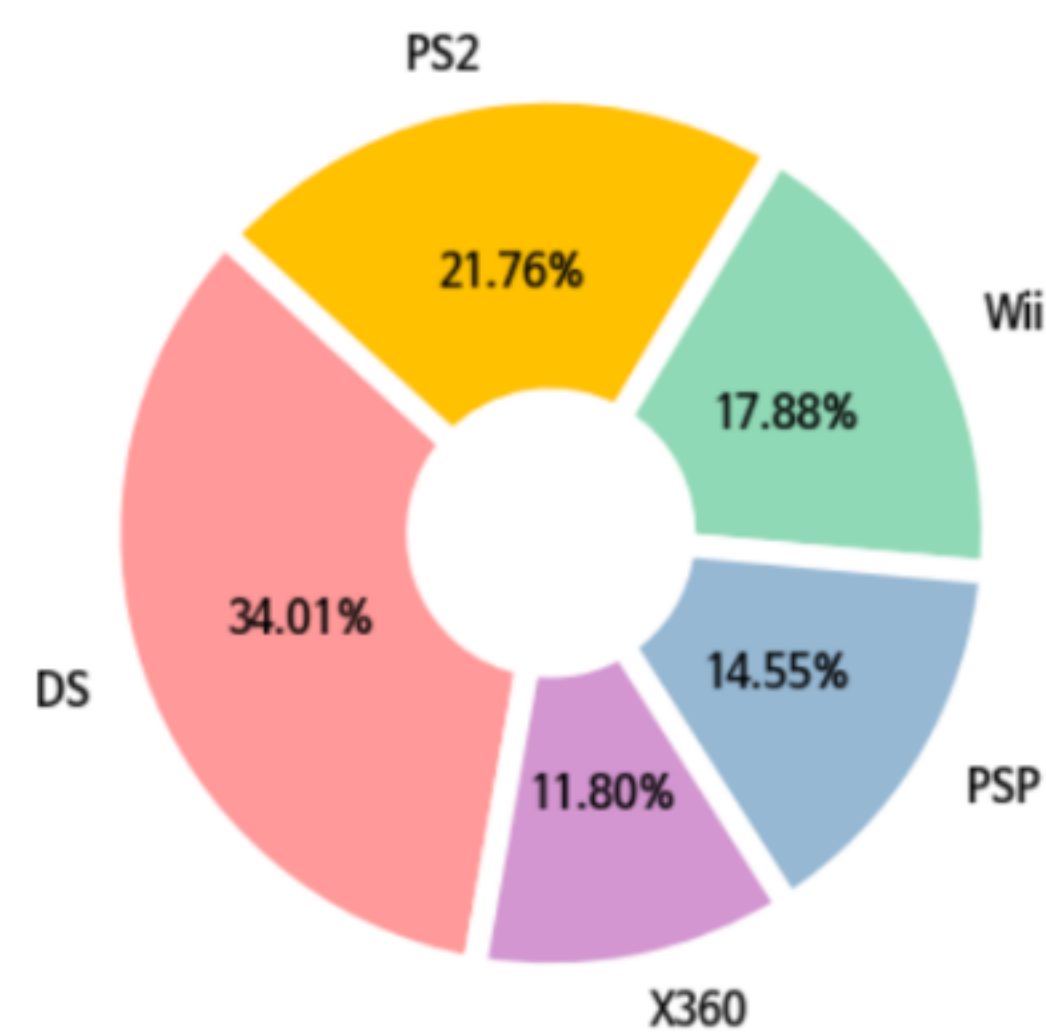
1995-1999년 판매된 플랫폼 분포



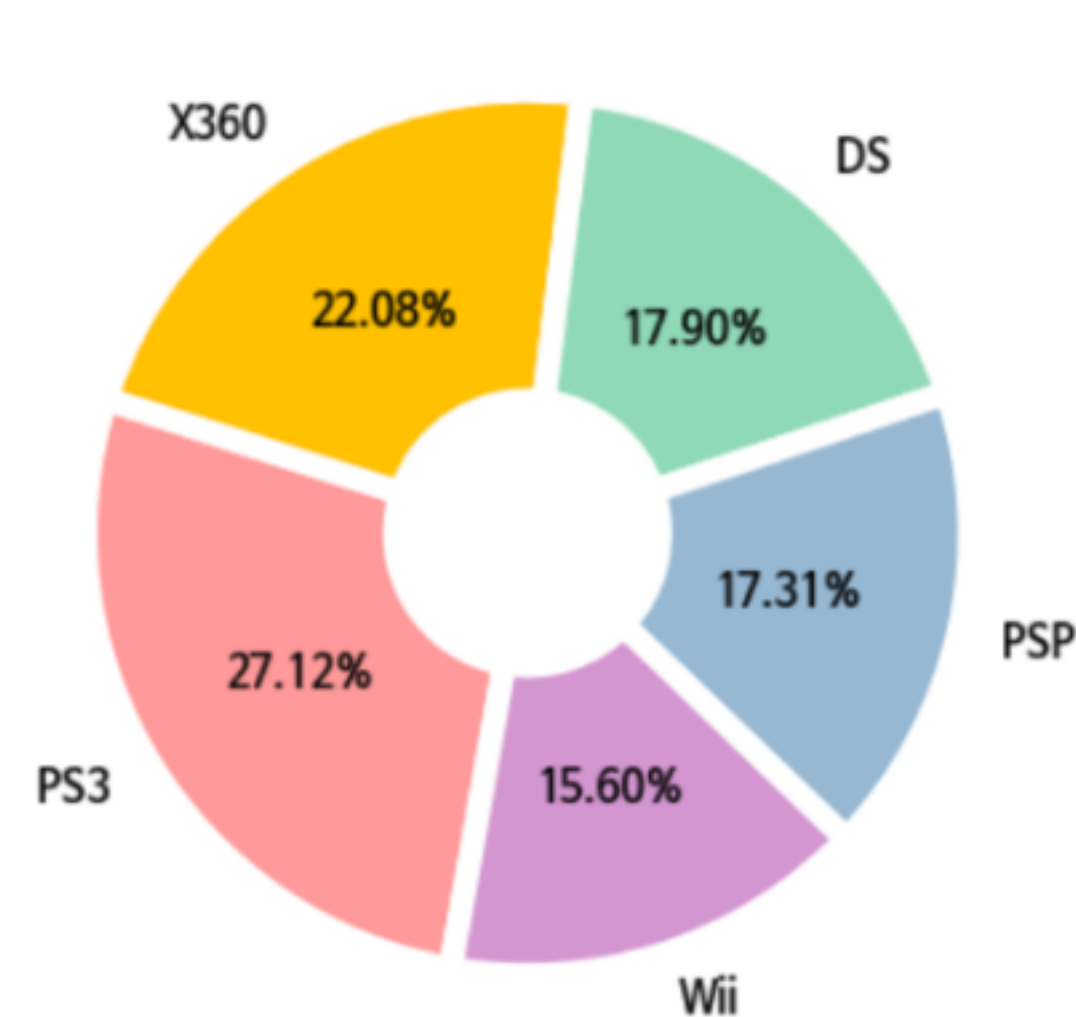
2000-2004년 판매된 플랫폼 분포



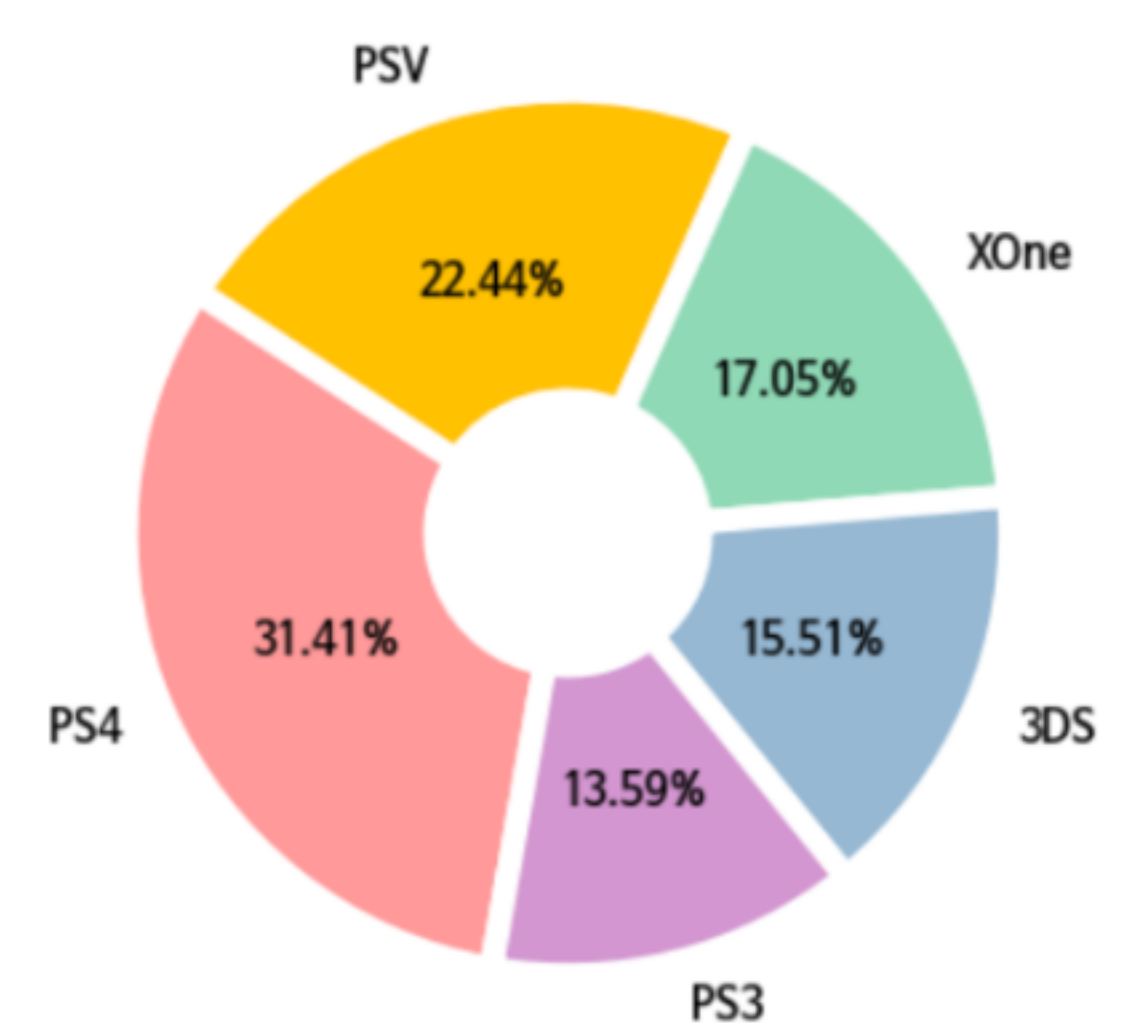
2005-2009년 판매된 플랫폼 분포



2010-2014년 판매된 플랫폼 분포



2015-2020년 판매된 플랫폼 분포

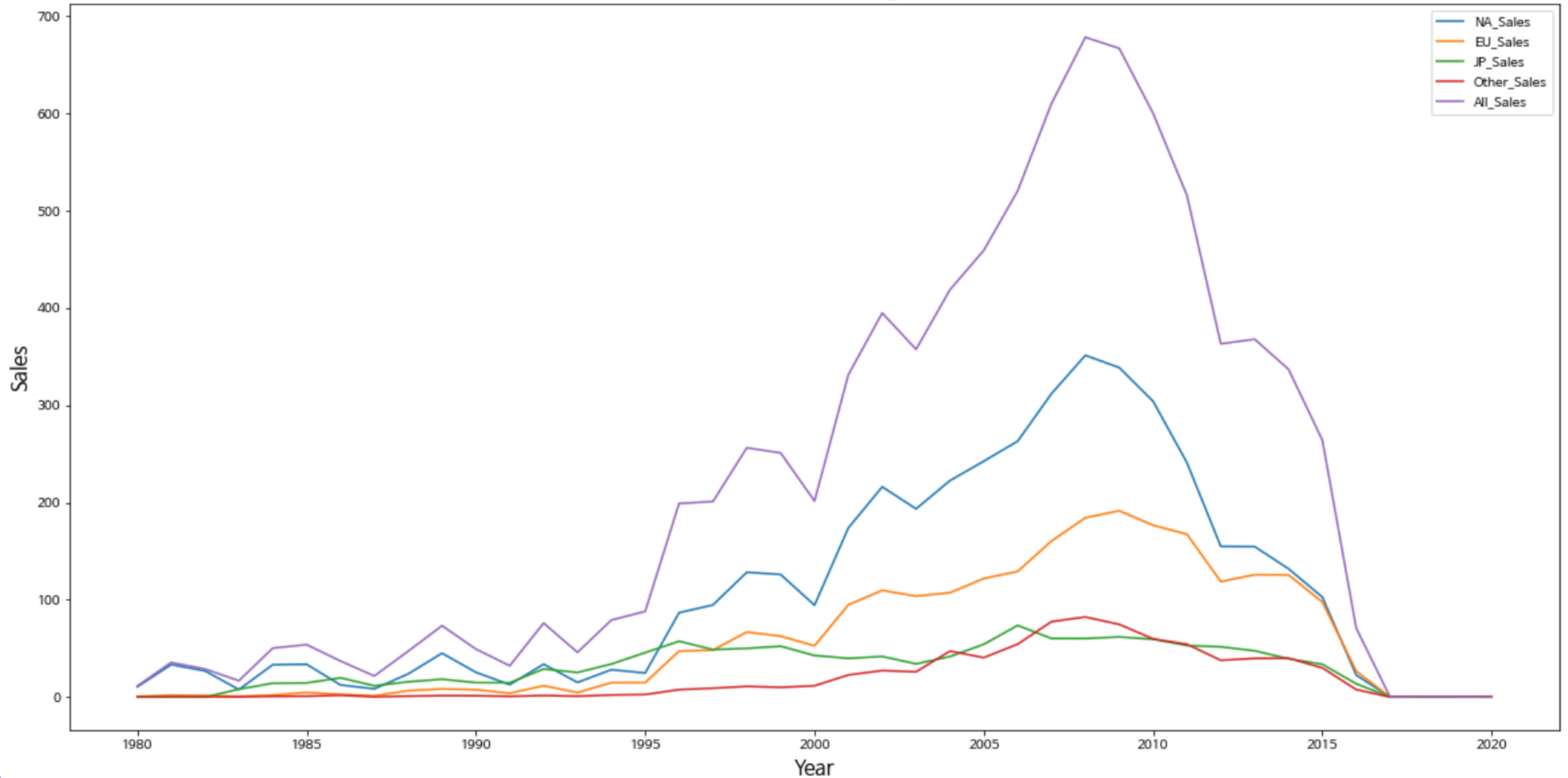




# 연도별 게임의 트렌드

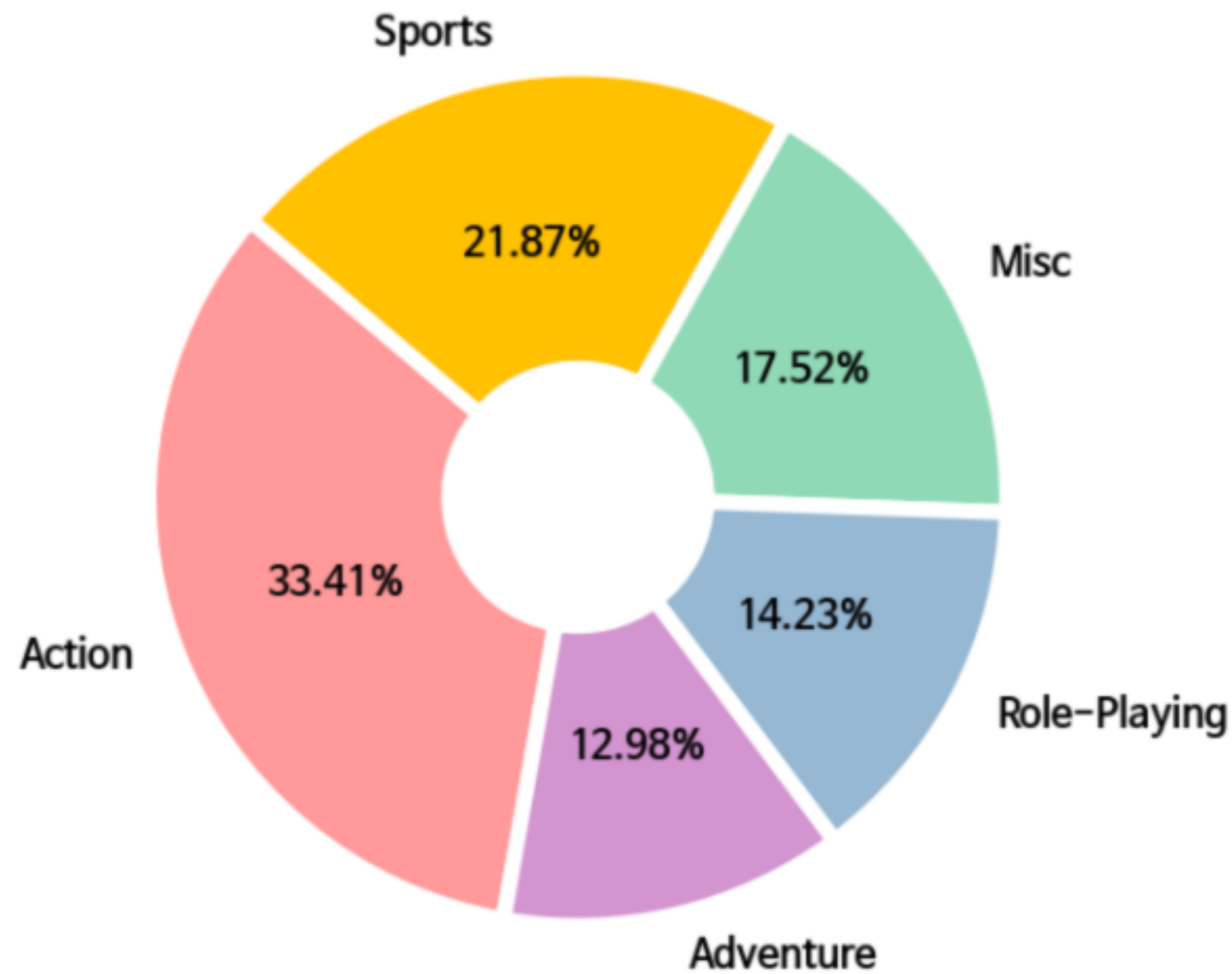
연도별 판매량 변화

연도별 판매량

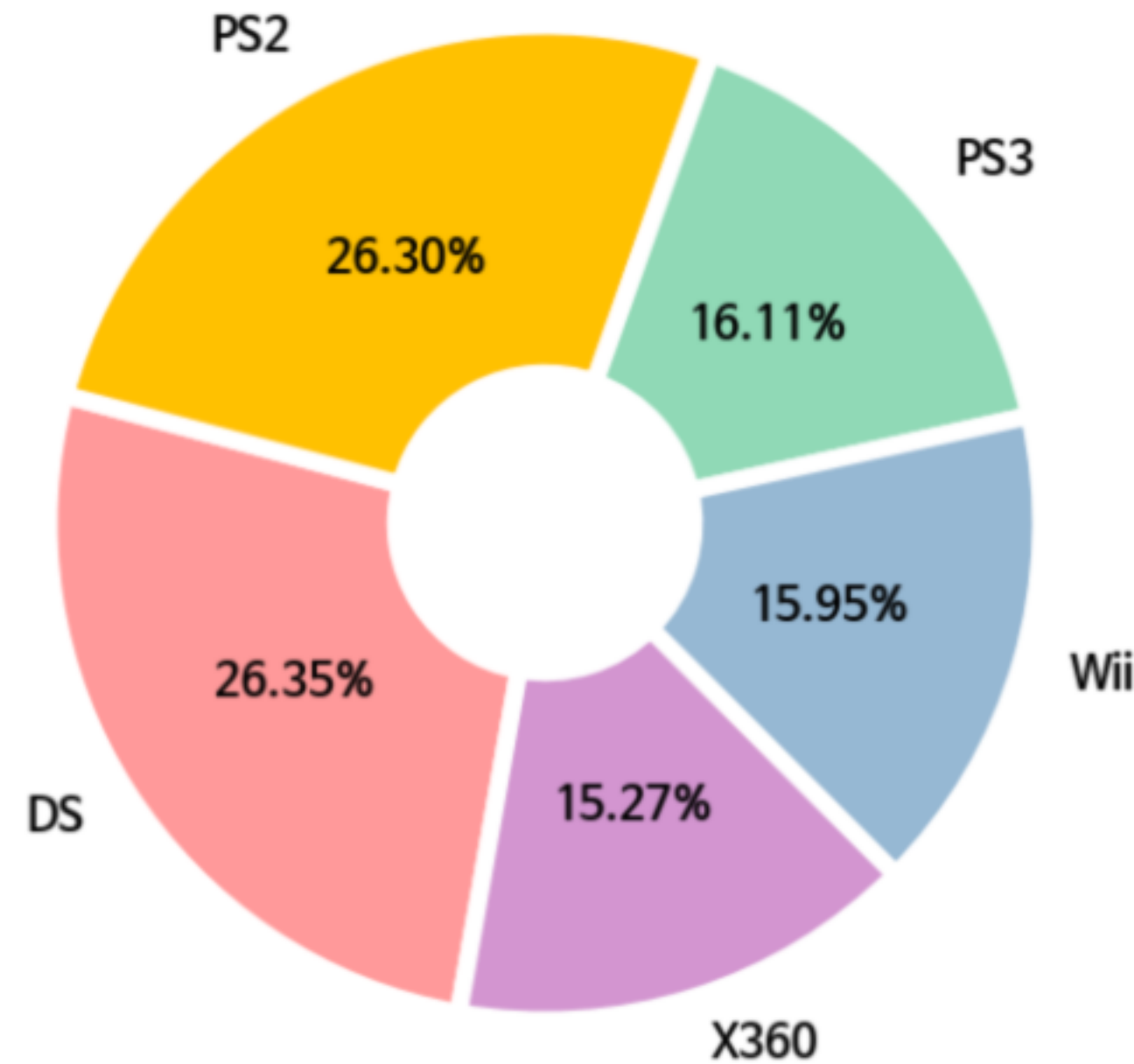


# 인기가 많은 게임 분석

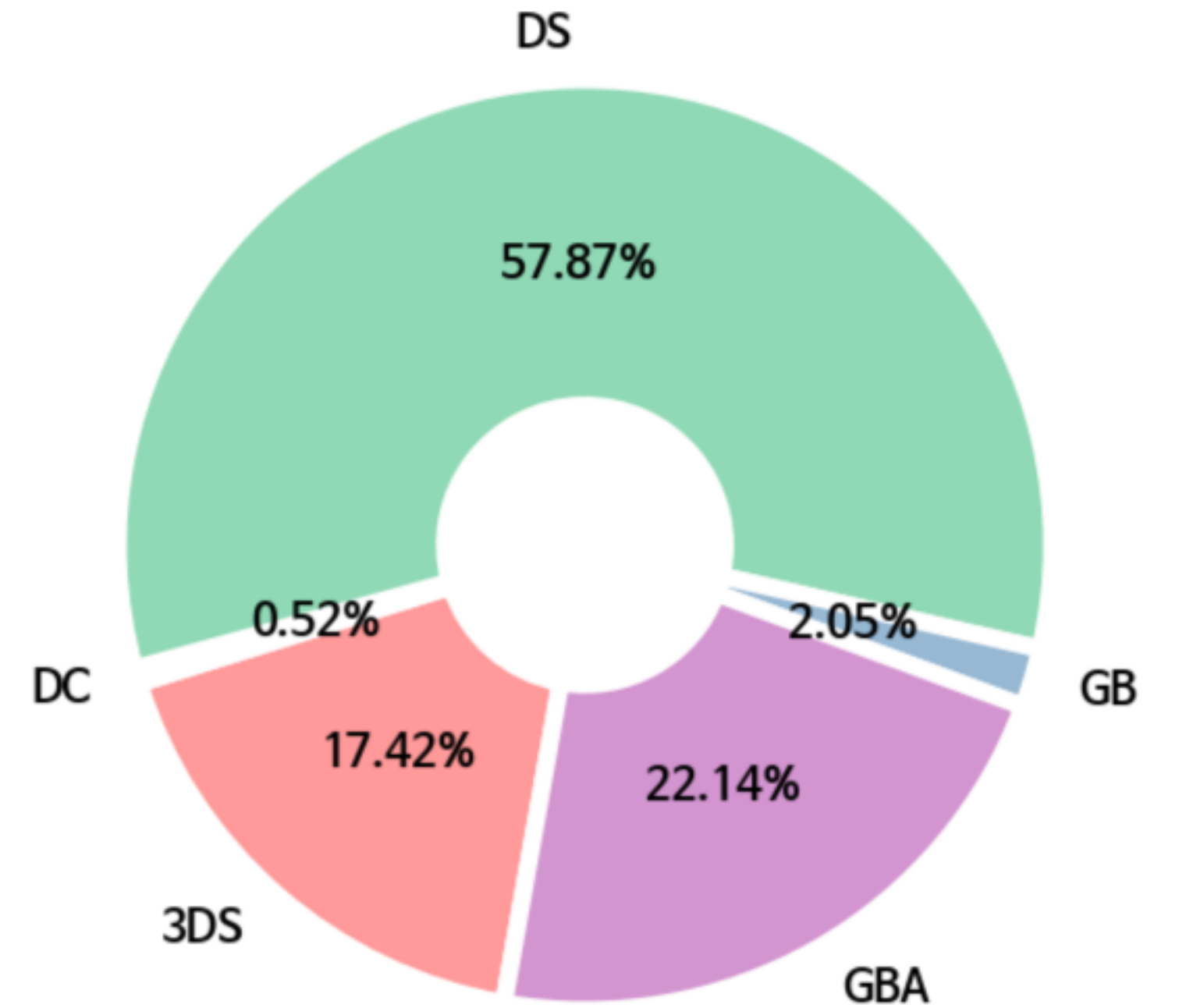
20년대 판매된 장르 분포



20년대 판매된 플랫폼 분포



20년대 판매된 플랫폼 분포 (판매량)



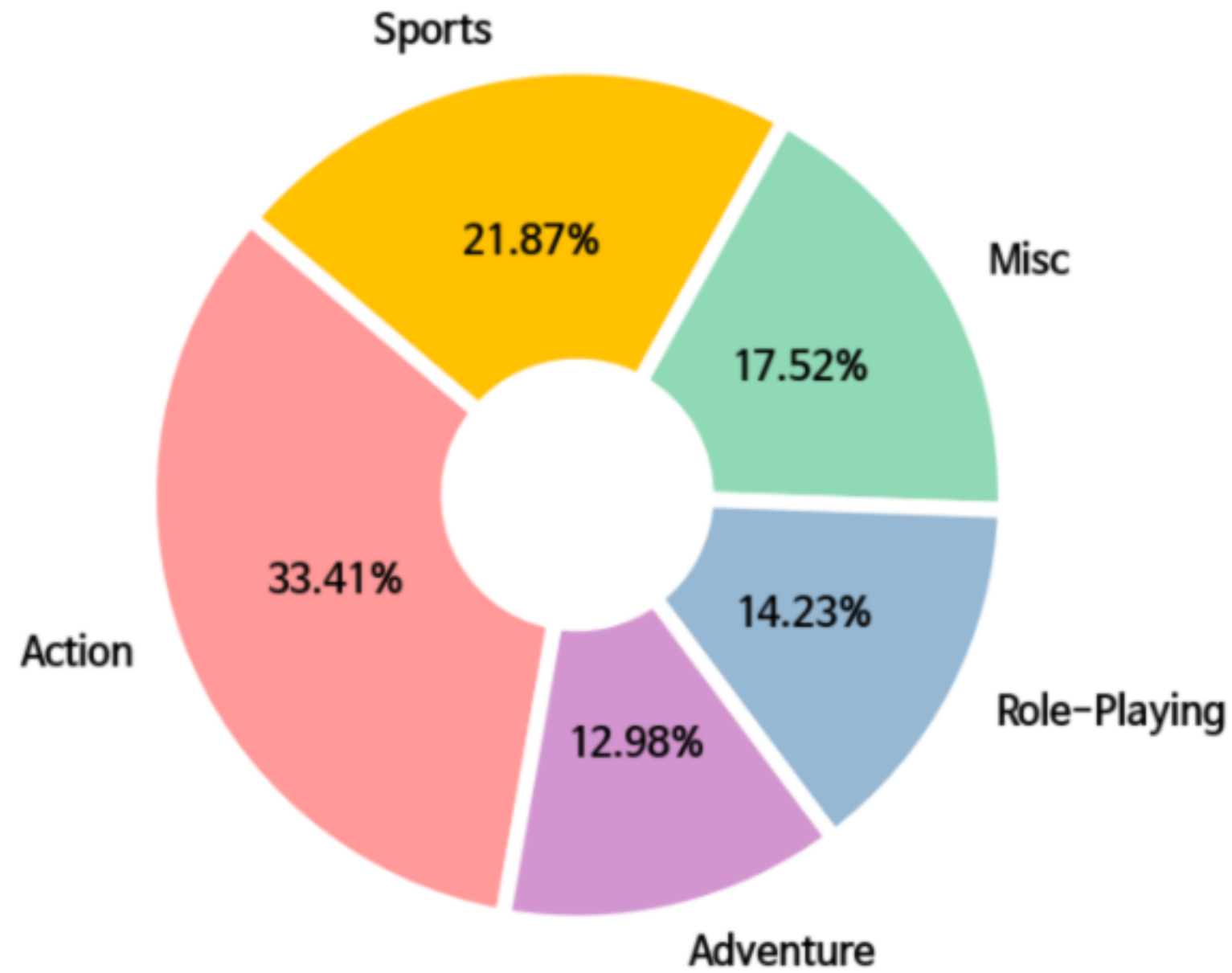
회사에서 배출한 게임의 장르와 플랫폼의 **개수** 비율

판매된 플랫폼의 **실제 판매량**의 비율

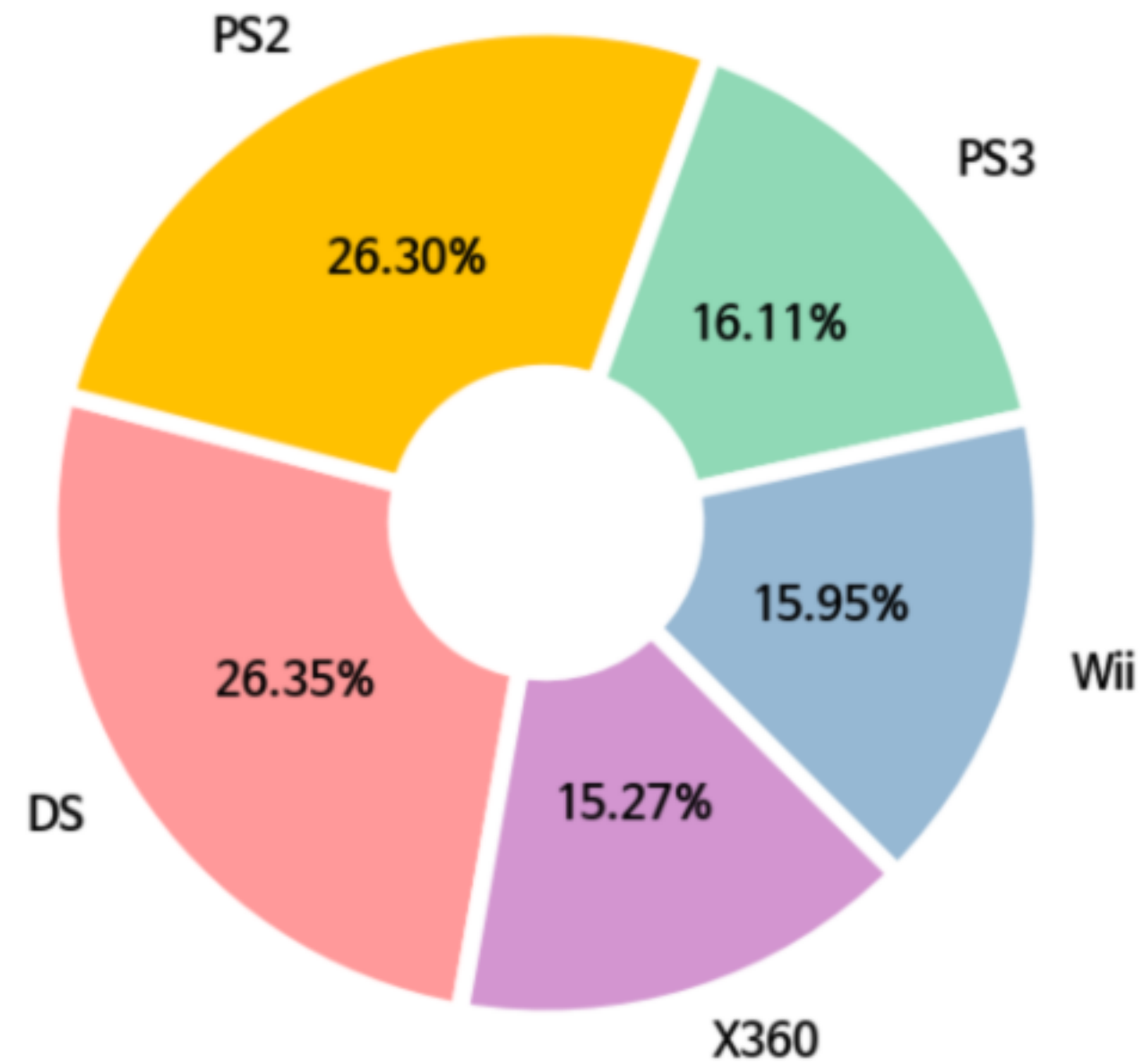
✓ 20년대에 배출된 게임의 플랫폼의 비율은 비슷함에도 불구하고  
실제 판매된 플랫폼의 판매량은 DS, GBA, 3DS가 상위권에 포함됨을 알 수 있다.

# 인기가 많은 게임 분석

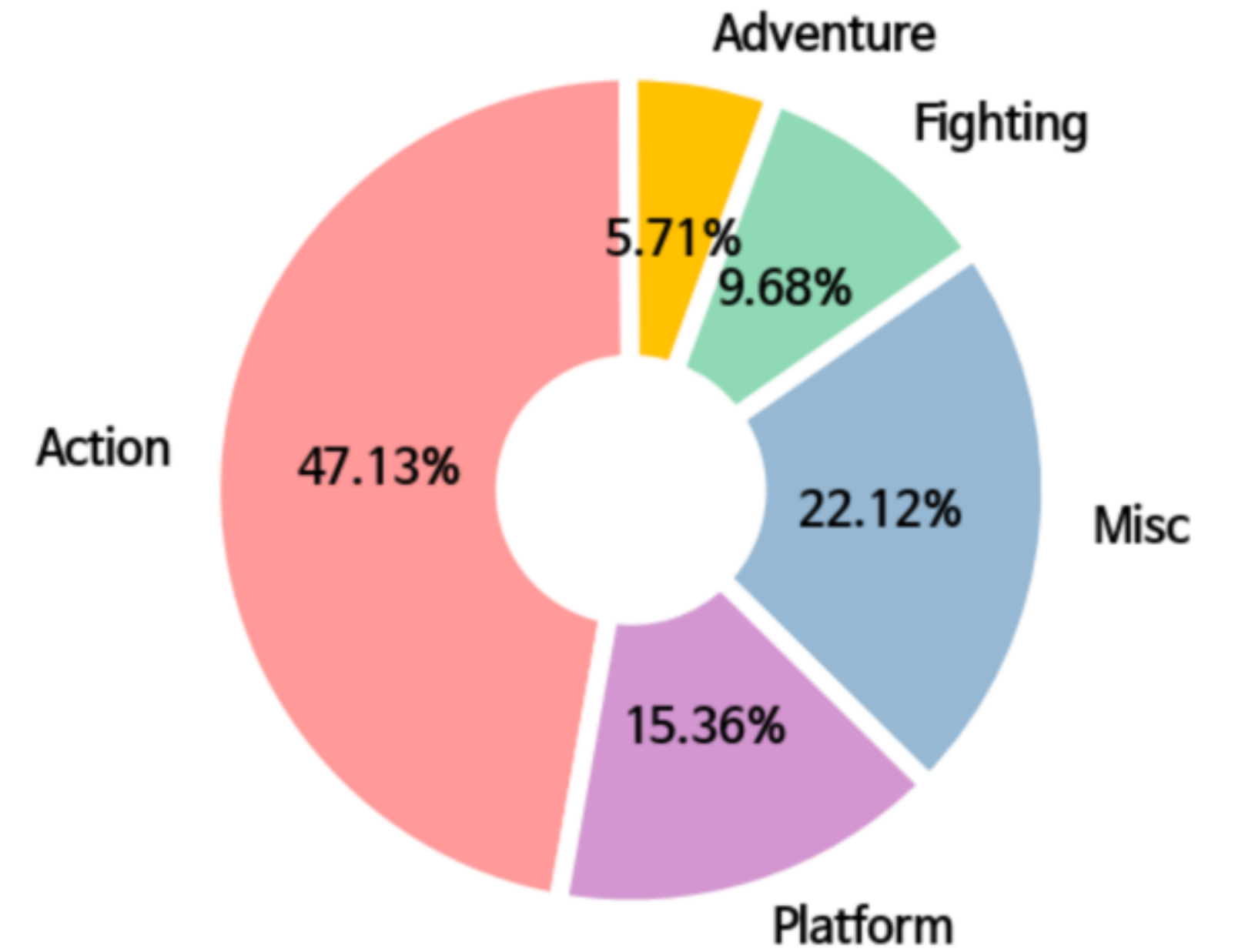
20년대 판매된 장르 분포



20년대 판매된 플랫폼 분포



20년대 판매된 장르 분포 (판매량)



회사에서 배출한 게임의 장르와 플랫폼의 **개수** 비율

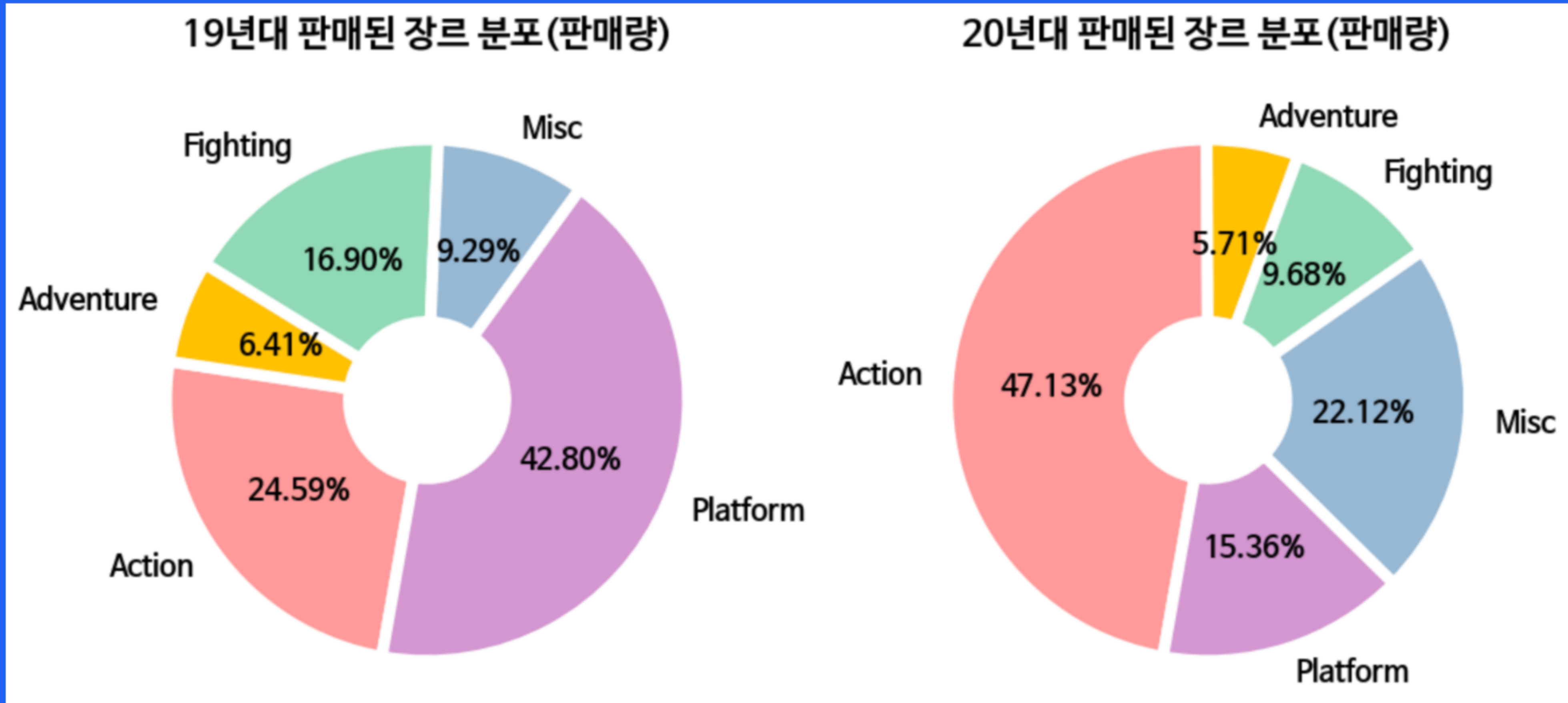
판매된 게임 장르의 **실제 판매량**의 비율



20년대에 배출된 게임 장르의 비율은 Action 장르가 살짝 우위에 있고  
다른 장르의 비율은 다소 비슷함에도 불구하고  
Action 이외에 실제 판매된 장르의 판매량은 장르마다 조금씩 다른 것을 알 수 있다.



# 인기가 많은 게임 분석



20년대로 넘어오면서 Action 장르의 판매량이 매우 높아지고 Platform 장르의 판매량은 낮춰진 것을 알 수 있다.



# 결론

데이터 시각화를 통해 얻은 인사이트 기반으로 결론 도출

장르는

Action

중심으로

게임을

기획하고

플랫폼은

가장

많았던

DS 플랫폼

이용하여

지역맞춤형

게임을

판매하는

방안이

좋다.

Action 게임을 DS 플랫폼을 바탕으로 하여 Action, Shooter, Sports 장르 위주로 기획하고 순차적으로 지역별로 인기가 많았던 장르를 순서로 기획하면 좋을 것 같다.

· SIMPLE PRESENTATION ·

# THANK YOU

AI\_18\_박찬선