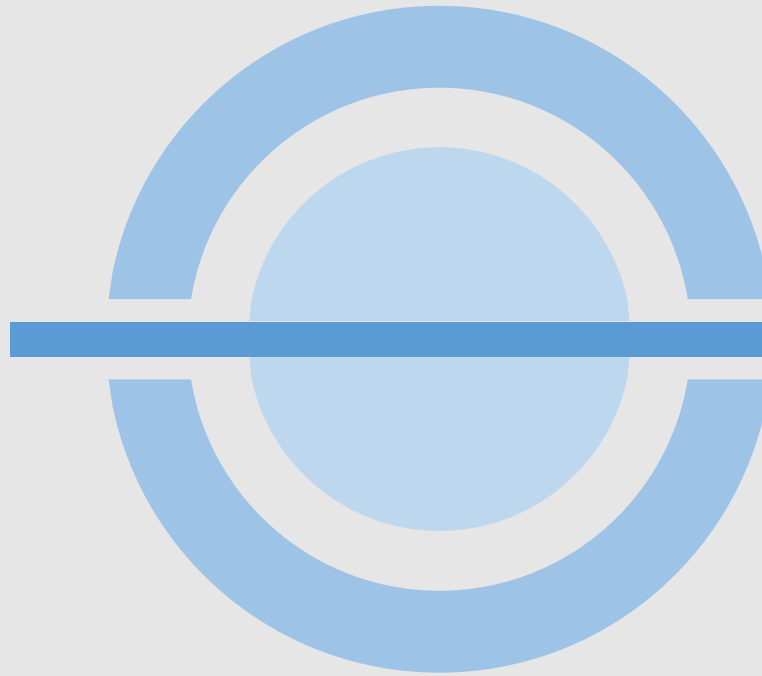


# Section1 Project

AIB\_12th\_이우림



데이터 소개

데이터 전처리

데이터 분석

결과

## 데이터 소개

01 각 게임 별 Platform, 출시한 연도, 장르, Publisher, 지역별 판매량이 나타나 있는 데이터

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
1	Candace K	DS	2008	Action	Destineer	0.04	0	0	0
2	The Munc	Wii	2009	Action	Namco Ba	0.17	0	0	0.01
3	Otome wa	PSP	2010	Adventure	Alchemist	0	0	0.02	0
4	Deal or No	DS	2010	Misc	Zoo Game	0.04	0	0	0
5	Ben 10 Ult	PS3	2010	Platform	D3Publishe	0.12	0.09	0	0.04
6	Power Pro	DS	2005	Sports	Konami Di	0	0	0.14	0
7	JASF: Jane	PC	2011	Simulation	Deep Silve	0.02	0.01	0	0
8	Samurai W	PSP	2005	Action	Tecmo Ko	0.07	0	0.11	0.01
9	Imabikisou	PS3	2007	Adventure	Sega	0	0	0.04	0
10	The Pengu	PS3	2011	Action	THQ	0.04	0.03	0	0.01
11	Ford Racin	PS	2001	Racing	Empire Int	480K	0.33M	0K	0.06
12	Harvest M	GBA	2003	Simulation	Ubisoft	0.14	0.05	0	0
13	Mount & I	PC	2008	Role-Playir	Paradox In	0	0.02	0	0

01 정확한 데이터 분석을 위해서는 우선 결측치를 제거해 주어야 한다.

```
df.isnull().sum()
```

```
Unnamed: 0      0
Name            0
Platform        0
Year           271
Genre           50
Publisher       58
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
dtype: int64
```

전처리 전

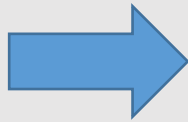
```
df = df.dropna(axis=0)
```

```
Unnamed: 0      0
Name            0
Platform        0
Year            0
Genre           0
Publisher       0
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
dtype: int64
```

전처리 후

02 정확한 데이터 분석을 위해서는 같은 column의 데이터의 형태가 같아야 한다.

2005	Ac
9	St
2010	Sp
1985	Pl
2000	Pl
2007	Pu
2008	Ra
2005	Ac
97	Ac
1997	St



```
def modify_year(year):  
    year = int(year)  
    if year <= 22:  
        return year + 2000  
    elif 22 < year < 1000:  
        return year + 1900  
    else:  
        return year  
  
df['Year'] = df['Year'].apply(modify_year)
```

03 정확한 데이터 분석을 위해서는 같은 column의 데이터의 형태가 같아야 한다.

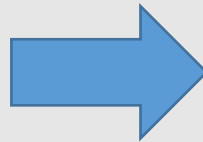
	0	0	0.04	0
	0.04	0.03	0	0.01
e Int	480K	0.33M	0K	0.06
ft	0.14	0.05	0	0
ov ln	0	0.02	0	0



```
def modify_sales(sales):  
    if 'M' in sales:  
        return float(sales.replace('M', ''))  
    elif 'K' in sales:  
        return float(sales.replace('K', '')) * 0.001  
    else:  
        return float(sales)  
  
df['NA_Sales'] = df['NA_Sales'].apply(modify_sales)  
df['EU_Sales'] = df['EU_Sales'].apply(modify_sales)  
df['JP_Sales'] = df['JP_Sales'].apply(modify_sales)  
df['Other_Sales'] = df['Other_Sales'].apply(modify_sales)
```

### 03 앞의 과정이 잘 진행되었는지 확인한다.

```
Unnamed: 0    int64
Name          object
Platform      object
Year          float64
Genre         object
Publisher     object
NA_Sales      object
EU_Sales      object
JP_Sales      object
Other_Sales   object
dtype: object
```



```
Unnamed: 0    int64
Name          object
Platform      object
Year          int64
Genre         object
Publisher     object
NA_Sales      float64
EU_Sales      float64
JP_Sales      float64
Other_Sales   float64
dtype: object
```

## 데이터 분석 – 지역에 따라서 선호하는 게임 장르가 다를까?

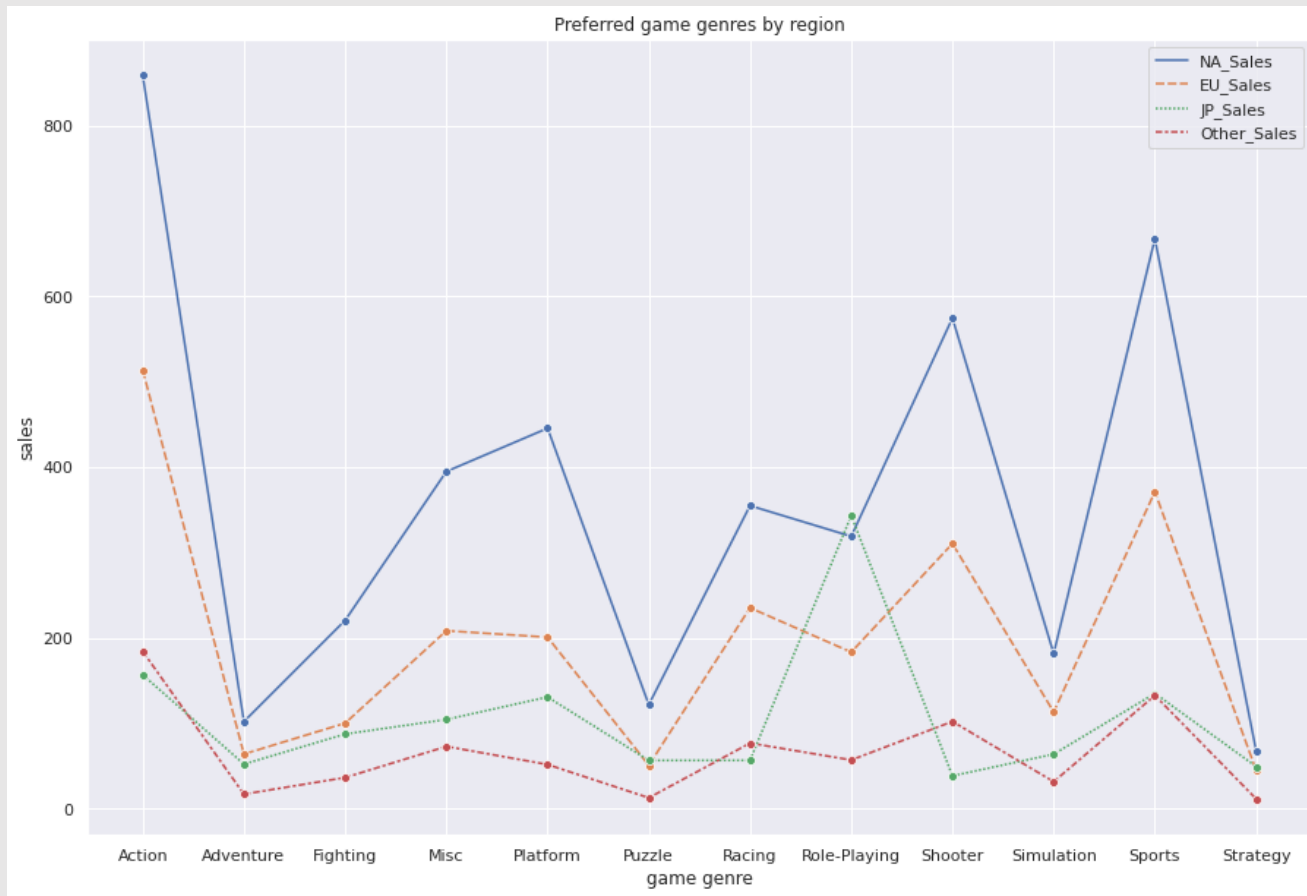
- 01 지역에 따라 선호하는 게임 장르가 다른지 알아보기 위해서 장르에 따른 지역 판매량을 나타내는 dataframe을 생성했다.

	NA_Sales	EU_Sales	JP_Sales	Other_Sales
Genre				
Action	858.75	513.64	156.60	183.99
Adventure	101.79	63.66	51.87	16.70
Fighting	220.35	99.90	87.14	36.15
Misc	394.73	208.15	104.18	72.64
Platform	445.38	200.63	130.53	51.45
Puzzle	121.58	50.47	56.24	12.44
Racing	354.70	235.02	56.32	76.51
Role-Playing	318.76	183.09	342.93	56.73
Shooter	574.48	310.23	38.18	101.86
Simulation	181.43	112.99	63.54	31.32
Sports	667.26	370.97	134.76	132.35
Strategy	67.83	44.82	49.10	11.23



# 데이터 분석 – 지역에 따라서 선호하는 게임 장르가 다를까?

## 02 데이터 시각화를 진행하였다.



- 북미, 유럽, 그 외 지역은 1위: Action, 2위: Sports, 3위: shooter로 선호하는 게임 장르가 비슷하다.
- 일본만 1위 게임장르가 Role-Playing이다.

## 데이터 분석 - 연도별 게임의 트렌드가 있을까?

- 01 연도별 게임의 트렌드를 확인하기 위해서 해당 연도에 출시된 게임의 장르와 그 게임의 총 판매량을 나타내는 dataframe을 생성했다.
- 02 해당 연도의 최고 판매량을 가진 장르만을 남겼다.

	Year	Genre	Total_Sales
0	1980	Action	0.34
1	1980	Fighting	0.77
2	1980	Misc	2.70
3	1980	Shooter	7.07
4	1981	Action	14.79
...	...	...	...
383	2016	Sports	14.63
384	2016	Strategy	0.52
385	2017	Action	0.01
386	2017	Role-Playing	0.04
387	2020	Simulation	0.29

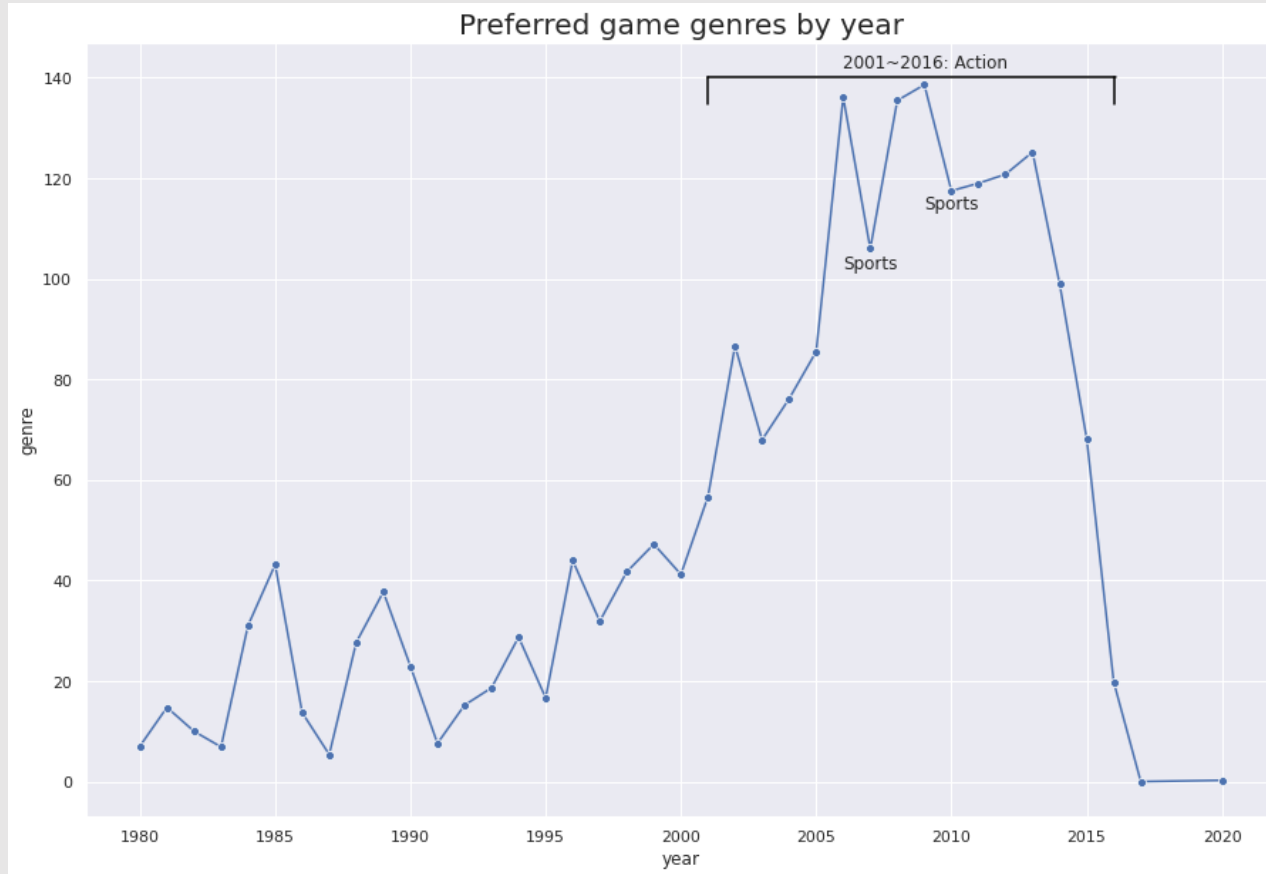


	Year	Genre	Total_Sales
0	1980	Shooter	7.07
1	1981	Action	14.79
2	1982	Puzzle	10.04
3	1983	Platform	6.93
4	1984	Shooter	31.10
5	1985	Platform	43.17
6	1986	Action	13.74
7	1987	Fighting	5.42
8	1988	Platform	27.73
9	1989	Puzzle	37.75
10	1990	Platform	22.98
11	1991	Platform	7.63
12	1992	Fighting	15.23
13	1993	Platform	18.68
14	1994	Platform	28.76
15	1995	Platform	16.69
16	1996	Role-Playing	43.99
17	1997	Racing	31.92

18	1998	Sports	41.75
19	1999	Role-Playing	47.24
20	2000	Sports	41.24
21	2001	Action	56.68
22	2002	Action	86.65
23	2003	Action	67.98
24	2004	Action	76.15
25	2005	Action	85.52
26	2006	Sports	136.12
27	2007	Action	105.99
28	2008	Action	135.46
29	2009	Sports	138.59
30	2010	Action	117.50
31	2011	Action	119.00
32	2012	Action	120.78
33	2013	Action	125.21
34	2014	Action	99.01
35	2015	Action	68.25
36	2016	Action	19.84
37	2017	Role-Playing	0.04
38	2020	Simulation	0.29

# 데이터 분석 – 지역에 따라서 선호하는 게임 장르가 다를까?

## 03 데이터 시각화를 진행하였다.



- 2001년~2016년에는 2006년과 2009년을 제외하고는 항상 Action 장르의 게임이 1위다.
- 그 이후에는 결측치로 인해 제거된 데이터가 많아 데이터가 정확하지 않다.

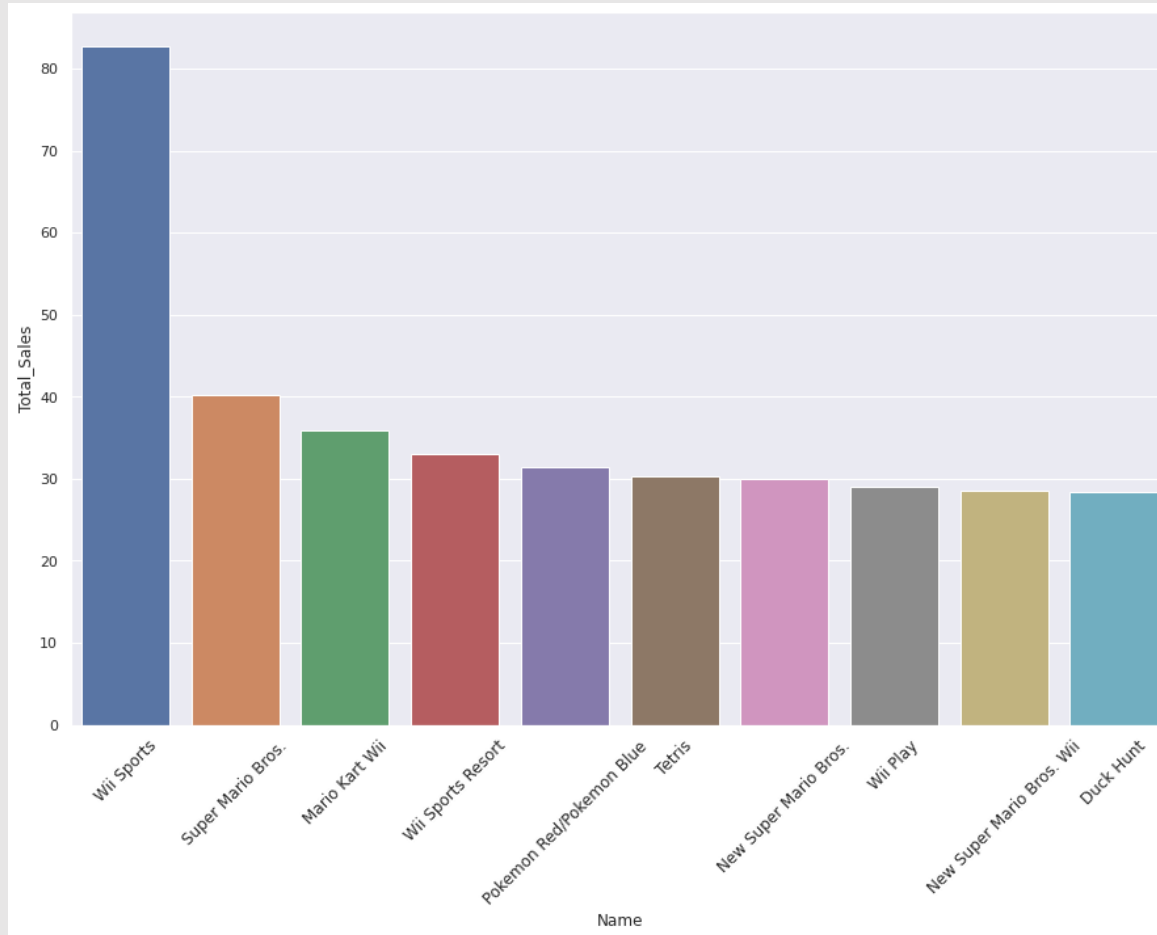
## 데이터 분석 – 출고량이 높은 게임에 대한 분석 및 시각화 프로세스

01 출고량이 높은 순서대로 10개의 게임을 나타내는 dataframe을 생성했다.

	Name	Platform	Publisher	Year	Genre	Total_Sales
0	Wii Sports	Wii	Nintendo	2006	Sports	82.74
1	Super Mario Bros.	NES	Nintendo	1985	Platform	40.24
2	Mario Kart Wii	Wii	Nintendo	2008	Racing	35.83
3	Wii Sports Resort	Wii	Nintendo	2009	Sports	33.00
4	Pokemon Red/Pokemon Blue	GB	Nintendo	1996	Role-Playing	31.38
5	Tetris	GB	Nintendo	1989	Puzzle	30.26
6	New Super Mario Bros.	DS	Nintendo	2006	Platform	30.01
7	Wii Play	Wii	Nintendo	2006	Misc	29.01
8	New Super Mario Bros. Wii	Wii	Nintendo	2009	Platform	28.61
9	Duck Hunt	NES	Nintendo	1984	Shooter	28.31

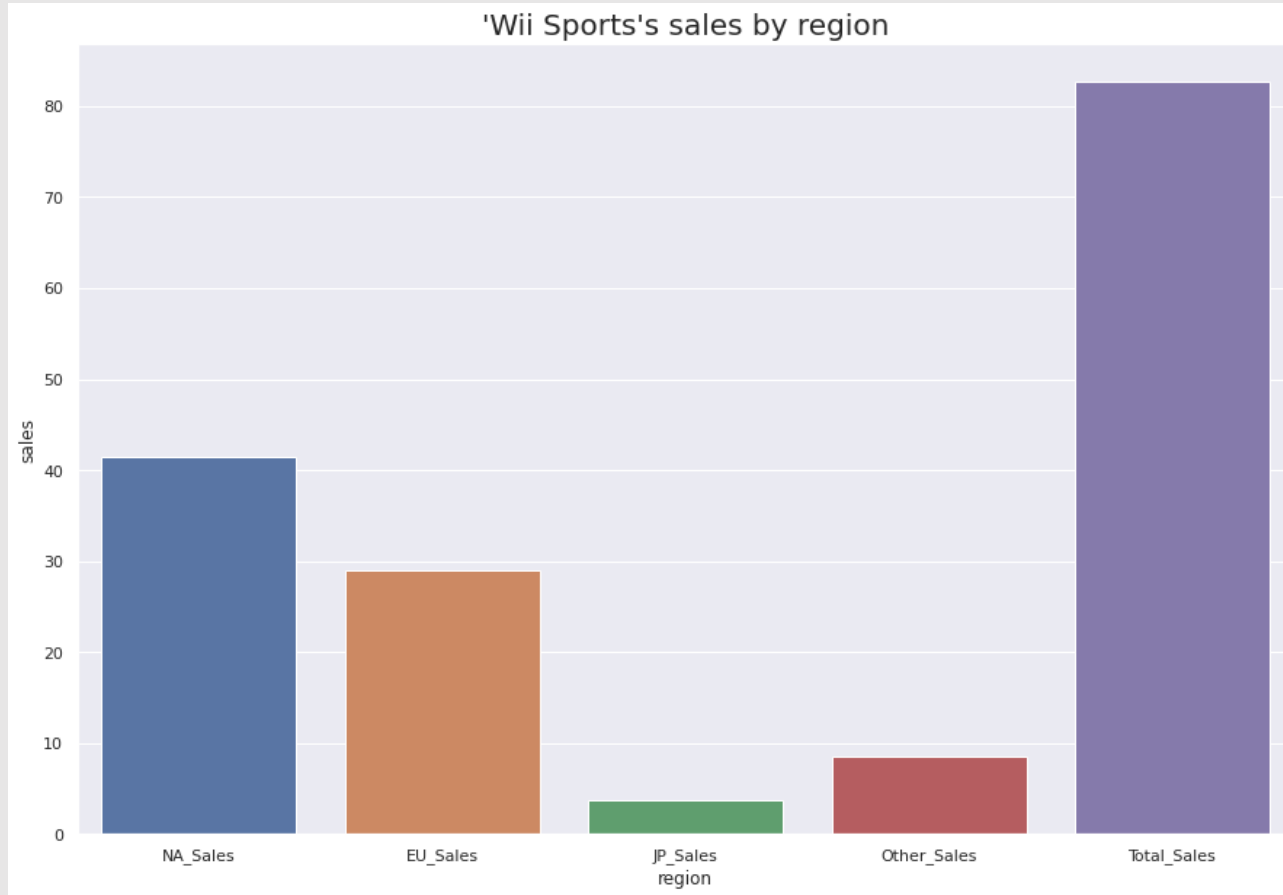
## 데이터 분석 - 출고량이 높은 게임에 대한 분석 및 시각화 프로세스

### 02 데이터 시각화를 진행하였다.



•시각화를 통해 판매량 1위는 압도적으로 Wii Sports인 것을 확인할 수 있다.

### 03 Wii Sports 지역별 판매량 분석 및 시각화

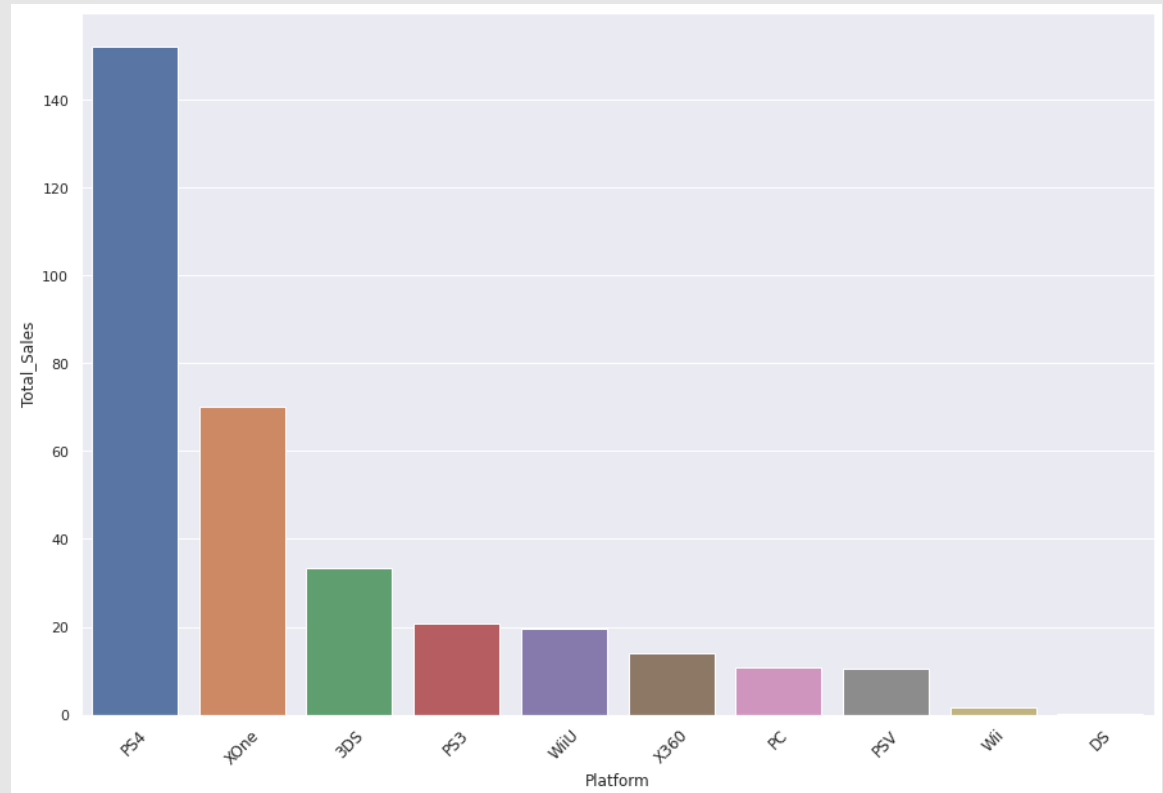
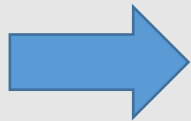


- 북미 지역의 판매량이 총 판매량의 절반 이상으로 다른 지역에 비해 높은 것을 알 수 있다.

## 데이터 분석 – 2015년 이후에 출시된 게임들의 플랫폼 별 판매량 비교

- 01 2015년 이후에 출시된 게임의 플랫폼과 판매량을 나타낸 dataframe을 생성하였다.
- 02 데이터 시각화를 진행하였다.

	Platform	Total_Sales
0	PS4	152.19
1	XOne	70.03
2	3DS	33.52
3	PS3	20.73
4	WiiU	19.65
5	X360	13.85
6	PC	10.63
7	PSV	10.44
8	Wii	1.56
9	DS	0.29



- PS4 플랫폼의 판매량이 압도적으로 높은 것을 확인할 수 있다.

## 데이터 분석 – $\chi^2$ -test 가설검정: 출시한 게임의 플랫폼과 판매량은 연관성이 없다.

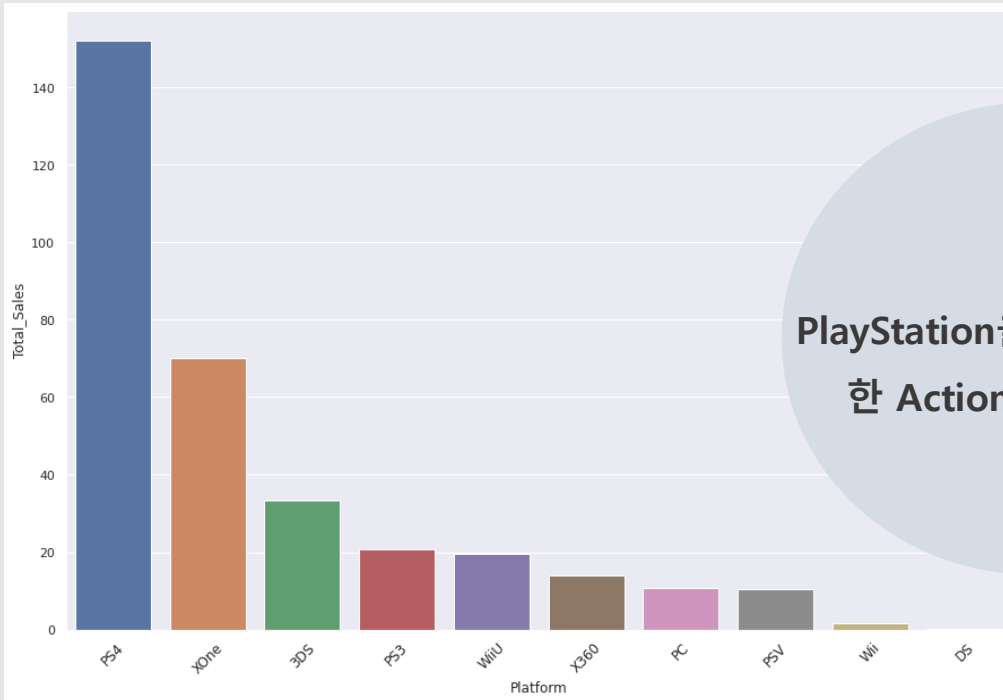
---

- 01
  - $H_0$ : 출시한 게임의 플랫폼과 판매량은 연관성이 없다.
  - $H_a$ : 출시한 게임의 플랫폼과 판매량은 연관성이 있다.
  
- 02
  - 신뢰구간: 0.05
  - P-value 값:  $7.1827567388788175e-236$
  
- 03
  - 결과: p-value 값이 0.05보다 작으므로 대립가설  $H_a$ 를 채택한다.

**• 출시한 게임의 플랫폼과 판매량은 연관성이 있다.**



## 결과 – 다음 분기에 어떤 게임을 설계해야 할까?



PlayStation을 Platform으로  
한 Action 장르의 게임



Thank you!