



## Section 1 presentation



AI\_14 최은지





전처리



데이터 분석



인사이트

---



## 전처리

## 분석 데이터 소개

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
...	...	...	...	...	...	...	...	...	...	...
16593	16594	Ice Age 2: The Meltdown	GC	2006.0	Platform	Vivendi Games	0.15	0.04	0	0.01
16594	16595	Rainbow Islands: Revolution	PSP	2005.0	Action	Rising Star Games	0.01	0	0	0
16595	16596	NBA 2K16	PS3	2015.0	Sports	Take-Two Interactive	0.44	0.19	0.03	0.13
16596	16597	Toukiden: The Age of Demons	PSV	2013.0	Action	Tecmo Koei	0.05	0.05	0.25	0.03
16597	16598	The King of Fighters '95	PS	1996.0	Fighting	Sony Computer Entertainment	0	0	0.16	0.01
16598 rows × 10 columns										



## 전처리

```
df.info()
```

```
<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      16548 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
dtypes: float64(1), int64(1), object(8)
memory usage: 1.3+ MB
```



문자형 -> 숫자형으로 변환

소문

조각의  
선장

확인

장소



## 전처리

```
df.isnull().sum() #결측값 확인
```

```
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
```



결측 값 행들 drop

소문

조각의  
선장

확인

장조



전처리

## 전처리

NA_Sales	EU_Sales	JP_Sales	Other_Sales
0.04	0	0	0
0.17	0	0	0.01
0	0	0.02	0
0.04	0	0	0
0.12	0.09	0	0.04
0	0	0.14	0
0.02	0.01	0	0
0.07	0	0.11	0.01
0	0	0.04	0

0.04	0.03	0	0.01
480K	0.33M	0K	0.06

```
def cng(x):  
    if str(x).find("K") != -1:  
        return float(str(x).split("K")[0].replace('.', ''))*1000  
    elif str(x).find("M") != -1:  
        return float(str(x).split("M")[0].replace('.', ''))*1000000  
    else:  
        return float(str(x).replace('.', ''))*1000000
```

문자열  
변환

확인

결과



NA\_Sales, EU\_Sales, JP\_Sales에 0.01M,  
0.01K 등의 문자열들이 있음 => 제거



## 전처리

PC	9	Strategy	Microsoft Game Studios
PS3	2010	Sports	Crave Entertainment
NES	1985	Platform	Nintendo
PS	2000	Platform	Activision
DS	2007	Puzzle	Crave Entertainment
Wii	2008	Racing	Takara Tomy
GBA	2005	Action	Nintendo
PS	97	Action	Namco Bandai Games

```
def year_cng(x):  
    if (x >= 0) & (x <= 20):  
        return int(x + 2000)  
    elif x <= 100:  
        return int(x + 1900)  
    else:  
        return int(x)  
  
df['Year'] = df['Year'].apply(year_cng)
```

원형

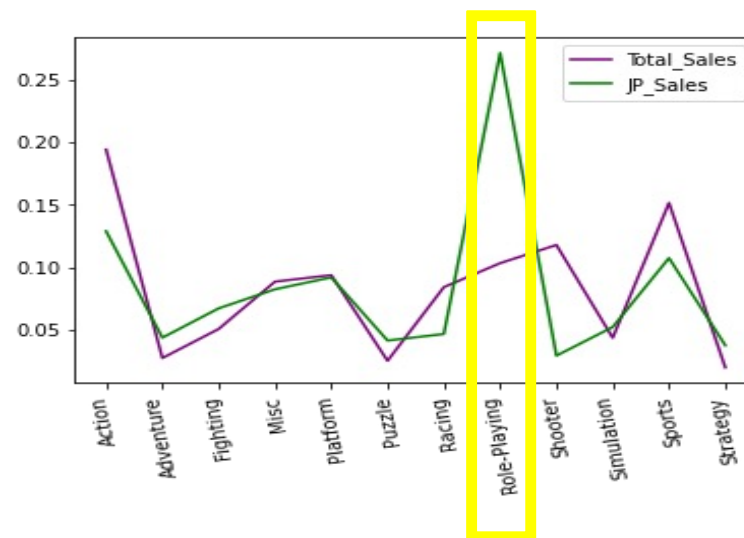
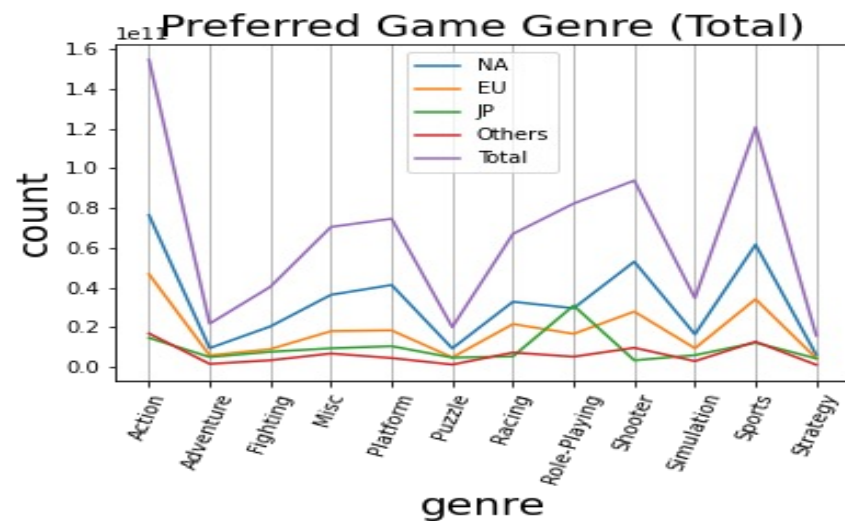
목적

결과



Year 속성에 4자리 년도가 아닌 2자리,  
1자리 숫자들도 있음  
=> 4자리 숫자로 변환

## 가설 1. 지역에 따라서 선호하는 게임 장르가 다를까?





## - 분석 결과

다른 지역과 다르게 일본 지역에서만 Role-Playing 장르가 가장 인기있는 것을 시각적으로 확인 가능

⇒ 일본의 경우만 예외적인 결과이므로 통계적으로 의미가 있는지 검증해볼 필요성이 있음



```
[p-value]
Total과 비교 : 0.0
NA와 비교 : 0.0
EU와 비교 : 0.0
Other과 비교 : 0.0
```

## - 가설 검증

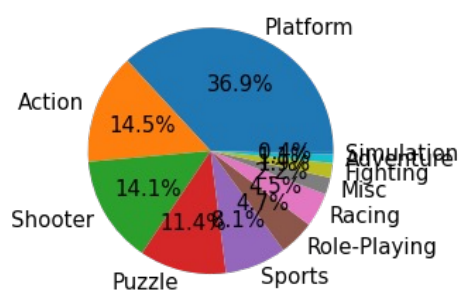
- 1) 귀무가설( $H_0$ ) : 전체 출고량의 분포와 일본지역 출고량의 분포의 차이가 없다
- 2) 대립가설( $H_1$ ) : 전체 출고량의 분포와 일본지역 출고량의 분포의 차이가 있다.

=> 결과를 보면 모두 p-value 값이 0이므로 귀무가설을 기각하고 대립가설을 채택할 수 있다

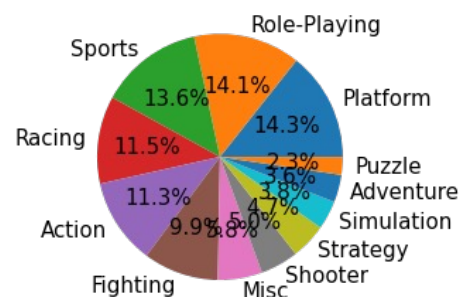
## 가설 2. 게임의 트렌드가 있을까?



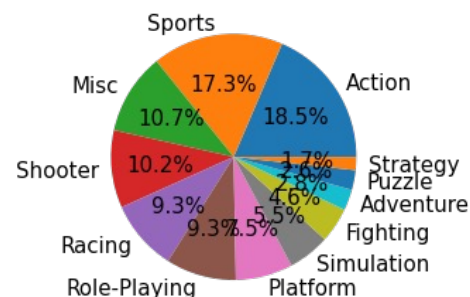
### 각 연도별 구간마다 장르별 출고량



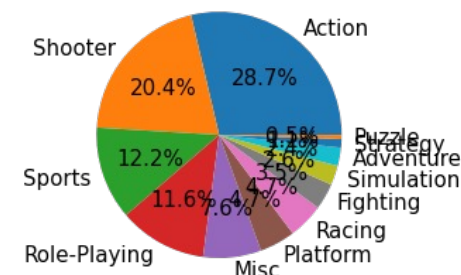
1980 - 1990



1990 - 2000



2000 - 2010



2010 - 2020

## - 분석 결과

1980년-1990년: Platform의 비중이 절반 이상 차지하고 있음

1990년-2000년: Platform의 비중이 낮아지고 다양한 장르에서 비중이 분산됨

2000년-2010년: Action과 Sports의 비중이 높아짐

2010년-2020년: Action, Shooter의 비중이 높지만, 연도별 전체 출고량을 볼 때, 감소하는 추세를 보임

=> 연대별로 비디오 게임의 트렌드가 변화한다는 것을 추론할 수 있음 -> 통계적으로 의미가 있는지 검증해볼 필요성이 있음



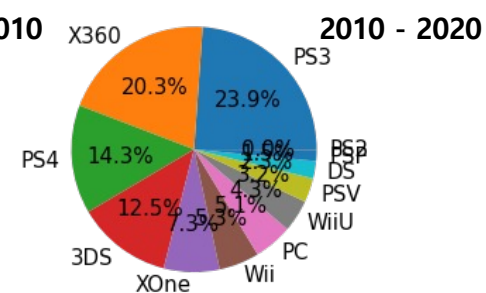
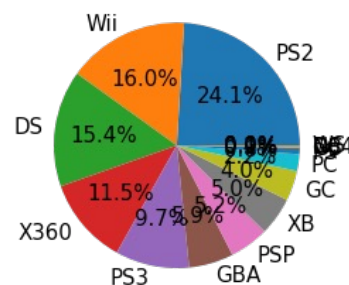
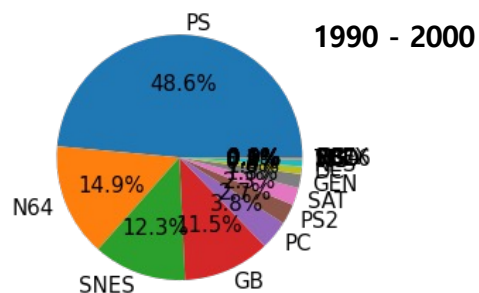
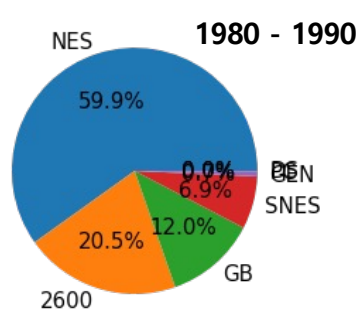
## - 가설 검증

1) 귀무가설( $H_0$ ) : 비디오 게임 장르의 트렌드는 변하지 않는다

2) 대립가설( $H_1$ ) : 비디오 게임 장르의 트렌드는 변한다

=> 비디오게임 시장에서 게임장르의 트렌드는 계속해서 변화하는 것으로 판단할 수 있음

## 각 연도별 구간마다 플랫폼별 출고량



### - 분석 결과

1980-2000: NES, PS 각각 하나의 플랫폼이 절반 이상의 비중을 차지하고 있음을 확인할 수 있다

2000-2010: 전보다는 고르게 분포하지만 PS2 플랫폼이 가장 많은 비율을 차지하고 있음을 확인할 수 있다

2010-2020: 플랫폼별로 차지하는 비율이 매우 고르게 분포한다고 볼 수 있다. PS3, X360이 비슷한 비중으로 가장 많은 비율을 차지하고 있음을 확인할 수 있다

## - 가설 검증

귀무가설( $H_0$ ): 플랫폼의 트렌드는 변하지 않는다

대립가설( $H_1$ ): 플랫폼의 트렌드는 변한다

=> 결과를 보면 p-value 값이 0.05보다 작으므로 기각역에 속하는 것을 알 수 있음

=> 즉 연도에 따라 플랫폼의 트렌드는 계속 변한다

### 가설 3. 출고량이 높은 게임에 대한 분석 및 시각화 프로세스?

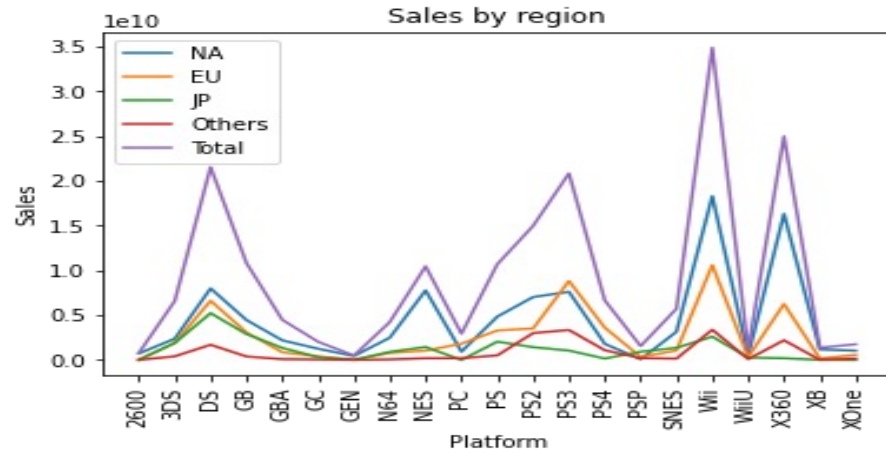
```
df['Total_Sales'] = df['NA_Sales'] + df['EU_Sales'] + df['JP_Sales'] + df['Other_Sales']

df['Sales_Rank'] = df['Total_Sales'].rank(method='min', ascending=False)
df_top = df.sort_values(by='Sales_Rank', ascending=True).head(200)
df_top = df_top.reset_index(drop=True)
df_top
```

=> 출고량 Top 200에 대해 genre, platform, publisher, year 시각화

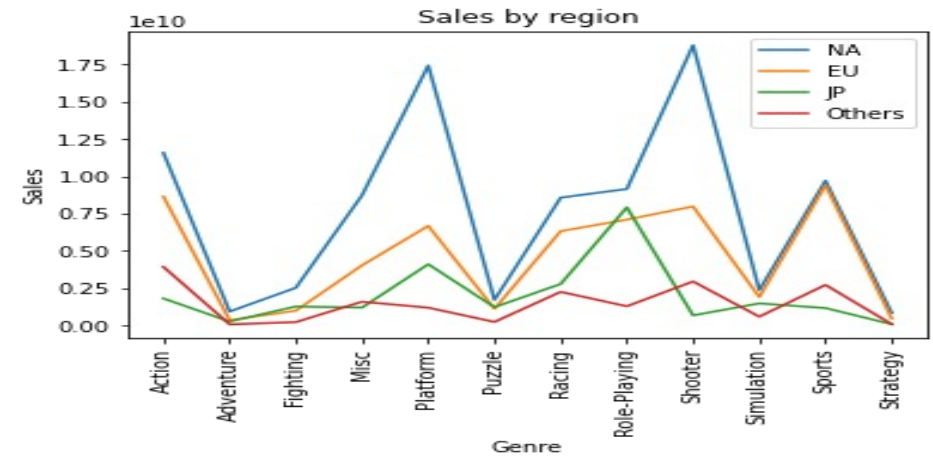
— ○ — 데이터 분석

각 지역별 platform별 판매량 top200 시각화



북미 : Wii, X360  
 유럽 : Wii, PS3  
 일본 : DS  
 기타 : Wii, PS2, PS3, X360

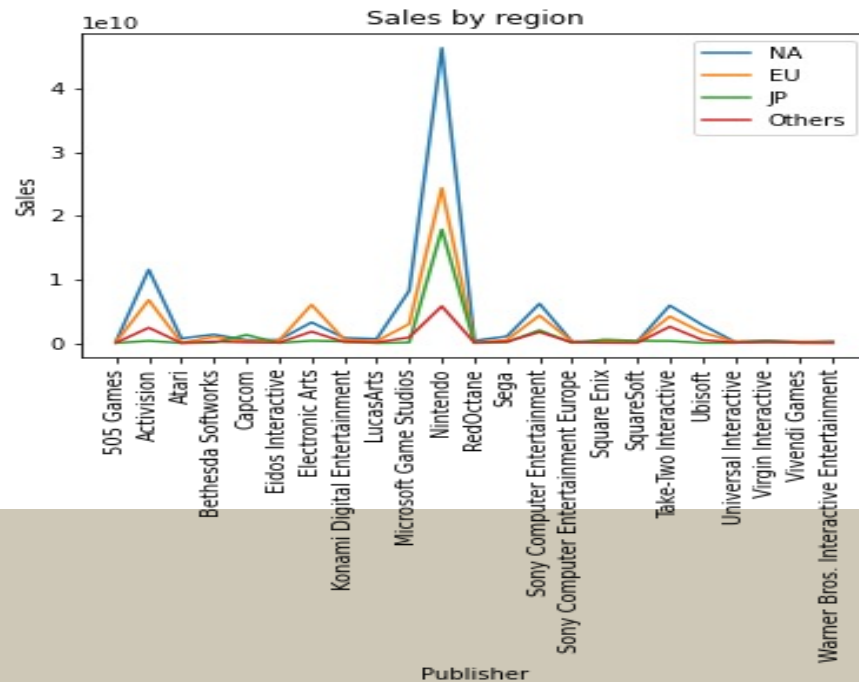
각 지역별 genre별 판매량 top200 시각화



북미 : Shooter, Platform  
 유럽 : Sports, Action, Shooter  
 일본 : Role-Playing  
 기타 : Action

— ○ — 데이터 분석

각 지역별 publisher별 판매량 top20 시각화



북미 : Nintendo

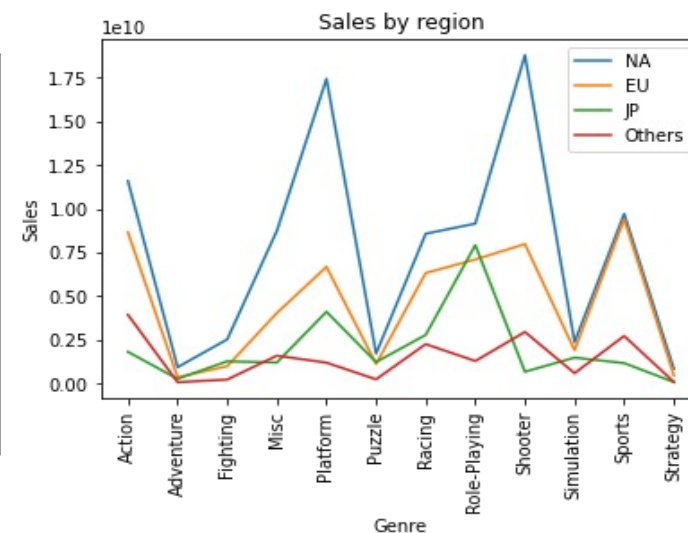
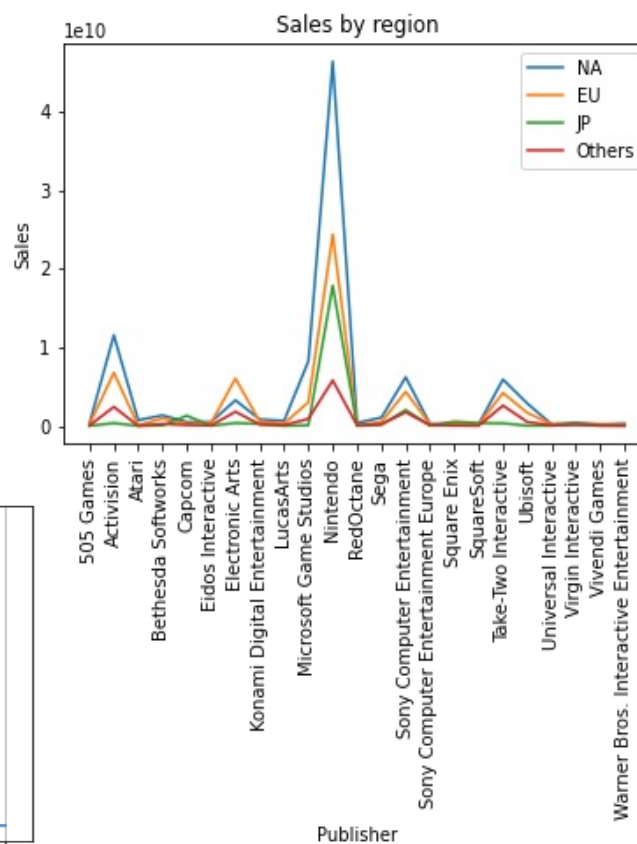
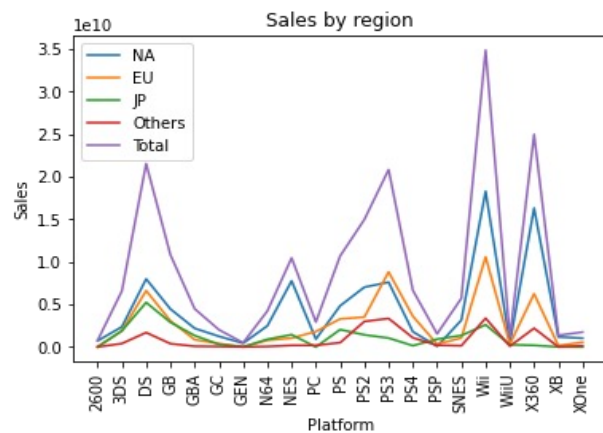
유럽 : Nintendo

일본 : Nintendo

기타 : Nintendo

=> 각 지역 모두 Nintendo가 압도적인 비중을 차지하고 있다







## 종합 결과

### -북미 :

- 1.Publisher: Nintendo
- 2.Platform: Wii
- 3.Genre: Shooter, Platform
- 4.출시시기: 최대한 빠르게

### -유럽 :

- 1.Publisher: Nintendo
- 2.Platform: Wii
- 3.Genre: Sports, Action, Shooter
- 4.출시시기: 최대한 빠르게

### -일본 :

- 1.Publisher: Nintendo
- 2.Platform: DS
- 3.Genre: Role-Playing
- 4.출시시기: 최대한 빠르게

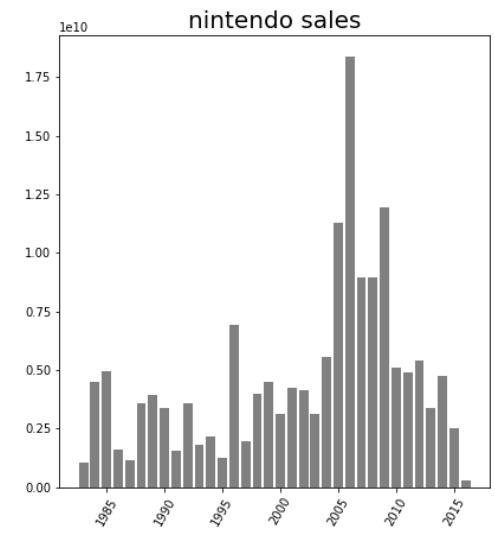
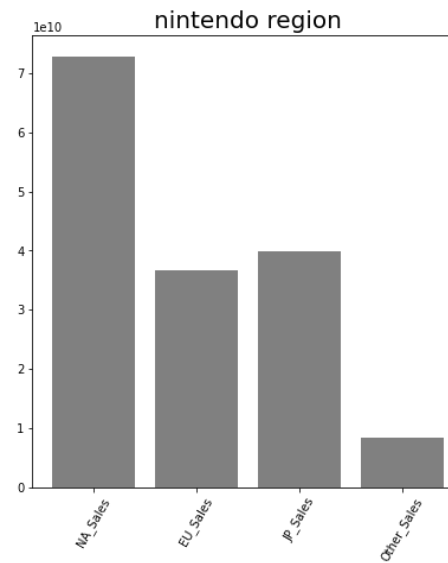
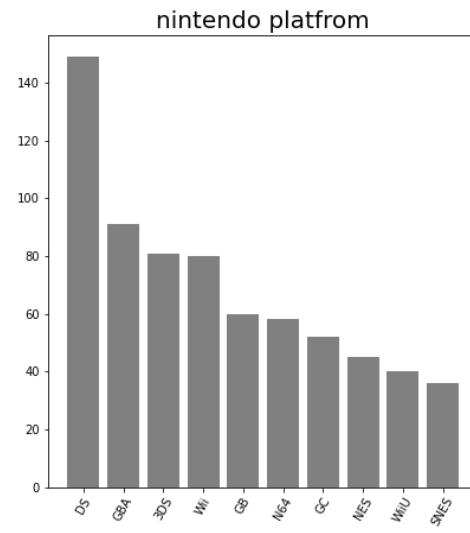
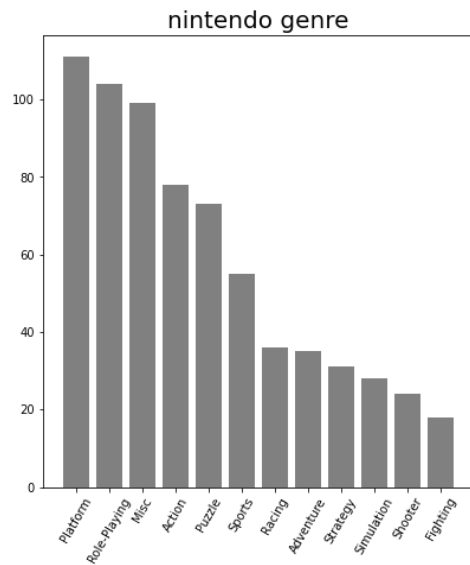
### -기타 :

- 1.Publisher: Nintendo
- 2.Platform: Wii
- 3.Genre: Action
- 4.출시시기: 최대한 빠르게



만약 닌텐도 출판 회사에서 **하나의 지역/플랫폼/장르만을** 선택해서 출시한다면?

## 닌텐도 출판회사와 플랫폼/장르/지역/출고량 간 시각화





인사이트

## **Publisher(출판회사)가 닌텐도인 경우 종합 결과**

1. Platform : DS

2. Region: 북미지역

3. Genre : Platform or Role-Playing or Misc

4. 시기

=> 전반적인 비디오 게임 수요가 줄어들고 있고, 닌텐도 수요도 이전보다 감소하는 추세를 보이고 있으므로 출시 시기는 빠르게 하는 것이 좋다고 판단

**Thank you**