



Codestates AI Bootcamp

다음 분기에 어떤 게임을 설계해야 할까?

Section 1 project

AI_Group | 김서인

Section 1 Project

INDEX

AI Coding Bootcamp 13th

- 데이터 전처리
- 국가별로 선호하는 게임 장르가 있는가
- 연도별 게임의 트렌드가 있을까
- 출고량이 높은 게임에 대한 분석 및 시각화 프로세스
- 결론

PREMIUM EDITION

데이터 전처리

Data Preprocessing



01

02

데이터 확인

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
0	Candace Kane's Candy Factory	DS	2008.0	Action	Destineer	0.04	0	0	0
1	The Munchables	Wii	2009.0	Action	Namco Bandai Games	0.17	0	0	0.01
2	Otome wa Oanesama Boku ni Koi Shiteru Portable	PSP	2010.0	Adventure	Alchemist	0	0	0.02	0
3	Deal or No Deal: Special Edition	DS	2010.0	Misc	Zoo Games	0.04	0	0	0
4	Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010.0	Platform	D3Publisher	0.12	0.09	0	0.04
...
16593	Ice Age 2: The Meltdown	GC	2006.0	Platform	Vivendi Games	0.15	0.04	0	0.01
16594	Rainbow Islands: Revolution	PSP	2005.0	Action	Rising Star Games	0.01	0	0	0
16595	NBA 2K16	PS3	2015.0	Sports	Take-Two Interactive	0.44	0.19	0.03	0.13
16596	Toukiden: The Age of Demons	PSV	2013.0	Action	Tecmo Koei	0.05	0.05	0.25	0.03
16597	The King of Fighters '95	PS	1996.0	Fighting	Sony Computer Entertainment	0	0	0.16	0.01

16598 rows × 9 columns

- 이름, 사용 플랫폼, 연도, 장르, 개발 회사
- 미국, 유럽, 일본, 기타 국가별 매출액
- 16,597개의 게임에 대한 데이터
- 연도는 1980 - 2020년도 까지 존재

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

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

≫≫ 결측치 삭제

- 결측치가 400개 정도
- 전체 16500개 대비 적은 숫자로 삭제하였음.

≫≫ 문자에서 숫자형으로 타입 변환

- 매출액 옆에 문자 및 특수 문자가 섞여 있음.
- 문자 및 특수 문자 삭제

ELDEN RING

DELUXE EDITION

국가별 선호도

Preference by Country

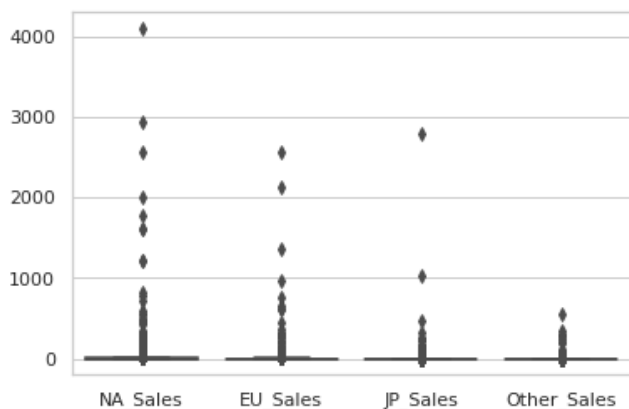


01

02

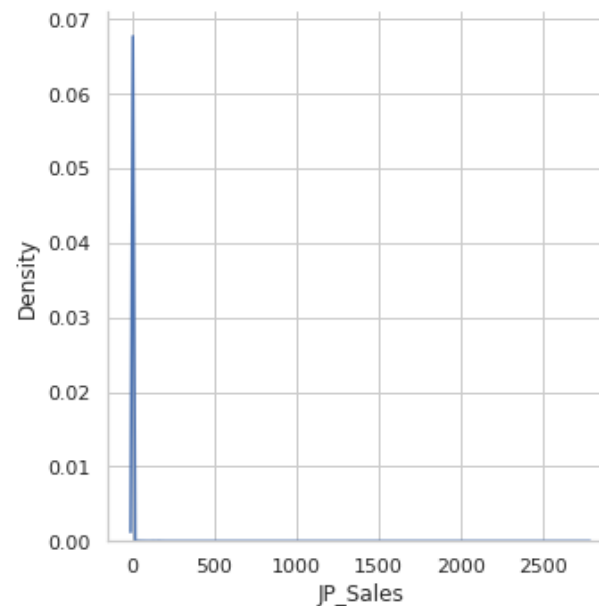
03

04



» 이상치 확인

- 값 확인 결과 1 미만도 많아서 이상치가 매우 극적인 상황
- 즉 게임 산업은 극적으로 성공한 경우와 아닌 경우로 나뉨.



» 일본 게임 매출 분포도 확인

- 밀도 1(100%)이 전체 합인데 매출이 1미만인 게임이 전체의 대부분.
- 그러나 2500을 넘는 게임도 있어서 이상치가 다른 게임을 압도하는 상황.

일본 Top 10 데이터 확인

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Total_Sales
3589	Final Fantasy IX	PS	2000.0	Role-Playing	SquareSoft	1.62	0.77	2780.0	0.14	2782.53
9464	Densha De Go!	PS	1997.0	Simulation	Taito	0.00	0.00	1020.0	0.07	1020.07
6167	Resident Evil Director's Cut	PS	1996.0	Action	Virgin Interactive	1.82	1.24	470.0	0.25	473.31
6342	Kirby's Dream Collection: Special Edition	Wii	2012.0	Action	Nintendo	0.34	0.00	330.0	0.02	330.36
9995	Virtua Racing	GEN	1994.0	Racing	Sega	0.00	0.00	260.0	0.00	260.00
5671	Summon Night 3	PS2	2003.0	Role-Playing	Banpresto	0.00	0.00	240.0	0.00	240.00
2318	Castlevania: Symphony of the Night	PS	1997.0	Platform	Konami Digital Entertainment	0.58	0.40	210.0	0.08	211.06
5583	Digimon World 2	PS	2000.0	Role-Playing	Namco Bandai Games	0.23	0.16	160.0	40.00	200.39
12506	Kidou Senshi Gundam: Giren no Yabou - Axis no ...	PSP	2008.0	Strategy	Namco Bandai Games	0.00	0.00	160.0	0.00	160.00
2400	SimEarth: The Living Planet	SNES	1991.0	Simulation	Imagineer	0.00	0.00	160.0	0.00	160.00

- 2000년에 Final Fantasy 9은 지금도 역작으로 남은 게임으로 일본 역대 전체 매출의 20% 차지
- Top100이 전체 게임 매출이 과반 수 이상의 영향을 행사하기에 게임 전체 집단에 대한 조사가 오히려 매출로만 봤을 때는 너무 넓은 모집단일 수 있음.
- 특정 회사 편중 X, 장르 편중 X, 연도 편중 X

One-way Anova Test

```
model_NA = ols('NA_Sales ~ C(Genre)', df1).fit()
anova_lm(model_NA)
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Genre)	11.0	2.478295e+04	2252.995022	0.70026	0.739621
Residual	16229.0	5.221471e+07	3217.370574	NaN	NaN

```
model_JP = ols('JP_Sales ~ C(Genre)', df1).fit()
anova_lm(model_JP)
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Genre)	11.0	1.009242e+04	917.492648	1.56788	0.10091
Residual	16229.0	9.496892e+06	585.180328	NaN	NaN

```
model_EU = ols('EU_Sales ~ C(Genre)', df1).fit()
anova_lm(model_EU)
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Genre)	11.0	9.846301e+03	895.118284	0.869009	0.570442
Residual	16229.0	1.671659e+07	1030.044362	NaN	NaN

```
model_NA = ols('JP_Sales ~ C(Genre)', scaled_df2).fit()
anova_lm(model_NA)
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Genre)	11.0	4.315548	0.392323	0.381323	0.963855
Residual	15684.0	16136.435127	1.028847	NaN	NaN

- Anova test로 '각 국가별 매출 평균이 장르마다 다르다' 가설을 세우고 검정.
- P-value가 5%미만은 되어 신뢰성이 있다고 할 수가 있음. 여기서 p-value는 100번 시도하면 5번은 틀린 가설을 인용할 확률이라고 쉽게 이해 가능.
- 미국, 유럽, 일본, 기타를 대상으로 검정했으나 일본만 10% 정도로 그나마 유의하다고 볼 수 있고 나머지는 유의하지 않아서 장르마다 국가별 매출이 다르다고 보기 어려움.

표준화 One-way Anova

	NA_Sales	EU_Sales	JP_Sales	Other_Sales
0	-0.037988	-0.033685	-0.022645	-0.038341
1	-0.035703	-0.033685	-0.022645	-0.036979
2	-0.038691	-0.033685	-0.021821	-0.038341
3	-0.037988	-0.033685	-0.022645	-0.038341
4	-0.036582	-0.030889	-0.022645	-0.032894
...
16136	-0.036055	-0.032443	-0.022645	-0.036979
16137	-0.038516	-0.033685	-0.022645	-0.038341
16138	-0.030957	-0.027783	-0.021409	-0.020638
16139	-0.037813	-0.032132	-0.012344	-0.034256
16140	-0.038691	-0.033685	-0.016052	-0.036979

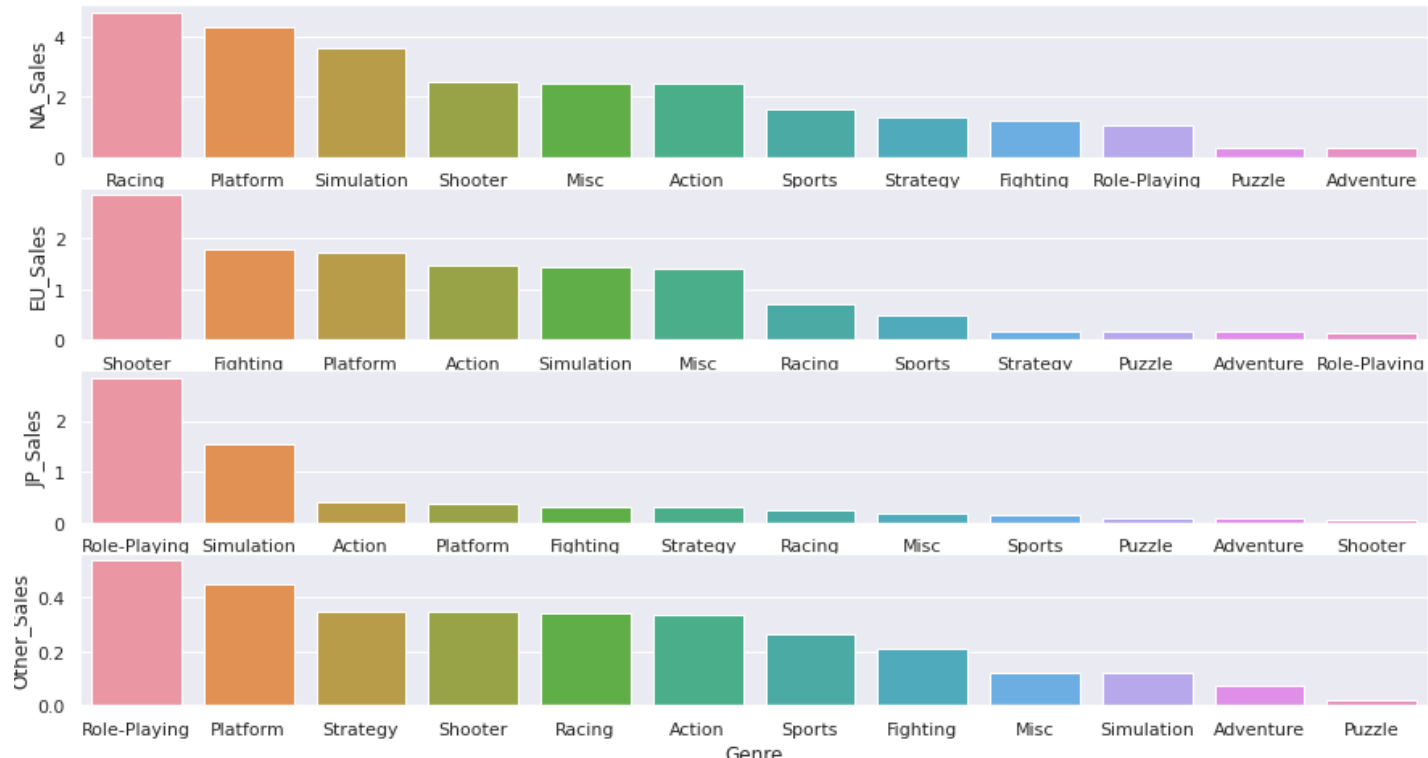
16141 rows × 5 columns

```
model_Other = ols('Other_Sales ~ C(Genre)', df1).fit()
anova_lm(model_Other)
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Genre)	11.0	304.185964	27.653269	0.51576	0.89421
Residual	16229.0	870142.524759	53.616521	NaN	NaN

- 일본은 성공한 게임들의 매출이 너무 커서 장르별 평균이 10% 유의 수준에 는 다를 수 있다라는 결론이 도출되었음.
- 다시 확인하기 위해 표준화 과정을 진행. 표준화 과정은 평균 0, 분산 1로 분 포를 전환하여 통계적 검정의 전제 조건인 표준정규 분포 가정을 만족시키 기 위해 사용.
- 그 결과 p-value는 매우 높아져 유의하지 않음. 이상치가 너무 커서 검정 자 체의 유의성이 떨어짐.

장르별 게임 당 평균 매출



- 그림에도 국가별 평균 매출액으로 Bar plot을 그려 도식화를 해보면
- 미국 : Racing, Platform, Simulation 순서대로 게임 당 평균 매출이 높다.
- 유럽 : Shooter, Fighting, Platform 순서대로 게임 당 평균 매출이 높다.
- 일본 : Role-playing, Simulation, Action 순서대로 게임 당 평균 매출이 높다.
- 기타 : Role-playing, Platform, Strategy 순서대로 게임 당 평균 매출이 높다.



연도별 트렌드

Preference by Years



01

02

03

```
Conditions=[
(df1['Year'] >=1976) & (df1['Year'] <=1980),
(df1['Year'] >=1981) & (df1['Year'] <=1985),
(df1['Year'] >=1986) & (df1['Year'] <=1990),
(df1['Year'] >=1991) & (df1['Year'] <=1995),
(df1['Year'] >=1996) & (df1['Year'] <=2000),
(df1['Year'] >=2001) & (df1['Year'] <=2005),
(df1['Year'] >=2006) & (df1['Year'] <=2010),
(df1['Year'] >=2011) & (df1['Year'] <=2015),
(df1['Year'] >=2016) & (df1['Year'] <=2020))

condition_values = [1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020]
```

5년 단위로 데이터를 끊어 줌.



- 개별 연도로 보면 패턴이 쉽게 나오지 않음.
- 1976-1980년 -> 1980으로 그룹화

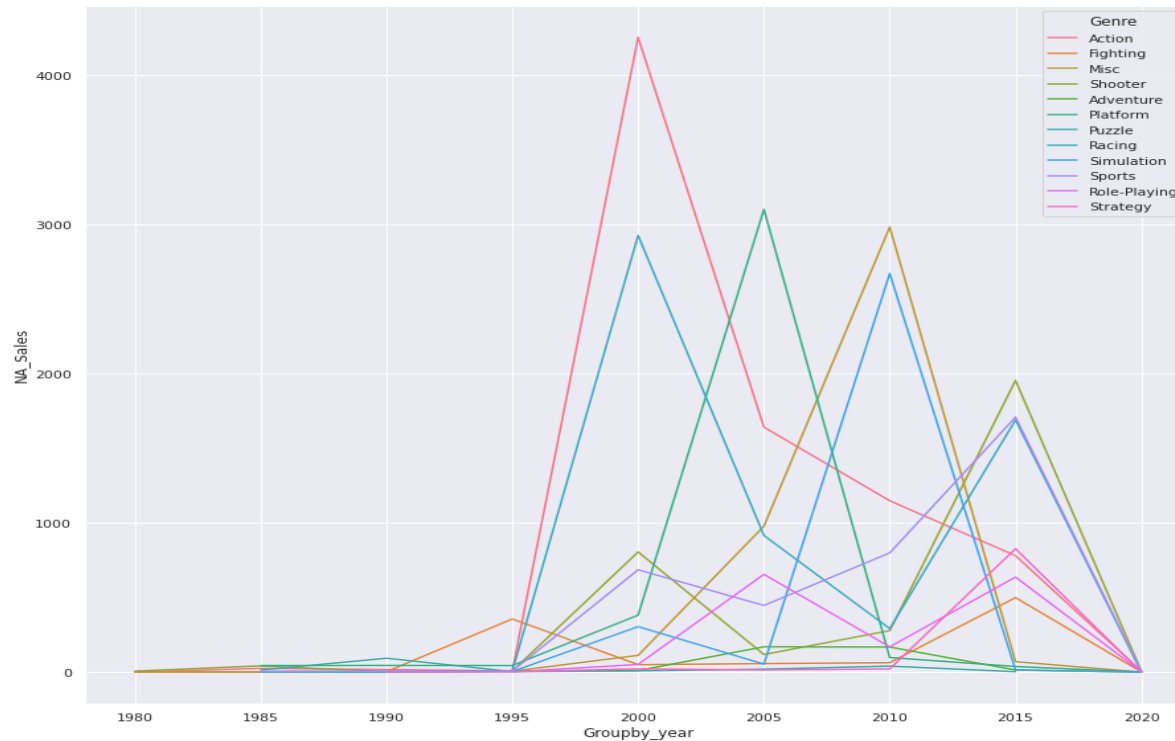
5개년 단위로 장르에 해당하는 수익을 모두 더하여 x축 기준으로 시계열로 나타냄.

01

02

03

미국 장르별 매출 합산액



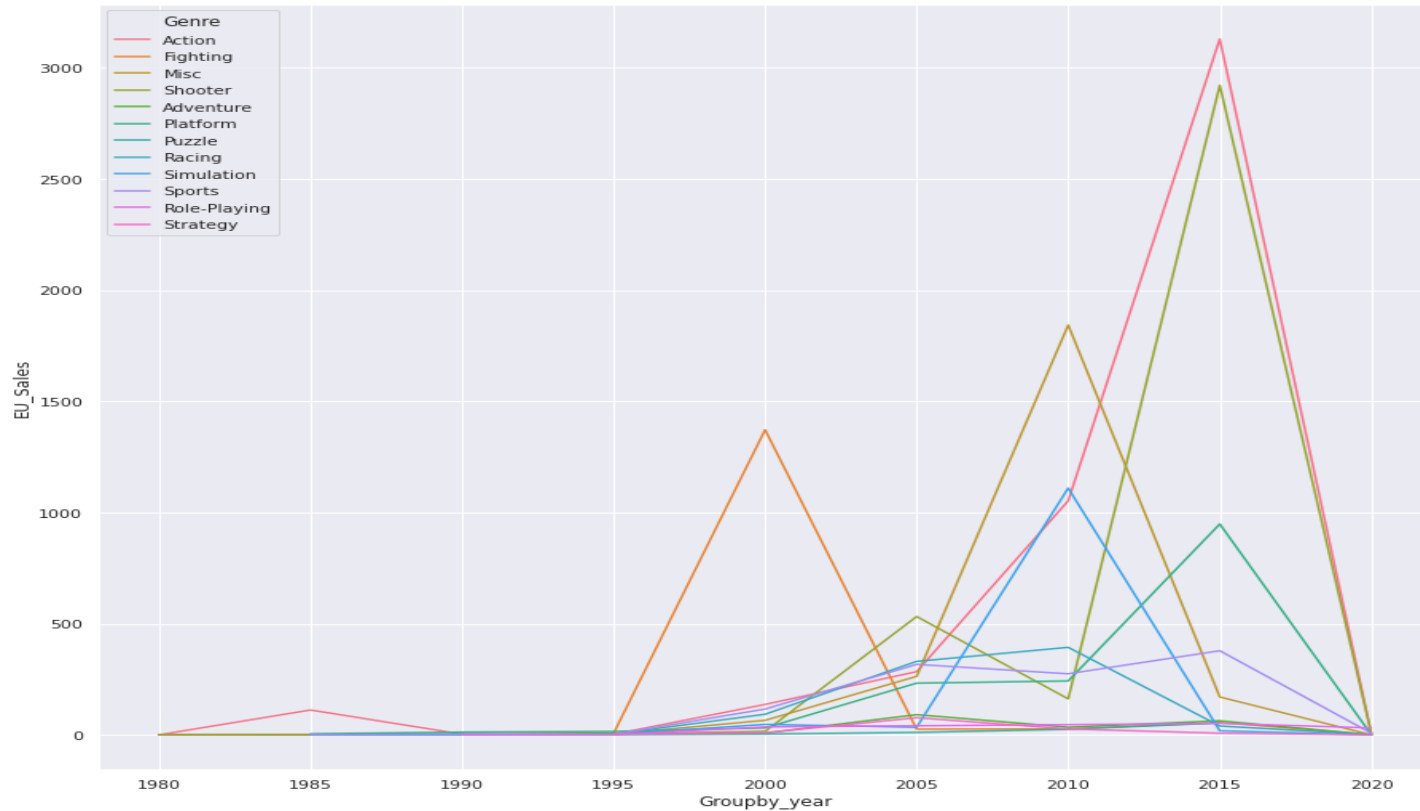
- 2000년도 Action, 2005년도 platform, 2010년 Misc, 2015년 Shooting 순으로 트렌드가 있었음.
- 매출액은 2000년도 피크를 찍고 꾸준히 감소 중
- 그럼에도 이상치가 전체 합에 큰 영향을 미치기에 그 부분을 항상 염두해야 함.

01

02

03

유럽 장르별 매출 합산액



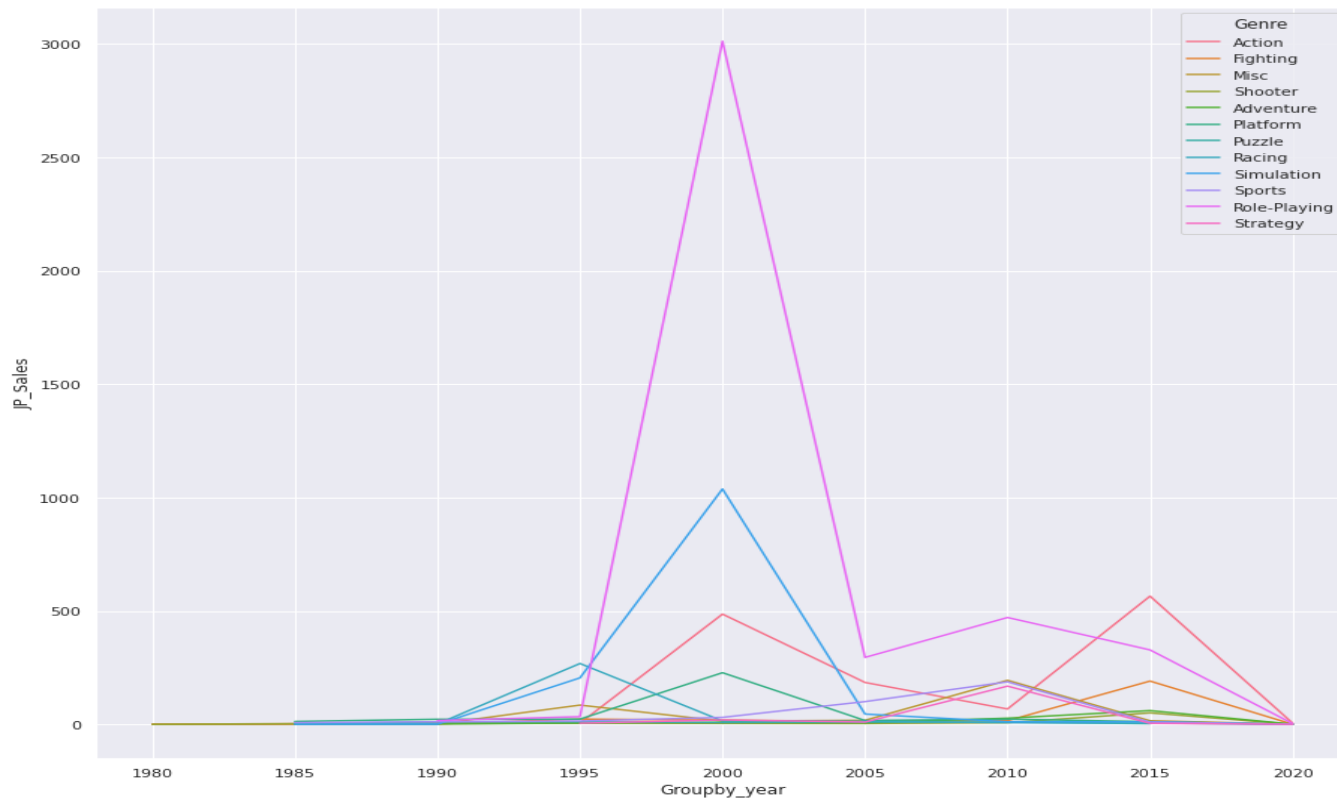
- 2015년에 Action과 Misc가 급격히 성장함.
- 2000년까지만 해도 Fighting 게임 매출 제외하고는 매출이 크지 않았는데 2015년에 와서는 다른 게임 장르들도 성숙하고 전체 매출액도 커지면서 게임 시장 자체가 성숙해진 느낌.

01

02

03

일본 장르별 매출 합산액



- 2000년 Role-playing게임이 매우 유행했고 다른 장르 게임도 매출액이 컸음.
- 그 이후에 매출이 급격히 감소하여 시장이 작아진 느낌.
- 2015년에는 Action 장르가 인기.

```
#2 samples t-test (두 집단간 검정)
```

```
ttest1= df1[(df1['Genre']=='Role-Playing') & (df1['Groupby_year']==2000)]['JP_Sales']  
ttest2= df1[(df1['Genre']=='Action') & (df1['Groupby_year']==2005)]['JP_Sales']
```

```
#등분산 아님 Welch t test
```

```
stats.ttest_ind(ttest1, ttest2, equal_var=False)
```

```
Ttest_indResult(statistic=1.0683610429429826, pvalue=0.2872362077613213)
```

```
#정규성 검정 shapiro test
```

```
stats.shapiro(ttest1), stats.shapiro(ttest2)
```

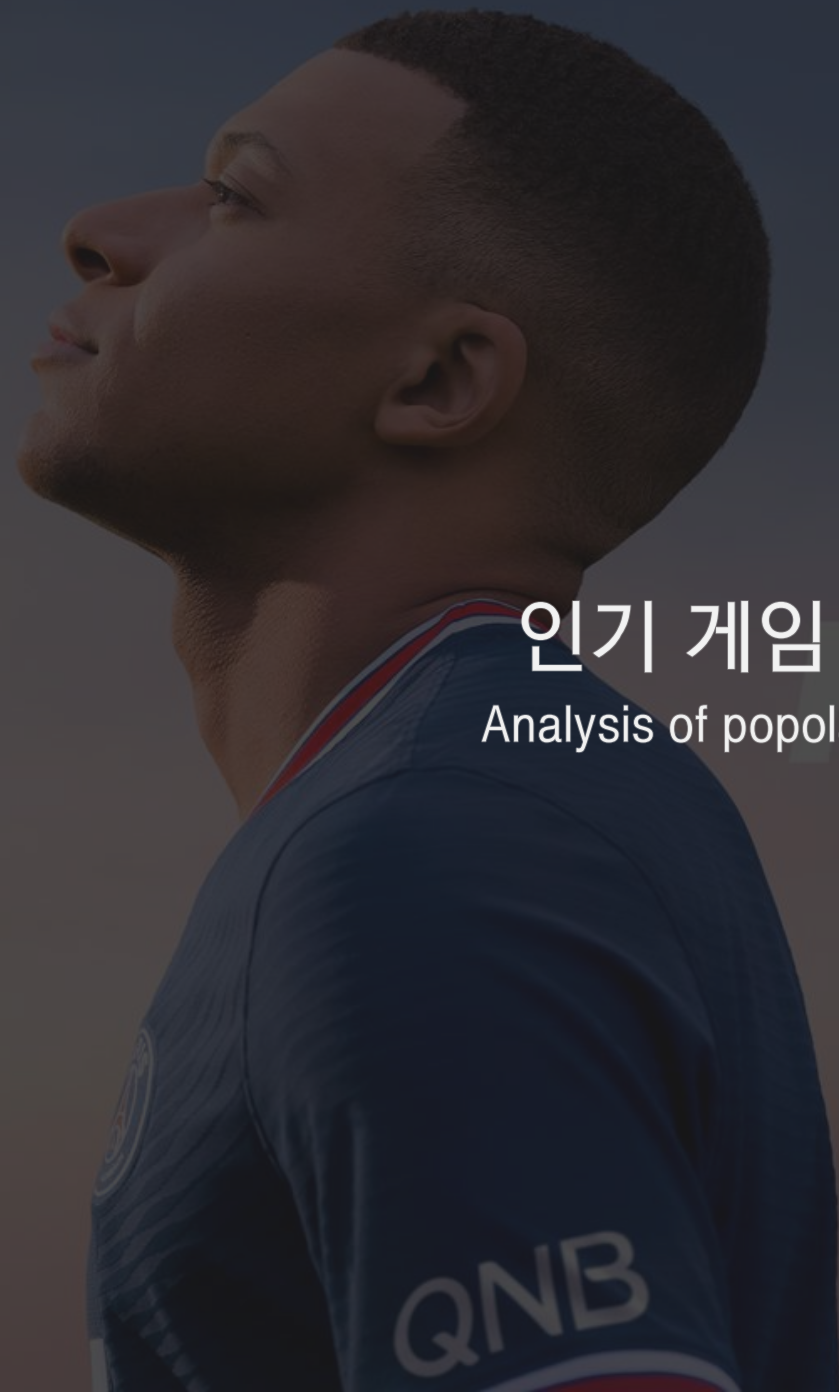
```
((0.06572341918945312, 9.82919640199491e-26), (0.030212998390197754, 0.0))
```

➤➤ 2-sample t-test 검정

- 2000년 일본 Role-playing 장르 게임 매출 평균과 2005년 일본 Action 장르 게임 매출 평균이 다른가에 대한 검정.
- 평균이 크게 다름에도 불구하고 이상치 제외하고는 차이가 유의하게 나지 않아 검정 결과는 매출 평균이 유의하게 다르지 않다는 결론.

➤➤ Shapiro test

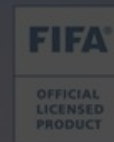
- T test를 하려면 정규성 증명이 되어야 하는데 Shapiro test는 p-value가 5% 이상일 시에는 정규 분포라는 결론.
- P-value가 0에 가까움으로 정규 분포가 아님.
- T 검정 수행이 어려움.



인기 게임 분석

Analysis of popular games

FIFA 22

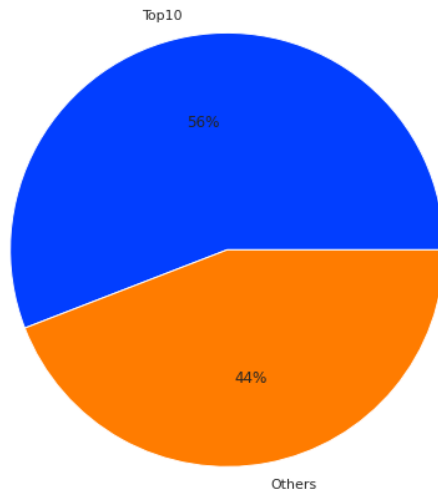


01

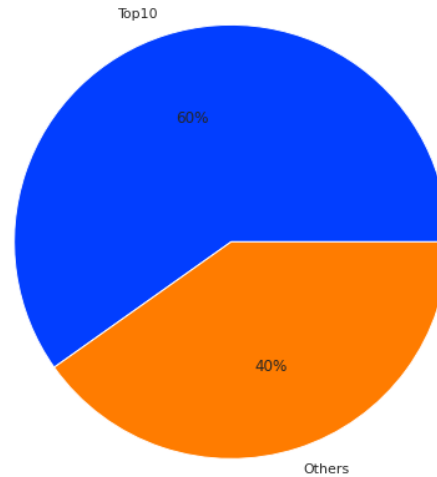
02

03

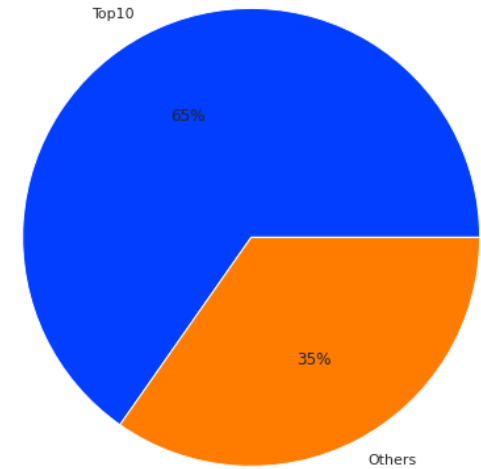
미국



유럽



일본



Top10의 전체 매출 비중

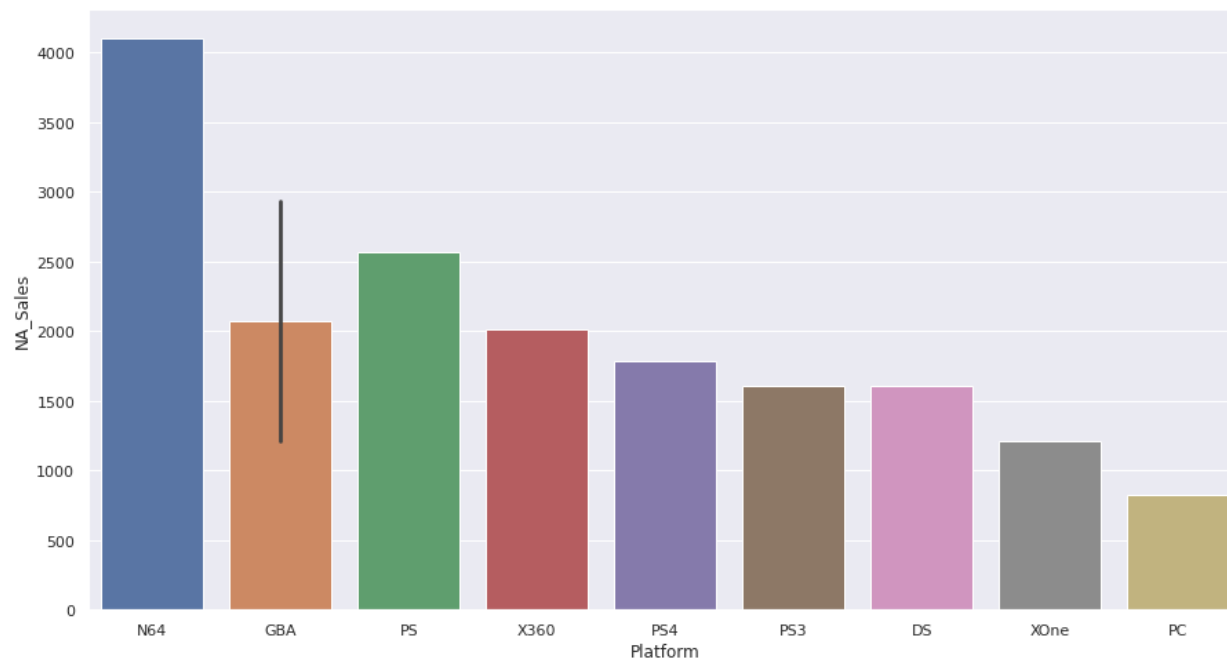
- 56-65%로 매우 높음.

01

02

03

미국 플랫폼 별 매출 합산액



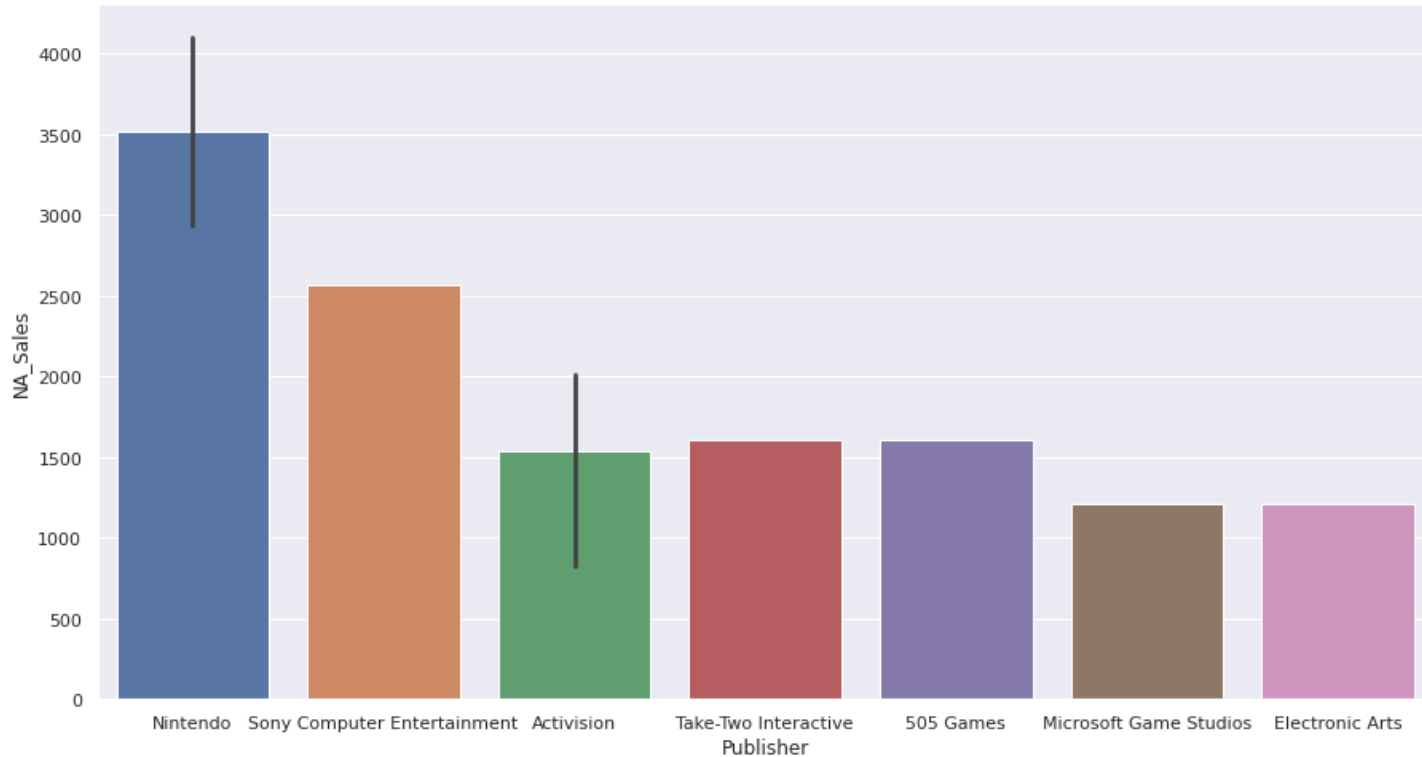
- 과거 매출이 컸던 N64와 GBA. 그 당시에는 플랫폼 간 게임 전환이 불가능했음.
- 현재는 PS, Xbox 골고루 높음. 기술의 발달로 플랫폼과 전환이 쉬워졌고, 개발 당시부터 게임이 모든 플랫폼을 공략해서 나와서 플랫폼이 이제는 크게 상관 없음.

01

02

03

미국 회사별 매출 합산액



- 미국은 특정 회사가 이제 게임 생태계를 지배하는 것이 아닌 일본 계열 회사인 Nintendo, Sony 그리고 미국 회사 Activision, Microsoft 등 다양한 회사가 경쟁하고 있음.
- 경쟁 상 타국의 게임회사도 유연하게 들어올 수 있는 곳. 반면에 일본은 자국 게임에 대한 선호도가 매우 높아서 그렇지 않음.

A composite image featuring several Overwatch characters in a dynamic battle scene. In the center is the massive, grey and yellow mechanical titan, Reinhardt, with his iconic shield and a determined expression. To his left, the agile, blue and white robot, D.Va, is shown in mid-air. Below Reinhardt, the character Tracer is visible, wearing her signature blue and white outfit and goggles. To the right, the character Genji is depicted in his sleek, silver and black samurai-inspired armor, holding his katana. In the upper right, the character Mercy is seen with her wings spread, holding her staff. The background is a hazy, orange-tinted sky, suggesting a battlefield environment. The overall style is a detailed, stylized illustration typical of Overwatch promotional art.

결론

Result

≫ 어떤 게임을 만들어야 하나?

- 장르는 상관없이 없고, 우수한 게임을 만든다.
- 연도별로 트렌드는 존재하지만 그 트렌드는 몇몇 매우 인기 있는 게임에 의해 돌아가기에 트렌드가 게임 시장을 주도하는 것이 아니라, 게임이 시장을 주도하므로 원인과 결과를 반대로 생각하면 안 된다. 잘 만들어서 소비자가 좋아하는 것이 중요하다.
- 매출이 높은 게임에서 플랫폼은 중요했으나 이제는 플랫폼 간 경계가 없어져 중요하지 않다.
- 게임 회사는 미국, 유럽 시장에서는 안 중요하나 일본 시장에서는 중요하기에 일본 시장은 한국기업으로서 피한다.