

Secetion 1. Project 1

< 목차 >

데이터 전처리

- 데이터 전처리 과정

메인가설

- **지역에 따라서 선호하는 게임 장르가 다를까?**
 - 지역과 게임장르의 관계가 독립적인지 확인 하기 위해 2-Sampled 카이 제곱 검정을 사용
 - 귀무가설: 지역과 장르는 독립적이다
 - 대립가설: 지역과 장르는 독립적이지 않다
- **연도별 게임의 트렌드가 있을까?**
 - 연도별 장르 출고량의 그래프를 시각화
 - 연도별 플랫폼 출고량의 그래프를 시각화

시각화

- **출고량이 높은 게임에 대한 분석 및 시각화 프로세스**
 - 최근 10년간 게임, 장르, 플랫폼 출고량 순위 시각화
 - 최근 10년간 각 지역 별 게임, 장르, 플랫폼 출고량 순위 시각화
 - 1980-2020 까지 게임 장르 플랫폼 출고량 순위 시각화

결론

< 다음 분기에 어떤 게임을 설계 해야 할까? >

- 최근 10년간 게임 트렌드 파악
- 각 지역별 게임 트렌드 파악

1. 데이터 전처리



Year 에 100보다 작은 수로 쓰인 곳이 있음

0 으로 바꾼후 년도를 포함한 데이터 분석에 사용하지 않았음

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
75	Age of Empires III: Complete Collection	PC	0.0	Strategy	Microsoft Game Studios	0.01	0.12	0	0.03
82	Treasures of the Deep	PS	0.0	Action	Namco Bandai Games	0.1	0.07	0	0.01
214	NASCAR Unleashed	PS3	0.0	Racing	Activision	0.09	0	0	0.01
305	Kung Fu Panda: Showdown of Legendary Legends	PS3	0.0	Action	Little Orbit	0	0.02	0	0
769	NASCAR Thunder 2004	PS2	0.0	Racing	Electronic Arts	0.53	0.41	0	0.14
...
15986	Sugoi Hebereke	SNES	0.0	Fighting	Sunsoft	0	0	0.02	0
16104	Fab 5 Soccer	DS	0.0	Sports	Destineer	0.01	0	0	0
16150	Dragon Ball Z: Burst Limit	X360	0.0	Fighting	Atari	0.24	0.18	0.03	0.05
16395	Point Blank 3	PS	0.0	Shooter	Namco Bandai Games	0.01	0.01	0	0
16483	Super Robot Taisen F Kanketsuhen	SAT	0.0	Strategy	Banpresto	0M	0	0.55	0

100 rows × 9 columns



Platform에 이름이 아닌 숫자 2600이 쓰여 있음

0 으로 바꾸고 플랫폼 데이터를 분석 할때 사용하지 않았음

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
38	Beamrider	0	1984.0	Racing	Activision	0.26	0.01	0	0
211	Mountain King	0	1982.0	Action	CBS Electronics	0.29	0.02	0	0
665	Smurf: Rescue In Gargamel's Castle	0	1981.0	Action	Coleco	0.55	0.03	0	0.01
821	Fantastic Voyage	0	1981.0	Action	20th Century Fox Video Games	0.32	0.02	0	0
878	Deadly Duck	0	1981.0	Shooter	20th Century Fox Video Games	0.21	0.01	0	0
...
16211	Crazy Climber	0	1981.0	Action	Atari	0.21	0.01	0	0
16264	Frogs And Flies	0	1981.0	Action	Mattel Interactive	0.33	0.02	0	0
16318	Polaris	0	1982.0	Action	Tigervision	0.26	0.02	0	0
16364	Pitfall II: Lost Caverns	0	1983.0	Platform	Activision	1.22	0.07	0	0.02
16473	Custer's Revenge	0	1981.0	Action	Mystique	0.76	0.05	0	0.01

115 rows × 9 columns



Sales에 M,K 로 숫자 단위 표현이 되어 있는 곳이 있음

단위를 M 으로 통일 하여 숫자로 표현 함, ex) 10K → 0.01M → 0.01

```
#함수 적용
df['NA_Sales']=df['NA_Sales'].apply(ch_sles)
df['EU_Sales']=df['EU_Sales'].apply(ch_sles)
df['JP_Sales']=df['JP_Sales'].apply(ch_sles)
df['Other_Sales']=df['Other_Sales'].apply(ch_sles)
# 확인
df
```

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales
1	Candace Kane's Candy Factory	DS	2008.0	Action	Destineer	0.04	0.00	0.00	0.00
2	The Munchables	Wii	2009.0	Action	Namco Bandai Games	0.17	0.00	0.00	0.01
3	Otome wa Oanesama Boku ni Koi Shiteru Portable	PSP	2010.0	Adventure	Alchemist	0.00	0.00	0.02	0.00
4	Deal or No Deal: Special Edition	DS	2010.0	Misc	Zoo Games	0.04	0.00	0.00	0.00
5	Ben 10 Ultimate Alien: Cosmic Destruction	PS3	2010.0	Platform	D3Publisher	0.12	0.09	0.00	0.04
...
16594	Ice Age 2: The Meltdown	GC	2006.0	Platform	Vivendi Games	0.15	0.04	0.00	0.01
16595	Rainbow Islands: Revolution	PSP	2005.0	Action	Rising Star Games	0.01	0.00	0.00	0.00
16596	NBA 2K16	PS3	2015.0	Sports	Take-Two Interactive	0.44	0.19	0.03	0.13
16597	Toukiden: The Age of Demons	PSV	2013.0	Action	Tecmo Koei	0.05	0.05	0.25	0.03
16598	The King of Fighters '95	PS	1996.0	Fighting	Sony Computer Entertainment	0.00	0.00	0.16	0.01

16241 rows × 9 columns

2. 메인가설

1. 지역에 따라서 선호하는 게임 장르가 다를까?

- 지역과 게임장르의 관계가 독립적인지 확인 하기 위해 2-Sampled 카이 제곱 검정을 사용

각 행 마다 가장 출고량이 높았던 나라 찾아서 새로운 열을 만들어 (max_coun)저장

Genre와 max_coun 열을 조합하여 crosstab!

```
result= pd.crosstab(df['max_coun'],df['Genre']) # 가장 큰나라 + 장르 조합 해서 만듦
```

Genre	Action	Adventure	Fighting	Misc	Platform	Puzzle	Racing	Role-Playing	Shooter	Simulation	Sports	Strategy
max_coun												
EU	581	174	54	197	94	103	272	130	274	153	368	196
JP	667	667	298	442	94	98	67	765	106	196	391	229
NA	1969	426	478	1025	681	365	873	567	899	494	1525	243
OT	23	1	5	17	4	0	7	3	1	3	15	1



→ **Two - Sample 카이제곱 검정 진행!**

: 데이터 타입은 범주형, 변수는 '장르' '가장 출고량이 많은 지역' 2개
이 두 변수 사이에 연관성이 있는지 없는지 판별 가능!

가설 검정

H0: 지역과 장르는 독립적이다

H1: 지역과 장르는 독립적이지 않다

```
# 귀무가설: 지역과 장르는 독립적이다
# 대립 가설: 지역과 장르는 독립적이지 않다
from scipy.stats import chi2_contingency
chi2_contingency(result, correction=False)
```

→ 실행 했을시 결과

```
(2137.30693937281,
 0.0,
 33,
 array([[ 517.88929253,  202.68013053,  133.46838249,  268.69503109,
          139.54239271,   90.47078382,  194.8478542 ,  234.16907826,
          204.59823902,  135.22664861,  367.47761837,  106.93454837],
        [ 801.97032202,  313.85752109,  206.68062311,  416.08398498,
          216.08644788,  140.09728465,  301.72895758,  362.61929684,
          316.82778154,  209.40336186,  569.05239825,  165.5920202 ],
        [1904.18077705,  745.21642756,  490.73794717,  987.94070562,
          513.07093159,  332.64392587,  716.41863186,  860.99532049,
          752.26894896,  497.20275845, 1351.14555754,  393.17806785],
        [ 15.9596084 ,   6.24592082,   4.11304723,   8.28027831,
           4.30022782,   2.78800566,   6.00455637,   7.21630441,
           6.30503048,   4.16723108,  11.32442584,   3.29536359]]))
```

카이제곱 검정 결과

chi-square 값: 2137.30693937281

pvalue: 0.0

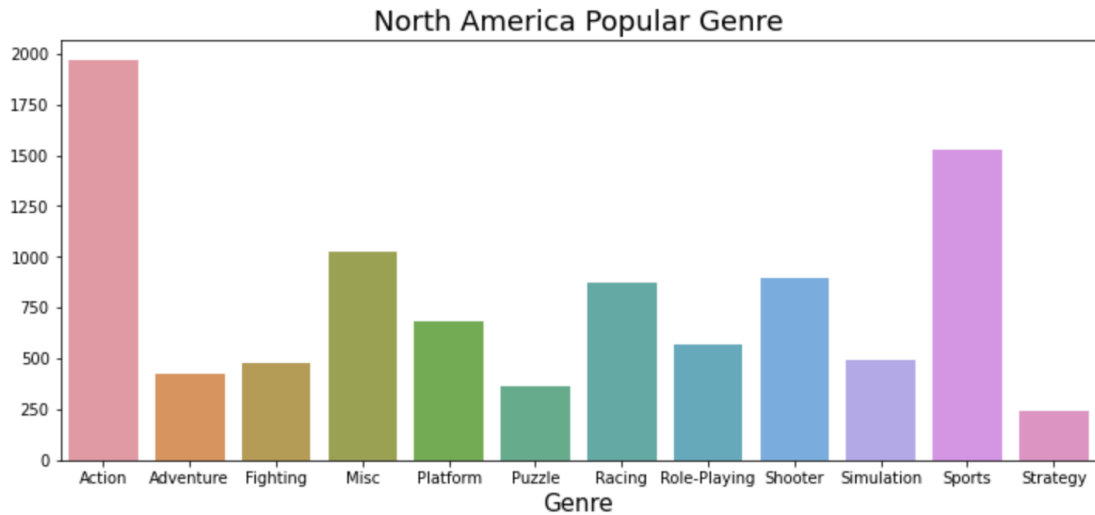
df(degree of freedom): 33

P-value 가 0.05 보다 낮다 : 귀무가설 기각, 대립가설 채택

그러므로 지역과 장르는 독립적이지 않으면 연관성이 있을 확률이 높다!

+ 추가 각 지역별 선호 장르 그래프 (카이 제곱 검정에 활용한 데이터 기반)

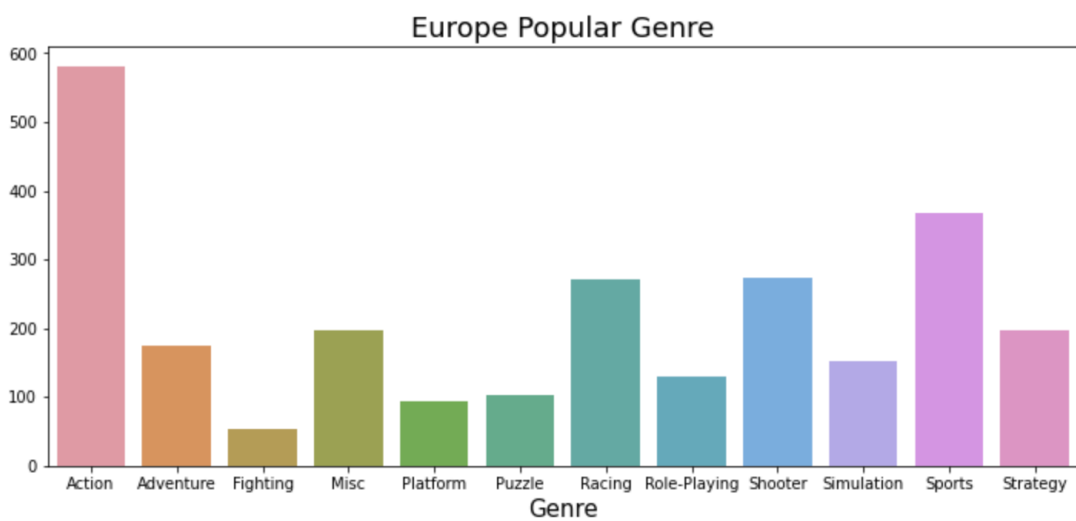
< 북미 >



북미 지역 선호 장르 그래프

Action, Sports, Misc 순서

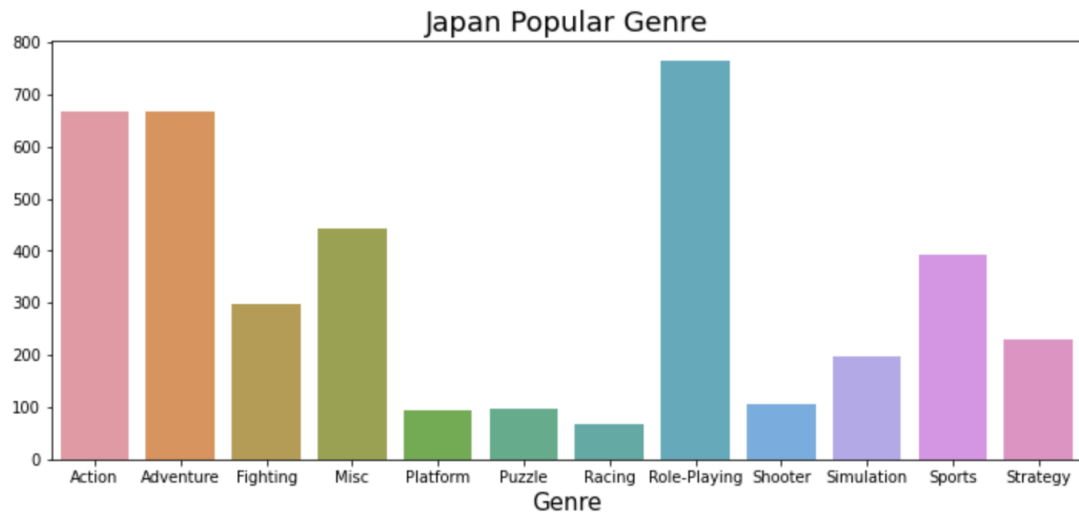
< 유럽 >



유럽 지역 선호 장르 그래프

Action, Sports, Shooter 순서

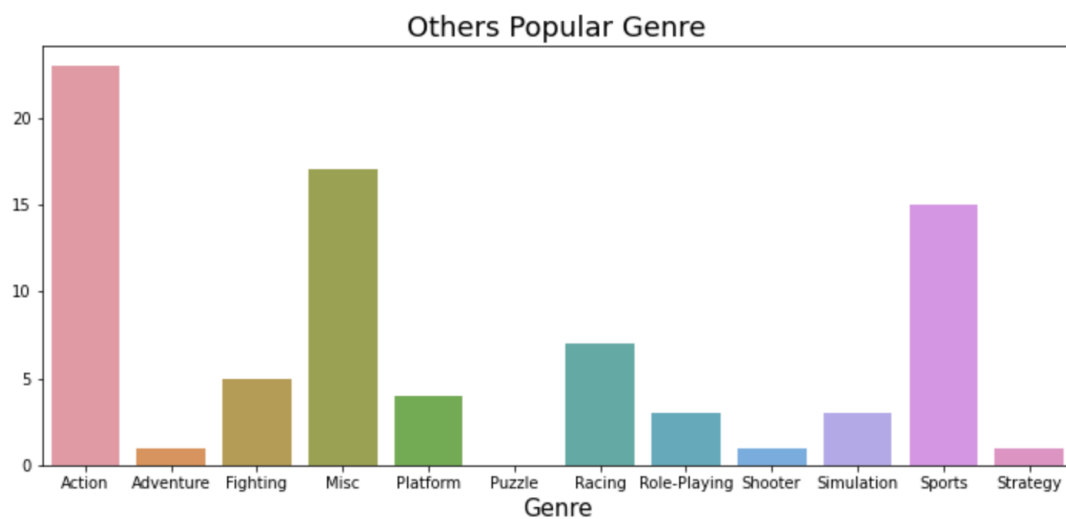
< 일본 >



일본 지역 선호 장르 그래프

Role-Playing, Adventure, Action 순서

< 다른 지역 >



다른 지역 선호 장르 그래프

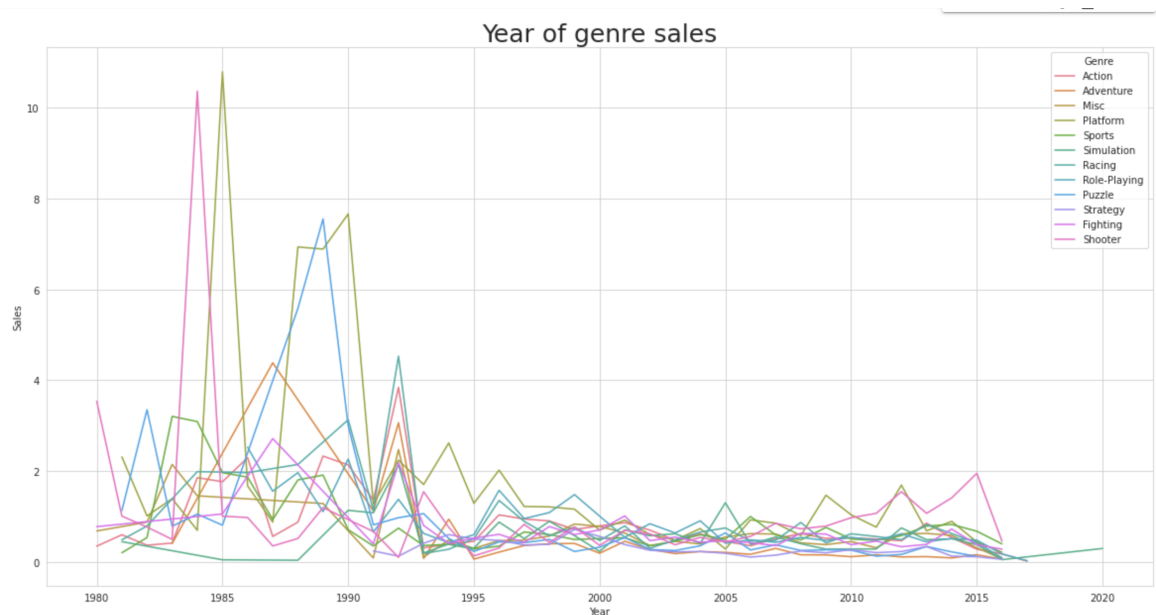
Action, Misc, Sports 순서

카이제곱 검정과 지역별 그래프를 봤을 때 지역별로 장르의 선호가 다른것을 알 수 있음

2. 연도별 게임의 트렌드가 있을까 ?

데이터 프레임 '년도' '장르' 사용

< 연도별 장르 출고량의 line plot >

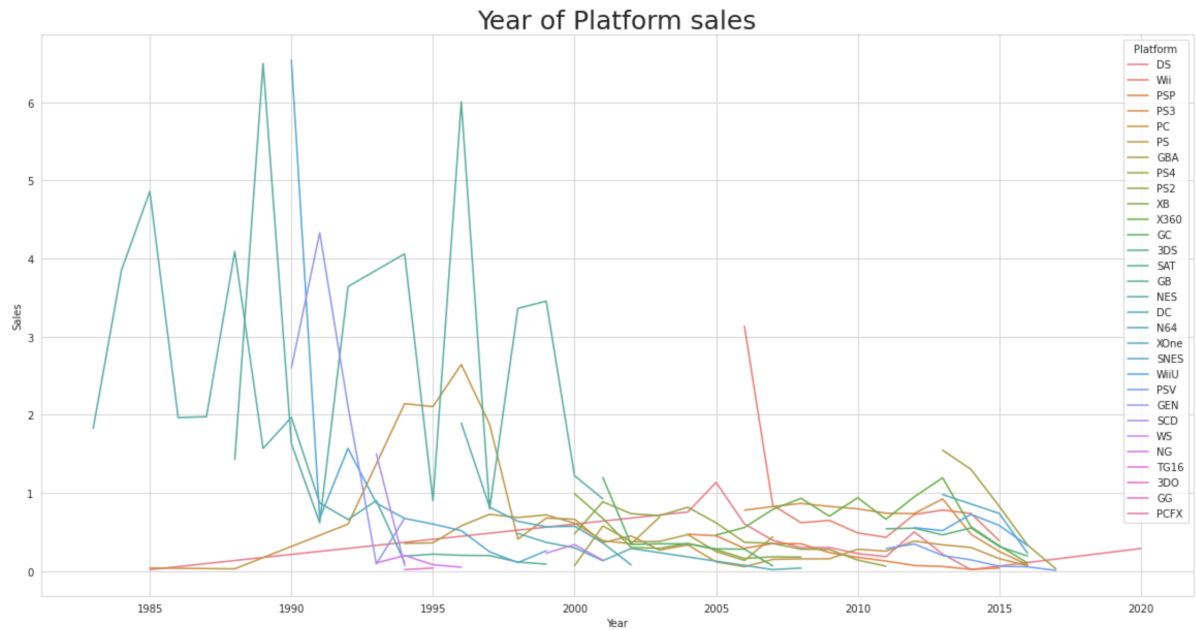


- 모든 장르의 기울기가 10 년 사이에 변화율이 높다, 특히 가장 인기 많던 장르들이 변화율이 높은것으로 관찰된다
- 1980-90 대의 인기 장르가 2010-20 대에서 이전 만큼의 인기를 보여주지 않음

즉, 연도별로 기간별로 장르의 트렌드가 변동한다

데이터 프레임 '년도' '플랫폼' 사용

< 연도별 플랫폼 출고량의 line plot >



- 플랫폼의 변화율이 5년의 기간 사이에서도 변동이 높은걸로 관찰
- 일정 기간 이후 사라진 플랫폼도 존재(예측 이유: 사람들의 선호도 하락)

즉, 연도별로 기간별로 플랫폼의 트렌드가 변동한다



기간별 장르와 플랫폼의 변화율이 높은걸로 관찰 되기 때문에 연도별 게임의 트렌드는 존재할 확률이 높다

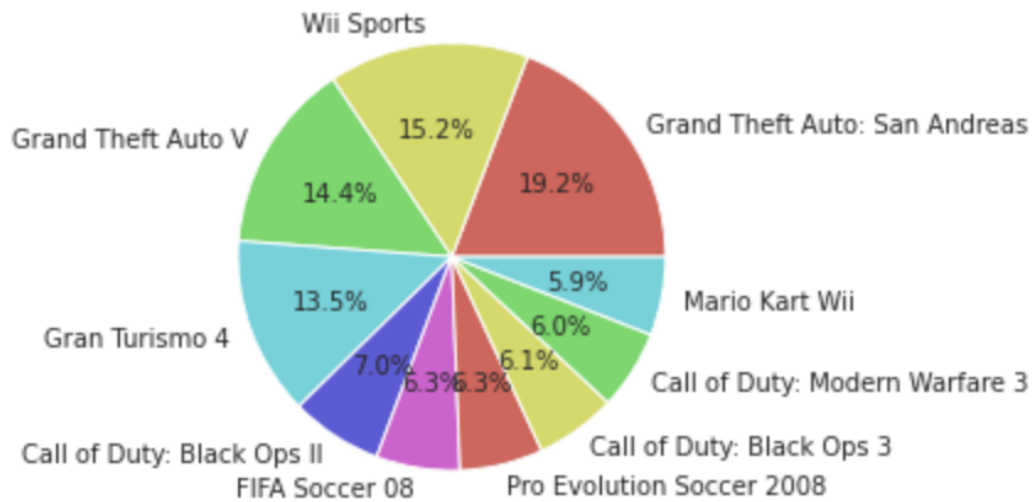
3. 시각화

출고량이 높은 게임에 대한 시각화 와 분석

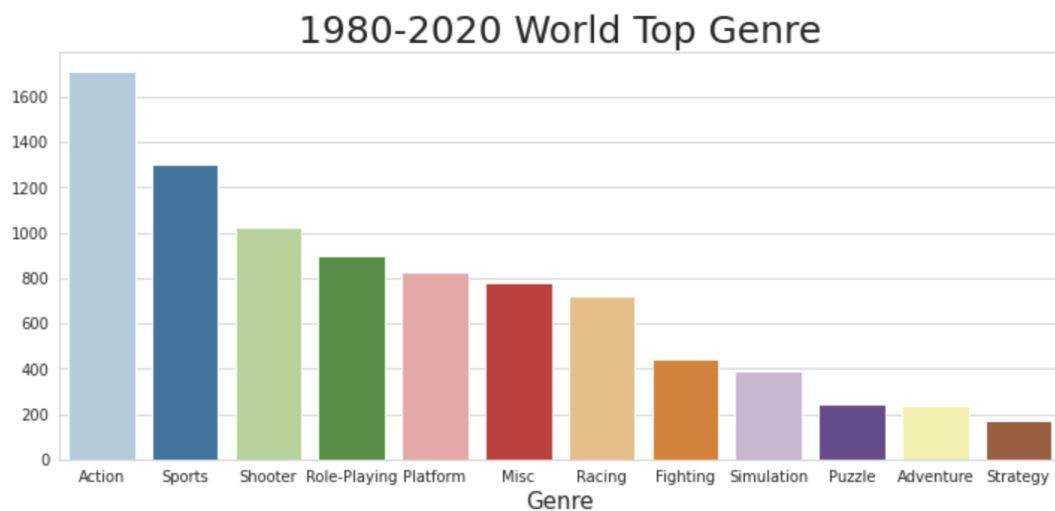


1980-2020 까지 게임 장르 플랫폼 출고량 순위

- 게임 출고량 순위 top 10

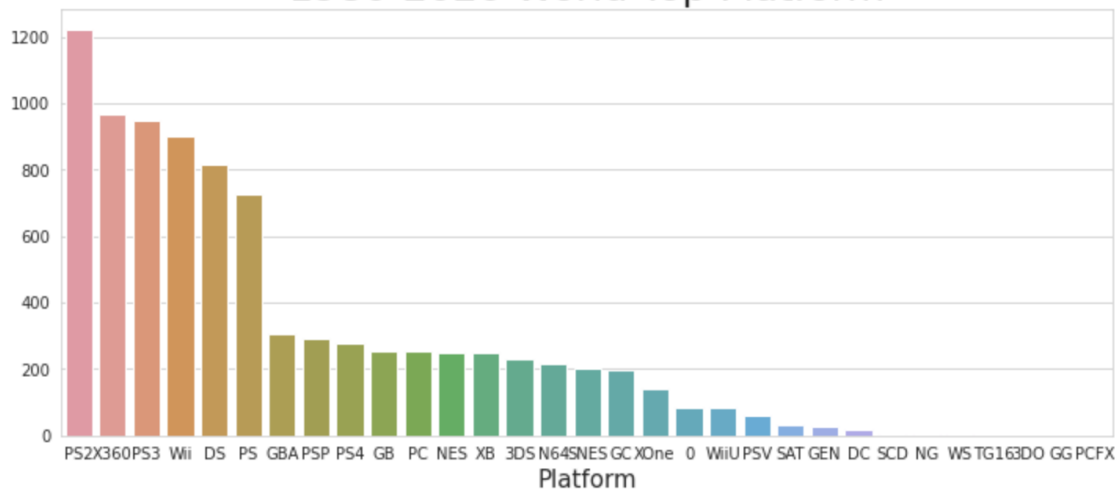


- 장르 순위



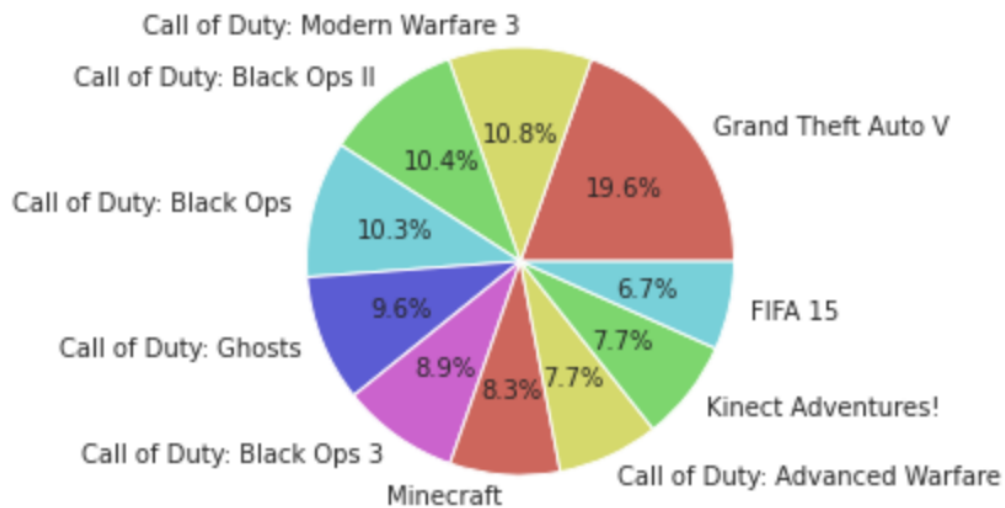
- 플랫폼 순위

1980-2020 World Top Platform



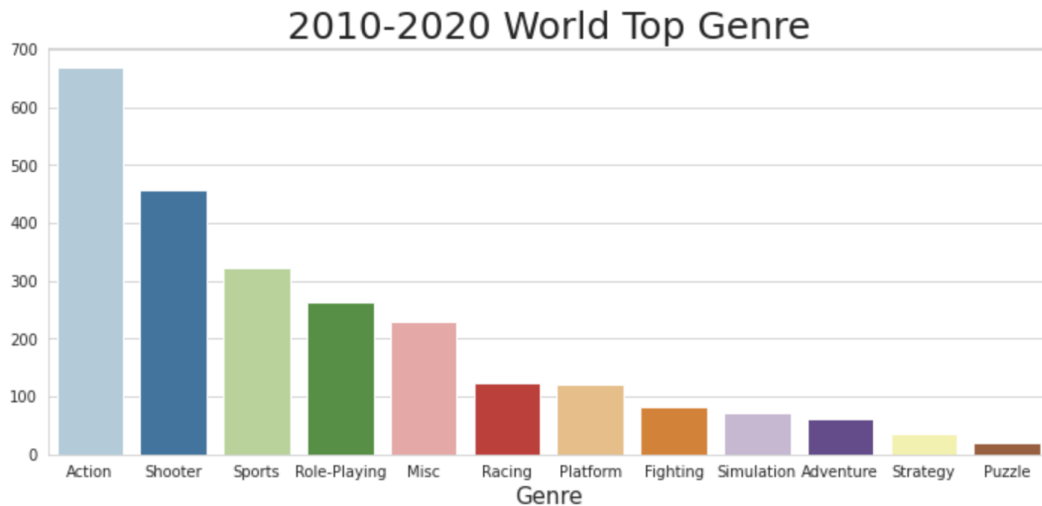
최근 10년간 게임, 장르, 플랫폼 출고량 순위

• 최근 10년간 게임 top 10



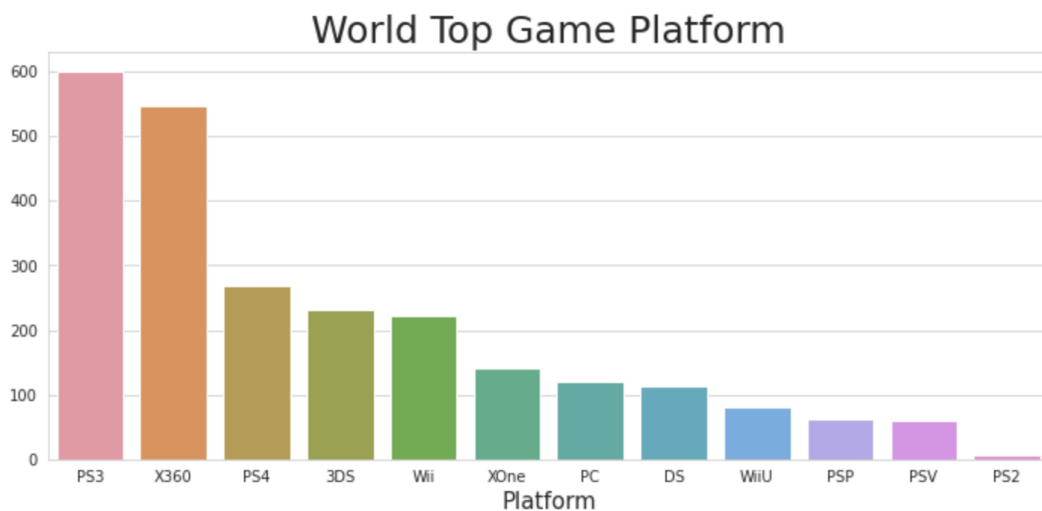
Top 3 : Grand Theft Auto V, Call of Duty: Modern Warfare, Call of duty: Black Ops II

• 최근 10년간 장르 순위



Top 3: Action, Shooter, Sports

- 최근 10년간 플랫폼 순위



Top3: PS3, X360, PS4



최근 10년간 각 지역 별 게임, 장르, 플랫폼 출고량 순위

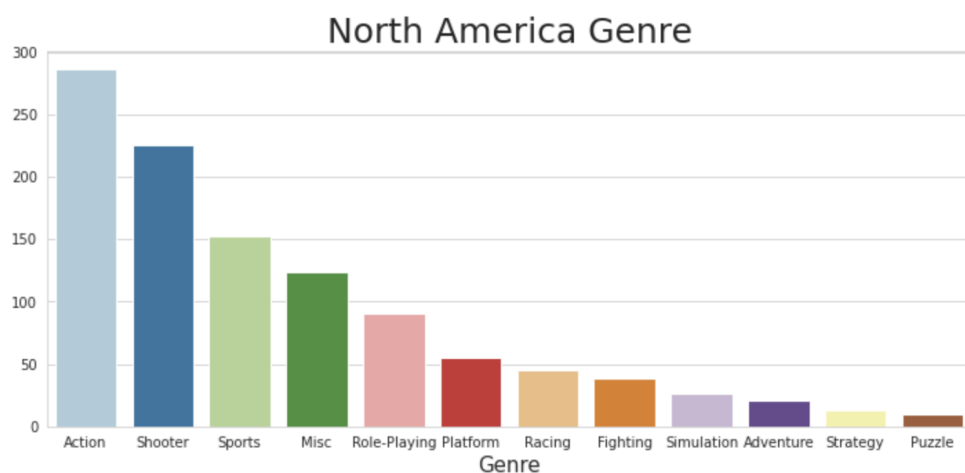
- 북미 지역

- 최근 10년간 게임 top 10



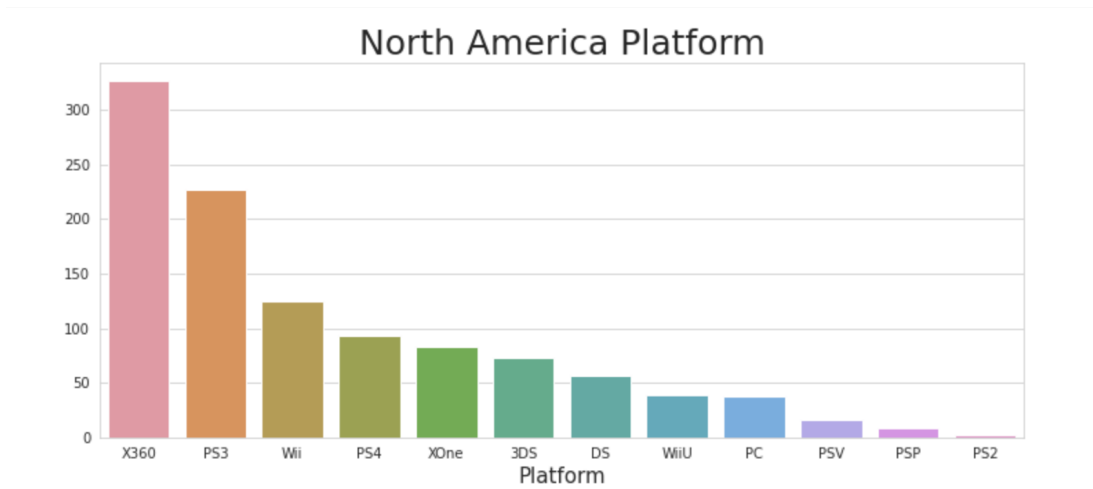
Top3: Grand Theft Auto V, call of Duty: Black Ops, Call of Duty: Modern Warfare 3

◦ 최근 10년간 장르 순위



Top3: Action, Shooter, Sports

◦ 최근 10년간 플랫폼 순위



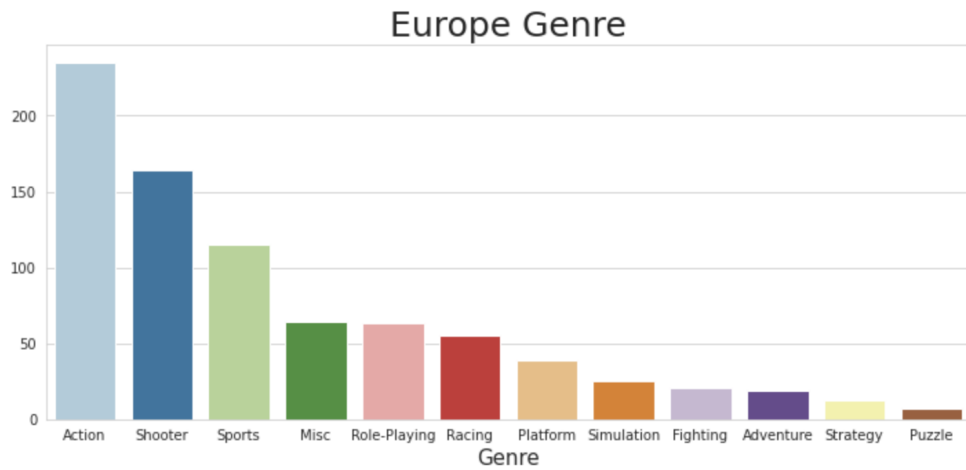
Top3: X360,PS3,Wii

- 유럽 지역
 - 최근 10년간 게임 top 10



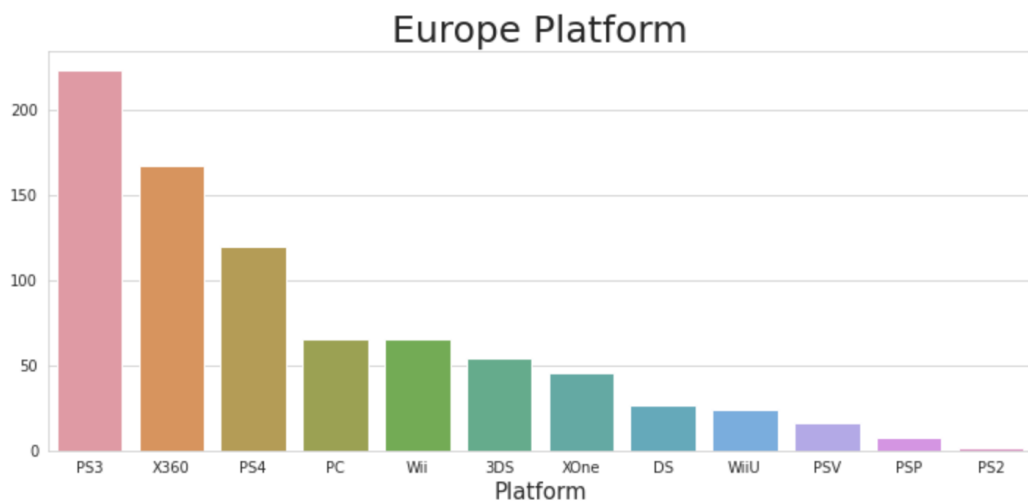
Top3: Grand Theft Auto V, FIFA 15, Call of Duty: Modern Warfare 3

- 최근 10년간 장르 순위



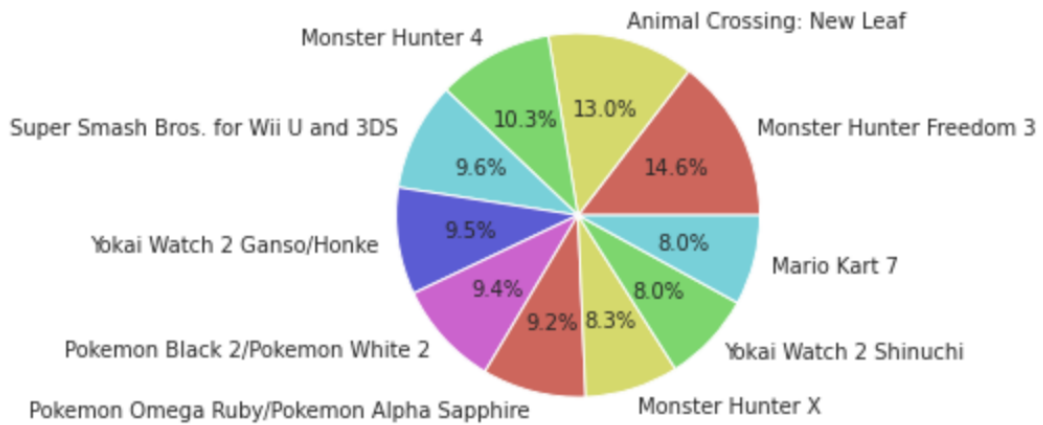
Top 3: Action, Shooter, Sports

- 최근 10년간 플랫폼 순위



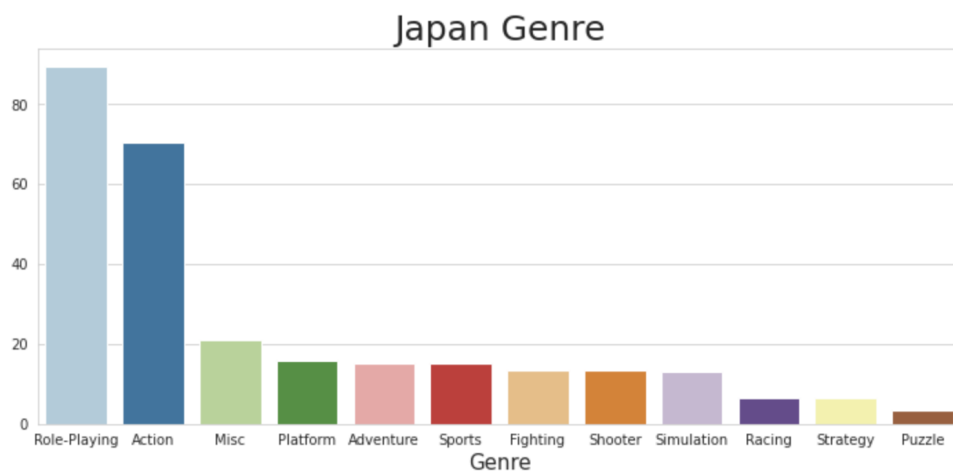
Top3: PS3, X360, PS4

- 일본 지역
 - 최근 10년간 게임 top 10



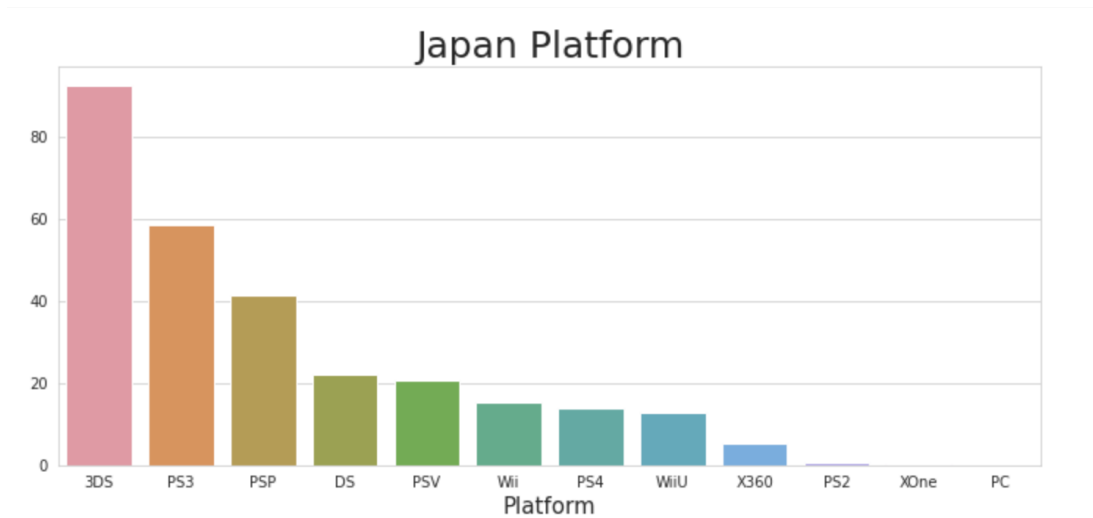
Top 3: Monster Hunter Freedom 3,Animal Crossing: New Leaf,Monster Hunter 4

- 최근 10년간 장르 순위



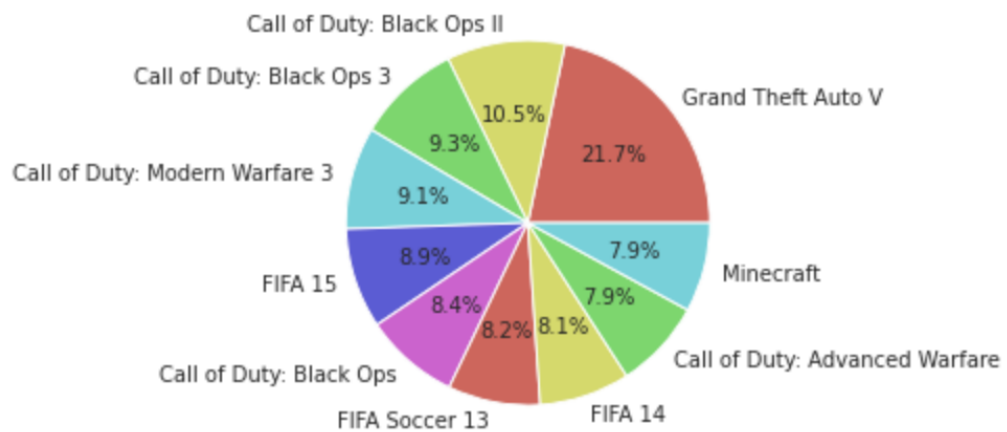
Top 3: Role-Playing,Action,Misc

- 최근 10년간 플랫폼 순위



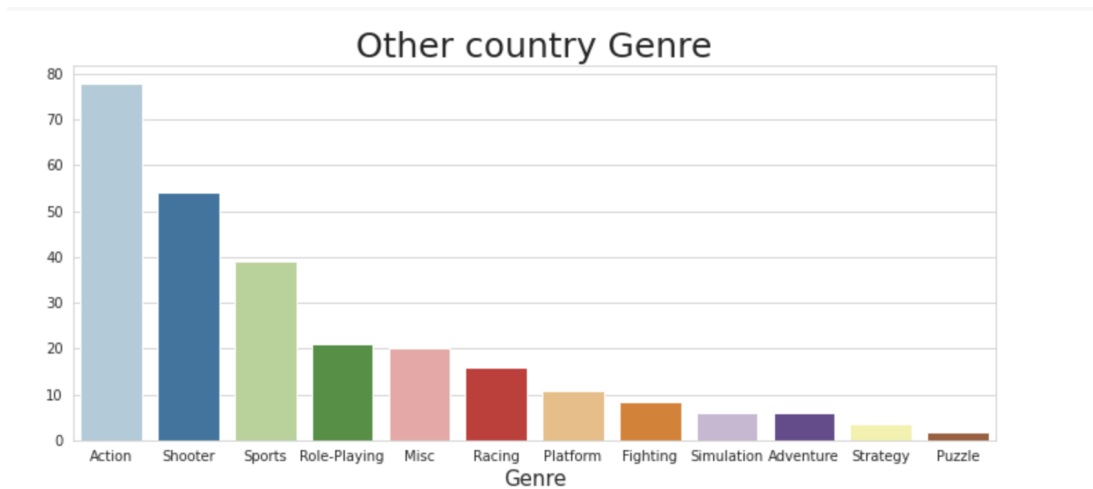
Top 3: 3DS,PS3,PSP

- 이 외 지역
 - 최근 10년간 게임 top 10



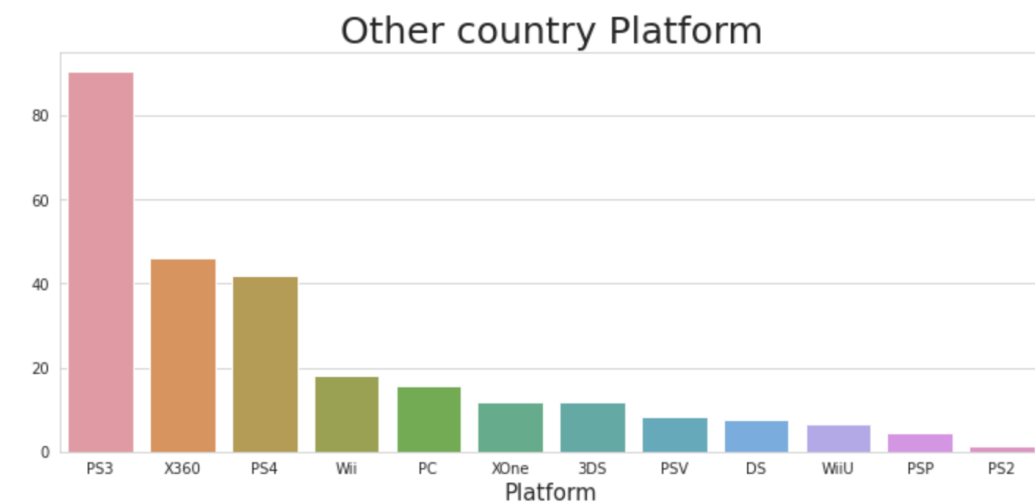
Top 3: Grand Theft Auto V, Call of Duty: Black Ops II, Call of Duty: Black Ops 3

- 최근 10년간 장르 순위



Top 3: Action,Shooter,Sports

- 최근 10년간 플랫폼 순위



Top 3: PS3,X360,PS4

결론

최근 10년간 게임 트렌드 파악

연도별 게임의 트렌드가 있을까 ? 라는 가설 검정에서 게임의 트렌드가 5-10 주기 안으로 변동이 큰걸로 관찰 되기 때문에, 최근 10년의 데이터를 조합하여 게임의 트렌드를 파악 해야 한다!

- 최근 10년간 게임 top 10

Top 3 : Grand Theft Auto V, Call of Duty: Modern Warfare, Call of duty: Black Ops ||

- 최근 10년간 장르 순위

Top 3 : Action, Shooter, Sports

- 최근 10년간 플랫폼 순위

Top3: PS3, X360,PS4

최근 10년간의 트렌드를 파악 하자면, 장르는 Action, Shooter가 선호가 높고 플랫폼으로는 PS3가 트렌드 입니다

각 지역별 게임 트렌드 파악

지역에 따라서 선호하는 게임 장르가 다를까?라는 가설 검정에서 지역과 장르는 독립적이지 않은것으로 보였기 때문에 지역 별 게임 트렌드를 파악 해야 한다!

<북미>

- 최근 10년간 게임 top 10

Top3: Grand Theft Auto V, call of Duty: Black Ops, Call of Duty: Modern Warfare 3

- 최근 10년간 장르 순위

Top3: Action,Shooter,Sports

- 최근 10년간 플랫폼 순위

Top3: X360,PS3,Wii

북미 지역의 장르 트렌드는 ‘Action’,’Shooter’ 이고, 플랫폼은 ‘X360’이 선호가 높습니다

<유럽>

- 최근 10년간 게임 top 10

Top3: Grand Theft Auto V, FIFA 15, Call od Duty: Modern Warfare 3

- 최근 10년간 장르 순위

Top3: Action, Shooter, Sports

- 최근 10년간 플랫폼 순위

Top3: PS3, X360, PS4

유럽 지역의 장르 트렌드는 'Action', 'Shooter' 이고, 플랫폼은 'PS3' 이 선호가 높습니다

<일본>

- 최근 10년간 게임 top 10

Top3: Monster Hunter Freedom 3, Animal Crossing: New Leaf, Monster Hunter 4

- 최근 10년간 장르 순위

Top3: Role-Playing, Action, Misc

- 최근 10년간 플랫폼 순위

Top3: 3DS, PS3, PSP

일본 지역의 장르 트렌드는 'Role-Playing', 'Action' 이고, 플랫폼은 '3DS'가 선호가 높습니다

<이 외 지역>

- 최근 10년간 게임 top 10

Top3: Grand Theft Auto V, Call of Duty: Black Ops II, Call of Duty: Black Ops 3

- 최근 10년간 장르 순위

Top3: Action, Shooter, Sports

- 최근 10년간 플랫폼 순위

Top3: PS3, X360, PS4

이 외 지역의 장르 트렌드는 'Action', 'Shooter' 이고, 플랫폼은 'PS3'가 선호가 높습니다

< 다음 분기에 어떤 게임을 설계 해야 할까? >

- 전체 지역 출시용으로는 전 지역에서 최근 10년내에 선호도가 높았던 장르와 플랫폼을 고려하여 장르는 '**Action**' 과 플랫폼은 '**PS3**'으로 출시 해야 판매율이 높을 것으로 예상
- **북미** 지역 출시용으로는 전 지역에서 최근 10년내에 선호도가 높았던 장르와 플랫폼을 고려하여 장르는 '**Action**', '**Shooter**' 과 플랫폼은 '**X360**', '**PS3**'으로 출시 해야 판매율이 높을 것으로 예상
- **유럽** 지역 출시용으로는 전 지역에서 최근 10년내에 선호도가 높았던 장르와 플랫폼을 고려하여 장르는 '**Action**', '**Shooter**' 과 플랫폼은 '**PS3**', '**X360**'으로 출시 해야 판매율이 높을 것으로 예상
- **일본** 지역 출시용으로는 전 지역에서 최근 10년내에 선호도가 높았던 장르와 플랫폼을 고려하여 장르는 '**Role-Playing**', '**Action**' 과 플랫폼은 '**3DS**', '**PS3**'으로 출시 해야 판매율이 높을 것으로 예상