

Game Data Analysis

2021

START →

AI Bootcamp 1st Project

2021.10.12.TUE

Codestates AI Bootcamp 08th 정유경

00 INDEX

01

들어가기 전에

분석의 전체적인 과정과
최종적인 목표를 소개합니다.

02

데이터 전처리

데이터 세트를 확인하고,
전처리를 진행한 과정을 설명합니다.

03

지역별 선호 장르?

북미, 일본, 유럽 등 지역에 따른
장르 선호도 차이가 존재할지 탐색합니다.

04

연도별 게임 트렌드?

1980년대부터 2010년대까지,
시대별 게임의 트렌드는 어떠했을까요?

05

많이 팔리는 타이틀?

출고량 상위 10개 타이틀,
공통적으로 어떤 특징이 있었을까요?

06

결론

다음 분기에 우리는
어떤 게임을 설계해야 할까요?

07

한계

본 분석의 한계를 언급하고
보완의 가능성을 제시합니다.

01 들어가기 전에

분석의 전체적인 과정과
최종적인 목표를 소개합니다.

게임 데이터, 분석하는 과정은?

1. 데이터 확인
2. 데이터 전처리
3. 가설 설정
4. 가설 검증
5. 결론 도출

데이터 분석, 무엇을 위해서?

- ? 지역에 따라서 선호하는 게임 장르가 다를까?
- ? 연도별 게임의 트렌드가 있을까?
- ? 출고량이 높은 게임은 어떤 특징을 가질까?
-

→ 다음 분기에는

어떤 게임을 만들어야 할까?



02 데이터 전처리

데이터 세트를 확인하고,
전처리를 진행한 과정을 설명합니다.

데이터 로드, 데이터 개형 확인

데이터의 전체적인
형태를 확인합니다.

데이터 클렌징, 피처 추가 등 전처리

본격적인 분석에 앞서
데이터를 다듬고,
필요한 피처를 추가합니다.



데이터 로드, 데이터 개형 확인

데이터의 전체 형태를 파악합니다.

컬럼별 결측치 확인, 카테고리 확인

빠져 있는 값은 무엇인지,
장르/플랫폼의 종류는 어떤지
확인해 봅니다.

데이터 로드 및 개형 확인

```
[5] #데이터 로드
df = pd.read_csv('vgames2.csv')
df.shape
df.head()
```

총 16598개 데이터

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

```
[4] #컬럼별 결측치의 수 확인:
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

Year 발매연도: 271개 결측치 존재

Genre 장르: 50개 결측치 존재

Publisher 유통사: 58개 결측치 존재

```
[20] df['Genre'].unique()

array(['Action', 'Adventure', 'Misc', 'Platform', 'Sports', 'Simulation',
       'Racing', 'Role-Playing', 'Puzzle', 'Strategy', 'Fighting',
       'Shooter', nan], dtype=object)
```

```
[25] df['Platform'].unique()

array(['DS', 'Wii', 'PSP', 'PS3', 'PC', 'PS', 'GBA', 'PS4', 'PS2', 'XB',
       'X360', 'GC', '3DS', '2600', 'SAT', 'GB', 'NES', 'DC', 'N64',
       'XOne', 'SNES', 'WiiU', 'PSV', 'GEN', 'SCD', 'WS', 'NG', 'TG16',
       '3DO', 'GG', 'PCFX'], dtype=object)
```

데이터 확인

Features: 변수

01	Name	타이틀명	게임의 타이틀(영문명 기준)
02	Platform	플랫폼	타이틀의 발매 기종(PC 포함 총 35종류로 구분)
03	Year	발매연도	타이틀 발매 연도(1980~2020)
04	Genre	장르	게임의 장르 (Action, Adventure, Misc, Platform, Sports, Simulation, Racing, Role-Playing, Puzzle, Strategy, Fighting, Shooter의 12종류로 구분)
05	Sales	판매고	NA_Sales 북미에서 해당 타이틀이 판매된 수 JP_Sales 일본에서 해당 타이틀이 판매된 수 EU_Sales 유럽에서 해당 타이틀이 판매된 수 Other_Sales 북미, 일본, 유럽 외의 지역에서 해당 타이틀이 판매된 수

피처 추가 1: 전세계 판매량 컬럼

지역별 트렌드와 전세계의 트렌드를
비교분석하기 위해
'total_Sales' 컬럼을 추가합니다.

1. total_Sales: 전세계 판매량 컬럼을 추가하기 위해 컬럼값을 합산

```
[15] df['total_Sales'] = df['NA_Sales']+df['EU_Sales']+df['JP_Sales']+df['Other_Sales']  
df.sort_values('total_Sales', ascending=False)
```

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	total_Sales
2908	Wii Sports	Wii	2006.0	Sports	Nintendo	41490000	29020000	3770000	8460000	82740000
4199	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29080000	3580000	6810000	770000	40240000
12579	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15850000	12880000	3790000	3310000	35830000
3352	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15750000	11010000	3280000	2960000	33000000
7829	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11270000	8890000	10220000	1000000	31380000
...
14265	Nitroplus Blasterz: Heroines Infinite Duel	PS3	2015.0	Fighting	Nitroplus	0	0	10000	0	10000
2311	Shinseiki Evangelion: Koutetsu no Girlfriend 2...	PSP	2009.0	Adventure	CyberFront	0	0	10000	0	10000
11376	Rail Simulator	PC	2007.0	Simulation	Electronic Arts	0	10000	0	0	10000
3571	Hyperdimension Neptunia mk2	PS3	2011.0	Action	Nippon Ichi Software	0	10000	0	0	10000
5598	SCORE International Baja 1000: The Official Game	PS2	2008.0	Racing	Activision	0	0	0	0	0

16598 rows x 10 columns

피처 추가 2: '콘솔' 컬럼

플랫폼을 거치형(가정용) 콘솔과
휴대용 콘솔, PC로 구분합니다.

2. Consoles: Platform 컬럼의 값을 가정용 콘솔(housing), 휴대형 콘솔(Handheld), PC의 세 가지로 그룹핑.

```
[37] #플랫폼 그룹핑 함수 작성
```

```
def cons(x) :  
    if x == 'PC' :  
        return x  
    elif x == 'GBA' or x == 'PSP' or x == '3DS' or x == 'DS' or x == 'PSV' or x == '3DS' or x == 'GB' or x == 'GG' or x == 'WS' :  
        return "Handheld"  
    else :  
        return "Housing"  
  
df["Consoles"] = df["Platform"].apply(lambda x : cons(x))  
df
```

	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	total_Sales	Consoles
0	Candace Kane's Candy Factory	DS	2008.0	Action	Destineer	40000	0	0	0	40000	Handheld
1	The Munchables	Wii	2009.0	Action	Namco Bandai Games	170000	0	0	10000	180000	Housing
2	Otome wa Oanesama Boku ni Koi Shiteru Portable	PSP	2010.0	Adventure	Alchemist	0	0	20000	0	20000	Handheld
3	Deal or No Deal: Special Edition	DS	2010.0	Misc	Zoo Games	40000	0	0	0	40000	Handheld

03 지역별 선호장르

북미, 일본, 유럽 등 지역에 따른
장르 선호도 차이가 존재할지 탐색합니다.

EUROPE

JAPAN

NORTH AMERICA

? 지역에 따라
선호하는 게임 장르가 다를까?

? 전세계 판매량 추이와
가장 유사한 판매량 추이를 보이는
지역은 어디일까?

지역에 따라 선호하는 게임 장르가 다른가?

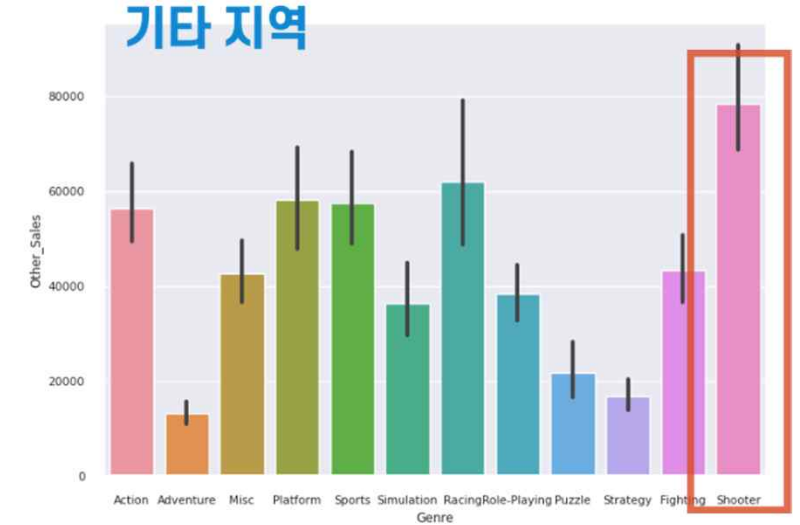
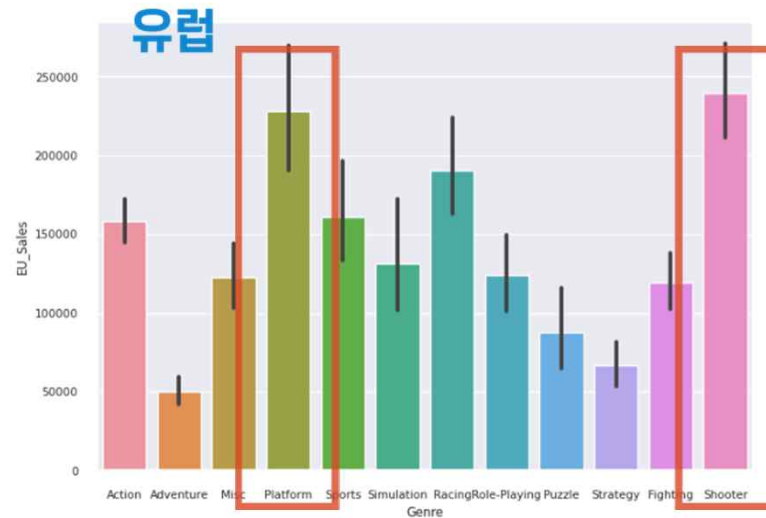
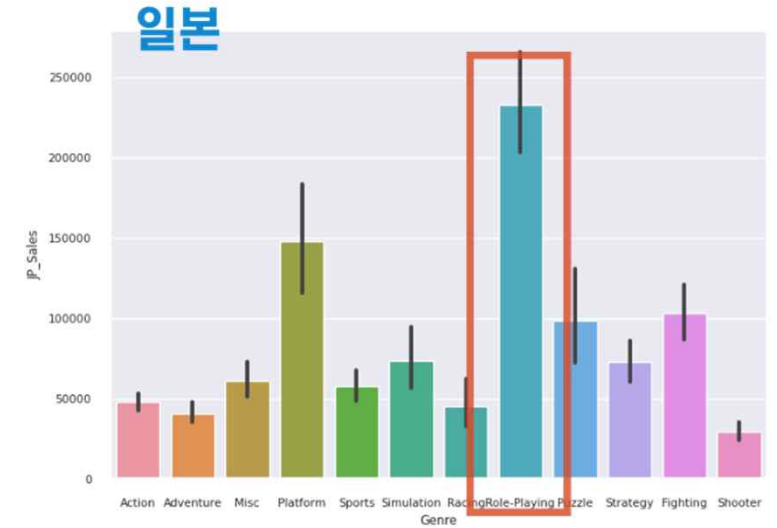
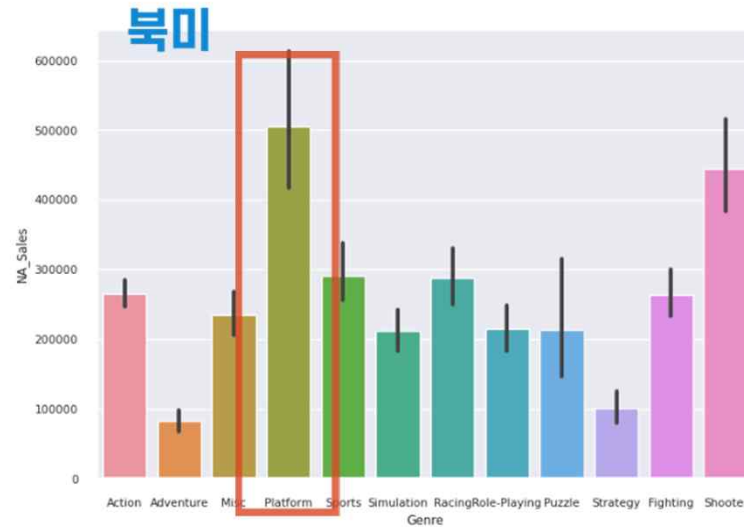
북미, 일본, 유럽, 그밖의 지역
네 곳에서의 타이틀 출고량을
장르별로 비교한 결과...

북미: 플랫폼(점프)

일본: 롤 플레이

유럽: 플랫폼, 슈팅

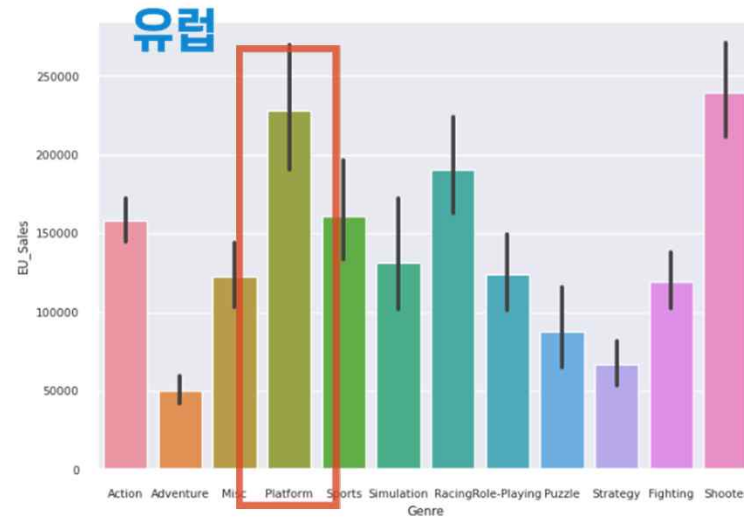
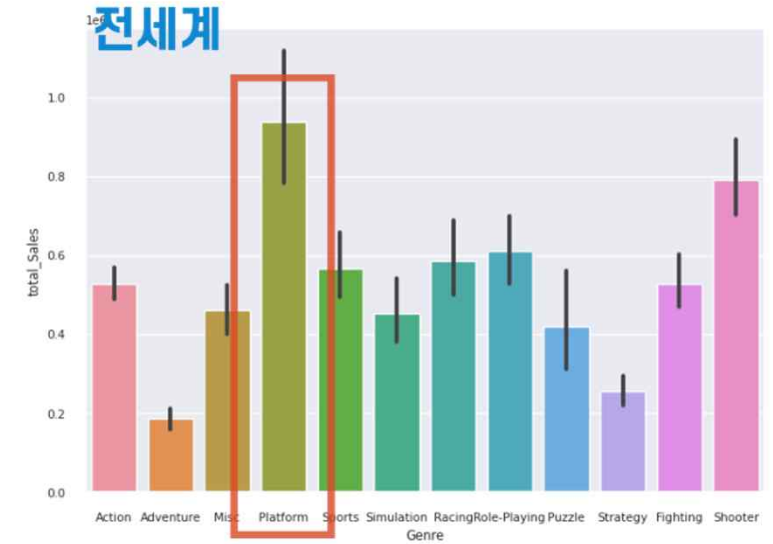
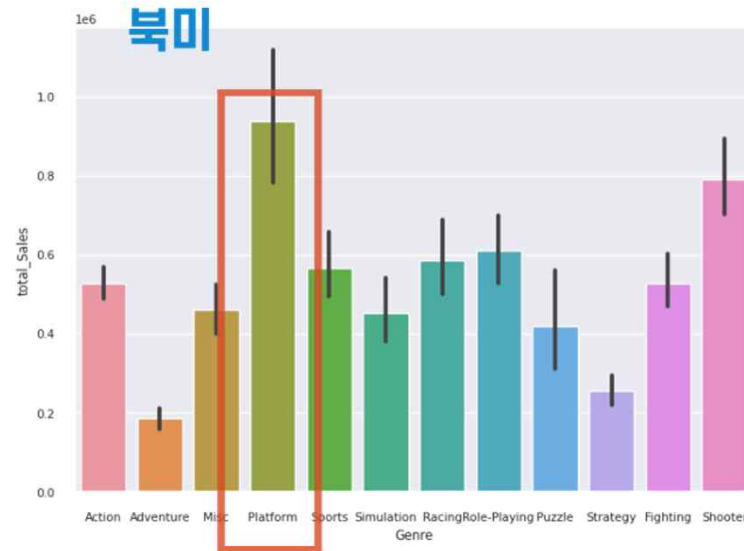
기타 지역: 슈팅



전세계 판매량 패턴과 가장 유사한 판매량 패턴을 보이는 지역은 어디일까?

북미, 일본, 유럽, 그밖의 지역
네 곳에서의 타이틀 출고량과
전세계 타이틀 출고량 간의
상관관계를 비교한 결과...

미국, 유럽의 판매량 패턴과
전세계 판매량 패턴의 상관계수가
각각 0.94, 0.90으로,
높은 상관관계 존재



df.corr()

	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	total_Sales
Year	1.000000	-0.091402	0.006014	-0.169316	0.041058	-0.074714
NA_Sales	-0.091402	1.000000	0.767727	0.449787	0.634737	0.941050
EU_Sales	0.006014	0.767727	1.000000	0.435584	0.726385	0.902839
JP_Sales	-0.169316	0.449787	0.435584	1.000000	0.290186	0.611821
Other_Sales	0.041058	0.634737	0.726385	0.290186	1.000000	0.748351
total_Sales	-0.074714	0.941050	0.902839	0.611821	0.748351	1.000000

04 연도별 게임 트렌드

1980년대부터 2010년대까지,
시대별 게임의 트렌드는 어떠했을까요?

시대와 장르 간의 상관관계

시대별 인기 장르는 무엇인지,
게이머들에게 꾸준히 사랑받는
스테디셀러는 무엇인지 알아봅니다.

시대, 장르, 플랫폼 간의 상관관계

1977년부터 1992년까지
시대를 풍미한 '아타리2600'부터
최근의 휴대기인 DS, 거치기 Wii까지.
콘솔 게임 플랫폼의 트렌드는
어떻게 변해왔을까요?

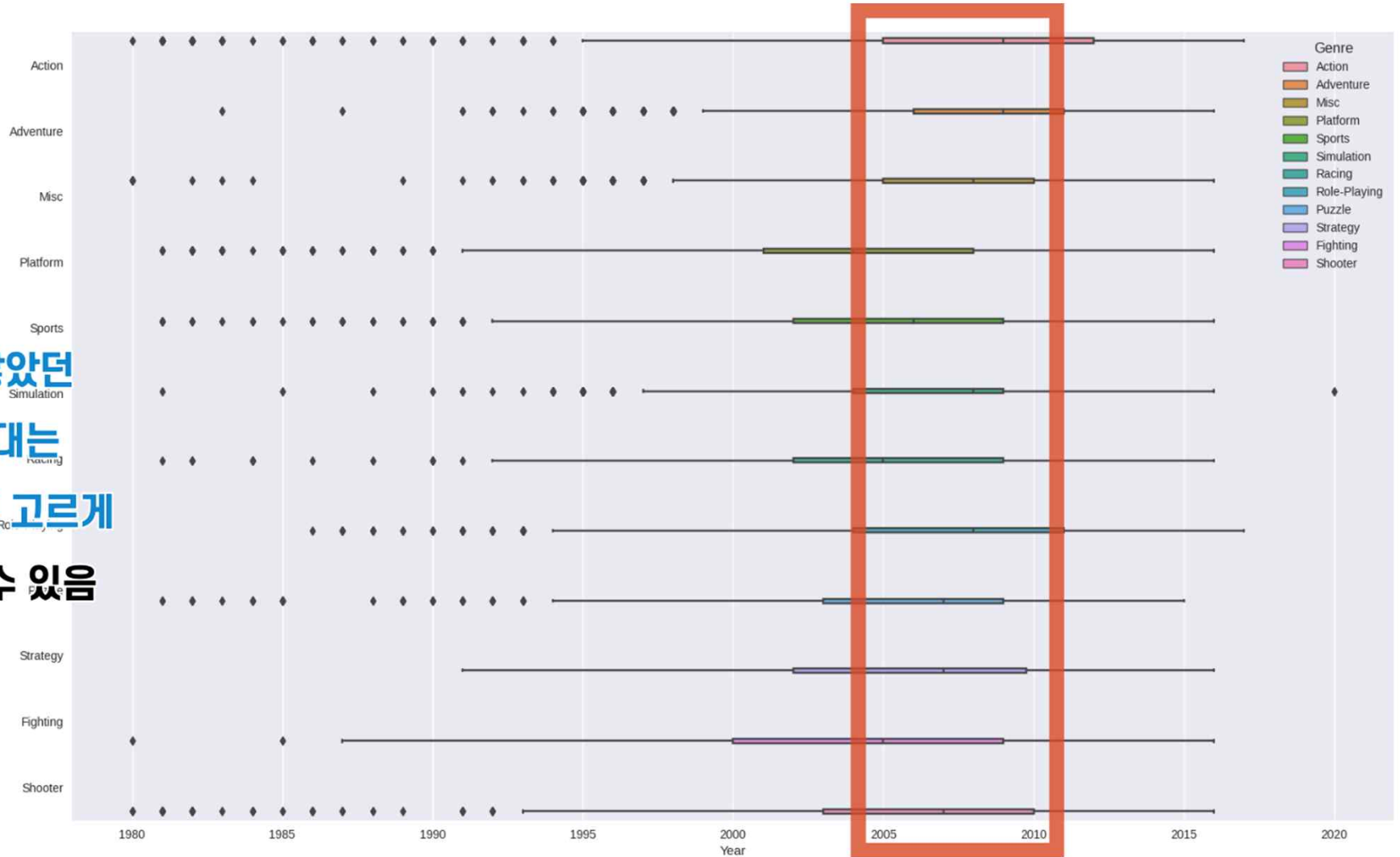


history of video games

인기 게임 장르는 시대에 따라 변할까?

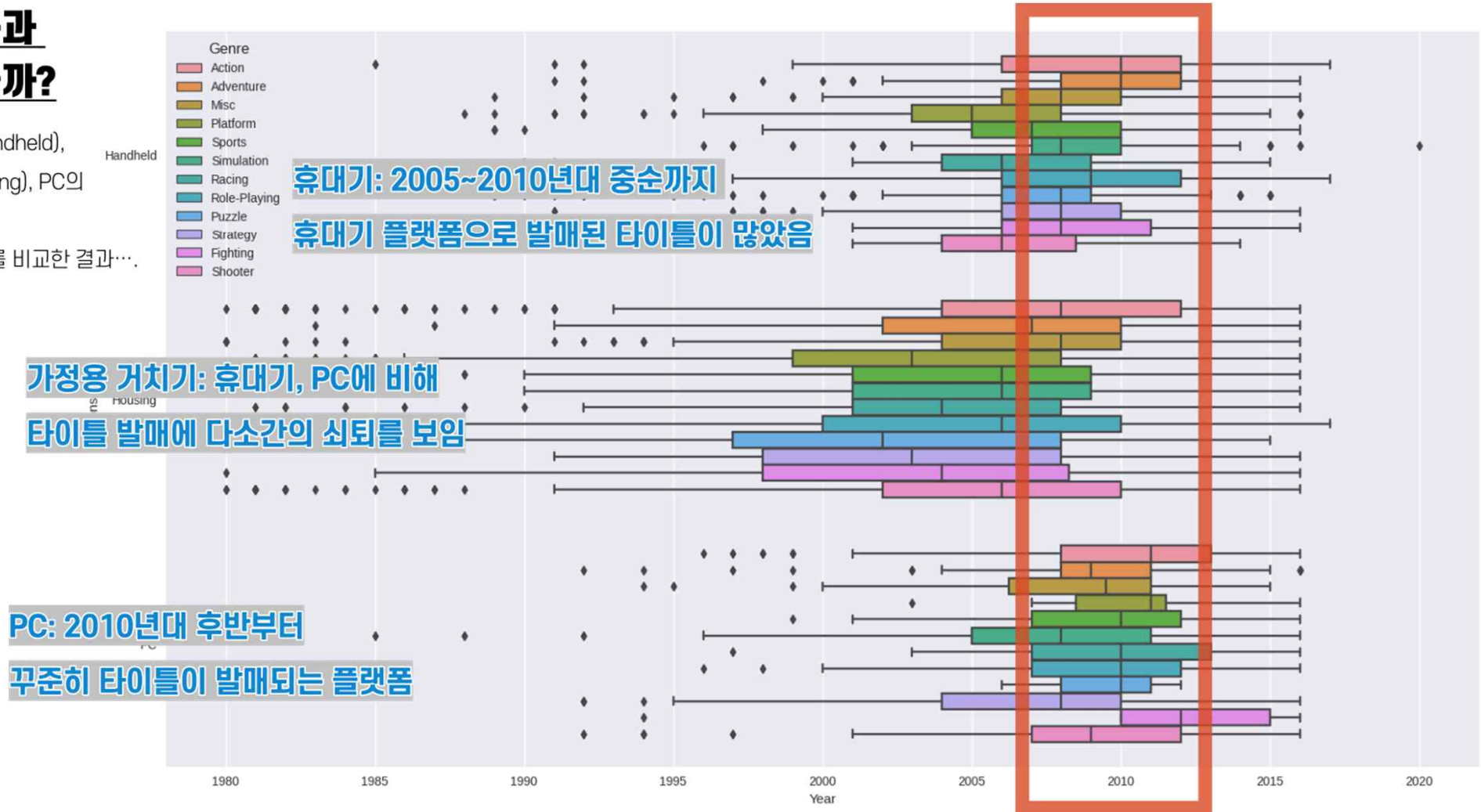
북미, 일본, 유럽, 그밖의 지역
네 곳에서의 타이틀 출고량과
전세계 타이틀 출고량 간의
상관관계를 비교한 결과...

데이터가 가장 많았던
2005~2010년대는
모든 장르가 비교적 고르게
사랑받았음을 알 수 있음



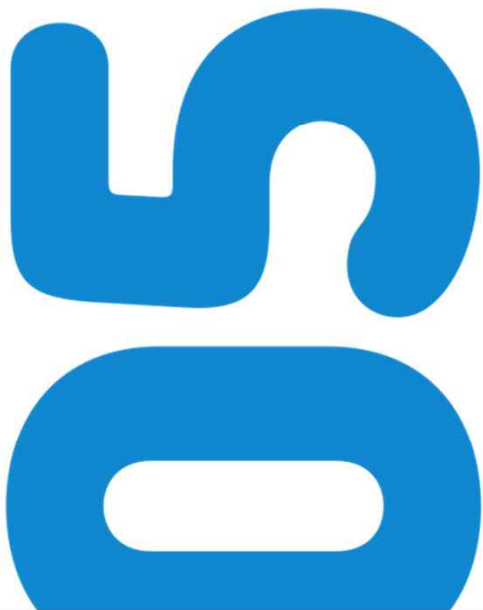
게임의 트렌드는 발매 플랫폼과 관련이 있을까?

플랫폼을 휴대기(handheld),
가정용 거치기(housing), PC의
세 가지로 그룹화해
발매된 타이틀의 수를 비교한 결과...



05 많이 팔리는 타이틀

출고량 상위 10개 타이틀,
공통적으로 어떤 특징이 있었을까요?



데이터 분석

최근 10년 최다 판매고 타이틀

Grand Theft Auto V (PS3, 2013) Action, 청소년 이용불가 등급의 범죄 액션물	21,389,999
Kinect Adventures! (X360, 2010) Misc, XBOX360·XBOXONE 주변기기의 번들 소프트웨어	21,820,000
Wii Sports Resort (WII, 2009) Sports, Wii 번들 소프트웨어(일본 외), Wii Sports의 후속작	33,000,000

전 기간 최다 판매고 타이틀

Wii Sports (WII, 2006) Sports, Wii 번들 소프트웨어(일본 외), 모션 인식 활용	82,740,000
Super Mario Bros. (NES, 1985) Platform, 슈퍼 마리오 시리즈의 첫 작품	40,240,000
Mario Kart Wii (WII, 2008) Racing, 마리오 카트 시리즈의 Wii 버전	35,830,000

전세계 각지에서는 어떤 게임들이 많이 팔려왔을까?

북미, 일본, 유럽, 그밖의 지역
네 곳에서의 역대 판매량 10위권
타이틀을 비교한 결과...

북미

샌드박스형, 퍼즐형 타이틀이 인기
수명이 긴 스테디셀러형 작품들이 다수
거치기로 출시된 타이틀이 많음

[65] #1. 지역별 출고량 상위 10작품을 확인
df.sort_values('NA_Sales', ascending=False).head(10)

	Name	Platform	Year	Genre	Publisher	NA_Sales	Consoles
2908	Wii Sports	Wii	2006.0	Sports	Nintendo	41490000	Housing
4199	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29080000	Housing
10488	Duck Hunt	NES	1984.0	Shooter	Nintendo	26930000	Housing
1203	Tetris	GB	1989.0	Puzzle	Nintendo	23200000	Handheld
12579	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15850000	Housing
3352	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15750000	Housing
4343	Kinect Adventures!	X360	2010.0	Misc	Microsoft Game Studios	14970000	Housing
16531	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14590000	Housing
3014	Wii Play	Wii	2006.0	Misc	Nintendo	14030000	Housing
13111	Super Mario World	SNES	1990.0	Platform	Nintendo	12780000	Housing

[64] df.sort_values('JP_Sales', ascending=False).head(10)

	Name	Platform	Year	Genre	Publisher	JP_Sales	Consoles
7829	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	10220000	Handheld
9416	Pokemon Gold/Pokemon Silver	GB	1999.0	Role-Playing	Nintendo	7200000	Handheld
4199	Super Mario Bros.	NES	1985.0	Platform	Nintendo	6810000	Housing
6937	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	6500000	Handheld
10555	Pokemon Diamond/Pokemon Pearl	DS	2006.0	Role-Playing	Nintendo	6040000	Handheld
10621	Pokemon Black/Pokemon White	DS	2010.0	Role-Playing	Nintendo	5650000	Handheld
10917	Pokemon Ruby/Pokemon Sapphire	GBA	2002.0	Role-Playing	Nintendo	5380000	Handheld
3273	Animal Crossing: Wild World	DS	2005.0	Simulation	Nintendo	5330000	Handheld
511	Brain Age 2: More Training in Minutes a Day	DS	2005.0	Puzzle	Nintendo	5320000	Handheld
1847	Monster Hunter Freedom 3	PSP	2010.0	Role-Playing	Capcom	4870000	Handheld

일본

포켓몬스터 시리즈의 인기
그 외에도 동물의 숲, 슈퍼마리오 등
닌텐도의 퍼스트 파티 타이틀이 압도적 비중
휴대기로 출시된 타이틀이 많음


```
df.sort_values('EU_Sales', ascending=False).head(10)
```

유럽

복미와 유사한 패턴

갑작스러운(?) 닌텐독스의 등장

거치기와 휴대기의 비중이 비슷함

	Name	Platform	Year	Genre	Publisher	EU_Sales	Consoles
2908	Wii Sports	Wii	2006.0	Sports	Nintendo	29020000	Housing
12579	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	12880000	Housing
3352	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	11010000	Housing
11137	Nintendogs	DS	2005.0	Simulation	Nintendo	11000000	Handheld
3483	Grand Theft Auto V	PS3	2013.0	Action	Take-Two Interactive	9270000	Housing
13229	Brain Age: Train Your Brain in Minutes a Day	DS	2005.0	Misc	Nintendo	9260000	Handheld
6937	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	9230000	Handheld
3014	Wii Play	Wii	2006.0	Misc	Nintendo	9200000	Housing
7829	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	8890000	Handheld
10899	Wii Fit Plus	Wii	2009.0	Sports	Nintendo	8590000	Housing

```
df.sort_values('Other_Sales', ascending=False).head(10)
```

기타 지역

유럽과 유사한 패턴을 보임

거치기 타이틀의 비중이 높음

타 국가에 비해 상대적으로 판매량이 낮음

	Name	Platform	Year	Genre	Publisher	Other_Sales	Consoles
14639	Grand Theft Auto: San Andreas	PS2	2004.0	Action	Take-Two Interactive	10570000	Housing
2908	Wii Sports	Wii	2006.0	Sports	Nintendo	8460000	Housing
14635	Gran Turismo 4	PS2	2004.0	Racing	Sony Computer Entertainment	7530000	Housing
3483	Grand Theft Auto V	PS3	2013.0	Action	Take-Two Interactive	4139999	Housing
12579	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	3310000	Housing
3352	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	2960000	Housing
8213	Pro Evolution Soccer 2008	PS2	2007.0	Sports	Konami Digital Entertainment	2930000	Housing
6937	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	2900000	Handheld
3014	Wii Play	Wii	2006.0	Misc	Nintendo	2850000	Housing
11137	Nintendogs	DS	2005.0	Simulation	Nintendo	2750000	Handheld

PLATFORM:

PC,
or multi-platform

PC 타이틀의 판매고 성장세를 고려하여
PC / 닌텐도 서브파티 멀티 플랫폼으로의 개발 추진
스팀 덱(Steam Deck) 발매 이슈도 참고



GENRE:

PLEASE ENTER
THE TITLE

어드벤처를 기반으로 한 복합 장르
샌드박스 요소/오픈월드 설계 도입
후속작으로의 연계 가능성 확보

07 한계

본 분석이 가지는 한계와
보완 가능성을 알아봅니다.

결측치 처리의 문제

- 발매연도 결측치: 처리하지 않았음
- 장르 결측치: 구글링을 통해 직접 분류했으나,
장르의 중복이 존재하는 상황

한정된 그룹핑

- 플랫폼을 가정용/휴대용/PC로만 나누어 분석
→ 콘솔 각각의 특징이나 발매 역사 등을
고려하지 못함

퍼블리셔 분석 미흡

유통사(publisher)와 게임의 특성 간
상관관계를 고려하지 않고
분석을 진행



THANK
Let's make a cool GAME!
YOU