

Section 1

Project

: Game Data

AI Bootcamp_16_황윤상

Part 0. 데이터 전처리

데이터 가정 / 결측치 처리 / Tidy data

■ 원본 데이터

```
<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       16540 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
```

■ 결측치 처리

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16360 entries, 0 to 16359
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Unnamed: 0      16360 non-null  int64
1   Name            16360 non-null  object
2   Platform        16360 non-null  object
3   Year            16360 non-null  float64
4   Genre           16360 non-null  object
5   Publisher       16360 non-null  object
6   NA_Sales        16360 non-null  object
7   EU_Sales        16360 non-null  object
8   JP_Sales        16360 non-null  object
9   Other_Sales     16360 non-null  object
dtypes: float64(1), int64(1), object(8)
memory usage: 1.2+ MB
```

■ 분석 데이터

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 65440 entries, 0 to 65439
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Unnamed: 0      65440 non-null  object
1   Name            65440 non-null  object
2   Platform        65440 non-null  object
3   Year            65440 non-null  int64
4   Genre           65440 non-null  object
5   Publisher       65440 non-null  object
6   Country_Sales   65440 non-null  object
7   Sales_Amount    65440 non-null  int64
dtypes: int64(2), object(6)
memory usage: 4.5+ MB
```

• 출고량 : 백만장(M) 을 기준으로 가정

- * 출고량 이상치는 특별히 성공한 게임으로 보아 제거하지 않음
(1천만장이상의 비중 0.02%)

• 연도 : 50이하의 숫자는 2000년대로, 50이상의 숫자는 1900년대로 가정

- * 게임 출시년도 확인결과 데이터는 모두 2016년 이전 게임임

• 결측치 : 게임명이 같은 게임은 동일한 게임으로 보고,

출시연도, 장르, 배급사를 작성함

=> 동일한 게임명이 없는 결측치에 대해서는 제거함 (전체대비 1.4%)

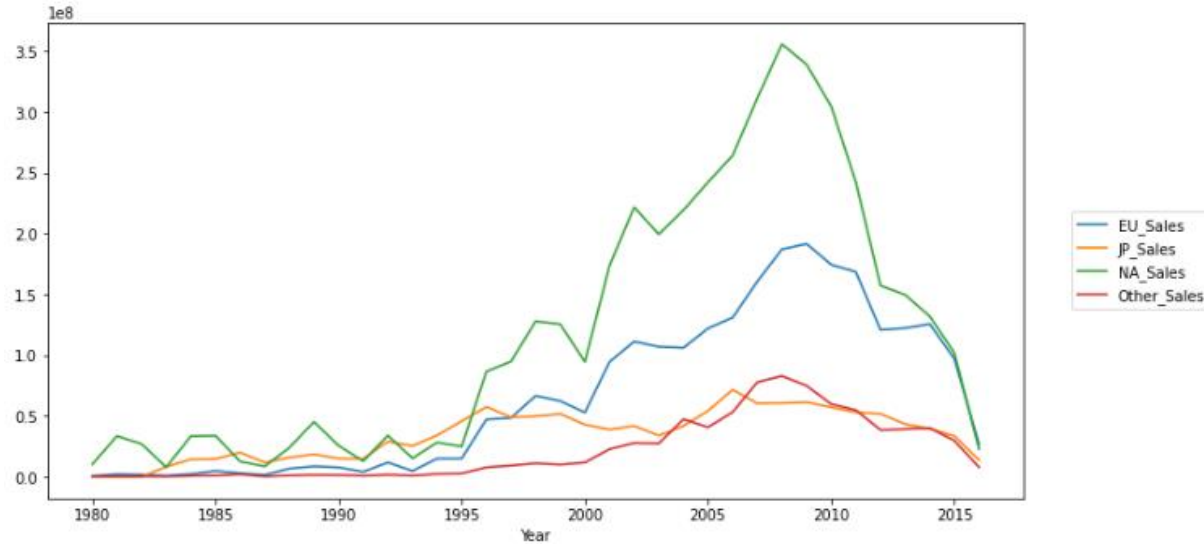
• 각 국가별 출고량을 하나의 출고량 컬럼으로 정리

- 각 국가에 대한 데이터를 별도의 컬럼으로 작성함
=> Tidy data로 작성

Part 1. 게임 판매 현황

시장규모 / 선호 장르 및 게임 플랫폼 / 지역별 선호 장르

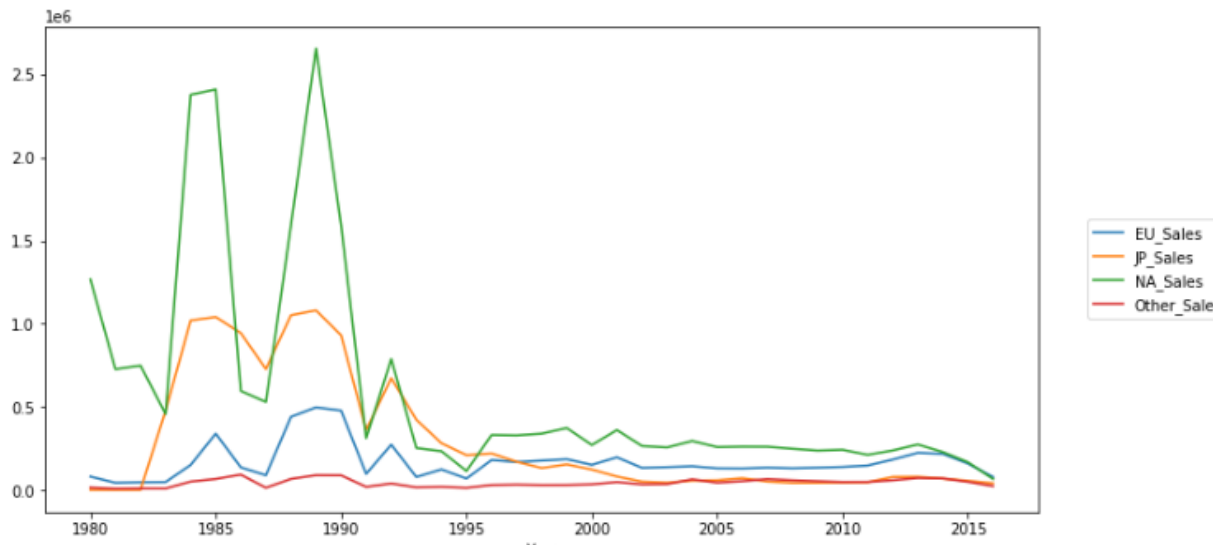
총
출고
량
(정)



■ 지역별 게임 출고량 추이

- 1995년 이후, 게임 전체 출고량은 크게 증가함
- 지역별 게임 시장의 규모는 북미가 가장 크고, 그 다음이 유럽, 일본 순임
- 3개의 시장을 제외한 기타 국가의 시장은 일본시장과 유사한 규모임

게임
당
출고
량
(정)

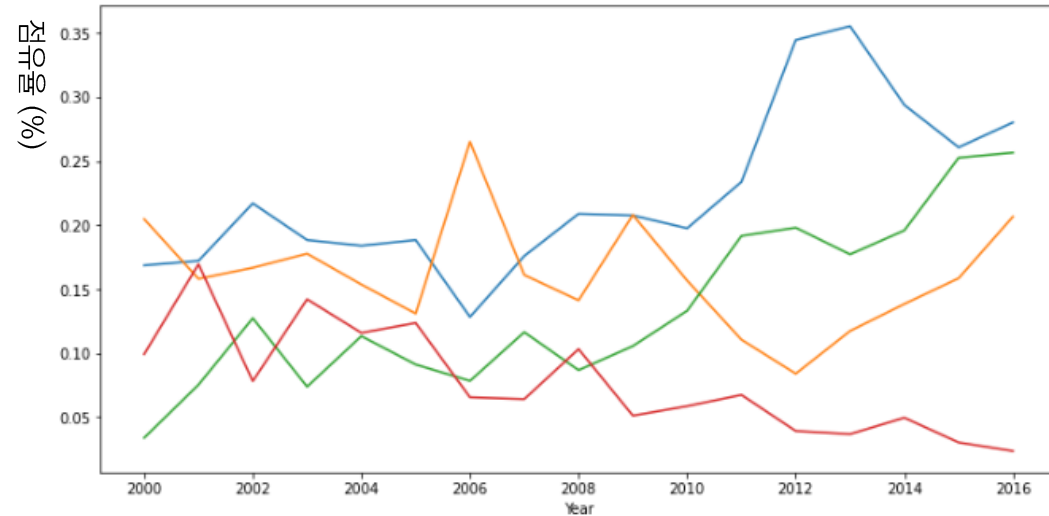


■ 1개의 게임당 출고량 추이

- 게임 1개당 출고량은 1995년 이후에도 큰 변화 없음
- 이를 통해 1995년 이후, 다수의 게임이 출시되어 전체 게임시장의 규모가 증가한 것을 알 수 있음
- 2000년대 이전의 시장은 게임 판매량이 충분하지 않아, 이후 현황분석시에는 제외함

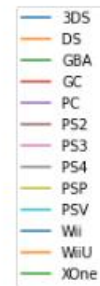
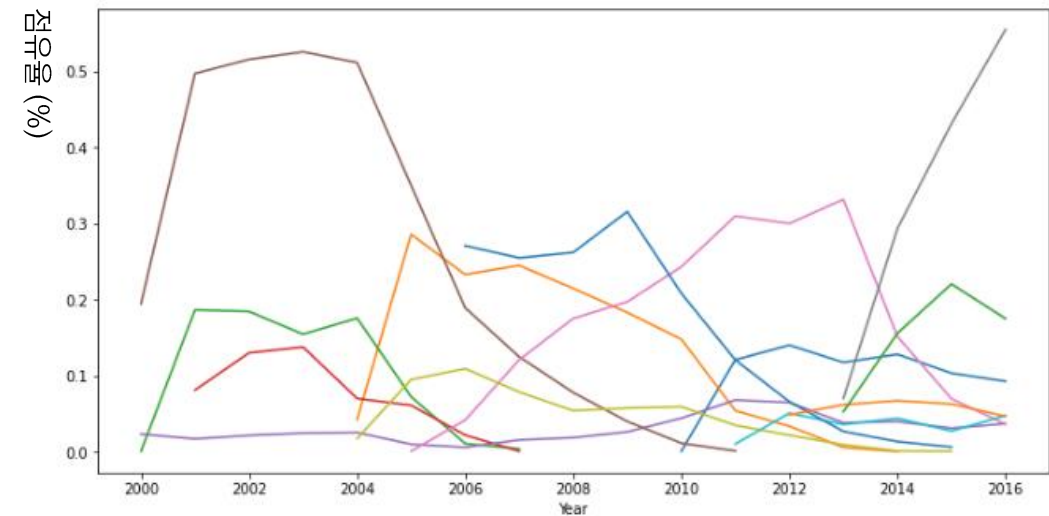
Part 1. 게임 판매 현황

시장규모 / 선호 장르 및 게임 플랫폼 / 지역별 선호 장르



■ 선호하는 게임장르 변화 추이

- 2000년 이후 가장 인기 있는 게임은 Action 장르임
- Shooter장르의 게임은 시장점유율이 지속적으로 증가하여 2010년 이후 점유율 2순위를 기록함
- Sports 장르의 게임은 점유율이 10% 이상인 지속적으로 인기있는 게임임
- Racing 게임은 2000년대 초반에는 인기있었으나, 점점 인기가 줄어들고 있음

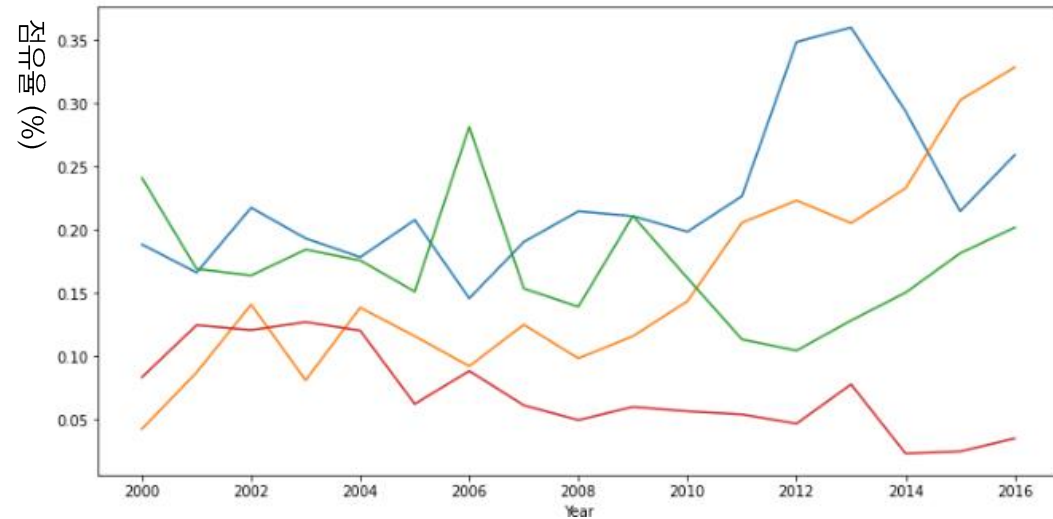


■ 게임플랫폼 당 출고량 추이

- PS플랫폼은 다음 버전이 출시되면 이후의 출시량이 줄어들며 세대가 교체됨
- PS2, PS4가 각 연도의 40%이상의 점유율을 보인데 반해, 2005년 출시된 PS3는 Wii 플랫폼과 DS플랫폼의 강세에 30%이하의 점유율을 보임

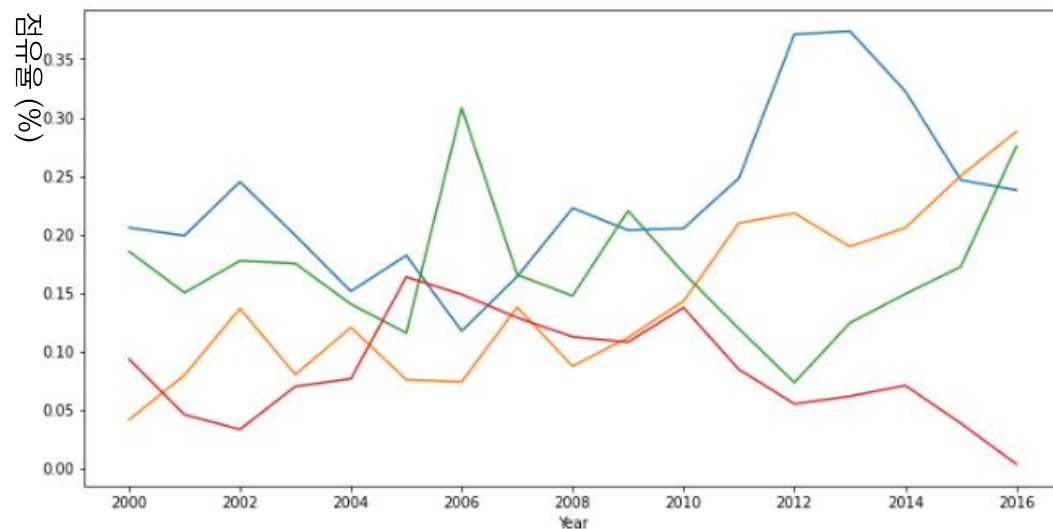
Part 1. 게임 판매 현황

시장규모 / 선호 장르 및 게임 플랫폼 / 지역별 선호 장르



■ 북미지역 선호 게임장르 변화 추이

- 지속적으로 Action 장르가 가장 높은 점유율을 보였으나, 2015년 Shooter 장르가 역전함
- Sports 장르는 점유율 3순위로 높은 인기가 있음
- 2000년대 초반에 인기있었던 Platform 장르의 인기는 줄어듦

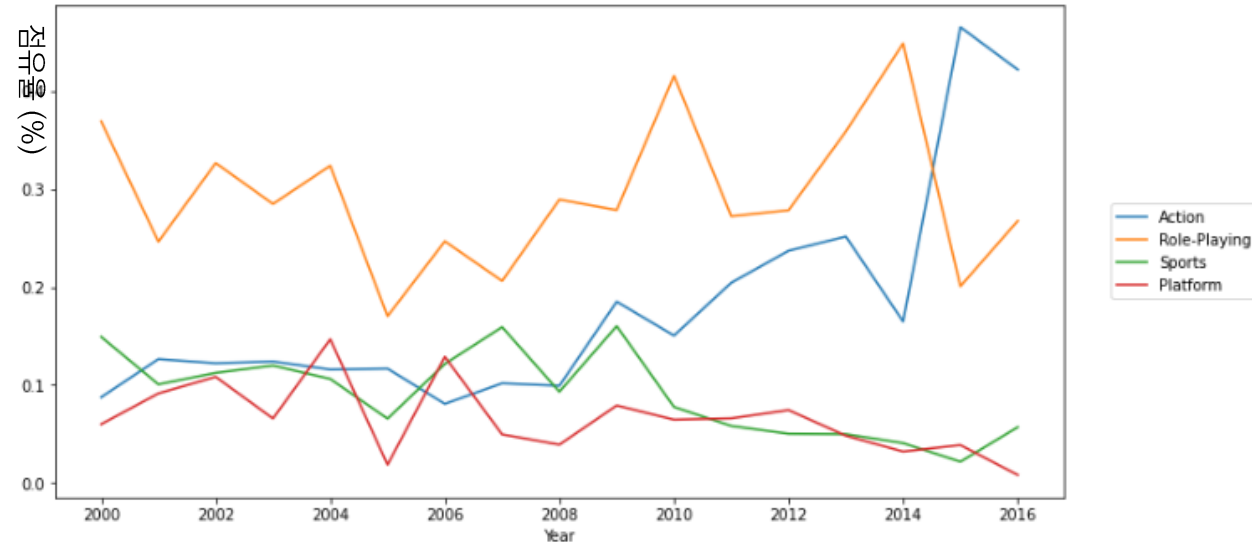


■ 유럽지역 선호 게임장르 변화 추이

- Action 장르, Shooter 장르, Sports 장르가 인기임
- 2016년 Shooter 장르, Sports 장르가 1,2순위임
- 타 지역에 비해 Sports 장르의 인기가 높음
- 2000년대 중반에 인기있었던 Misc 장르의 인기는 줄어듦

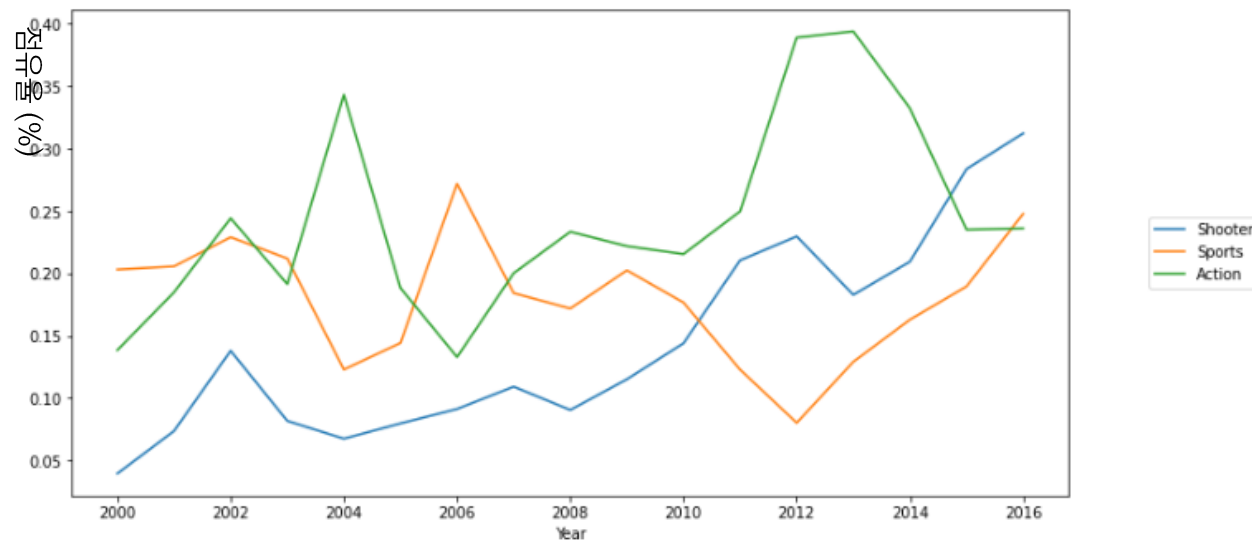
Part 1. 게임 판매 현황

시장규모 / 선호 장르 및 게임 플랫폼 / 지역별 선호 장르



■ 일본지역 선호 게임장르 변화 추이

- 지속적으로 Role-Playing 장르가 가장 높은 점유율을 보였으나, 2015년 Action 장르가 역전함
- Sports, Platform 장르의 인기는 줄어들고 있음



■ 기타지역 선호 게임장르 변화 추이

- Shooter 장르, Sports 장르가, Action 장르를 역전하여 2016년 점유율 1,2순위임
- 2000년대 초반 타 지역에 비해 Sports 장르의 인기가 높았음

Part 2. 게임 설계 방향

분석가정 및 가설설정 / 가설검정 / 게임 설계 방향 제시

■ 현황 분석 정리

- ▶ 전세계적으로 슈팅게임의 시장 점유율이
액션게임의 점유율을 역전하는 현상을 보이고 있음

■ 분석 가정

- ▶ 게임 출시로 분석하는 연도는 2017년으로 가정
 - 데이터는 모두 2016년 이전 데이터임. 그에 가장 최신 데이터를 모았다는 가정으로 분석 연도를 2017년으로 가정함
- ▶ 게임을 제작하는 회사는 북미에 위치하고 있는 것으로 가정
 - 현황 분석 결과, 지역별로 유행하는 장르는 상이함. 그에 지역을 가정함
 - 게임 시장이 가장 큰 북미시장을 가정하여 분석을 진행함

- ▶ 2017년도 하반기에 슈팅게임을 출시하는 게 적절한지에 대한 검증 진행
 - › 2016년도 각 슈팅게임의 출고량이 다른 게임들의 출고량보다 높았는지를 통해 검증 진행

- 귀무가설 : 북미지역에서 2016년 발매된 슈팅게임의 게임당 출고량은 다른 게임들의 게임당 출고량보다 작거나 같다.
- 대립가설 : 북미지역에서 2016년 발매된 슈팅게임의 게임당 출고량은 다른 게임들의 게임당 출고량보다 크다.

* 2016년 북미에서 발매된 게임 개수 : 345개

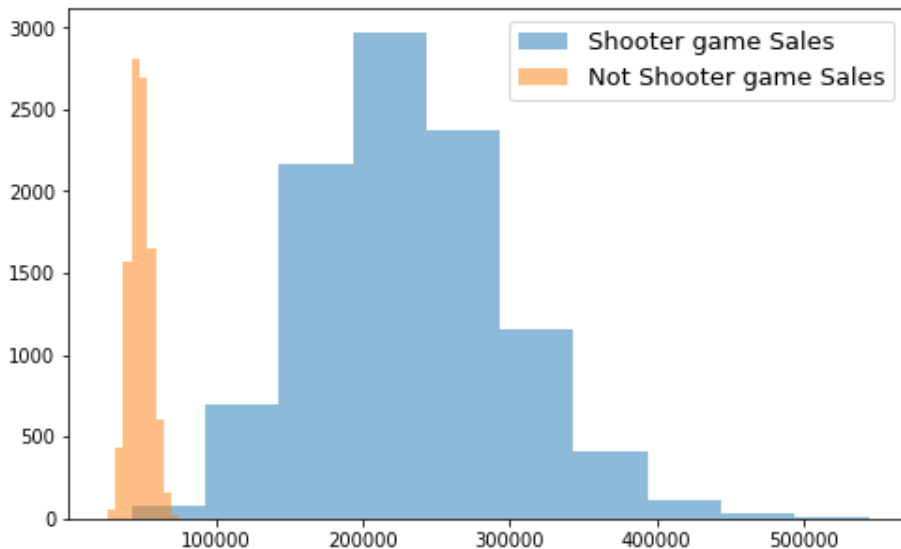
* 2016년 북미에서 발매된 슈팅게임 개수 : 32개

Part 2. 게임 설계 방향

분석가정 및 가설설정 / 가설검정 / 게임 설계 방향 제시

■ 신뢰구간

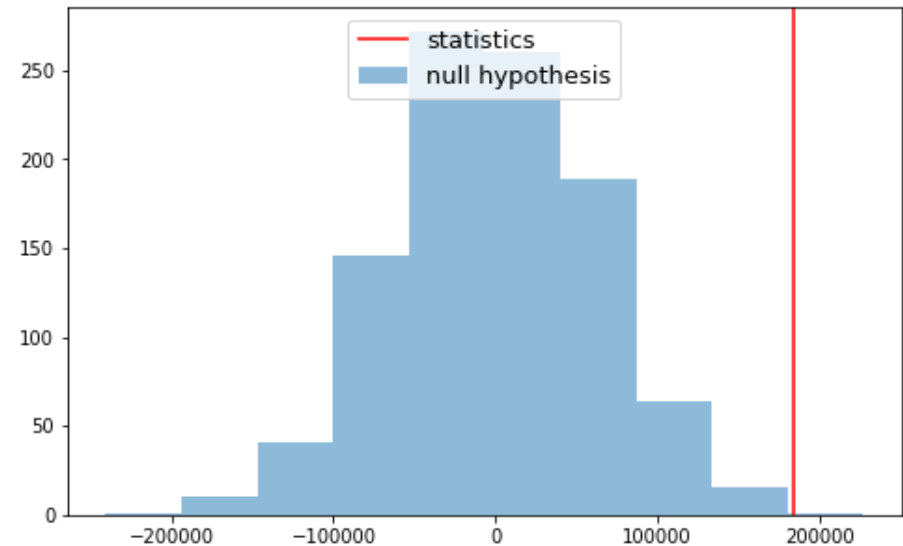
- ▶ 슈팅게임이 아닌 다른 게임들의 출고량 분포가 슈팅게임의 출고량 분포보다 작음



- * 전체 게임시장의 출고량 차이여부를 확인하고자 하기 위하여, 소수의 데이터를 통해 전체 시장을 추정할 수 있는 부트스트랩 기법을 사용함
- * 중복추출을 허용하여 10,000개의 샘플을 추출함

■ p-value

- ▶ 두집단의 평균값 차이가 우측 끝에 위치하고 있음
- ▶ 평균값 차이의 선이후의 면적은 0에 가까운 값임 => 귀무가설 기각

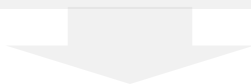


- 귀무가설 : 북미지역에서 2016년 발매된 슈팅게임의 게임당 출고량은 다른 게임들의 게임당 출고량보다 작거나 같다. => 기각
- 대립가설 : 북미지역에서 2016년 발매된 슈팅게임의 게임당 출고량은 다른 게임들의 게임당 출고량보다 크다. => 채택

Part 2. 게임 설계 방향

분석가정 및 가설설정 / 가설검정 / 게임 설계 방향 제시

- 북미지역에서 2016년 발매된 슈팅게임의 게임당 출고량은 다른 게임들의 게임당 출고량보다 크다는 대립가설이 채택됨
- 슈팅게임의 시장점유율은 2015년 액션게임의 시장점유율을 역전한 이후 1순위를 유지 하고 있음



- 게임시장의 현황을 보았을때 슈팅게임이 성장하고 있고, 2016년 타게임들에 비해 슈팅게임이 성공하였음을 확인함
- 그에 2017년 북미시장을 대상으로 발매할 슈팅게임을 개발하는 것이 좋을 것으로 판단됨