

발표 슬라이드

게임 회사의 데이터 팀에 합류.

다음 분기에 어떤 게임을 설계해야 할까?

회사의 목표는 결국 **많은 이익창출!**

따라서 요즘 트렌드에 맞는 게임을 만들어야 사람들이 많이 이용하고 이익이 생길 것.

따라서 요즘 트렌드를 알아보자.

1. 장르

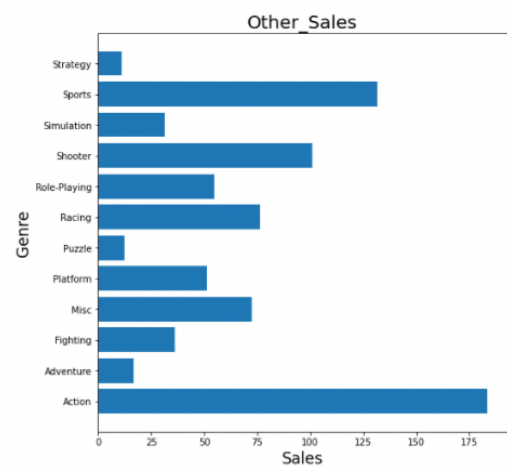
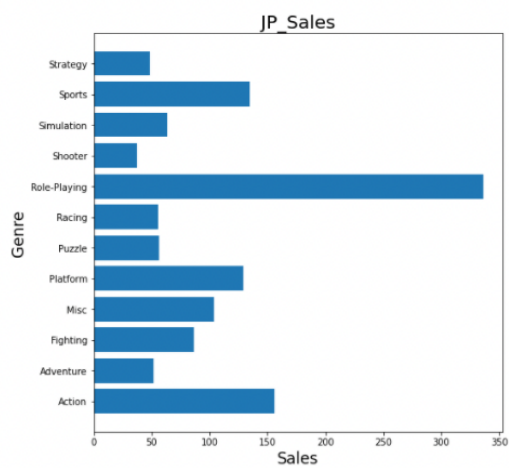
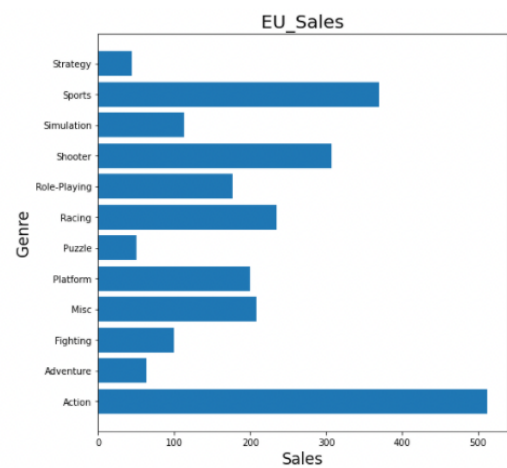
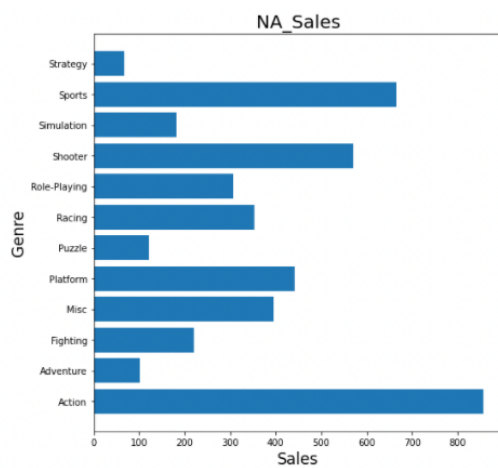
우선, 게임을 설계하기 위해 어떤 장르의 게임을 만들지를 결정해야 함.

레이싱, RPG, 격투, 액션 등 여러 장르가 있다.

이런 여러 장르 중 트렌드에 맞는 장르를 알아보고 선택하자!

- 지역에 따라 선호하는 게임의 장르에 대해서

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Total_Sales
Genre					
Action	855.96	512.52	155.73	183.64	1707.85
Sports	664.33	369.69	134.58	131.73	1300.33
Shooter	570.89	307.26	37.34	100.82	1016.31
Role-Playing	307.08	176.31	336.25	54.71	874.35
Platform	442.46	199.40	129.22	51.15	822.23
Misc	394.29	207.68	103.71	72.54	778.22
Racing	353.57	234.20	55.42	76.23	719.42
Fighting	220.11	99.72	86.39	36.10	442.32
Simulation	181.04	112.68	63.46	31.24	388.42
Puzzle	121.00	50.39	56.19	12.39	239.97
Adventure	101.33	63.32	51.79	16.64	233.08
Strategy	67.82	44.70	48.13	11.20	171.85



장르에 따른 지역별 출고량의 합계에 대해서 알아보았다.

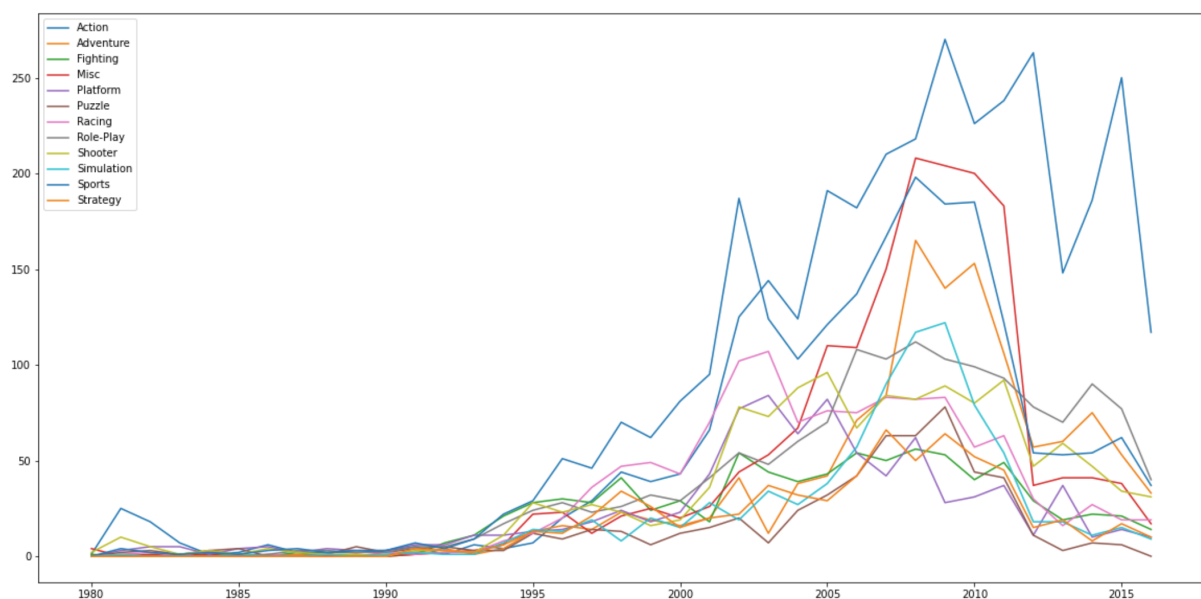
- 북미지역(NA)에서는 'Action'장르의 게임
- 유럽지역(EU)에서는 'Action'장르의 게임
- 일본지역(JP)에서는 'Role-Playing'장르의 게임
- 다른지역(Other)에서는 'Action'장르의 게임

위와 같이 각 지역별로 높은 출고량을 갖는 게임 장르를 확인할 수 있었다.

지역별 선호하는 게임 장르를 알아본 결과 'Action'장르가 인기가 많다는 것을 알 수 있다.

따라서, 게임을 설계할 때 'Action'장르의 게임을 설계하는 것이 바람직하다고 생각한다.

• 연도별 게임의 장르에 대해서



연도별 게임 장르 종류에 대해서 알아보았다.

(2017년과 2020년 자료는 너무 극단적으로 수치가 빈약하여 그래프를 표시할 때는 2016년의 자료까지만 표시하였다.)

그래프를 보면 '1995년' 이후로 모든 장르의 게임이 출시되었다는 것을 알 수 있고 '2015년' 이후로 모든 장르의 게임이 하락하는 것을 알 수 있다.

그래도 다음 분기 게임을 설계하기 위해서 장르를 선택한다면 ‘2003년’부터 꾸준히 인기가 있던 ‘Action’장르로 게임을 설계하는 것이 바람직하다고 생각한다.

2. 플랫폼

그 다음은 게임을 설계할 때, 어떠한 플랫폼을 바탕으로 게임을 개발할 것인지를 알아야 한다.

예를 들어, ‘PS’의 플랫폼의 수요가 많다고 한다면, 다음 게임도 ‘PS’에서 이용할 수 있게 해 준다면 많은 소비자들이 게임을 이용할 것이다.

(‘X360’은 ‘X box’를 의미, ‘PS2’, ‘PS3’, ‘PS4’는 ‘Play Station’을 의미, ‘DS’는 ‘Nintendo DS’를 의미...)

따라서 트렌드에 맞는 플랫폼을 알아보고 선택하자!

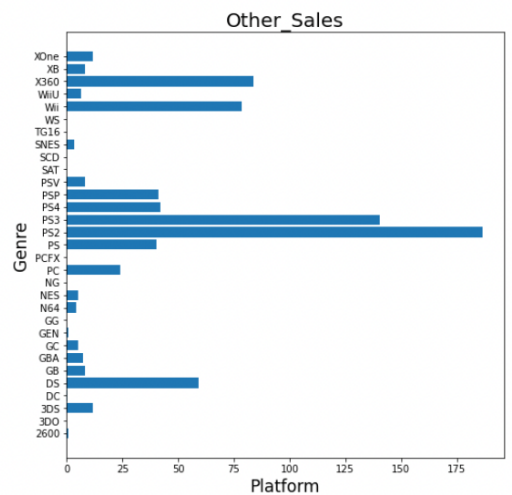
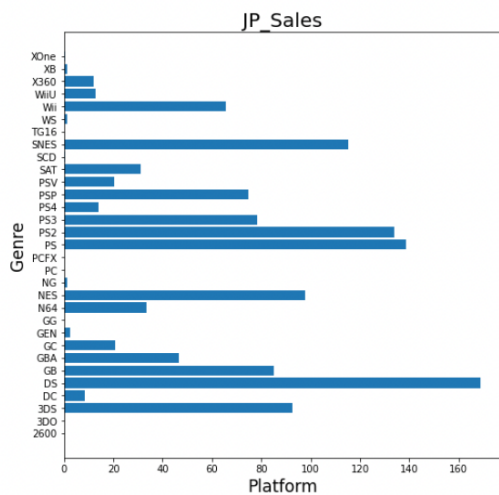
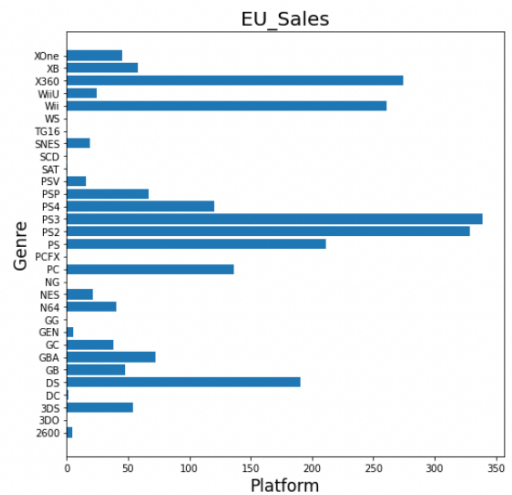
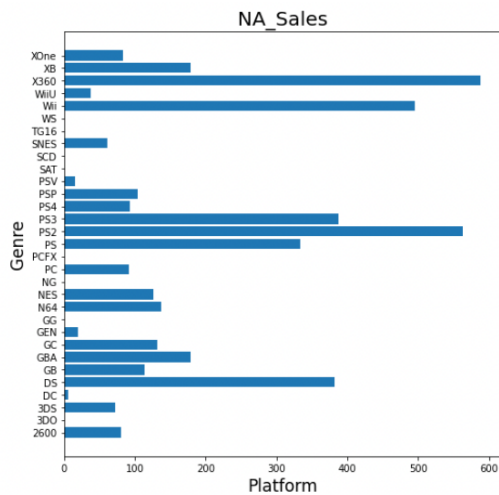
- 지역에 따라 선호하는 플랫폼에 대해서

NA_Sales : X360

EU_Sales : PS3

JP_Sales : DS

Other_Sales : PS2



플랫폼에 따른 지역별 출고량 합계에 대해서 알아보았고 그 중 지역별 가장 높은 플랫폼을 알아보았다.

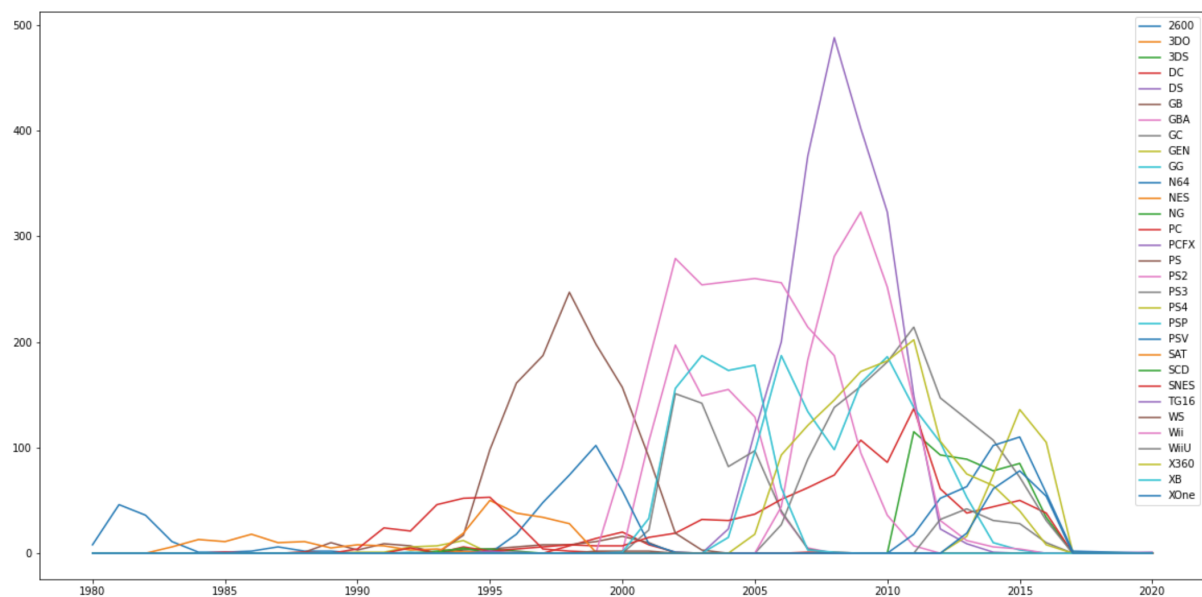
- 북미지역(NA)에서는 'X360' 플랫폼 → 587.58 millions
- 유럽지역(EU)에서는 'PS3' 플랫폼 → 339.40 millions
- 일본지역(JP)에서는 'DS' 플랫폼 → 168.73 millions
 - 일본지역2위 플랫폼은 'PS' → 138.67 millions
- 다른지역(Other)에서는 'PS2' 플랫폼 → 186.74 millions

이익적인 측면에서 게임을 설계한다면 북미지역과 유럽지역, 다른 지역을 노려 'X360'과 'PS' 플랫폼에 맞는 게임을 설계하는 것이 바람직하다고 생각한다.

(비록 일본에서는 출고량 1위 플랫폼은 'DS'이지만 2위 플랫폼인 'PS2'의 출고량도 많기 때문에 'PS'플랫폼으로 설계하는 것이 일본 시장을 포기하는 것은 아니라고 생각.)

또한 위의 결과로 알 수 있는 것은 확실히 지역 게임(닌텐도DS는 일본 기업 게임)이 인기가 많다는 사실도 알 수 있다.

• 연도별 플랫폼에 대해서



Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	...	2009	2010	2011	2012	2013	2014	2015	2016	2017	2020
Platform																					
PS4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	16	74	136	105	1	0
PSV	0	0	0	0	0	0	0	0	0	0	...	0	0	18	52	63	102	110	57	2	0
XOne	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	19	61	78	54	0	0
PC	0	0	0	0	0	1	0	0	1	0	...	107	86	137	61	38	44	50	38	0	0
3DS	0	0	0	0	0	0	0	0	0	0	...	0	0	115	93	89	78	85	34	0	0
PS3	0	0	0	0	0	0	0	0	0	0	...	158	181	214	147	127	107	72	31	0	0
WiiU	0	0	0	0	0	0	0	0	0	0	...	0	0	0	32	42	31	28	10	0	0
X360	0	0	0	0	0	0	0	0	0	0	...	172	182	202	106	75	64	40	8	0	0
Wii	0	0	0	0	0	0	0	0	0	0	...	323	252	143	31	12	6	4	0	0	0
PSP	0	0	0	0	0	0	0	0	0	0	...	161	186	138	105	53	10	3	0	0	0
DS	0	0	0	0	0	1	0	0	0	0	...	402	323	149	23	9	1	0	0	0	1
2600	8	46	36	11	1	1	2	6	2	2	...	0	0	0	0	0	0	0	0	0	0
3DO	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
DC	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
GB	0	0	0	0	0	0	0	0	1	10	...	0	0	0	0	0	0	0	0	0	0
GBA	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
GC	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
GEN	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
GG	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
N64	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
NES	0	0	0	6	13	11	18	10	11	5	...	0	0	0	0	0	0	0	0	0	0
NG	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
PCFX	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
PS	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
PS2	0	0	0	0	0	0	0	0	0	0	...	95	36	7	0	0	0	0	0	0	0
SAT	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
SCD	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
SNES	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

그래프에서 '2009년' 즈음 'DS' 플랫폼의 인기가 급 상승했던 적이 있다. 하지만 'DS' 플랫폼의 인기는 오래 가지 못한것을 확인할 수 있다.

다음 분기 게임의 설계를 위해서는 최근 가장 많이 사용했던 플랫폼에 대해서 알아봐야 한다.

'2016~2014년'의 플랫폼의 사용량에 대해서 알아보았다.

- '2016년'에는 'PS4'
- '2015년'에는 'PS4'
- '2014년'에는 'PS3'

최근 플랫폼 트렌드를 알아본 결과 많은 게임들이 'PS'를 기반으로 만들어진다는 것을 알 수 있다.

따라서 게임을 개발할 때 **'Play Station' 플랫폼 기반으로 먼저 개발**하는 것이 바람직하다고 생각한다.

(추후, 업데이트 시 'X box' 플랫폼도 사용할 수 있도록 해준다면 더욱 많은 이익을 볼 수 있을 것이라고 생각.)

3. 장르와 플랫폼의 관계

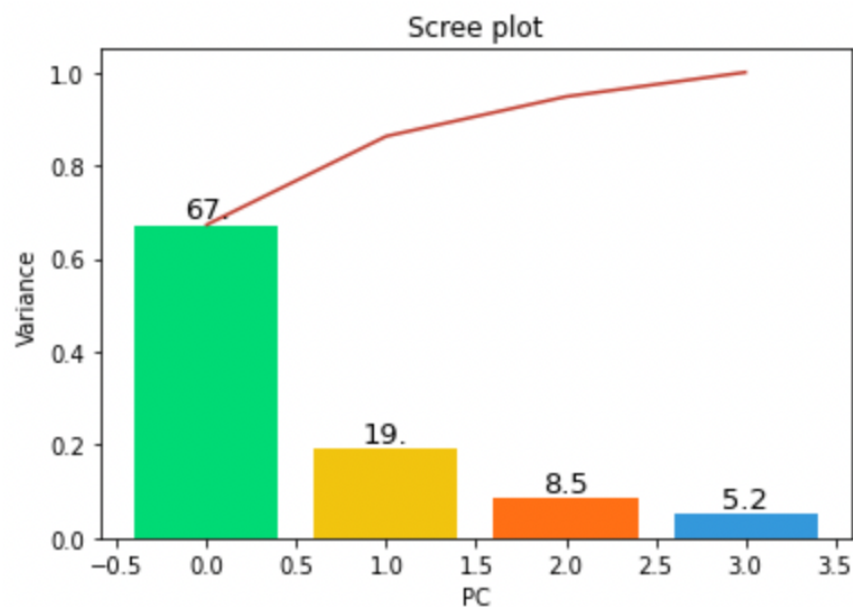
위의 분석으로 알 수 있는 것은 게임을 설계할 때, 장르는 'Action'으로 / 플랫폼은 'PS'를 기반으로 개발을 하면 되겠다는 결과가 나왔다.

그러면 과연 장르와 플랫폼은 연관이 있을까?

(다시말해, 'PS'플랫폼 기반으로 'Action'장르 게임을 만드는게 최적화가 잘 되나..? 이런 게임과 플랫폼의 연관)

→ 출고량의 데이터 분포를 통해 관계 파악 (PCA)

출고량 데이터들의 분포를 알아보기전에 PCA를 통해 차원 축소 후 시각화를 하려고 한다.



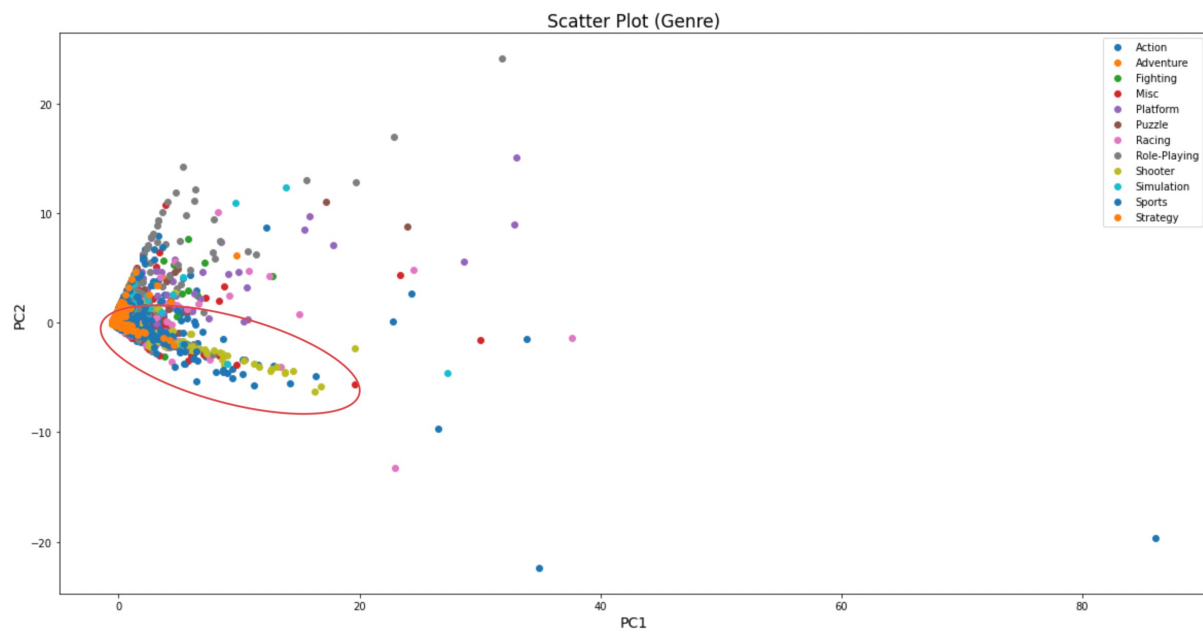
scree plot으로 알아본 결과 2차원으로 축소를 하여도 누적 분산 설명력이 86.27%%를 가지는 것을 확인할 수 있었다.

따라서 2차원으로 차원 축소 진행하였다.

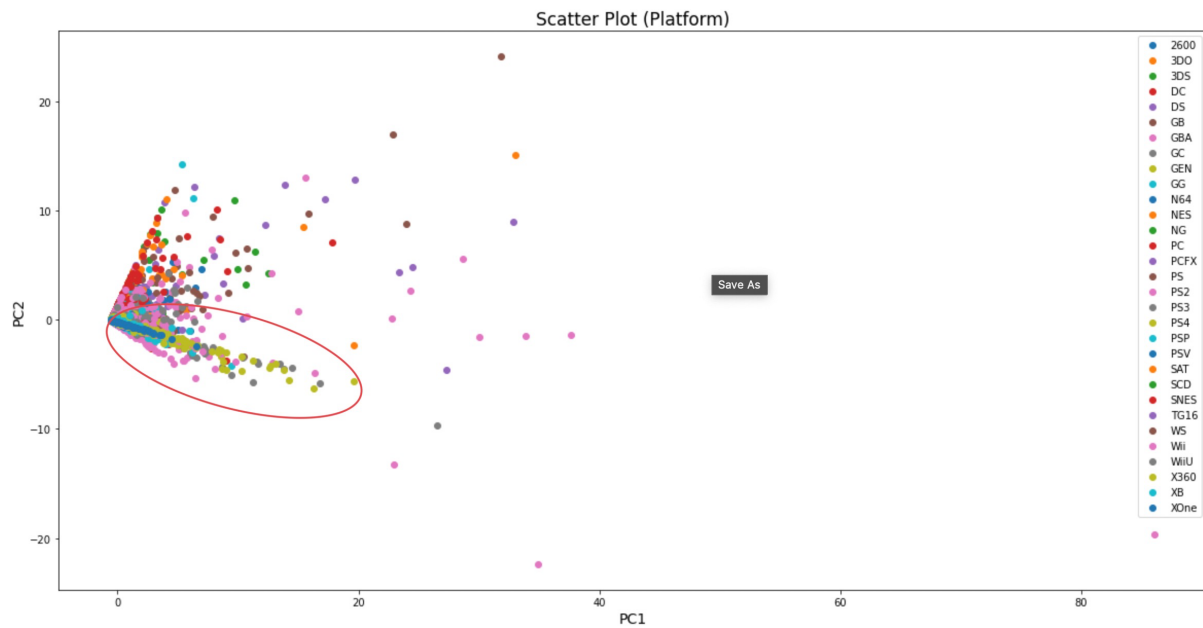
	PC1	PC2	Genre	Platform
0	-0.533115	-0.058868	Action	DS
1	-0.420815	-0.090041	Action	Wii
2	-0.535250	0.002269	Adventure	PSP
3	-0.533115	-0.058868	Misc	DS
4	-0.274947	-0.180500	Platform	PS3
...
16136	-0.390040	-0.101299	Platform	GC
16137	-0.552883	-0.056711	Action	PSP
16138	0.321884	-0.344309	Sports	PS3
16139	-0.088889	0.587329	Action	PSV
16140	-0.339056	0.388273	Fighting	PS

16141 rows x 4 columns

그리고 PC1, PC2를 축으로 하여 장르와 플랫폼에 대해서 scatter plot을 그려보았다.



장르에 대한 scatter plot



플랫폼에 대한 scatter plot

scatter plot → 'Action' 과 'Shooter' 장르의 게임은 'PS4'와 'X360' 플랫폼의 분포와 상당한 일치를 보인다.

가설 검정을 통해 통계적인 수치로 파악해보자!

따라서 **Two Sample Chi-Squared test**를 진행해 보았다.

H0: 장르와 플랫폼은 연관성이 없다.

H1: 장르와 플랫폼은 연관성이 있다.

Chi saure: 5772.691201706838
P-value: 0.0

그 결과 P-value값이 유의수준인 0.05보다 낮게 나왔다.

즉, 귀무가설을 기각.

장르와 플랫폼은 연관성이 있다는 결론이 나왔다.

4. 추가 분석

1) Numerical Data (출고량 데이터)의 평균 비교 → 아쉬운점!!!

비교해야 할 그룹 수가 2개 이상이므로 ANOVA를 사용!

<https://partrita.github.io/posts/ANOVA-python/>

ANOVA (ANalysis Of VAriance; 분산 분석)은 여러 그룹간의 평균의 차이가 통계적으로 유의미 한지를 판단하기 위한 시험법입니다. T-test 처럼 말이죠. 하지만 그룹의 수가 2개 이상일때 사용한다는 점이 다른점이지요.

분산분석에는 다음과 같은 3가지의 조건이 필요해요.

- 정규성: 각각의 그룹에서 변인은 정규분포.
- 분산의 동질성: Y의 모집단 분산은 각각의 모집단에서 동일.
- 관찰의 독립성: 각각의 모집단에서 크기가 각각인 표본들이 독립적으로 표집.

그런데 ANOVA는 위와 같은 조건이 필요하다.

따라서 각 feature들이 조건에 맞는지 먼저 확인을 해 보았다.

- 정규성 검정

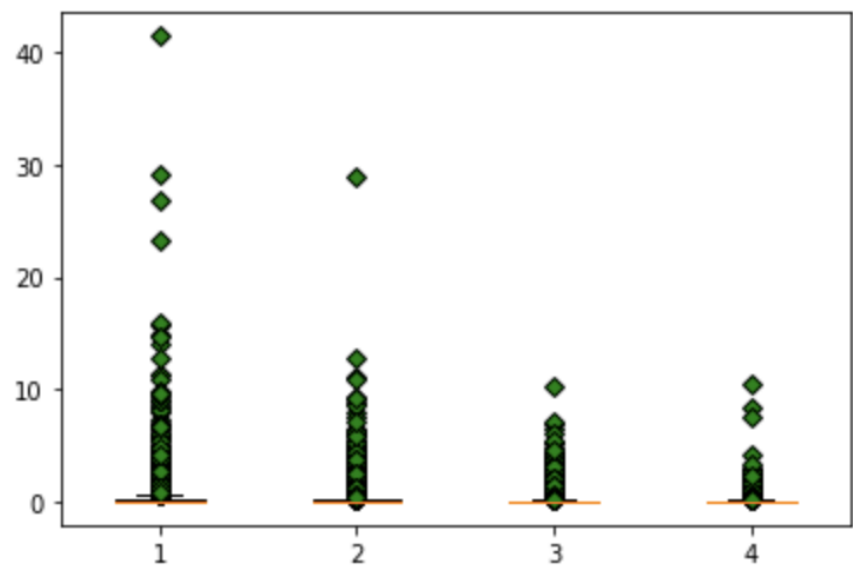
```
Test-statistics: 0.26422691345214844, p-value: 0.0
Test-statistics: 0.25336194038391113, p-value: 0.0
Test-statistics: 0.24638605117797852, p-value: 0.0
Test-statistics: 0.20084506273269653, p-value: 0.0
```

각 feature들이 정규성을 만족하는지 정규성 검정을 진행해 보았다. (scipy의 Shapiro검정 사용)

그런데 P-value값들이 0.05보다 작아 정규성을 만족하지 못한다는 결론이 나왔다.

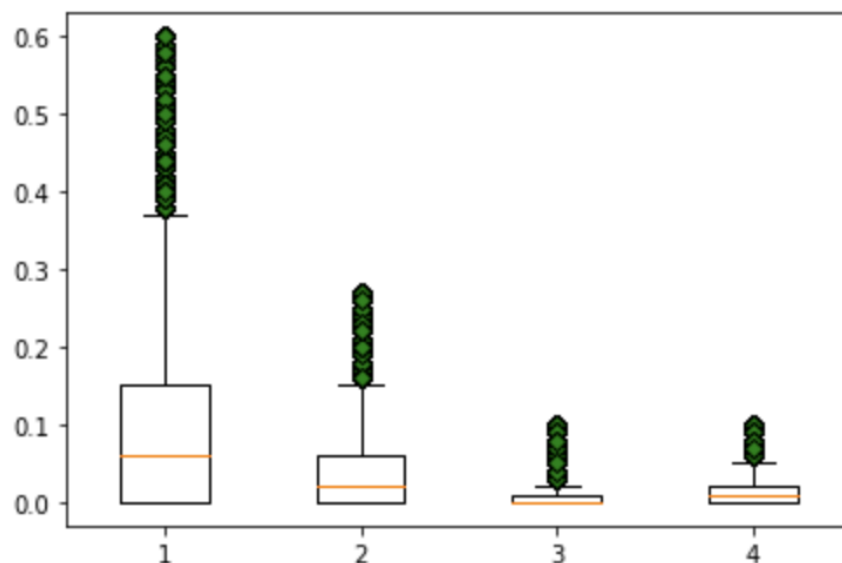
왜 그럴까?

box-plot으로 데이터 분포에 대해서 알아보자.



box-plot에서 보다시피 이상치가 너무 많이 존재한다.

따라서 이상치를 제거해야 겠다고 생각.



이상치를 제거 후 다시 그린 box-plot이다. (그런데 제대로 이상치가 제거된 것 같아 보이지 않는다..)

이상치를 제거 후 정규화를 진행했다. 정규화 방법은 다음과 같은 방법을 사용했다.

- MinMaxScaler
- StandardScaler
- RobustScaler

정규화 후 정규성 검정을 다시 실시해 보았다.

```
Test-statistics: 0.8025397062301636, p-value: 0.0
Test-statistics: 0.729623556137085, p-value: 0.0
Test-statistics: 0.538764238357544, p-value: 0.0
Test-statistics: 0.7223619222640991, p-value: 0.0
```

MinMaxScaler

```
Test-statistics: 0.8025411367416382, p-value: 0.0
Test-statistics: 0.729607343673706, p-value: 0.0
Test-statistics: 0.5387471914291382, p-value: 0.0
Test-statistics: 0.7223609685097827, p-value: 0.0
```

StandardScaler

```
Test-statistics: 0.8025386929512024, p-value: 0.0
Test-statistics: 0.7296082973480225, p-value: 0.0
Test-statistics: 0.538764238357544, p-value: 0.0
Test-statistics: 0.7223640084266663, p-value: 0.0
```

RobustScaler

그런데 위의 3가지 방법으로 정규화한 데이터의 정규성 검정 P-value가 모두 0.05를 넘지 못하였다.

따라서 분산 분석의 조건 중 하나인 정규성을 만족하지 못하기 때문에 ANOVA를 진행할 수 없었다..

2) 총 출고량 상위 500개 데이터를 통해 clsutering (PCA후 → K-Means Clustering)

각 지역별 출고량을 합친 feature, 총 출고량에 대해 내림차순으로 정렬 했을 때, 상위 500개 데이터를 가지고 군집을 해보려고 한다.

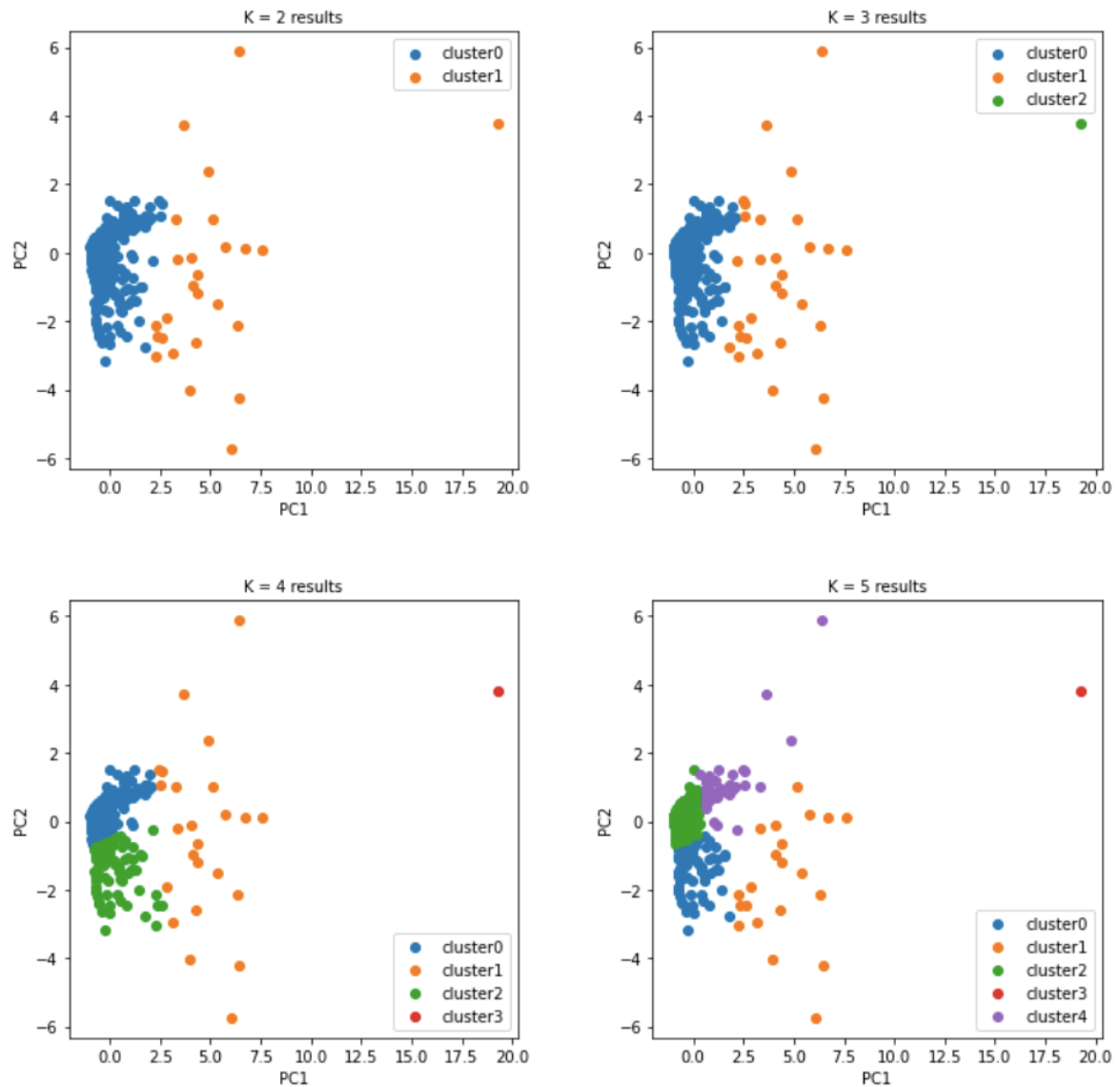
우선 위와 같이 PCA를 먼저 진행 후 군집을 하려함.

```
0    0.584629
1    0.810106
2    0.922939
3    1.000000
dtype: float64
```

누적 분산 설명력이 80.71%인 2차원으로 차원 축소

그 다음 K-Means Clustering을 진행

K값을 2, 3, 4, 5로 바꿔가며 clustering을 해보았다.



차원 축소 후 scatter plot을 그려본 결과 데이터가 한 덩어리로 뭉쳐있는 형태로 보인다.
따라서 군집을 하기에는 좋은 데이터가 아니다라는 생각이 들었다.

결과

'Genre: Action / Platform: 'PS'

위의 결과를 종합하면 북미 지역과 유럽 지역을 타겟으로 'Action'장르의 게임을 설계한다!

플랫폼은 'Play Station' 기반으로 먼저 게임을 설계 후 출시한다.

다만, 추후 'X box' 플랫폼 기반도 설계해 출시한다면 북미 지역의 유저들의 Needs도 만족할 것이라고 생각한다.