

빅데이터 최종보고서

Team_09

201411165 홍예린

201511017 김경택

201511054 김진모

프로젝트명 : 게임 평론과 판매량 사이의 관계 분석 & 게임 판매량 분석을 통한 고객 타겟팅

Introduction

시장에 어떠한 제품을 출시하여 성공적으로 판매하기 위해서는, 고객들의 수요가 높다고 판단되는 것을 내놓는 것이 하나의 전략일 것이다. 이에 본 프로젝트에서는 게임을 사용하는 고객과 게임 사이의 관계를 분석하는 것을 주 목표로 하였다. 첫번째는 고객의 반응과 게임의 판매량 사이에 실제로 어떤 관계가 있는가를 분석하고자 하였으며, 두번째로는 게임의 판매량을 토대로 특정 고객층에 맞는 게임을 찾을 방법론을 제시하고자 하였다.

이를 위해 우리는 게임의 Metacritic Score와 해당 게임의 판매량을 나타낸 데이터를 사용하였으며, 또한 출처가 다른 고객들의 평가 사이의 관계를 알아보기 위해 IGN 평론을 모은 데이터 역시 분석에 이용하였다. 아래의 Table 1과 Table 2는 각 데이터의 첫 다섯열을 나타낸다.

<Figure 01. IGN평론 Data example>

	score_phrase	title	platform	score	genre	editors_choice	release_year	release_month	release_day
0	Amazing	LittleBigPlanet PS Vita	PlayStation Vita	9.0	Platformer	Y	2012	9	12
1	Amazing	LittleBigPlanet PS Vita -- Marvel Super Hero E...	PlayStation Vita	9.0	Platformer	Y	2012	9	12
2	Great	Splice: Tree of Life	iPad	8.5	Puzzle	N	2012	9	12
3	Great	NHL 13	Xbox 360	8.5	Sports	N	2012	9	11
4	Great	NHL 13	PlayStation 3	8.5	Sports	N	2012	9	11

<Figure 02. Metacritic 평론과 game sales ranking data example>

index	title	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
0	Wii Sports	Wii	2006	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53	76.0	51.0	8	322.0	Nintendo	E
1	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	NaN	NaN	NaN	NaN	NaN	NaN
2	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	82.0	73.0	8.3	709.0	Nintendo	E
3	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	80.0	73.0	8	192.0	Nintendo	E
4	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	NaN	NaN	NaN	NaN	NaN	NaN

위에서 제시한 목표와 분석을 위해 찾은 데이터를 토대로, 우리는 본 연구의 토대가 될 두 가지 Question을 설정하였다. 첫 번째는 “게임 평론과 판매량 사이에는 어떤 관계가 있는가”이다. 또한, 두 번째는 “어떤 장르의 게임을, 어떤 시장에, 어떤 플랫폼으로 제출하는 것이 가장 높은 판매량을 끌어낼 수 있는가”이다.

Methodology

본 프로젝트에서 제시한 첫 번째 Question을 위한 분석은 다음과 같은 단계를 거쳐 진행되었다.

0. 메타크리틱 스코어가 유효한 row만을 모아서 이를 전체 데이터셋으로 설정한다.(약 6,800 rows)
1. 메타크리틱 데이터셋에 있는 평론가 점수와 일반 유저 점수 각각과 게임 판매량 사이의 상관계수를 알아본다.
2. 데이터를 1000개씩, 6개의 구간으로 나누어 각 구간의 산포도를 추출한다.
3. 각 구간마다 describe를 실행하여, 평균, 표준편차, 최소값, 최대값의 증감 추세를 알아본다.
4. 각 구간마다 점수의 분포를 하나의 bar graph에 나타내어 본다.
5. 높은 점수면 높은 판매량을 보인다는 명제에 대한 오류 여부를 확률적으로 검증해 본다.
6. 메타크리틱의 critic 점수, 메타크리틱의 user 점수, ign 평론 점수를 통해 서로에 대한 상관관계 변수 r 을 추출한다.

이어서 본 프로젝트에서 제시한 두 번째 Question을 위한 분석 방법은 다음과 같았다.

1. 연도에 따라 가장 많이 출시된 장르와 그 비율을 추출한다.
2. 실제로 해당 연도에서 가장 많이 팔린 장르와 그 연도의 총 판매량의 비율을 추출한다.
3. 해당 분석 두개를 merge
4. Platform마다 판매량이 가장 높은 장르를 찾는다.
5. 2~4번까지의 분석을 각 시장별로(NA, EU, JP)별로 진행하여 게임 시장별 데이터를 추출한다.
6. 1~5번까지의 분석을 연도의 scope를 달리하여 다시 한번 진행한다. (2010~2016년)

Results

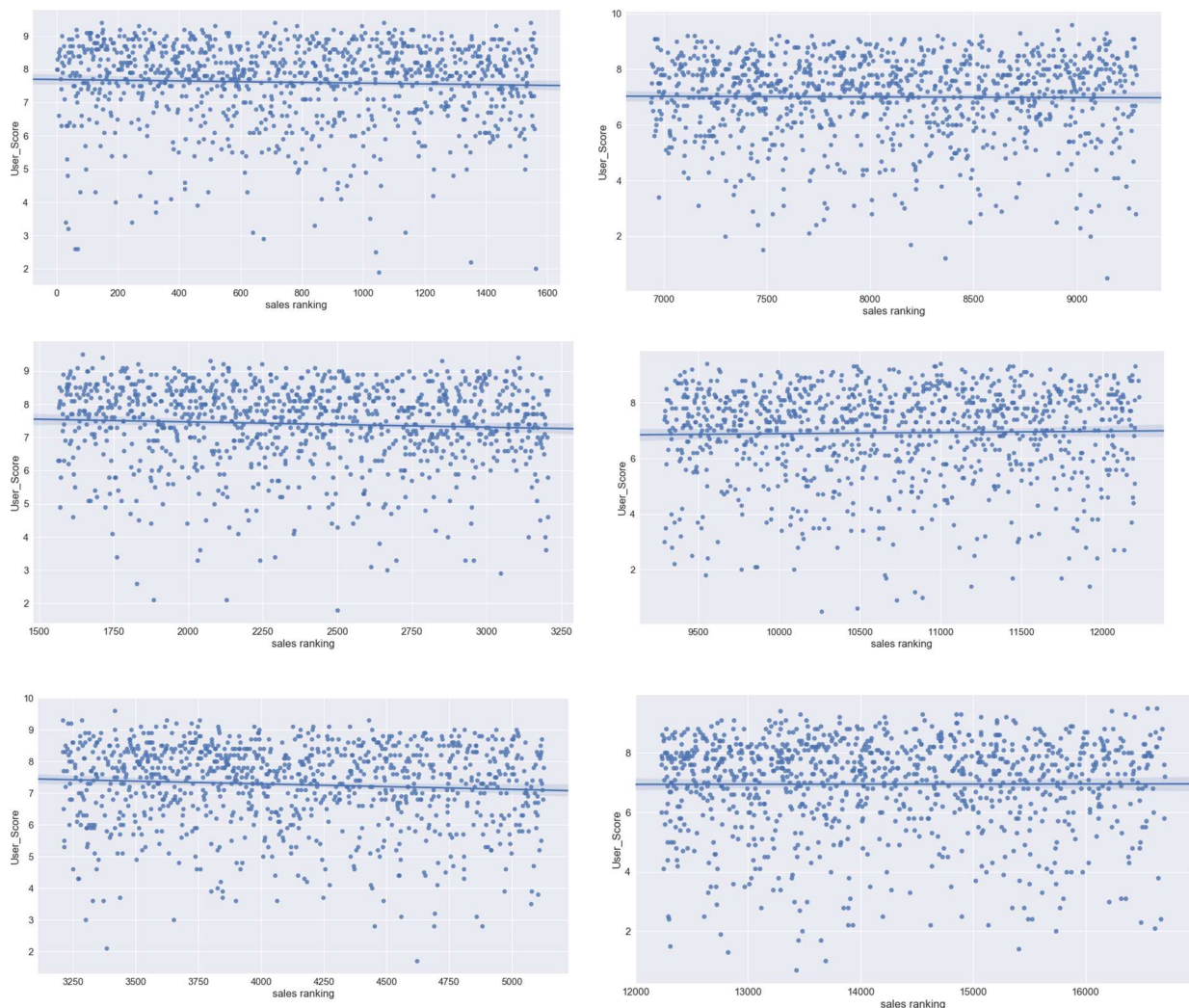
본 프로젝트에서 제시한 첫 번째 Question을 위해 분석한 결과는 다음과 같다.

1. 메타크리틱 데이터셋에 있는 평론가 점수와 게임 판매량 사이의 상관계수를 알아본다.

```
In [47]: 1 np.corrcoef(game_meta_valid['Global_Sales'],game_meta_valid['Critic_Score'])
          2 #looks like r value is really low... so, should we abandon this?

Out[47]: array([[ 1.          ,  0.23755572],
                [ 0.23755572,  1.          ]])
```

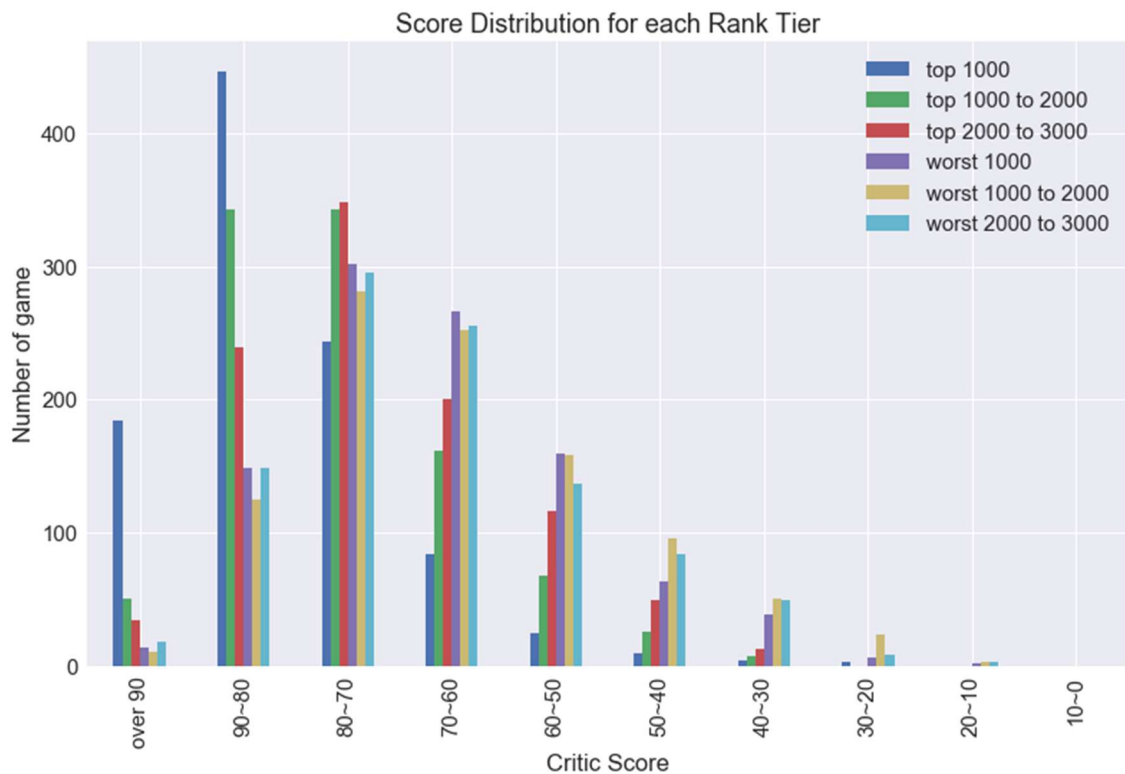
2. 데이터를 1000개씩, 6개의 구간으로 나누어 각 구간의 산포도를 추출한다.



3. 각 구간마다 describe를 실행하여, 평균, 표준편차, 최소값, 최대값의 증감 추세를 알아본다.

	Count	Mean	Std	Min	max
Top1000	1000	80.602	10.59	20	98
1000~2000	1000	74.824	11.251	32	97
2000~3000	1000	71.131	12.107	32	96
Worst 2000~3000	1000	66.547	13.243	19	96
Worst 2000~1000	1000	64.106	14.752	17	94
Worst 1000	1000	65.944	14.3317	13	94

4. 각 구간마다 점수의 분포를 하나의 bar graph에 나타내어 본다.



5. 높은 점수면 높은 판매량을 보인다는 명제에 대한 오류 여부를 확률적으로 검증해 본다.

*이때 등수는 판매량 순위를 뜻함.

90점 이상이고 1000등 이내	90점 이상이고 1000등 밖
184	141
90점 이하인데 1000등 안	90점 이하이고 1000등 밖
816	5684

80점 이상이고 1000등 이내	80점 이상이고 1000등 밖
631	1294
80점 이하인데 1000등 안	80점 이하이고 1000등 밖
369	4531

6. 메타크리틱의 critic 점수, 메타크리틱의 user 점수, ign 평론 점수를 통해 서로에 대한 상관관계 변수 r을 추출한다.

```
In [42]: 1 np.corrcoef(same_plat_compare['Critic_Score'],same_plat_compare['score'])
          2 ##결정계수 강한 양적 선형관계를 띠고있다.
```

```
Out [42]: array([[ 1.          ,  0.89287366],
                 [ 0.89287366,  1.          ]])
```

IGN평론 – 메타크리틱 비평가 평론 = 0.89

```
In [48]: 1 same_plat_compare['User_Score'] = same_plat_compare['User_Score'].astype(float)
          2 np.corrcoef(same_plat_compare['User_Score'],same_plat_compare['score'])
```

```
Out [48]: array([[ 1.          ,  0.52024364],
                 [ 0.52024364,  1.          ]])
```

IGN평론 – 메타크리틱 유저 평론 = 0.52

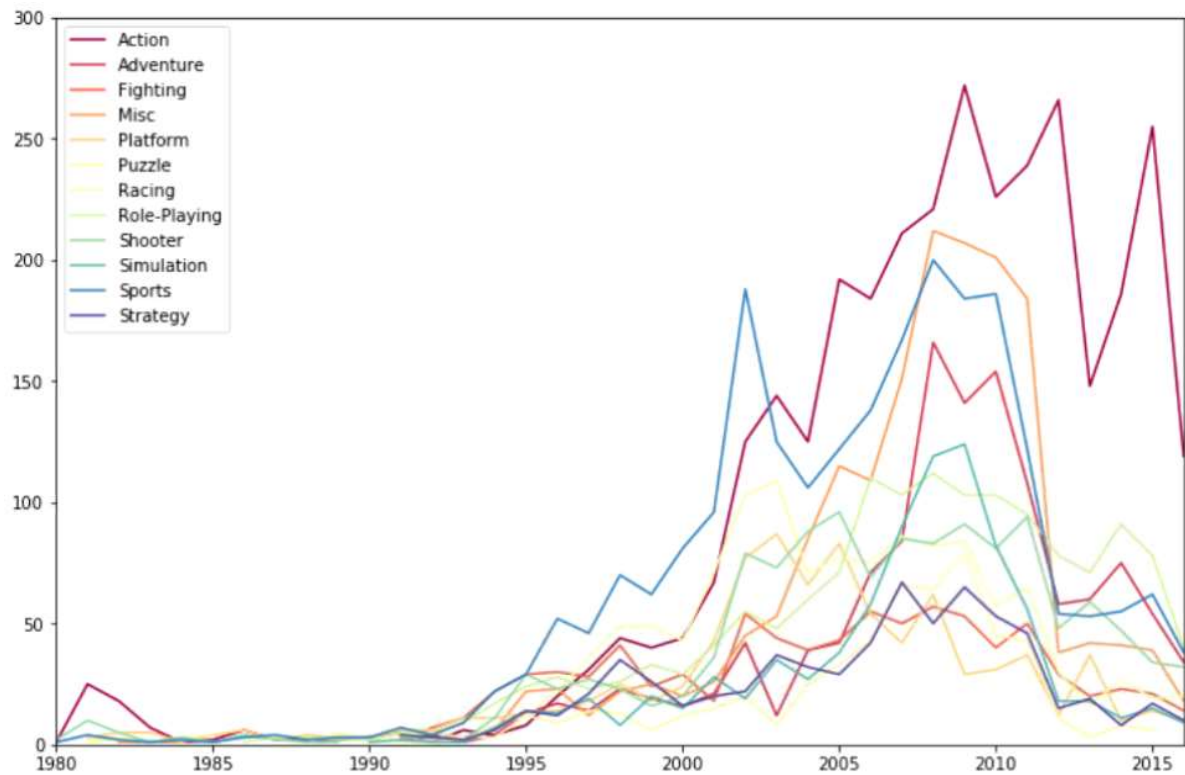
```
In [49]: 1 np.corrcoef(same_plat_compare['User_Score'],same_plat_compare['Critic_Score'])
```

```
Out [49]: array([[ 1.          ,  0.58186776],
                 [ 0.58186776,  1.          ]])
```

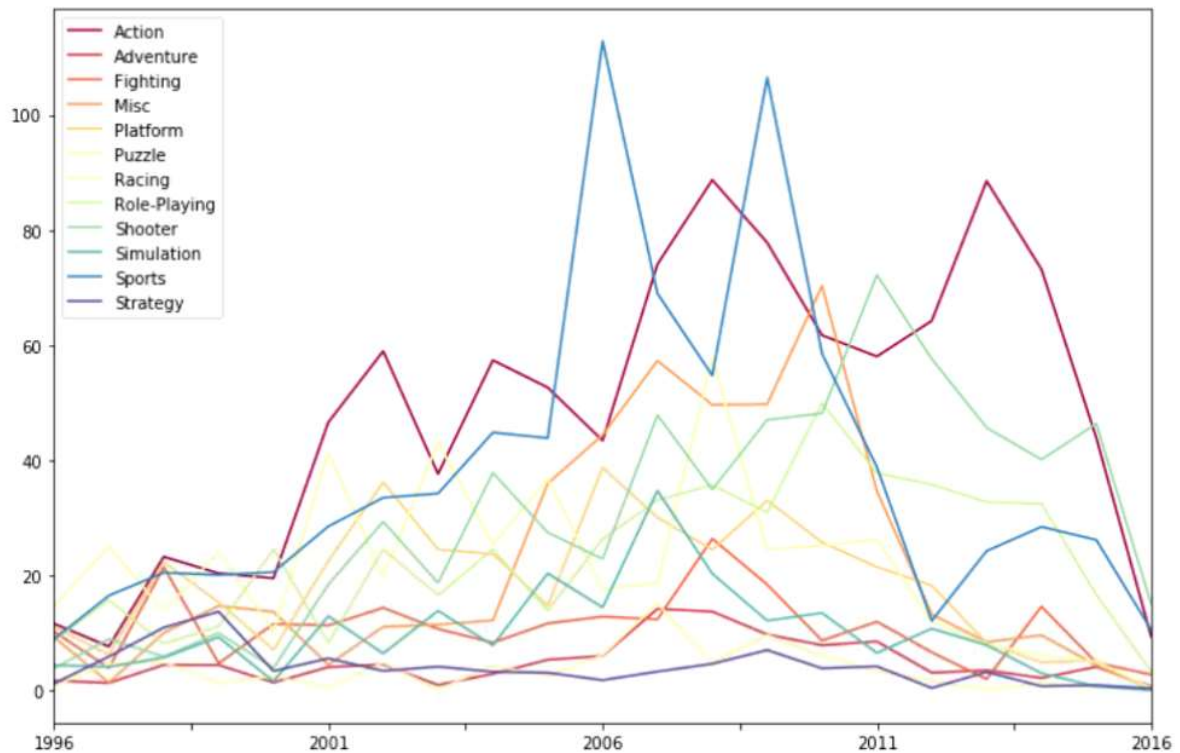
메타크리틱 비평가 평론 – 메타크리틱 유저 평론 = 0.58

두 번째 Question을 위해 분석한 결과는 다음 그래프와 표를 통해 설명할 수 있다.

1. 연도에 따라 가장 많이 출시된 장르와 그 비율



2. 실제로 해당 연도에서 가장 많이 팔린 장르와 그 연도의 총 판매량의 비율



3. 해당 분석 두개를 merge

Year	2010	2011	2012	2013	2014	2015	2016
Info							
Most_Genre	Action	Action	Action	Action	Action	Action	Action
Release_Ratio	17.9508	20.9833	40.4871	27.1062	31.9588	41.5309	34.593
Best_Genre	Misc	Shooter	Shooter	Action	Action	Shooter	Sports
Sales_Ratio	21.4198	16.7464	21.3708	26.0062	17.142	28.2429	24.3013

4. Platform마다 판매량이 가장 높은 장르 찾기.

	2600	3DO	DC	GG	Game Boy	Game Boy Advance	Game Cube	Genesis	NES	NG	Nintendo 3DS	Nintendo 64	Nintendo DS
Best Genre	Action	Adventure	Sports	Platform	Role-Playing	Platform	Action	Platform	Platform	Fighting	Role-Playing	Racing	Misc
Global Sales	29.34	0.06	3.66	0.04	88.24	78.3	37.84	15.45	95.78	1.42	75.74	40.09	137.76
Total Sales	97.08	0.1	15.97	0.04	255.45	318.5	199.36	28.36	251.07	1.44	247.46	218.88	822.49

5. 1~4번까지의 분석을 각 시장별로(NA, EU, JP)별로 진행하여 시장별 데이터를 추출.

	Global Genre	Global Sales	NA Genre	NA Sales	EU Genre	EU Sales	JP Genre	JP Sales	Total Sales
2600	Action	29.34	Action	27.39	Action	1.64	Action	0	97.08
3DO	Adventure	0.06	Adventure	0	Adventure	0	Adventure	0.06	0.1
DC	Sports	3.66	Sports	2.14	Platform	0.61	Fighting	1.83	15.97
GG	Platform	0.04	Platform	0	Platform	0	Platform	0.04	0.04
Game Boy	Role-Playing	88.24	Platform	29.4	Role-Playing	21.67	Role-Playing	34.55	255.45
Game Boy Advance	Platform	78.3	Platform	45.81	Platform	18.85	Role-Playing	21.21	318.5
Game Cube	Action	37.84	Action	26.11	Action	7.77	Fighting	3.69	199.36
Genesis	Platform	15.45	Platform	10.73	Platform	3.38	Platform	0.77	28.36
NES	Platform	95.78	Platform	53.64	Platform	10.19	Platform	30.26	251.07
NG	Fighting	1.42	Fighting	0	Fighting	0	Fighting	1.42	1.44
Nintendo 3DS	Role-Playing	75.74	Role-Playing	20.18	Role-Playing	13	Role-Playing	39.61	247.46
Nintendo 64	Racing	40.09	Racing	26.68	Racing	8.29	Platform	6.31	218.88
Nintendo DS	Misc	137.76	Simulation	68.56	Simulation	36.51	Role-Playing	54.55	822.49
PC	Simulation	51.91	Simulation	19.94	Simulation	27.37	Shooter	0.17	258.82
PCFX	Role-Playing	0.03	Role-Playing	0	Role-Playing	0	Role-Playing	0.03	0.03
PlayStation	Action	127.05	Sports	64.02	Action	45.13	Role-Playing	40.95	730.66

PlayStation 2	Sports	273.41	Sports	134.12	Action	71.44	Role-Playing	34.01	1255.64
PlayStation 3	Action	307.88	Action	121.16	Action	110.62	Action	29.59	957.84
PlayStation 4	Action	87.06	Action	29.7	Action	38.19	Action	5.41	278.1
PlayStation Portable	Action	64.72	Action	23.83	Action	17.68	Role-Playing	30.82	296.28
PlayStation Vita	Action	20.01	Action	4.95	Action	5.68	Role-Playing	7.54	61.93
SAT	Fighting	8.52	Fighting	0.51	Fighting	0.38	Fighting	7.59	33.59
SCD	Platform	1.5	Platform	1	Platform	0.36	Strategy	0.14	1.87
Super NES	Platform	65.65	Platform	32.71	Platform	10.74	Role-Playing	33.76	200.05
TG16	Adventure	0.14	Adventure	0	Adventure	0	Adventure	0.14	0.16
WS	Role-Playing	1.22	Role-Playing	0	Role-Playing	0	Role-Playing	1.22	1.42
Wii	Sports	292.06	Sports	151.77	Sports	94.99	Sports	18.25	926.71
Wii U	Platform	21.24	Platform	9.97	Action	6.59	Platform	3.65	81.86
Xbox	Shooter	63.55	Shooter	46.16	Shooter	15.25	Fighting	0.36	258.26
Xbox 360	Shooter	278.55	Shooter	174.22	Action	76.81	Shooter	3.1	979.96
Xbox One	Shooter	51.61	Shooter	31.78	Shooter	15.27	Shooter	0.16	141.06

본 분석을 통해서 게임 기획에 사용할 세 가지 기준을 발견할 수 있었다. 첫째는 게임을 출시할 시장, 둘째는 출시하고자 하는 게임 장르, 그리고 마지막은 출시할 게임의 플랫폼이다.

6. 1~5번까지의 분석을 연도의 scope를 달리하여 다시 한번 진행(2010~2016년)

이하 데이터는 ipylib참조.

Conclusion & Discussion

Question1 conclusion & Discussion

메타크리틱 평론가 점수 기준으로, 게임 판매 실적과 평론 사이의 뚜렷한 통계적 상관관계는 보이지 않는다. 그렇지만 구간별로 다시 분석을 진행해 본 결과 80~100점의 높은 점수와 판매 순위는 강한 positive correlation 을 보인다는 것을 알 수 있었다. 따라서 80점 이상의 좋은 평가를 보이는 게임들의 특징과, 높은 판매량을 보이는 게임들의 특징의 교집합을 찾는 작업은 유의미한 결과를 이끌어낼 것으로 추측할 수 있었다.

그러나 이러한 분석은 데이터 자체가 가지고 있는 한계점을 고려해야 한다. 메타크리틱 점수는 기본적으로 참여자들이 평가한 점수들의 평균값을 나타낸다. 따라서 평가를 시도한 사람의 수가 적을 경우에는 객관적으로 평가되었다고 보기 어려우며, 이는 User들의 평론이거나 판매 순위가

내려갈수록 크게 두드러지는 특징이다. 따라서 게임에 대한 평가를 시도한 사람의 수가 지나치게 낮거나, 낮은 판매량을 보이는 게임들 간에 어떠한 관계를 찾고자 하는 시도는 잘못된 결과를 이끌어낼 가능성이 높다.

Question2 conclusion & Discussion

게임을 출시할 때 보다 많은 판매량을 확보하기 위해 고려해야 할 세 가지 기준으로 시장, 장르, 플랫폼을 제시하였다. 이로부터 세 가지 기준 중 하나가 결정되었다면, 가장 많은 판매량을 기대할 수 있는 나머지 두 가지 기준을 선정할 수 있다. 예를 들어, 북미(NA, North America)에서 게임을 출시하여 판매하고 싶을 때 가장 많은 판매량이 기대되는 장르는 Shooter이며, 플랫폼은 Xbox 360으로 출시하는 것이 권장된다.

이러한 분석은 몇 가지 한계점을 갖고 있다. 가령, 일반적으로는 동일한 플랫폼의 새로운 세대로 결정하는 것이 맞지만 본 데이터는 이에 대한 정보 및 근거를 충분히 제시하지 못한다. 또한 특정 게임사에 대한 분석의 부재, 현재 가장 활발한 시장이라고 할 수 있는 모바일게임 시장에 대한 데이터를 고려하지 않았다. 만약 다양한 게임사에 대한 게임 출시 데이터가 확보된다면 거대 게임사의 경우와 상대적으로 규모가 작은 게임사의 경우를 따로 떼어서 분석하는 것도 가능할 것이다.

본 분석을 활용할 수 있는 방안.

게임 출시를 위한 기획에 있어서 최대의 판매량을 뽑기 위한 다각적 분석의 시발점으로써 본 분석을 활용할 수 있을 것이다. 작품 기획에 앞서 플랫폼, 게임 시장, 게임 장르를 결정하는 데에 있어서 판매량을 최대한으로 하기 위한 조건을 탐색하는 데에 본 분석법을 적용할 수 있다. 보고서에 포함된 데이터는 지금까지의 모든 게임에 대한 데이터이지만, 최근 3년, 혹은 1년 단위로 동일한 분석법을 적용시킴으로써 게임 트렌드를 다각적으로 분석할 수 있을 것이다.

기존의 분석과 다른 점

기존의 게임 시장 동향 분석은 굉장히 광범위한 분야에 걸쳐 소개되며, 작품을 기획하는 데에 필요한 데이터가 산발적으로 분포한다. 가령 한국콘텐츠진흥원에서 2017년 1월에 발간한 글로벌 게임산업 트렌드를 보면, 시장에 대한 데이터, 인기있는 플랫폼에 대한 데이터, 그리고 흥행하는 게임에 대한 데이터가 산발적으로 분포함을 알 수 있다. 반면 본 프로젝트에서 2번 Research Question을 해결하기 위해 분석한 표에서는 세 가지 데이터에 대한 정보가 모두 포함되어 있다.

분석 확장 방법

앞서 활용방안에도 소개했지만, 동일한 분석법을 짧은 기간 단위로 적용시켜보면서 최근 게임의 트렌드를 다각적으로 분석해보는 것이 가능할 것이다. 또한 본 프로젝트에서 제시한 것처럼 80점대 이상의 평론을 보이는 게임들의 공통점과, 지금까지 높은 판매량을 보인 게임들의 공통점을 추출하여 그 교집합을 분석하는 것으로 높은 판매량과 고객 평점을 동시에 만족시키기 위한 요소를 뽑을 수도 있을 것이다.