



비정형 데이터 분석

U2015524 정시원

INDEX.

1

**분석
개요**

2

**데이터
설명**

3

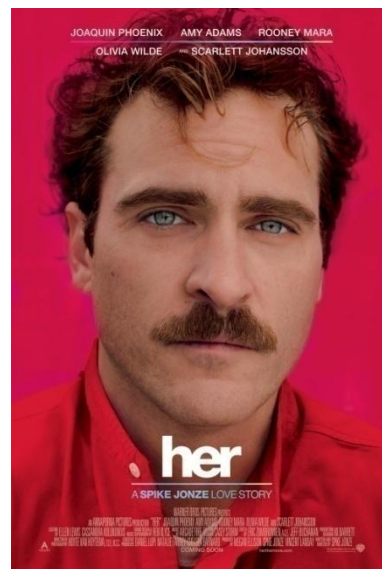
**실제
분석**

4

**결과
해석**

1 분석 개요

기대되는 영화였는데, 뚜껑을 열어보니 그닥 ...
 뭐 하나 기대될 것 없는 영화였는데, 실제로 개봉되고나니 대박 !!!



1 분석 개요

그렇다면 개봉전의 한줄평/ 개봉 후의 감상평을 비교해보면 어떨까?
(#.개봉일 이후로는 개봉전 평가 등록 불가 - 네이버 영화 기준)

영화 <정글북> 개봉전/후 140자평

개봉 전 ▾ | 개봉 후 ▾

기대지수 [?]

👍 보고싶어요

4,428

411

👎 글썽요

네티즌 평점 [?] 전체 ★★★★★ 9.47 참여 277명

★★★★★ 0 ▾

국내 개봉작입니다. 개봉 후에는 개봉전 평점을 등록할 수 없습니다.0/140

등록

140자 평 | 총 277건

[베스트](#) 평점 운영 정책 안내 [?]

✓ 호감순 ✓ 최신순

★★★★★ 10 [베스트](#) 외국에서 봤습니다. 저 40대 중반 남자이고 1년에 영화 한 20편쯤 보는데 최근 몇 년 간 본 영화 중 TOP3 안에 충분히 듭니다! 온 가족이 함께 볼 수 있는 영화로 강추.1000만 가뿐히 넘을 것으로 확신합니다. 가능하면 큰 화면에서 보세요.

개봉 전 ▾ | 개봉 후 ▾

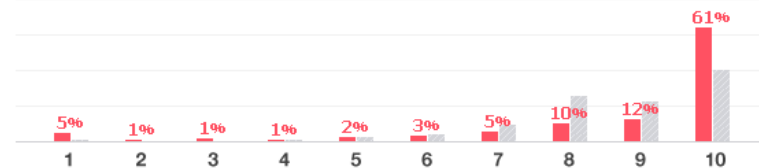
네티즌 평점 [?] ★★★★★ 8.66 3,122명

관람객 평점 [?] ★★★★★ 8.67 1,428명

점수 분포

남녀별·연령별

■ 네티즌 ■ 관람객



한줄분석이 영화를 가장 좋아하는 그룹은 20대 남자입니다.

그래프는 5분 주기로 업데이트 됩니다

★★★★★ 0 ▾

최대 한글 140자까지 가능하며, 스포일러는 삭제될 수 있습니다. 0/140

등록

140자 평 | 총 2,998건

[베스트](#) 평점 운영 정책 안내 [?]

✓ 호감순 ✓ 최신순 ✓ 평점 높은 순 ✓ 평점 낮은 순

☐ 관람객 평점만 보기 [?]

★★★★★ 10 [베스트](#) 호랑이 나올때마다 식겁했네유

비실미(chol****) | 2016.06.09 12:08 | 신고

2 데이터 설명

네이버 영화에서 140자평을 개봉전/개봉후에서 각각 추출함->csv파일
Csv파일에서 단어만 추출함->txt

성공한 영화 예시 : 아가씨 <2016>
기대와 달리 실패한 영화 예시: 디워 <2007>



2 데이터 설명 (아가씨)

네이버 영화에서 140자평을 개봉전/개봉후에서 각각 추출함->csv파일
Csv파일에서 단어만 추출함->txt

아가씨 개봉전 140자평 -> csv 파일

아가씨 개봉전 140자평 -> txt 파일

가치,각색,**감독**,감명,거기,광고,남성,노리개,**노출**,노출씬,당시,대감,대한,**동성애**,드라마,로
서,로써,**박찬욱**,**배우**,사람,**소설**,**스미스**,스토
리,여겨 던,여성,여한,영화,우리나라,**원작**
,일제강점기,입장,자극,저항,정도,정말,**주목**,폭
발,표현,**핑거**,**하정우**,해도,확정

1	star	comment
2		
3	10	T핑거스미스를 감명깊
4		
5	10	T원작 핑거스미스가 양
6		
7	5	T핑거스미스는 그제
8		
9	10	T하정우 배우 확정이라
10		
11	10	T핑거스미스 원작소설
12		
13	10	T핑거스미스를 감명깊
14		
15	10	T원작 핑거스미스가 양
16		
17	5	T핑거스미스는 그제
18		
19	10	T하정우 배우 확정이라
20		
21	10	T핑거스미스 원작소설
22		
23	10	T핑거스미스를 감명깊
24		
25	10	T원작 핑거스미스가 양
26		
27	5	T핑거스미스는 그제
28		
29	10	T하정우 배우 확정이라
30		
31	10	T핑거스미스 원작소설
32		
33	10	T핑거스미스를 감명깊

2 데이터 설명 (다워)

네이버 영화평 시스템의 변경 문제 발생

실제 스크래핑을 한 뒤 csv
파일을 확인하니 개봉전
140자 평임에도 예상과
달리 부정적인 단어가 압도
적으로 많아 확인해보니 8
년전 영화라 그런지 시스템
이 변경된 듯함.

현재는 개봉일 후로는 개봉전
140자평을 등록할 수 없
는데 이때는 개봉일이후
140자평이 존재함
(다워 개봉일 2007.08.01)

140자평 | 총9,211건

[베스트 평점 운영 정책 안내 ?](#)

✓ 호감순 ✓ 최신순

★★★★★ 1 뭐냐.. 장난하나
fpr1**** | 2007.08.08 19:47 | 신고

공감 4 비공감 1

★★★★★ 2 실망할거가튼데
up22**** | 2007.08.08 22:48 | 신고

공감 3 비공감 1

★★★★★ 2 전혀 기대 안했습니다.
exaq**** | 2007.09.25 05:27 | 신고

공감 4 비공감 2

★★★★★ 1 마케팅의 승리, 진실은 짱
binb**** | 2007.09.24 15:10 | 신고

공감 4 비공감 2

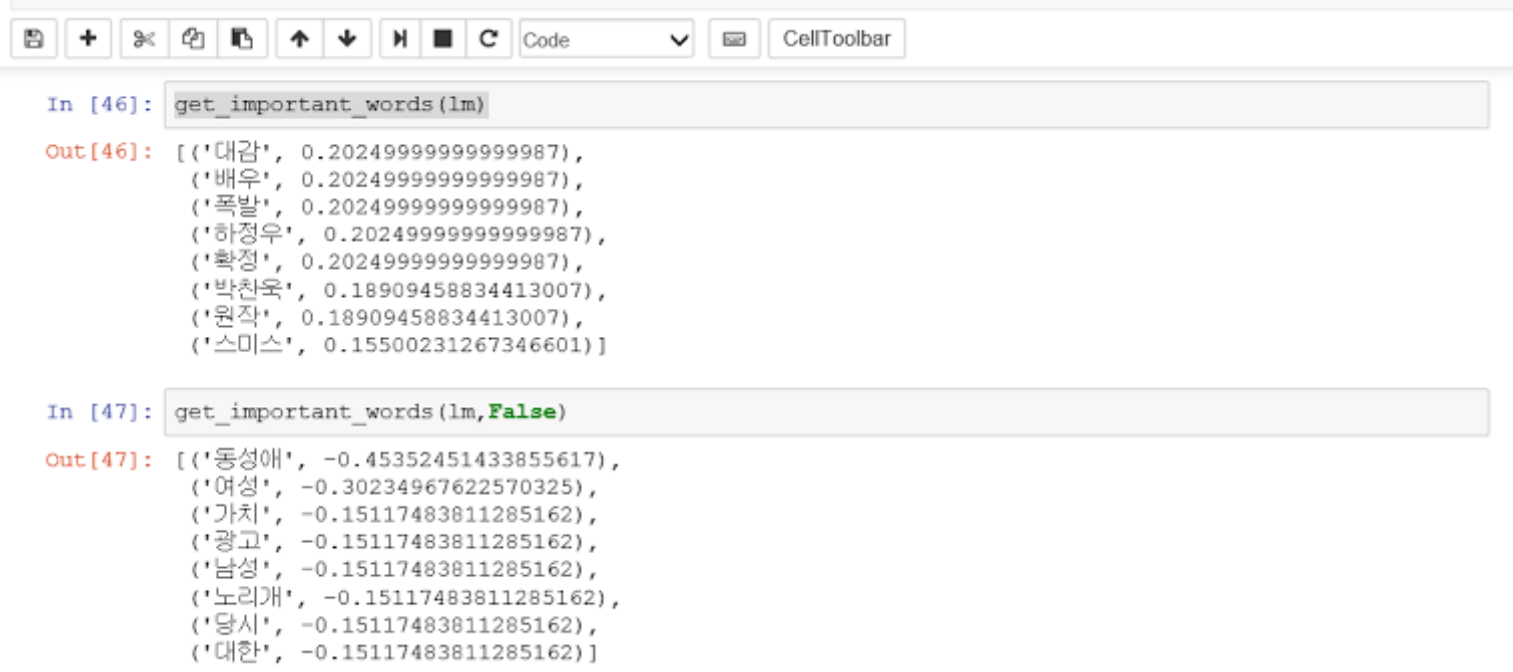
★★★★★ 1 이거 낯이면 바보
land**** | 2007.09.23 14:20 | 신고

공감 4 비공감 2

3 실제분석

개봉전 단어 추출한 txt파일에서 중요도 분석을 실행함

아가씨개봉전 단어중요도분석



```
In [46]: get_important_words(lm)

Out[46]: [('대감', 0.20249999999999987),
          ('배우', 0.20249999999999987),
          ('폭발', 0.20249999999999987),
          ('하정우', 0.20249999999999987),
          ('확정', 0.20249999999999987),
          ('박찬욱', 0.18909458834413007),
          ('원작', 0.18909458834413007),
          ('스미스', 0.15500231267346601)]

In [47]: get_important_words(lm, False)

Out[47]: [('동성애', -0.45352451433855617),
          ('여성', -0.30234967622570325),
          ('가치', -0.15117483811285162),
          ('광고', -0.15117483811285162),
          ('남성', -0.15117483811285162),
          ('노리개', -0.15117483811285162),
          ('당시', -0.15117483811285162),
          ('대한', -0.15117483811285162)]
```


3 실제분석

개봉후 단어 추출한 txt파일에서 중요도 분석을 실행함

아가씨개봉후단어중요도분석



```
In [72]: get_important_words(lm)
```

```
Out[72]: [('하정우', 0.380625000000000027),  
          ('김태', 0.345156250000000177),  
          ('리라', 0.345156250000000177),  
          ('신인', 0.345156250000000177),  
          ('정말', 0.345156250000000177),  
          ('관람객', 0.076124999999999207),  
          ('구원', 0.076124999999999207),  
          ('마코', 0.076124999999999207)]
```

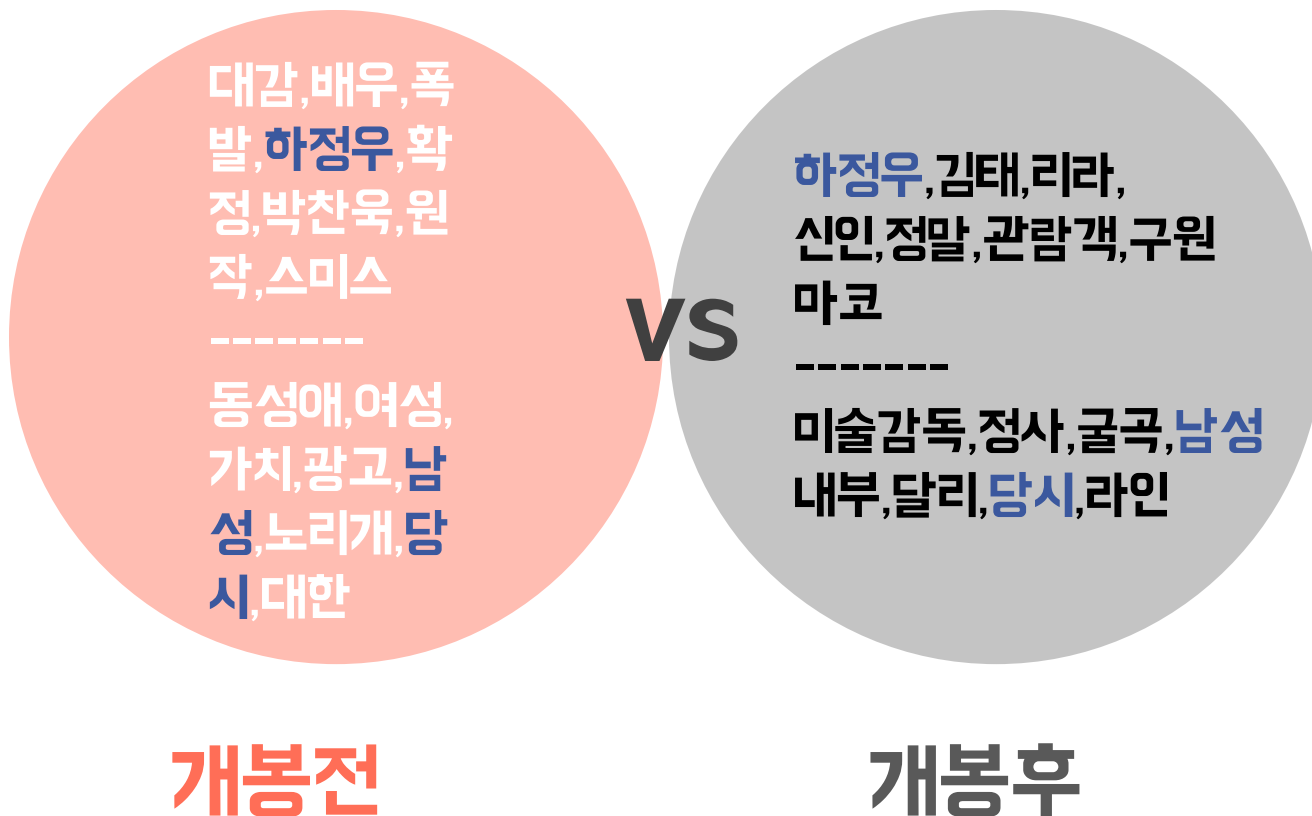
```
In [73]: get_important_words(lm, False)
```

```
Out[73]: [('미술감독', -1.6193749999999998),  
          ('정사', -0.058988095238130513),  
          ('굴곡', -0.029494047619065256),  
          ('남성', -0.029494047619065256),  
          ('내부', -0.029494047619065256),  
          ('달리', -0.029494047619065256),  
          ('당시', -0.029494047619065256),  
          ('라인', -0.029494047619065256)]
```

3 실제분석

개봉전후의 단어 중요도 비교

아가씨 단어중요도 분석



3 실제분석

개봉전 단어 추출한 txt파일에서 중요도 분석을 실행함

디워개봉전후단어중요도분석

```
In [25]: get_important_words(lm)
```

```
Out[25]: [('거가', 0.5862500000000005),  
          ('기대', 0.5862500000000005),  
          ('낮이', -0.20687500000000006),  
          ('바보', -0.20687500000000006),  
          ('승리', -0.20687500000000009),  
          ('장난', -0.20687500000000009),  
          ('진실', -0.20687500000000009),  
          ('하나', -0.20687500000000009)]
```

```
In [26]: get_important_words(lm, False)
```

```
Out[26]: [('승리', -0.20687500000000009),  
          ('장난', -0.20687500000000009),  
          ('진실', -0.20687500000000009),  
          ('하나', -0.20687500000000009),  
          ('낮이', -0.20687500000000006),  
          ('바보', -0.20687500000000006),  
          ('거가', 0.5862500000000005),  
          ('기대', 0.5862500000000005)]
```

3 실제분석

아가씨 영화가 개봉된 이후 양상이 조금 달라진 것을 감지 -> 분석
아가씨 개봉이후 기사 스크래핑

아가씨 개봉일 이후 네이버 기 사 스크래핑 (단어추출)

[('영화', 100), ('아가씨', 63), ('감독', 45), ('박찬욱', 27), ('이야기', 27), ('개봉', 23),
('하나', 22), ('조진웅', 20), ('관객', 19), ('영상', 19), ('원작', 19), ('하정우', 19),
('스크린', 18), ('경제', 16), ('사람', 16), ('금지', 15), ('기자', 15), ('무단', 15),
('배포', 15), ('전재', 15), ('다운로드', 14), ('소설', 14), ('공개', 13), ('뉴스', 13),
('의상', 13), ('돌파', 12), ('배우', 12), ('사진', 12), ('흥행', 12), ('김민희', 11),
('연기', 11), ('고객', 10), ('부분', 10), ('사냥', 10), ('스미스', 10), ('인기', 10),
('일본', 10), ('포스터', 10), ('핑거', 10), ('데코', 9), ('마음', 9), ('바흐', 9), ('보이', 9),
('작품', 9), ('캐릭터', 9), ('콘텐츠', 9), ('한국', 9), ('그때', 8), ('위해', 8), ('트렌드', 8)]

실제분석

아가씨 개봉이후 기사 스크래핑

아가씨 개봉일 이후 네이버 기사를 이용한 워드클라우드



3 실제분석

아가씨 영화가 개봉된 이후 양상이 조금 달라진 것을 감지 -> 분석
아가씨 개봉이후 기사 스크래핑

아가씨 개봉일 이후 기사 주요 언급단어 분석

('영화', 100), ('아가씨', 63), '영화 아가씨'로 검색했기 때문

('감독', 45), ('박찬욱', 27), 유명감독에 대한 언급이 가장 잦음

('조진웅', 20), ('하정우', 19), ('김민희', 11), ('배우', 12) : 유명 주연 배우에 대한 언급

('원작', 19), ('소설', 14), ('스미스', 10), ('핑거', 10),: 원작 '핑거 스미스'가 존재함

('연기', 11): 주연배우들의 뛰어난 연기력

('금지', 15), ('기자', 15), ('무단', 15), ('배포', 15), ('전재', 15), ('다운로드', 14): 인터넷 신문 기사를
클릭할 때 발생하는 부분 (영화와 무관)

4 결과해석

예시1. 영화 아가씨 (2016) 박찬욱

영화 아가씨

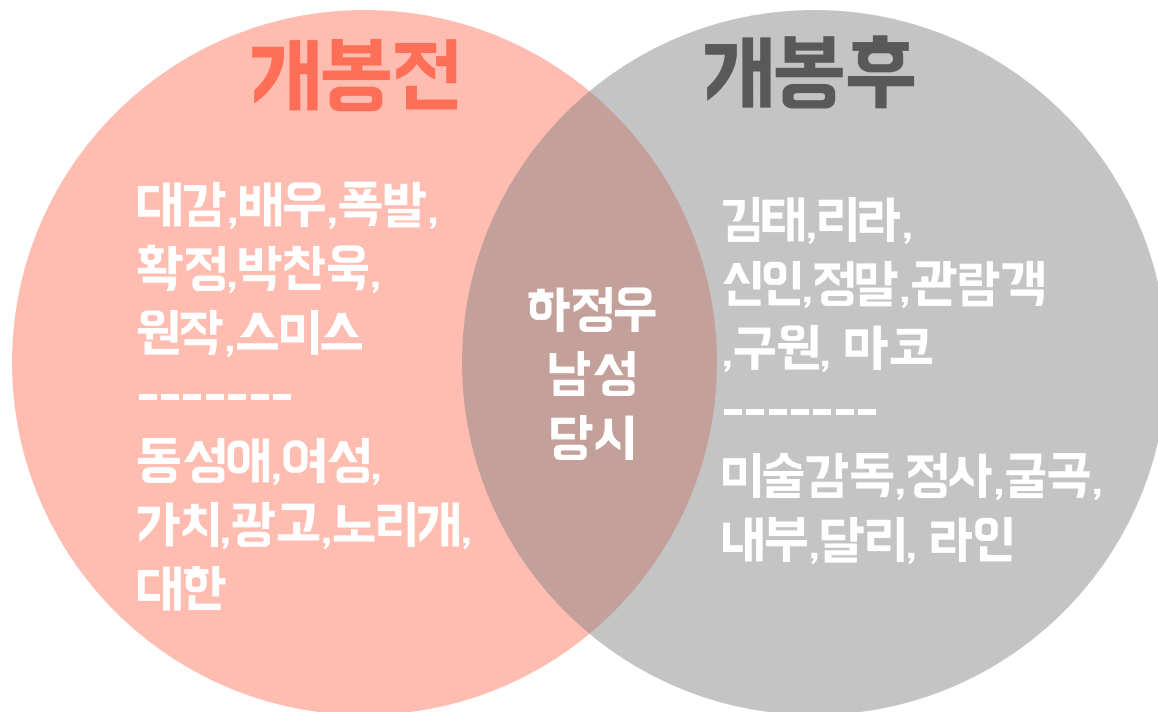
개봉전 평점 7.64 / 개봉후 7.32

주연배우

김민희, 김태리, 하정우, 조진웅

현재 성적

2016.6월 3주차 기준 예매율 1위



핵심 키워드는 크게 벗어나지 않음, 사람들이 생각했던 중요 키워드는 **대부분 일치**하는 것으로 판명. 단, **개봉 전에는 주연배우 중 하정우의 언급이 대다수였던 반면 개봉후에 눈에 띄게 신인 배우 김태리에 대한 언급이 급증함.**

새로 등장한 구원(아가씨 대사: 내인생을 망치려온 나의 구원자, 나의 타마코, 나의 숙희.), 신인 등도 김태리 관련 키워드임.

4 결과해석

예시1. 영화 아가씨 (2016) 박찬욱

영화 아가씨

개봉전 평점 7.64 / 개봉후 7.32

주연배우

김민희, 김태리, 하정우, 조진웅

현재 성적

2016.6월 3주차 기준 예매율 1위

신문기사

영화, 아가씨,
감독, 박찬욱,
조진웅, 하정우,
김민희, 배우,
원작, 소설,
스미스, 핑거, 연기

하지만 관람객들의 관람평과는 달리 실제 **신문기사는** 생각보다 **김태리에** 대한 언급이 많지 **않았음**. 실제로 김태리 단독 인터뷰 기사가 많지 않음.

워낙 유명한 감독인 박찬욱에 대한 언급과 원작이 있는 작품이라 원작 자체에 대한 언급, 그리고 비교적 유명배우인 하정우, 조진웅, 김민희에 대한 언급이 많았음.



감사합니다.
