



B조

# 국내 영화 관객수 예측 분석

---

로렌스 , 김용제 , 김종완 , 손병구

---

국내 영화 흥행 예측 분석

# INDEX

01

분석 Tool 소개

02

프로젝트 기간 및 일정

03

한국 영화 산업 소개

04

사용 데이터 & 분석방법

05

GUI구동

06

한계점 & 개선방안



01

# 분석 Tool 소개

## 분석 Tool 소개

IDE(개발도구)

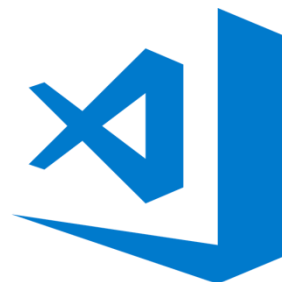
Jupyter Notebook , Visual Studio Code

프로그래밍 언어

Python 3

사용된 Python Library

Pandas, numpy, matplotlib, sklearn 등

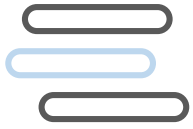




02

## 프로젝트 기간 및 일정

## 프로젝트 기간 및 일정



7월 29일 ~ 8월 2일 (5일)

수행방안 설계

7월 29일

데이터 수집 및  
정제

7월 30일 ~ 31일

데이터 분석과  
GUI

8월 1일

PPT작성

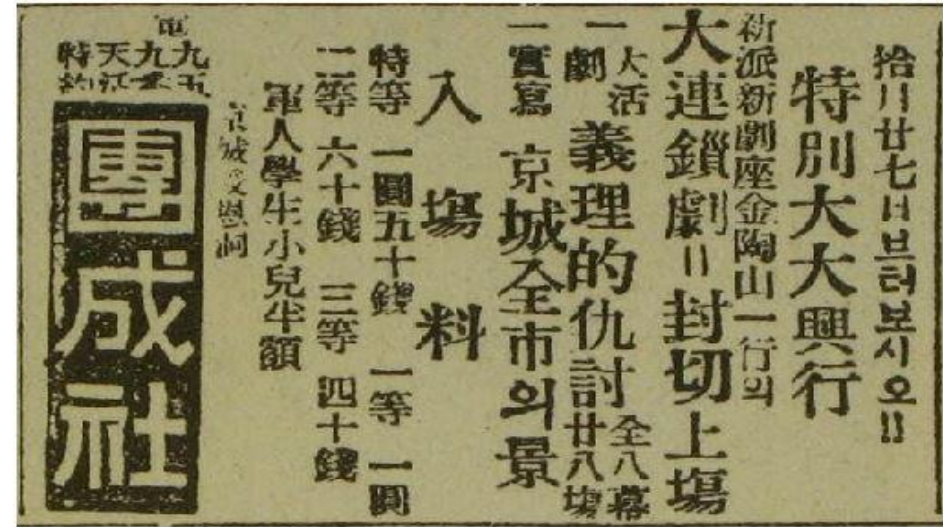
8월 2일



03

# 한국 영화 산업 소개

## 한국 영화 산업 소개



- 조선 말엽 움직이는 사진인 ‘활동사진’이 들어오면서 처음 소개.
- 한국인의 손으로 처음 활동사진이 제작된 것은 1923년
- 한국 최초로 만들어진 영화이자 연쇄극으로는 의리적 구투가 있다.



# 한국 영화 산업 소개

## 2017 Top 20 International Box Office Markets – All Films (US\$ Billions)

Source: IHS Markit, local sources

1. China	\$7.9	11. Brazil	\$0.9
2. Japan	\$2.0	12. Spain	\$0.7
3. U.K.	\$1.6	13. Italy	\$0.7
4. India	\$1.6	14. Taiwan	\$0.4
5. South Korea	\$1.6	15. Netherlands	\$0.3
6. France	\$1.5	16. Indonesia	\$0.3
7. Germany	\$1.2	17. Argentina	\$0.3
8. Russia	\$1.0	18. Poland	\$0.3
9. Australia	\$0.9	19. Turkey	\$0.2
10. Mexico	\$0.9	20. Hong Kong	\$0.2

## 세계 영화 시장

5위

## 2018 Top 20 International Box Office Markets – All Films (US\$ Billions)

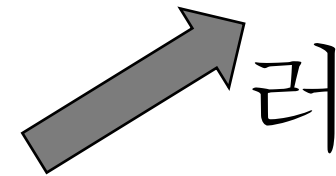
Source: IHS Markit, local sources

1. China <sup>7</sup>	\$9.0	11. Spain	\$0.7
2. Japan	\$2.0	12. Brazil	\$0.7
3. U.K.	\$1.7	13. Italy	\$0.7
4. South Korea	\$1.6	14. Netherlands	\$0.4
5. France	\$1.6	15. Indonesia	\$0.4
6. India	\$1.5	16. Taiwan	\$0.3
7. Germany	\$1.0	17. Poland	\$0.3
8. Australia	\$0.9	18. UAE	\$0.3
9. Mexico	\$0.9	19. Malaysia	\$0.3
10. Russia	\$0.9	20. Hong Kong	\$0.3

출처 : 미국영화협회(MPAA),

<https://www.mpa.org/wp-content/uploads/2019/03/MPAA-THEME-Report-2018.pdf>,  
<https://www.mpa.org/wp-content/uploads/2019/03/MPAA-THEME-Report-2017.pdf>

16억 달러



14억 달러



04

# 사용 데이터 & 분석방법

## 사용 데이터 & 분석방법

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
순위	영화명	매출액	누적매출액	관객수	누적관객수	스크린수	개별배급사	등급	개별장르	개별감독	개별배우	필	년	개별배급사_index	등급_index	개별장르_index	개별감독_index	개별배우_index
1	극한직업	1.39647E+11	1.39647E+11	16264806	16264806	2003	씨제이이엔엠(주)15세관람가	코미디	이병헌	류승룡		1	2019	0	0	0	0	0
1	극한직업	1.39647E+11	1.39647E+11	16264806	16264806	2003	씨제이이엔엠(주)15세관람가	코미디	이병헌	이하늬		1	2019	0	0	0	0	1
1	극한직업	1.39647E+11	1.39647E+11	16264806	16264806	2003	씨제이이엔엠(주)15세관람가	코미디	이병헌	진선규		1	2019	0	0	0	0	2
1	극한직업	1.39647E+11	1.39647E+11	16264806	16264806	2003	씨제이이엔엠(주)15세관람가	코미디	이병헌	이동휘		1	2019	0	0	0	0	3
1	극한직업	1.39647E+11	1.39647E+11	16264806	16264806	2003	씨제이이엔엠(주)15세관람가	코미디	이병헌	공명		1	2019	0	0	0	0	4
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	송강호		5	2019	0	0	1	1	5
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	이선균		5	2019	0	0	1	1	6
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	조여정		5	2019	0	0	1	1	7
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	최우식		5	2019	0	0	1	1	8
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	박소담		5	2019	0	0	1	1	9
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	장혜진		5	2019	0	0	1	1	10
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	이정은		5	2019	0	0	1	1	11
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	박명환		5	2019	0	0	1	1	12
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	현승민		5	2019	0	0	1	1	13
2	기생충	85567642365	85567642365	10047414	10047414	1948	씨제이이엔엠(주)15세관람가	드라마	봉준호	정현준		5	2019	0	0	1	1	14
3	돈	28869614062	28869614062	3389035	3389035	1431	(주)쇼박스 15세관람가	범죄	박누리	류준열		3	2019	1	0	2	2	15
3	돈	28869614062	28869614062	3389035	3389035	1431	(주)쇼박스 15세관람가	범죄	박누리	유지태		3	2019	1	0	2	2	16
3	돈	28869614062	28869614062	3389035	3389035	1431	(주)쇼박스 15세관람가	범죄	박누리	조우진		3	2019	1	0	2	2	17
3	돈	28869614062	28869614062	3389035	3389035	1431	(주)쇼박스 15세관람가	범죄	박누리	원진아		3	2019	1	0	2	2	18
3	돈	28869614062	28869614062	3389035	3389035	1431	(주)쇼박스 15세관람가	범죄	박누리	박명신		3	2019	1	0	2	2	19
3	돈	28869614062	28869614062	3389035	3389035	1431	(주)쇼박스 15세관람가	범죄	박누리	이화룡		3	2019	1	0	2	2	20
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)에이스메이커18세관람가	범죄	이원태	마동석		5	2019	2	1	2	3	21
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)에이스메이커18세관람가	액션	이원태	마동석		5	2019	2	1	3	3	21
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)키위미디어218세관람가	범죄	이원태	마동석		5	2019	3	1	2	3	21
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)키위미디어218세관람가	액션	이원태	마동석		5	2019	3	1	3	3	21
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)에이스메이커18세관람가	범죄	이원태	김무열		5	2019	2	1	2	3	22
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)에이스메이커18세관람가	액션	이원태	김무열		5	2019	2	1	3	3	22
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)키위미디어218세관람가	범죄	이원태	김무열		5	2019	3	1	2	3	22
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)키위미디어218세관람가	액션	이원태	김무열		5	2019	3	1	3	3	22
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)에이스메이커18세관람가	범죄	이원태	김성규		5	2019	2	1	2	3	23
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)에이스메이커18세관람가	액션	이원태	김성규		5	2019	2	1	3	3	23
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)키위미디어218세관람가	범죄	이원태	김성규		5	2019	3	1	2	3	23
4	악인전	29130540150	29130540150	3364360	3364360	1307	(주)키위미디어218세관람가	액션	이원태	김성규		5	2019	3	1	3	3	23

## 2003 ~ 2019 한국 영화

개별장르 : 한 영화 내의 장르를 분류  
 개별감독 : 한 영화 내의 감독을 분류  
 개별배우 : 한 영화 내의 배우를 분류

## 사용 데이터 & 분석방법

```
In [12]: 개별배급사_index = []

for i in range(len(csv)):
    for j in range(len(개별배급사)):
        if csv['개별배급사'][i] == 개별배급사[j]:
            개별배급사_index.append(j)
```

```
In [9]: 등급_index = []

for i in range(len(csv)):
    for j in range(len(등급)):
        if csv['등급'][i] == 등급[j]:
            등급_index.append(j)
```

```
In [10]: 개별장르_index = []

for i in range(len(csv)):
    for j in range(len(개별장르)):
        if csv['개별장르'][i] == 개별장르[j]:
            개별장르_index.append(j)
```

```
In [11]: 개별감독_index = []

for i in range(len(csv)):
    for j in range(len(개별감독)):
```

## 사용 데이터 & 분석방법

```
In [271]: # train data 불러오기  
train = train_data
```

```
In [272]: test = test_data
```

```
In [273]: categorical_feature_name = ['월', '개별배급사_index', '등급_index', '개별장르_index', '개별감독_index', '개별배우_index', '년', '일_index']  
for i in categorical_feature_name:  
    train[i] = train[i].astype('category')  
    test[i] = test[i].astype("category")
```

```
In [274]: # 필요한 Feature 구성  
feature_name = ['월', '개별배급사_index', '등급_index', '개별장르_index', '개별감독_index', '개별배우_index', '년', '일_index']
```

## 사용 데이터 & 분석방법

```
In [21]: from sklearn.neighbors import KNeighborsRegressor # KNN
model = KNeighborsRegressor(n_neighbors=1);
Y_train_log = np.log1p(train_label)
model.fit(X_train,Y_train_log)

knn_preds=model.predict(X_test)
knn_preds=np.exp(knn_preds)
knn_score = rmsle(test_label,knn_preds,False)
```

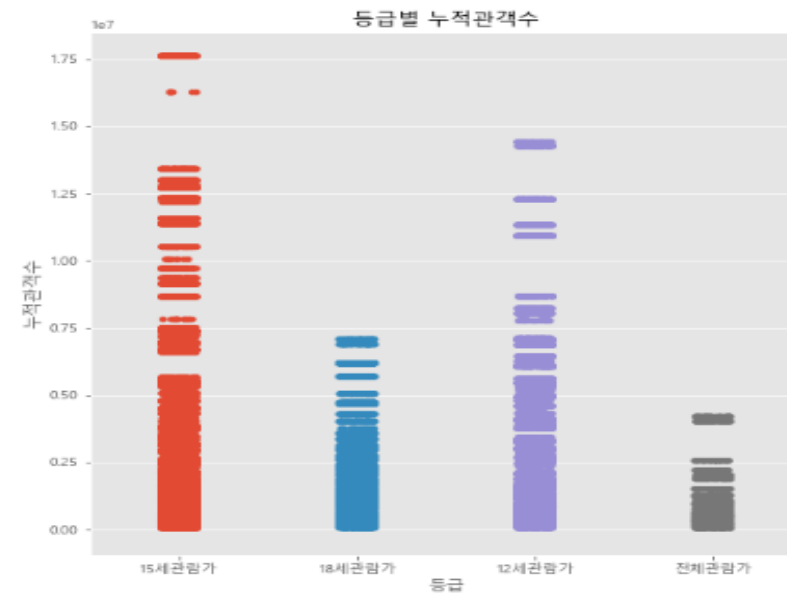
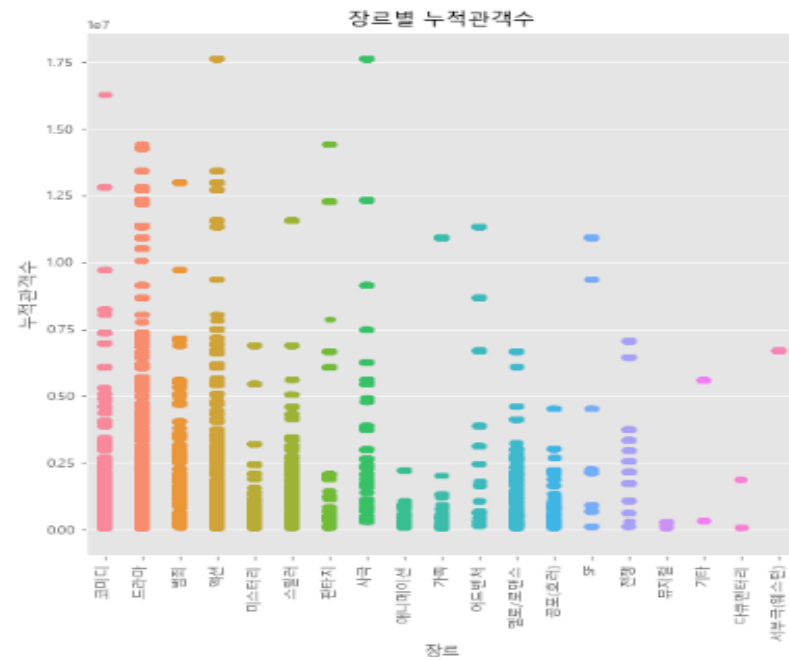
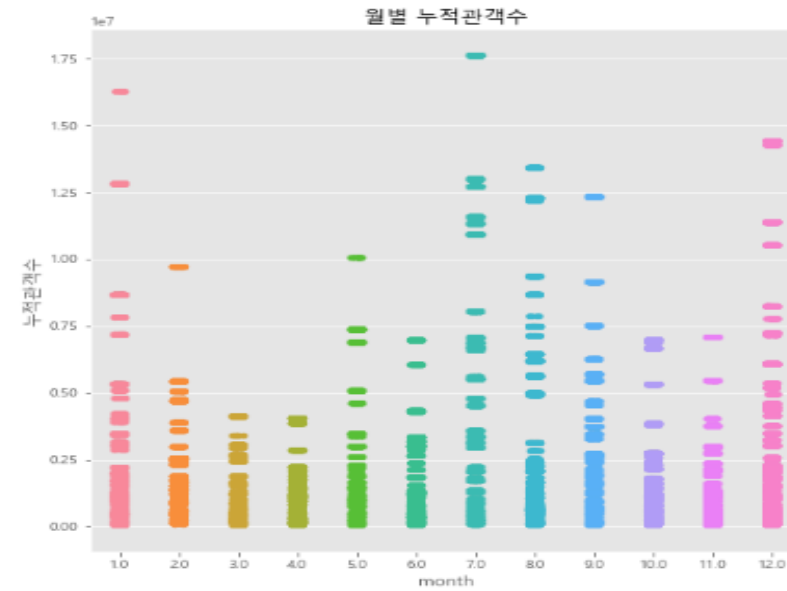
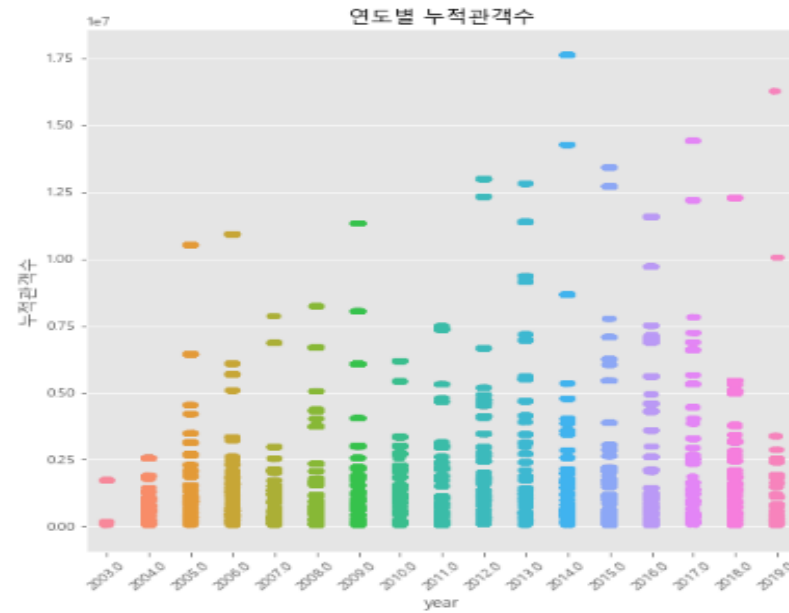
```
In [19]: csv2['KNN']=knn_preds
```

```
In [20]: csv2[['누적관객수','KNN']]
```

Out[20]:

	누적관객수	KNN
0	16264806	9351215.0
1	16264806	9351215.0
2	16264806	9351215.0
3	16264806	9351215.0
4	16264806	9351215.0
5	10047414	9351215.0
6	10047414	9351215.0

## 사용 데이터 & 분석방법



## 사용 데이터 & 분석방법



## 배급사 WordCloud





# 배우 사진



오달수



김의성



박진우



김윤석



류승룡



유승목



윤제문



이경영



장준녕



최진호





감독 WordCloud



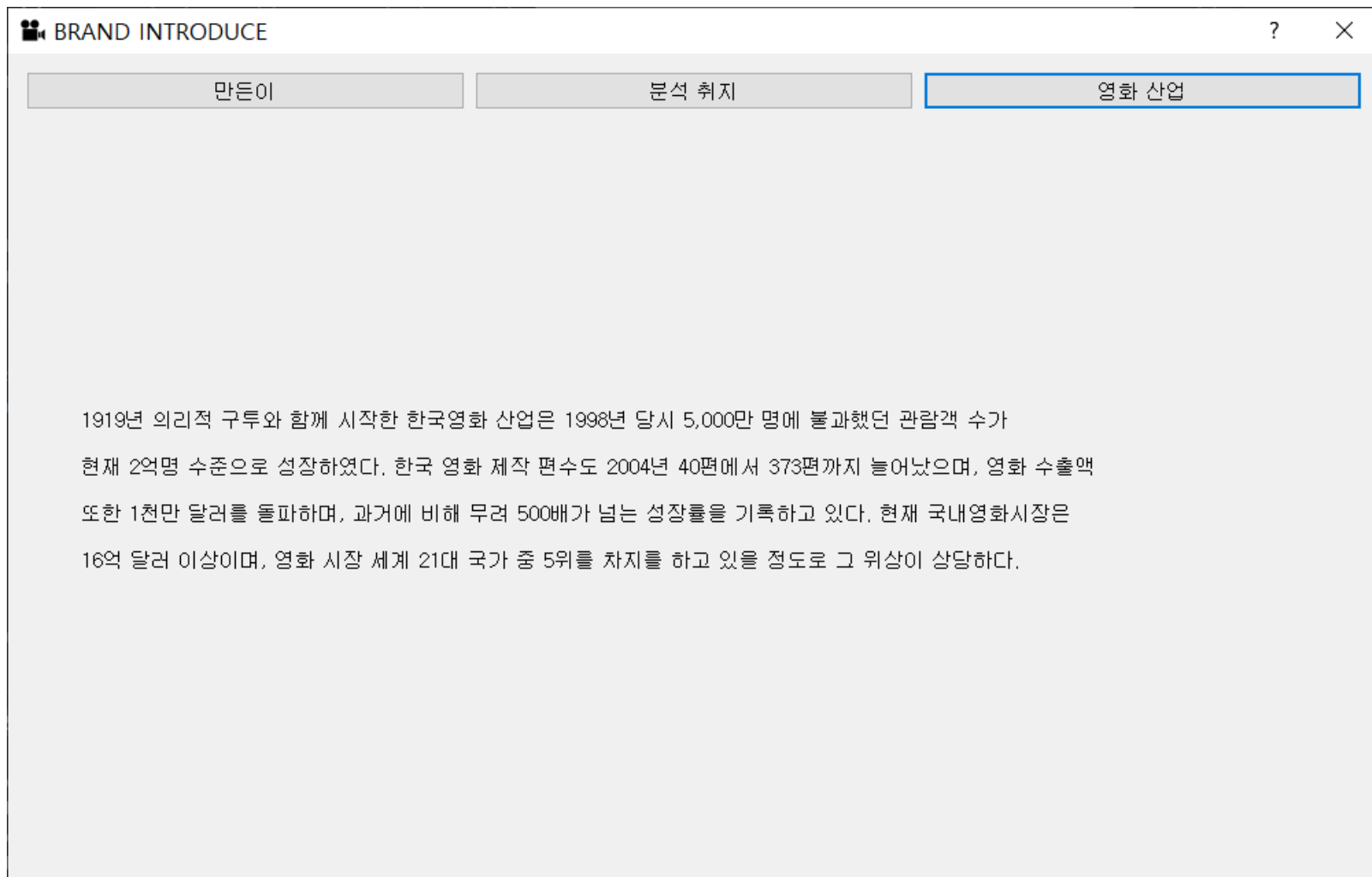
05

GUI 구동

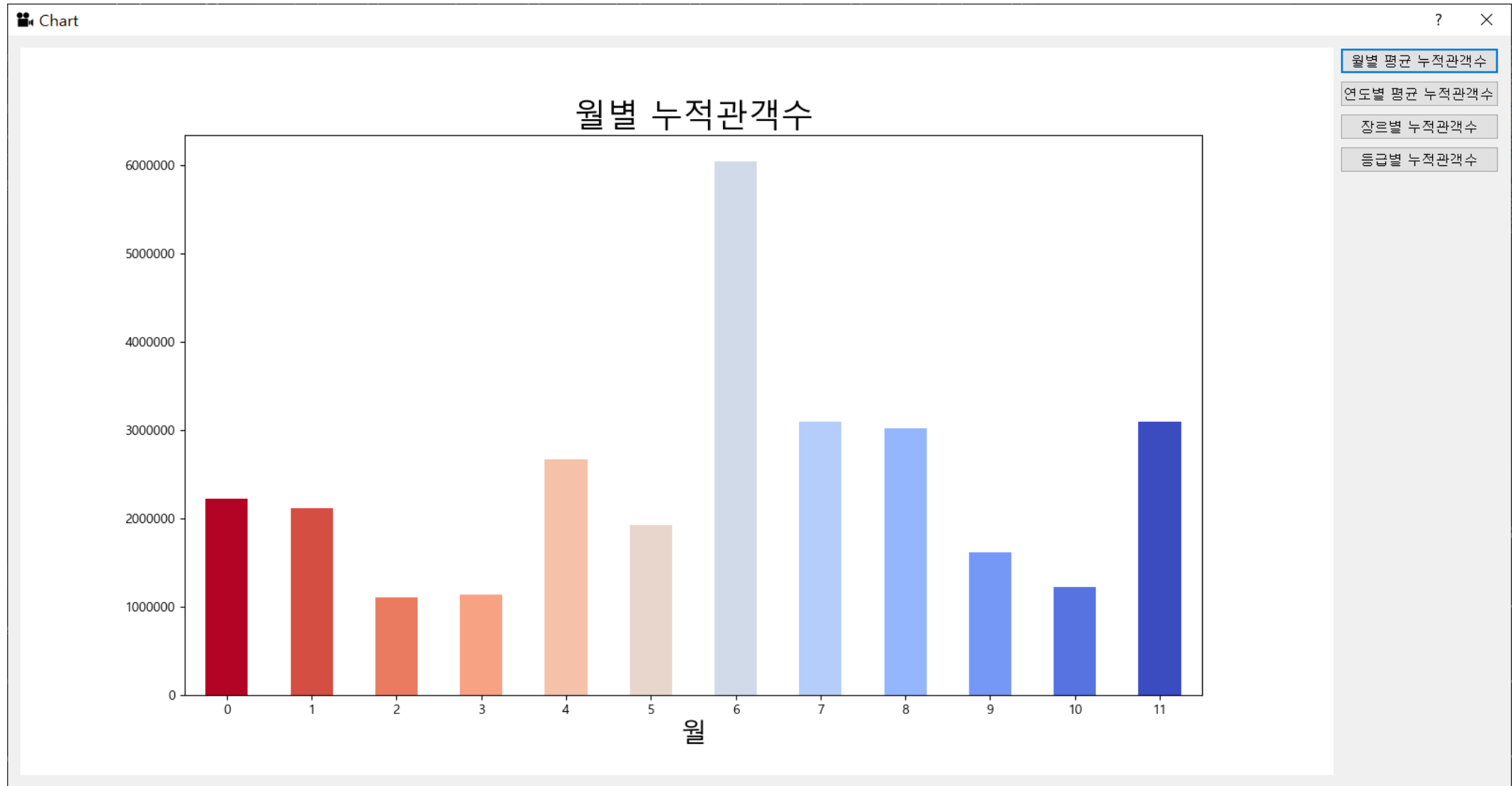
## GUI 구동 & 결과



## GUI 구동 & 결과



## GUI 구동 & 결과

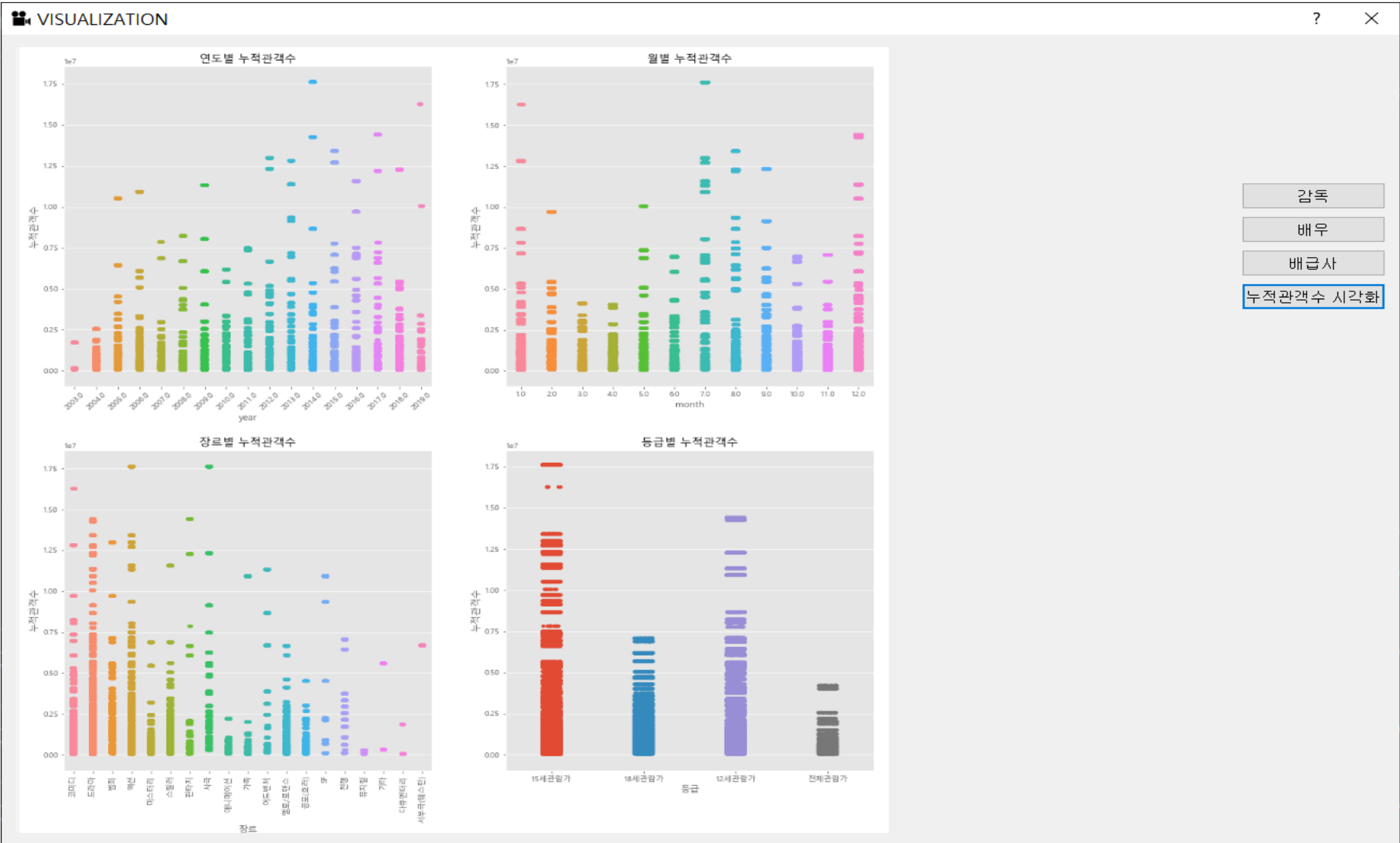


## GUI 구동 & 결과





# GUI 구동 & 결과

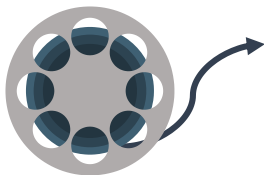




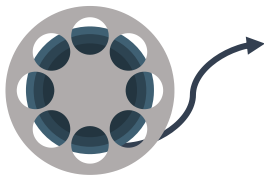
06

# 한계점 & 개선 방안

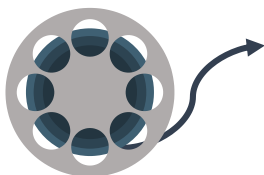
## 한계점 & 개선방안



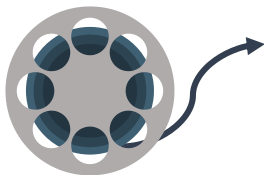
1. 주·조연과 배우를 구별하는 컬럼을 만들어 좀 더 정밀한 분석



2. 자료의 기간을 좀 더 확장시켜서 모델링



3. 순위파트에서 누적관객수가 중복값을 포함한 것 개정



4. 차트에서 x축을 숫자에서 컬럼명으로 변경.



감사합니다.