IMDB DataSet Analysis

Presented By: Deepthi V
EJ H
Hritik S
Sandhya B

What is your favourite movie?

Objectives

1. To analyse the perception of different movies across genres by analysing their rating, metascore, gross amount etc.
2. To identify top actors and top directors loved by audience.
3. To find insightful informations using data visualization.

Understanding The Dataset

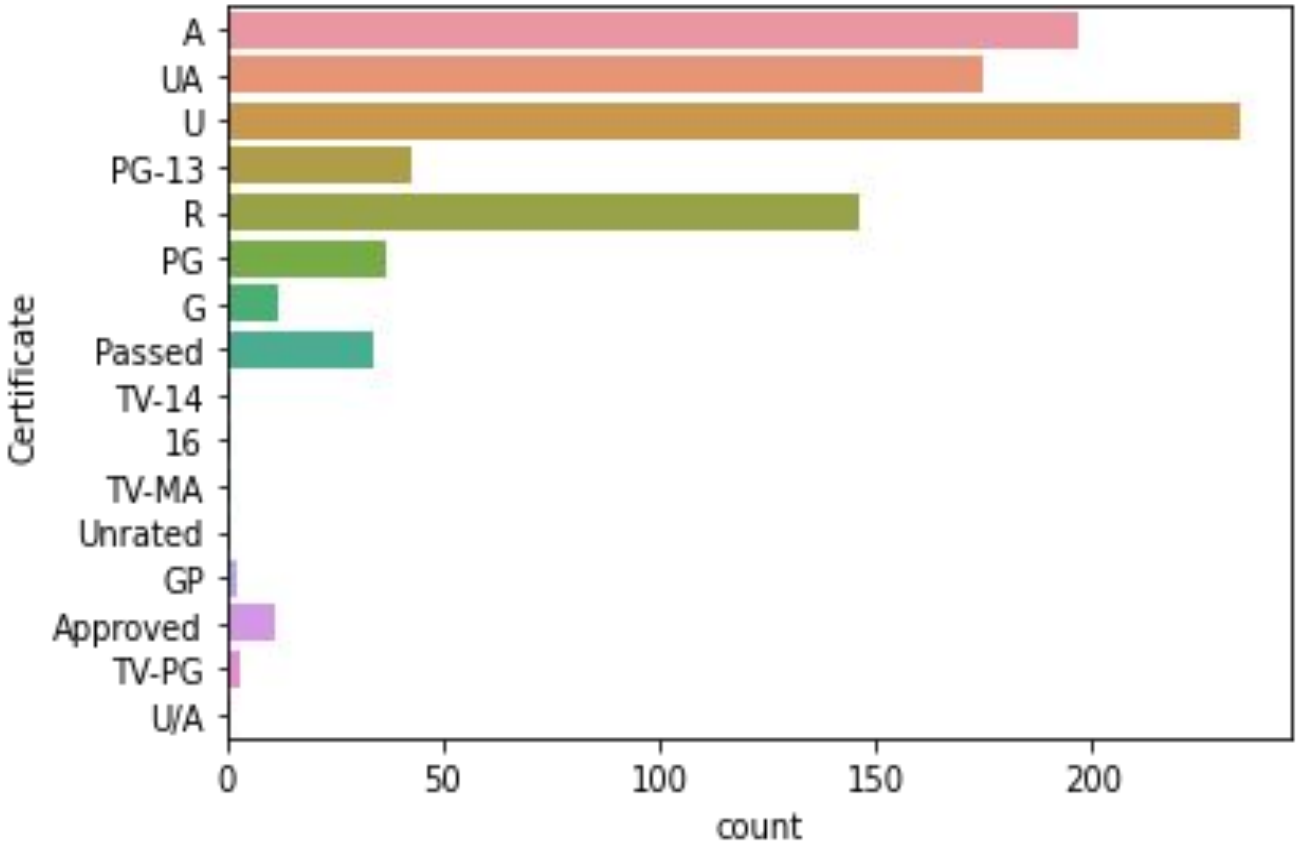
- IMDB Movies dataset : https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows
- Shape of dataset 1000x16
- Total 16 columns
- Average runtime is 122 min
- 202 unique genres

Column names along with there data types before and after cleaning are shown

Poster_Link	object
Series_Title	object
Released_Year	object
Certificate	object
Runtime	object
Genre	object
IMDB_Rating	float64
Overview	object
Meta_score	float64
Director	object
Star1	object
Star2	object
Star3	object
Star4	object
No_of_Votes	int64
Gross	object

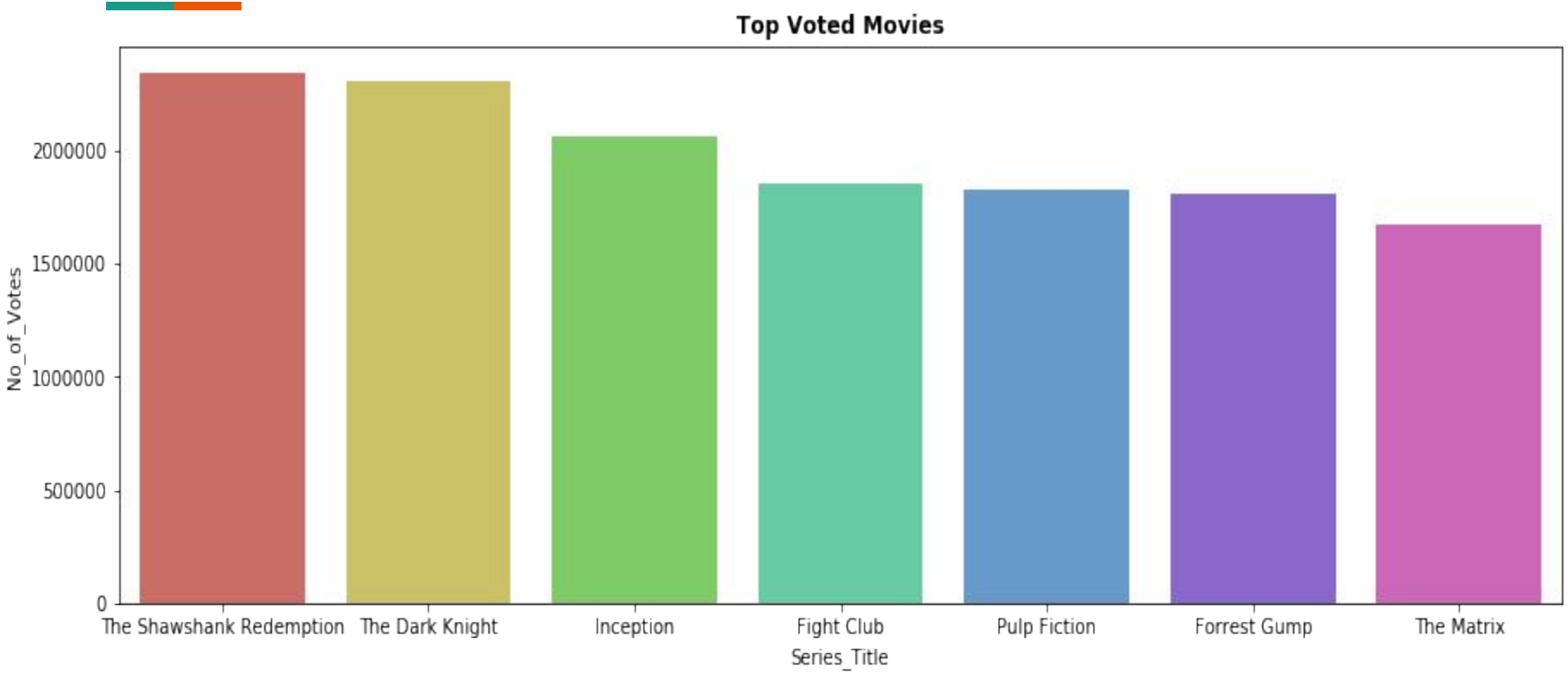
Poster_Link	object
Series_Title	object
Released_Year	object
Certificate	object
Runtime	int64
Genre	object
IMDB_Rating	float64
Overview	object
Meta_score	float64
Director	object
Star1	object
Star2	object
Star3	object
Star4	object
No_of_Votes	int64
Gross	float64

	Runtime	IMDB_Rating	Meta_score	No_of_Votes	Gross
count	1000.000000	1000.000000	1000.000000	1.000000e+03	1.000000e+03
mean	122.891000	7.949300	65.730000	2.736929e+05	5.653688e+07
std	28.093671	0.275491	30.570208	3.273727e+05	1.032382e+08
min	45.000000	7.600000	0.000000	2.508800e+04	0.000000e+00
25%	103.000000	7.700000	63.000000	5.552625e+04	4.457098e+05
50%	119.000000	7.900000	76.000000	1.385485e+05	1.070275e+07
75%	137.000000	8.100000	85.250000	3.741612e+05	6.153989e+07
max	321.000000	9.300000	100.000000	2.343110e+06	9.366622e+08

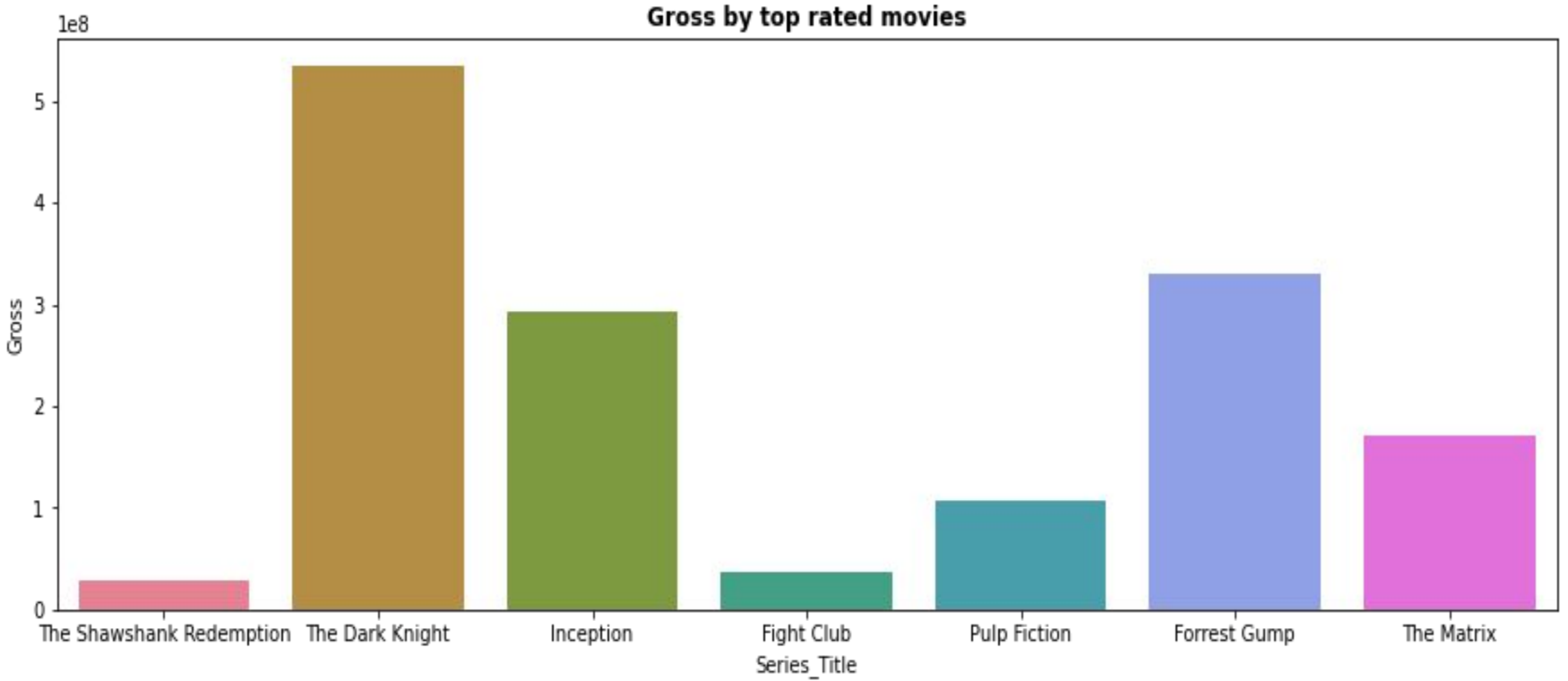


U	234
A	197
UA	175
R	146
PG-13	43
PG	37
Passed	34
G	12
Approved	11
TV-PG	3
GP	2
TV-14	1
16	1
TV-MA	1
Unrated	1
U/A	1

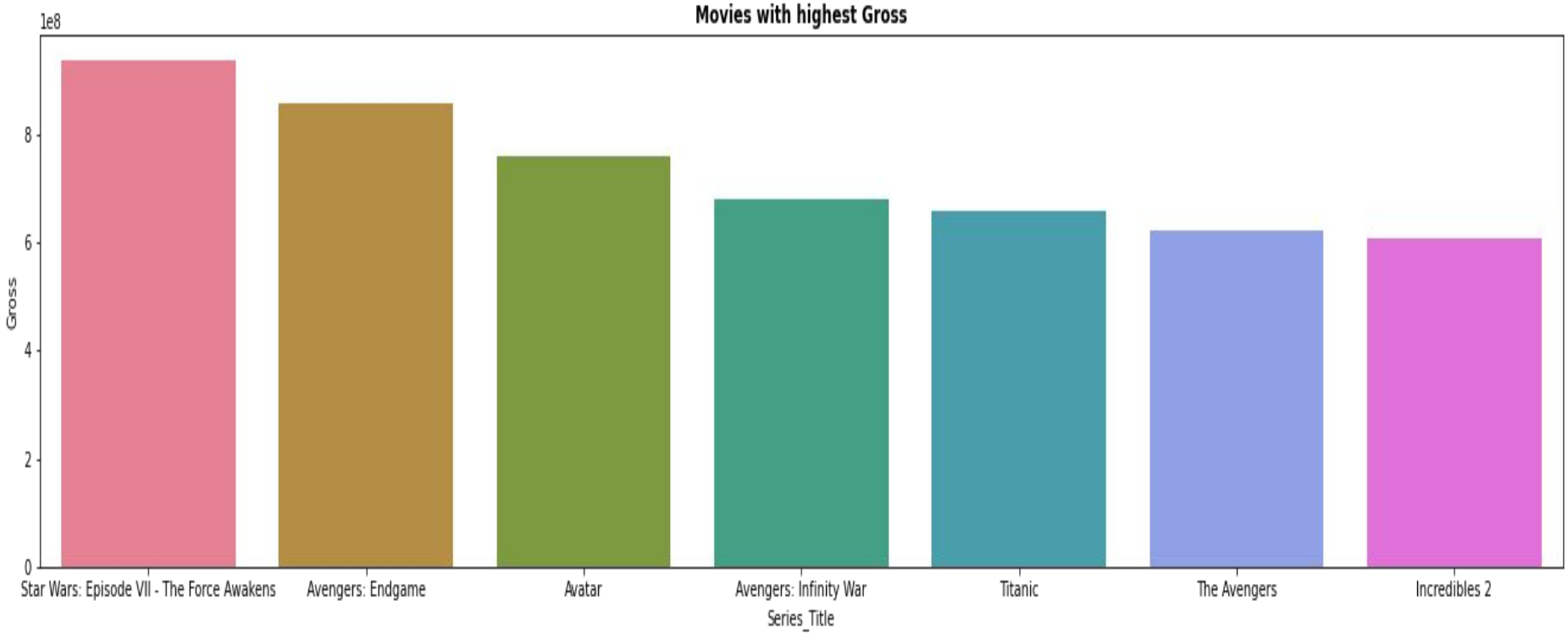
Top Voted Movies



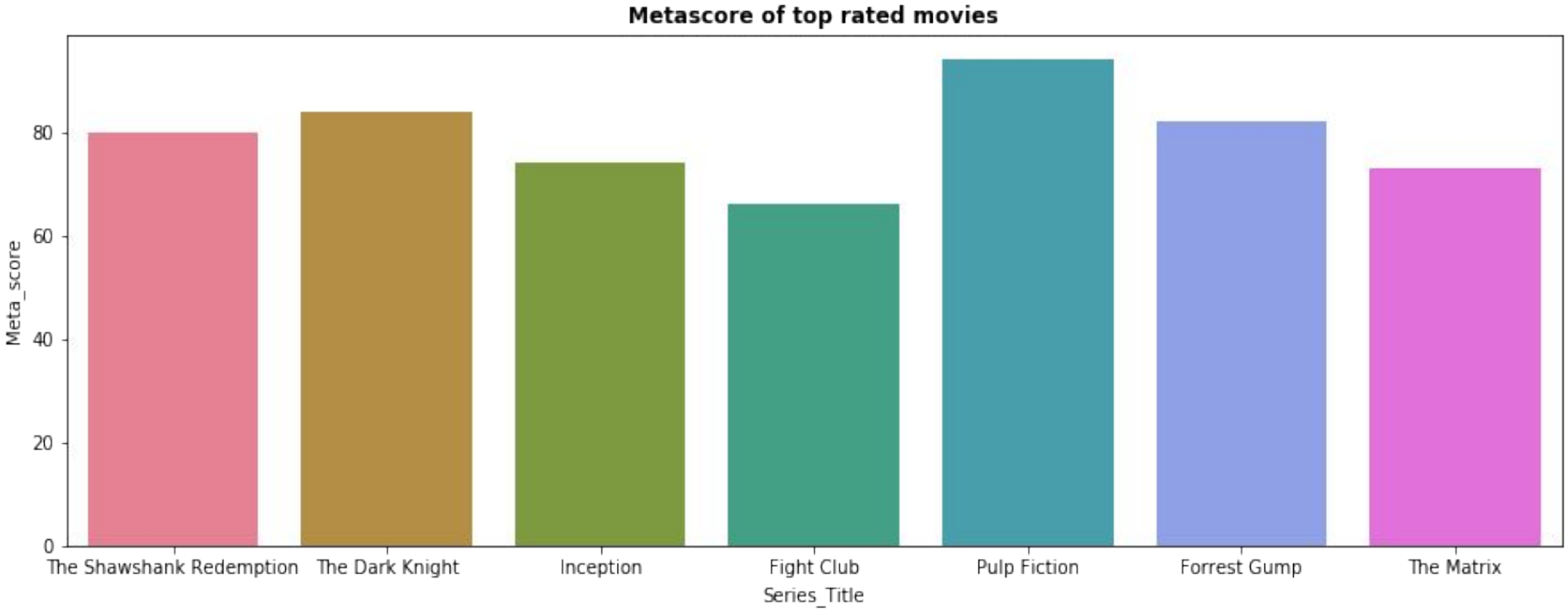
Gross by top rated Movies



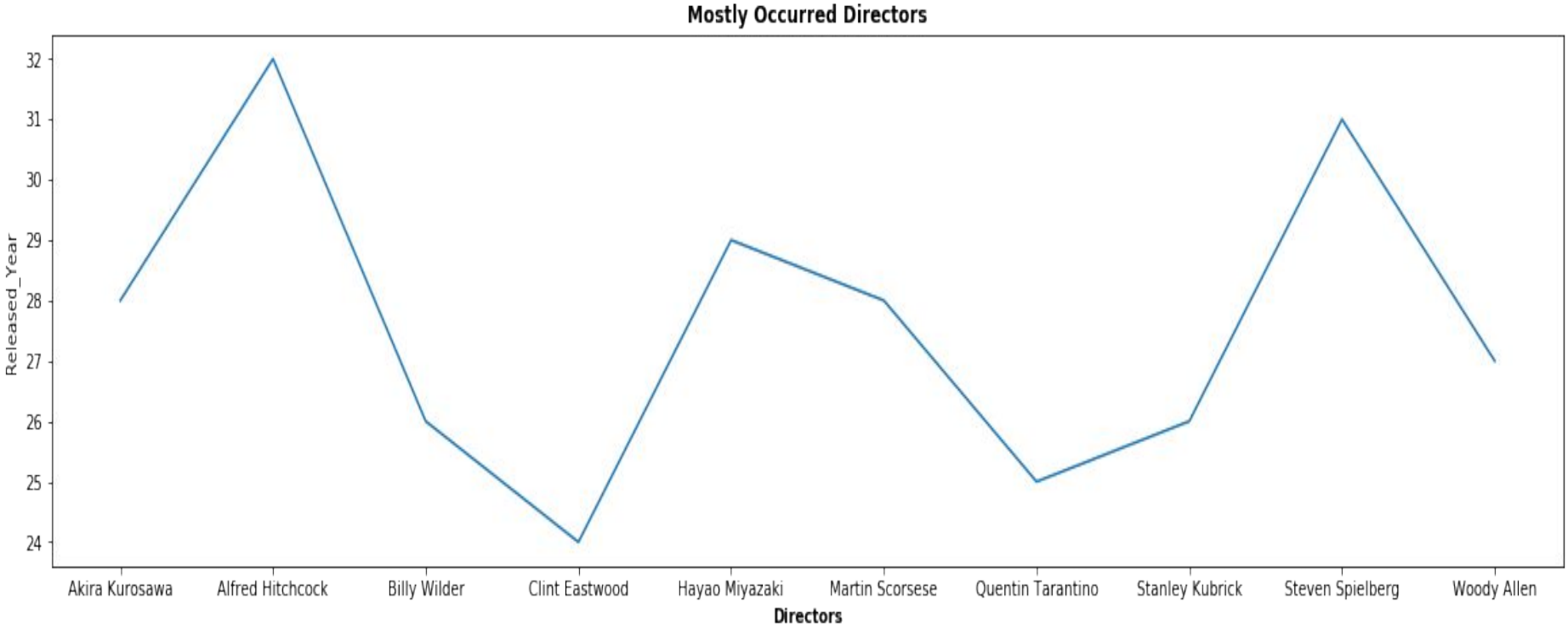
Movies with highest Gross



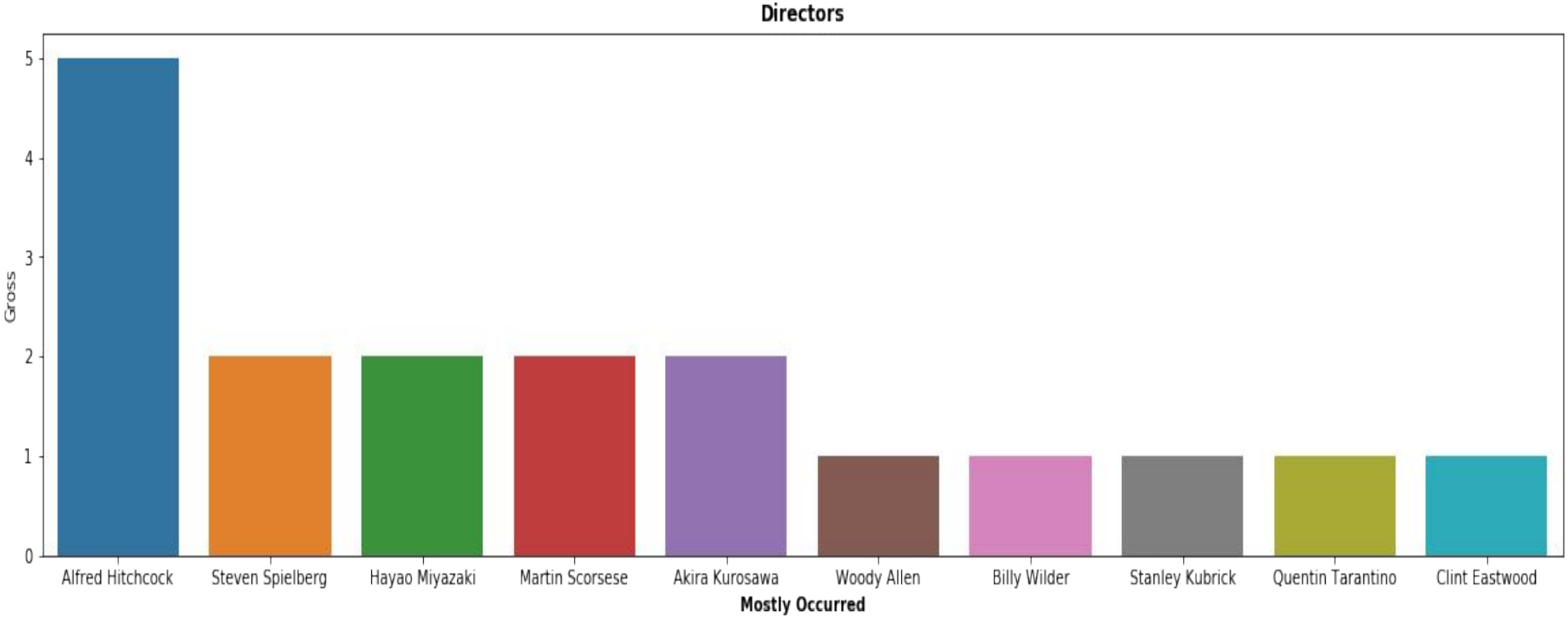
MetaScore of Top Rated Movies



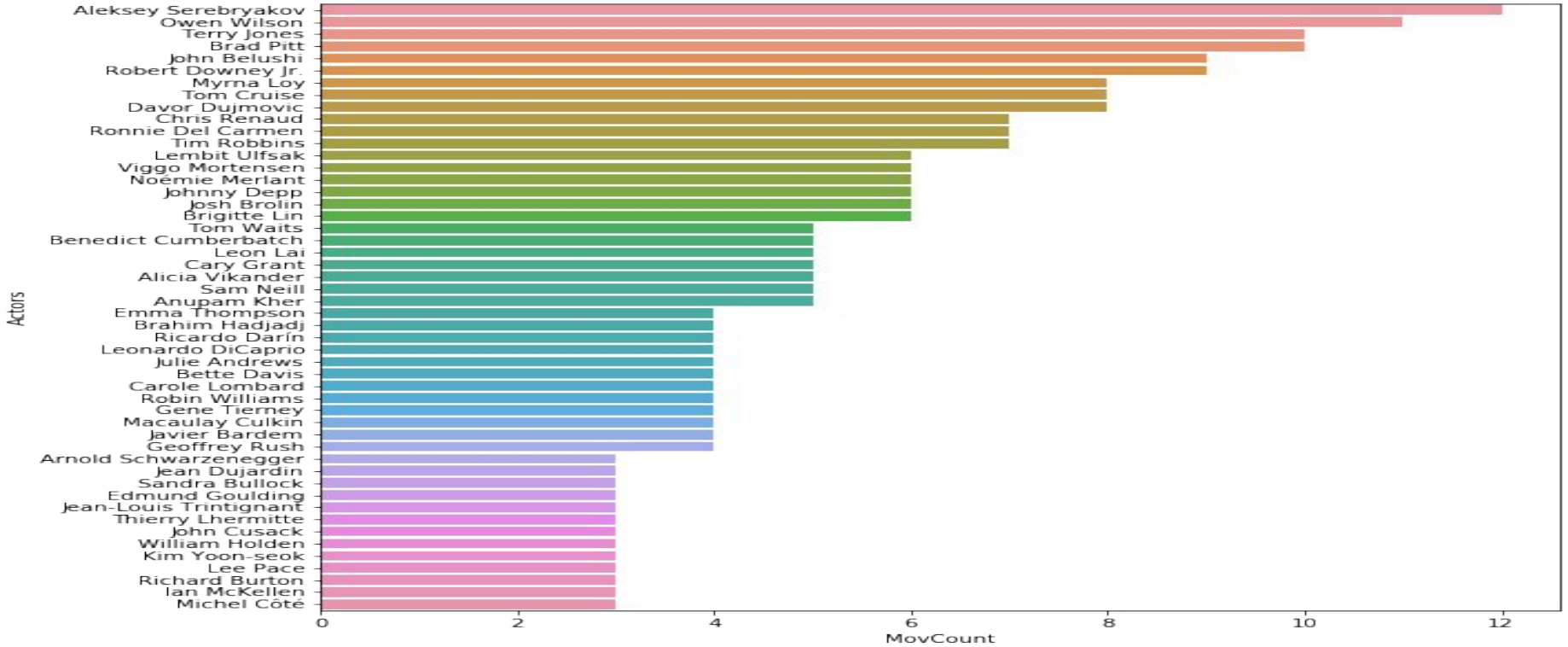
Release Year of Mostly Occurred Directors



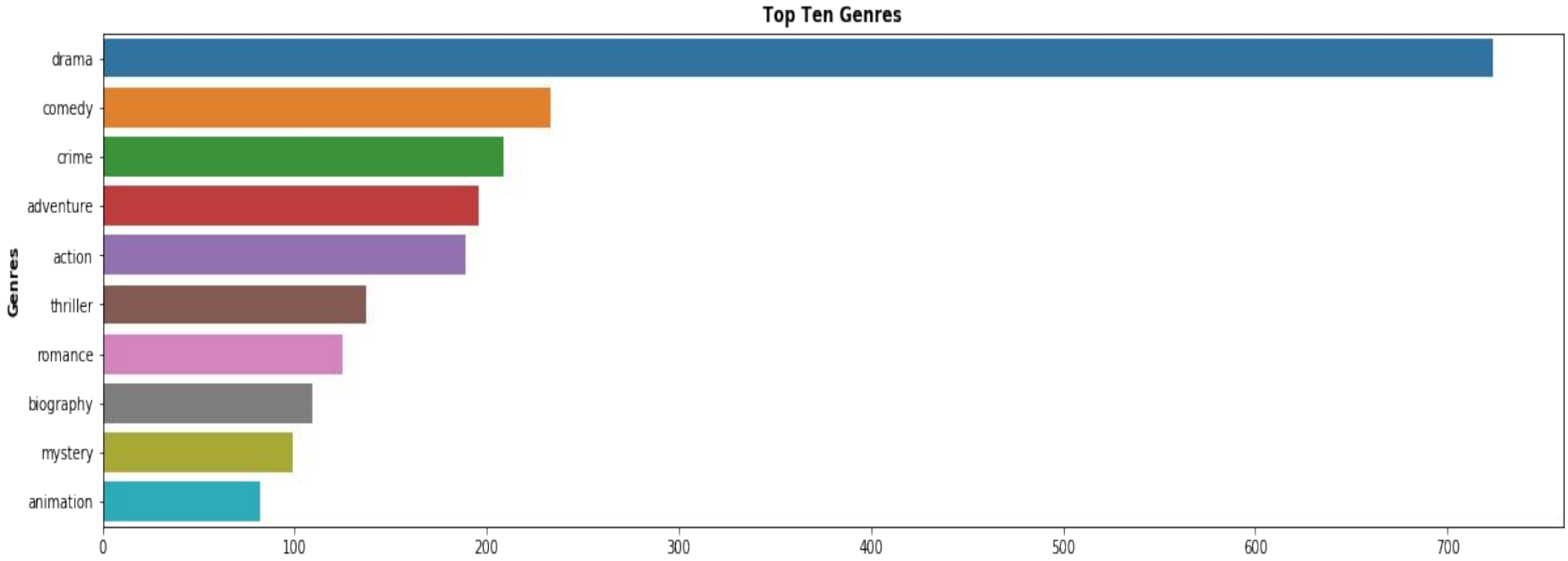
Mostly Occurred Directors and Gross Amount Collected



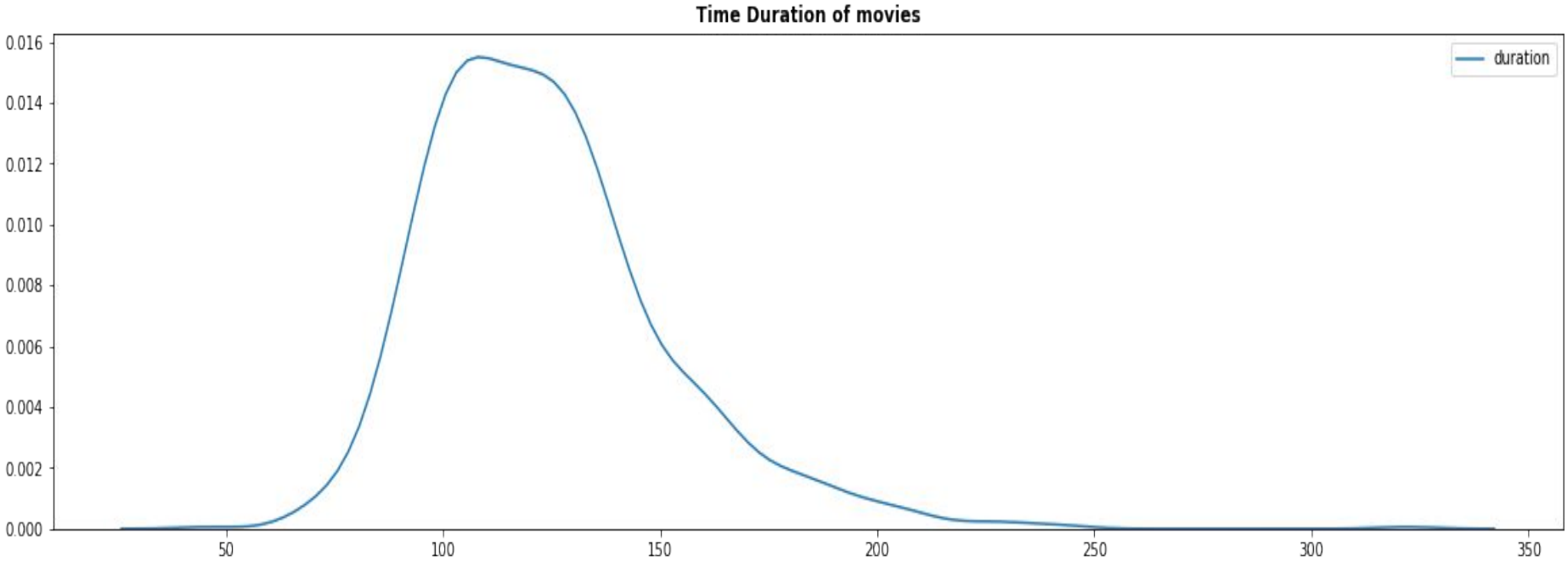
Actors and Their Movie Counts



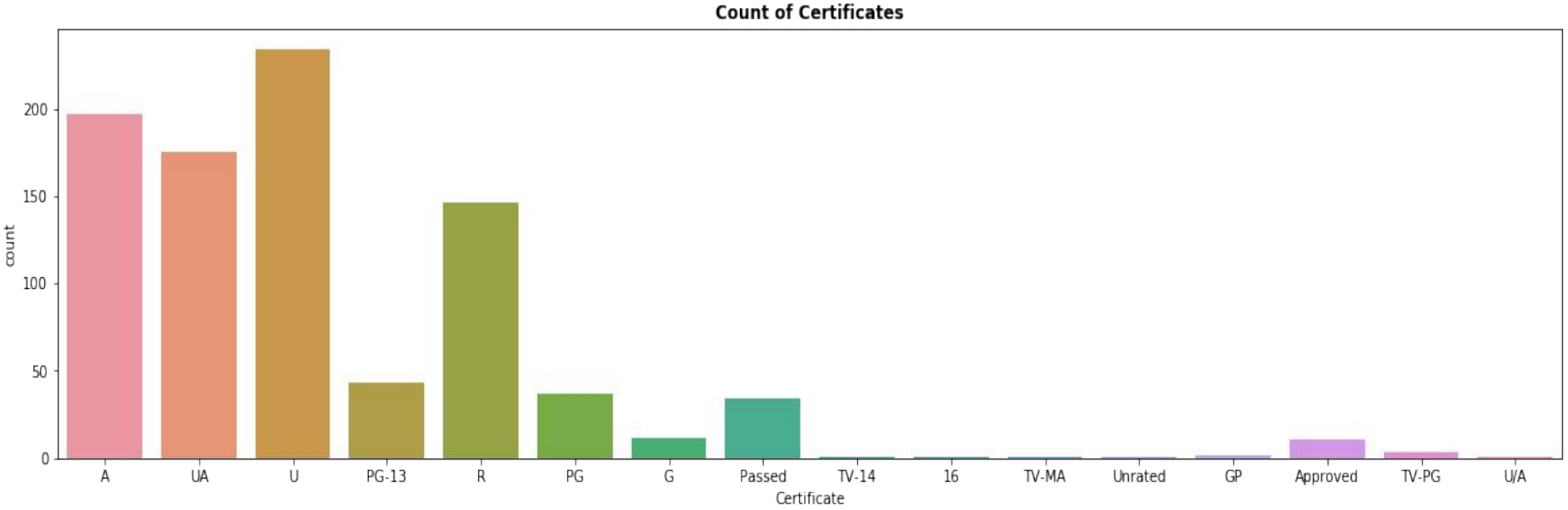
Top 10 Genres



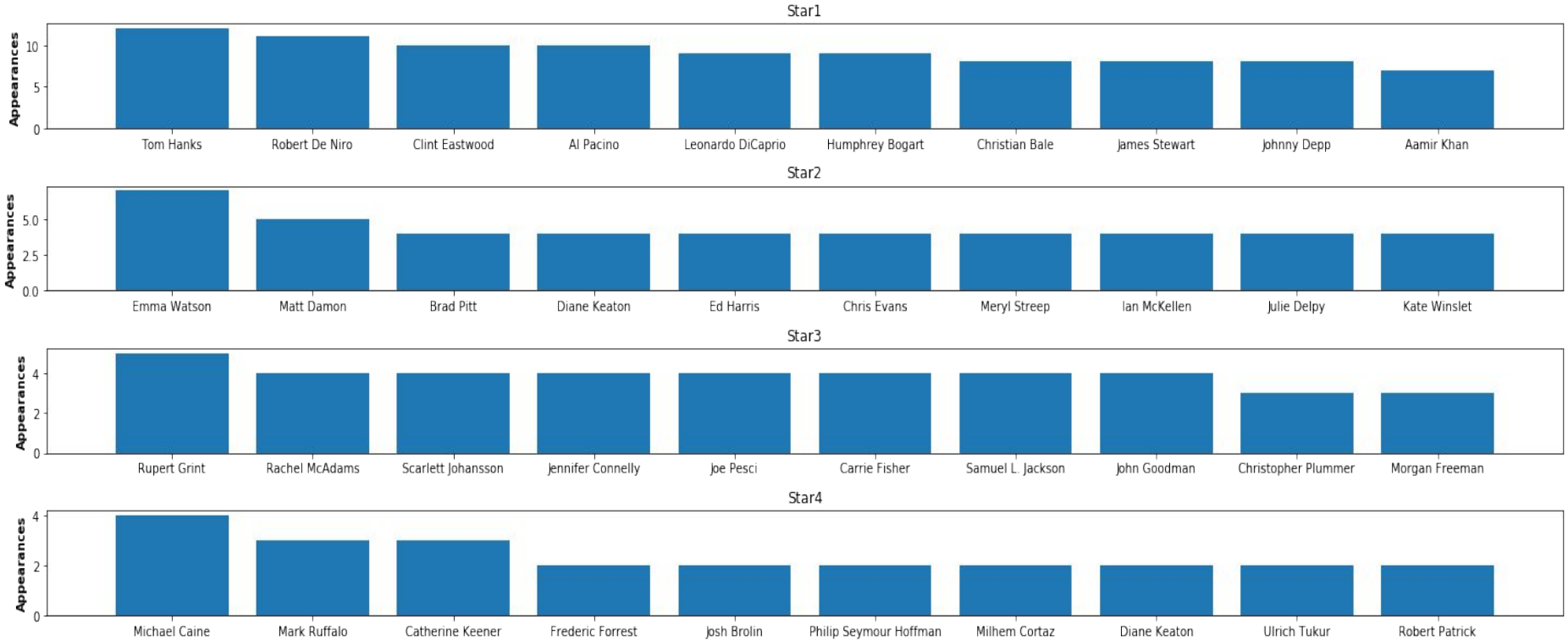
Time Duration of Movies



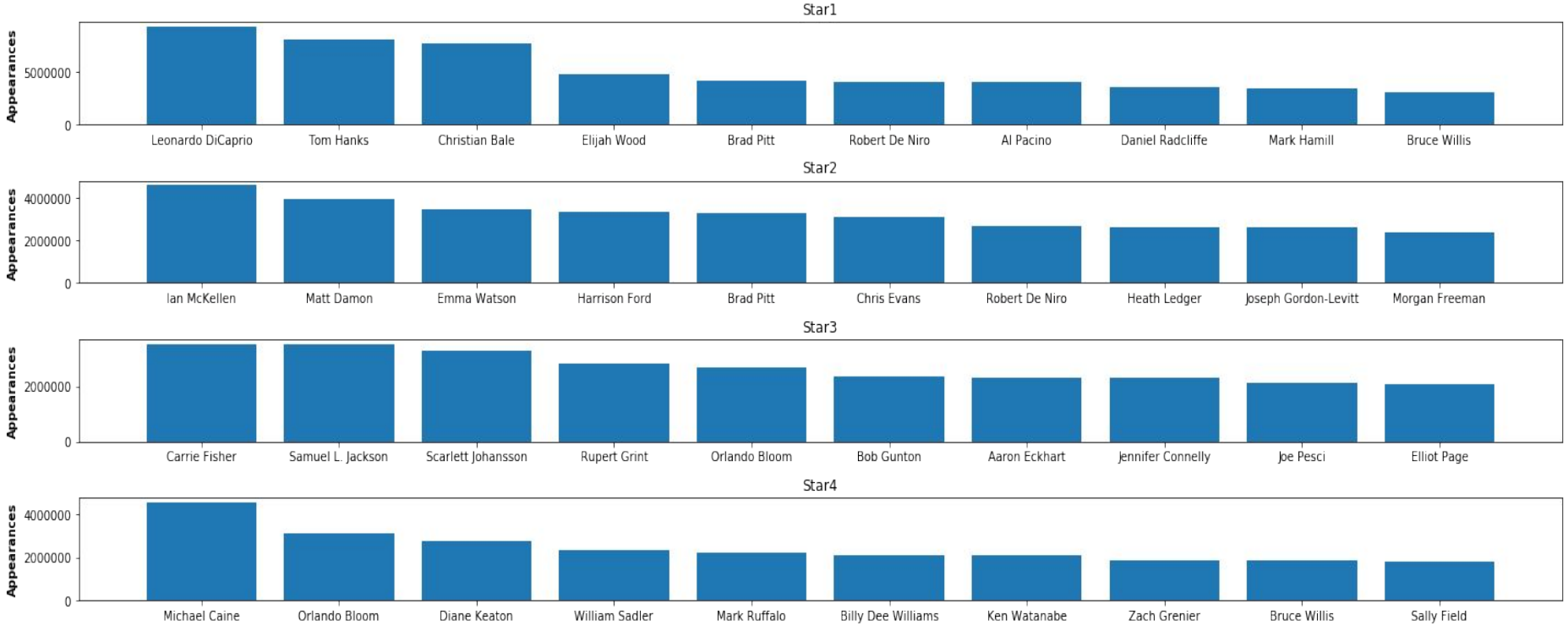
Count Of Certificates

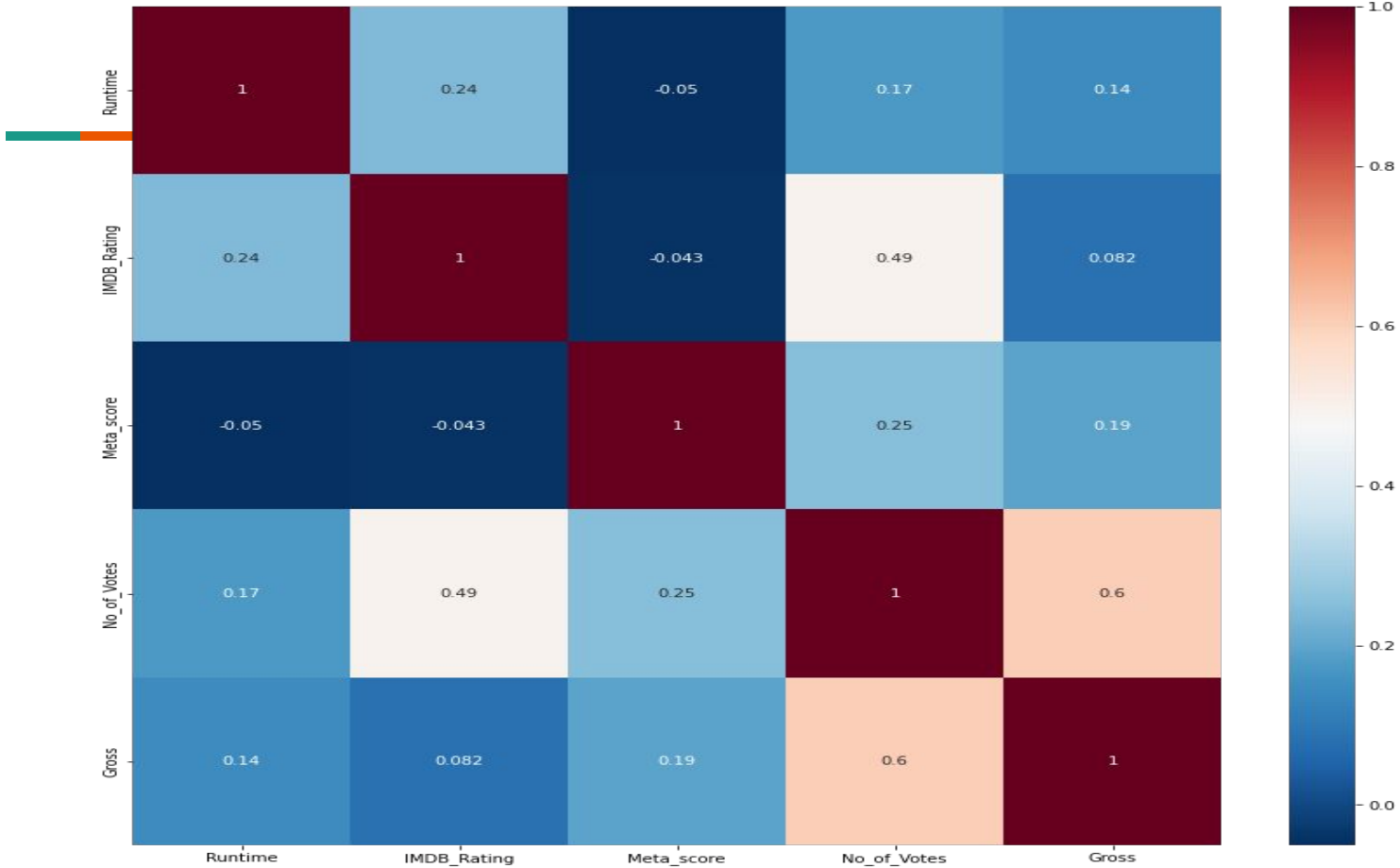


Stars with most occurrences in movies



Appearances Of Stars in Top voted movies




Runtime
IMDB_Rating
Meta_score
No_of_Votes
Gross
1
0.24
-0.05
0.17
0.14
0.24
1
-0.043
0.49
0.082
-0.05
-0.043
1
0.25
0.19
0.17
0.49
0.25
1
0.6
0.14
0.082
0.19
0.6
1
1.0
0.8
0.6
0.4
0.2
0.0

Insights

1. Drama is the most loved genre.
2. **Older films tend to have higher ratings IMDB_score and Meta_score.**
3. **Negative correlation between IMDB_rating and Meta_score.**
4. **Despite Shawshank redemption being the highly rated movie, it’s gross collection is too low.**

Conclusion

Based on the analysis, we can conclude that there are lot of parameters to consider while measuring success of a movie. For some movies, total vote count becomes important while for some gross collection is an important parameter. Same is the case with the presence of different actors and directors in a movie. Presence of certain actors and directors dont decide fate of the movie.Shawshank Redemption: Movie with highest votes does not have mostly occurred director or actor.

THANK YOU!