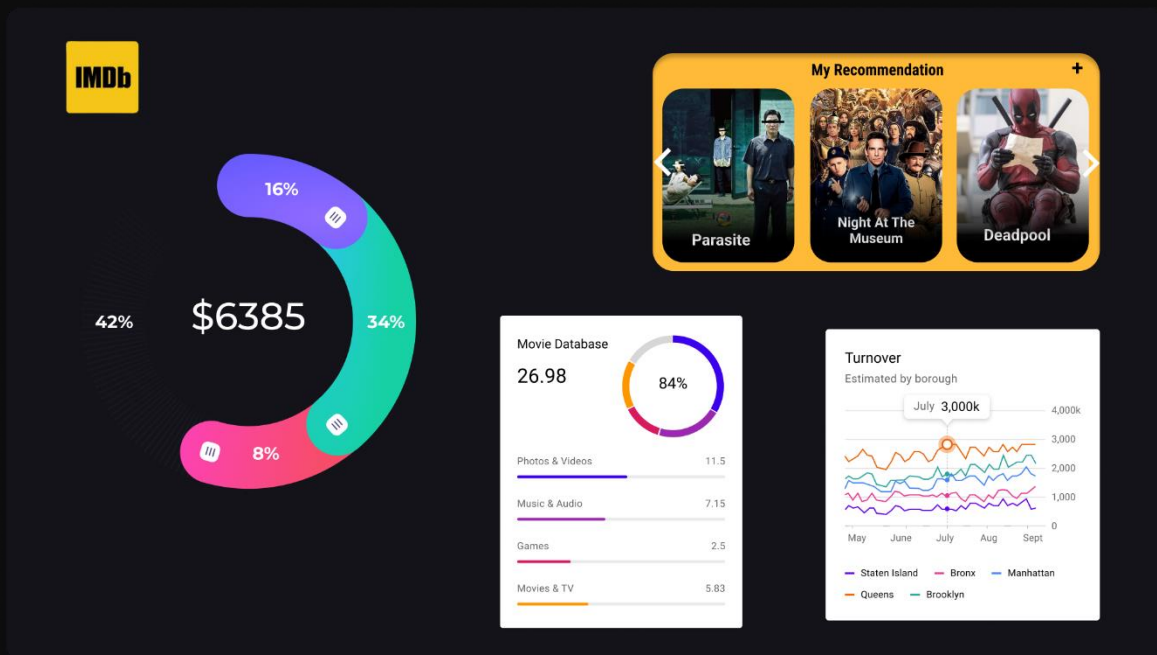




# Movie Analysis



# What is IMDB?

The Internet Movie Database (IMDb) is a website that serves as an online database of world cinema containing a large number of public data on films such as the title of the film, the year of release of the film, the genre of the film, the audience, budget, revenue, the rating of critics, the duration of the film, the summary of the film, actors, directors and much more.

The [Internet Movie Database](#) (IMDb) is the largest, most comprehensive movie database on the web. It offers an extensive database of movie, TV show, and cast information. The site was officially launched in 1990 and is now owned by Amazon.com.

IMDb is an extremely detailed and rich source of film data that features top movies, news, reviews, movie trailers, showtimes, DVD movie reviews, celebrity profiles, and more. If you've ever researched a movie or actor, you've probably landed on IMDb.

The site is truly a mammoth repository of movie information, and while it's great that most of it is completely free to access, it can seem cluttered and hard to sift through. Below are some tips on how to use IMDb, as well as a look at some of its features, including what you get with IMDbPro.

## Information about the data set:

This dataset contains data of over 4912 movies from 1916 to 2016 with their title, year of release, rating, runtime, genre, director, cast, critics, imdb Rating, imdb Votes, poster, plot, language of the movie, country, budget, gross, imdb link and much more. This vast dataset will help in performing various tasks like EDA & statistics.

The dataset contains 29 columns and 4912 records after deleting the duplicate records.

Purpose: Analyse the data available on IMDB for movies released between 1916 and 2016.

## Approach:

The dataset contains lots of information about the movies. So, I have to first understand the data and I have to clean the data.

The data contains many duplicate and null values so, I have removed all the duplicate data from the dataset, after cleaning up the dataset contains only 4912 records

After that, I must determine what the data is about, what categories of data exist, and what kinds of information are contained in the data.

After understanding the data, I have to do some sort of analysis on the datasets to get some useful insights.

I have used Excel 2019 to perform EDA operations on the dataset. Using the pivot table, excel formulas and many more functionality from the excel I got some useful insights.

## Tech Used:

### Microsoft Excel 2019

Microsoft Excel is a tool for organizing and performing calculations on data. It can analyse data, calculate statistics, generate pivot tables, and represent data as charts or graphs.

It enables users to format, organize and calculate data in a spreadsheet. It organizes data in an easy-to-navigate way. We need not to perform any complex mathematical functions. And it turns piles of data into helpful graphics and charts.

Analysis process: Used excel to have an overview of the datasets. The data has many types of categorical and statical records.

# Insights:

## 1. Cleaning the data:

The most important step before moving forward with the analysis is to clean the data by dropping columns, removing null values, removing duplicates etc.)

The dataset contains **246** duplicate rows.

So, to perform the analysis I have removed this duplicate row from this dataset.

## 2. Movies with highest profit:

I have created a new column called profit which contains the difference of the two columns: gross and budget. I have plotted profit (y-axis) vs budget (x- axis) and observed the outliers using the chart.

The most profitable movie is Avatar with amount of 523505 thousand dollars.

This is the list of top 10 movies with highest profit in thousand.



### 3. IMDB Top 250 movies:

The top 250 movies with the highest IMDb Rating corresponding to the imdb score and having the number of voted users greater than 25,000.

Rank	Movie Title	IMDB Score
1	The Shawshank Redemption	9.3
2	The Godfather	9.2
3	The Dark Knight	9
4	The Godfather: Part II	9
5	Fargo	9
6	Pulp Fiction	8.9
7	The Lord of the Rings: The Return of the King	8.9
8	Schindler's List	8.9
9	The Good, the Bad and the Ugly	8.9
10	12 Angry Men	8.9
11	Inception	8.8
12	Fight Club	8.8
13	Forrest Gump	8.8
14	The Lord of the Rings: The Fellowship of the Ring	8.8
15	Star Wars: Episode V - The Empire Strikes Back	8.8
16	Daredevil	8.8
17	It's Always Sunny in Philadelphia	8.8
18	The Matrix	8.7
19	The Lord of the Rings: The Two Towers	8.7
20	Star Wars: Episode IV - A New Hope	8.7
21	Goodfellas	8.7
22	One Flew Over the Cuckoo's Nest	8.7
23	City of God	8.7
24	Seven Samurai	8.7
25	Friday Night Lights	8.7
26	Se7en	8.6
27	Interstellar	8.6
28	The Silence of the Lambs	8.6
29	Saving Private Ryan	8.6
30	American History X	8.6
31	The Usual Suspects	8.6

32	Spirited Away	8.6
33	Casablanca	8.6
34	It's a Wonderful Life	8.6
35	Once Upon a Time in the West	8.6
36	Spartacus: War of the Damned	8.6
37	Hannibal	8.6
38	Modern Times	8.6
39	Luther	8.6
40	The Dark Knight Rises	8.5
41	Gladiator	8.5
42	Django Unchained	8.5
43	The Departed	8.5
44	Memento	8.5
45	The Prestige	8.5
46	The Green Mile	8.5
47	Terminator 2: Judgment Day	8.5
48	Back to the Future	8.5
49	Raiders of the Lost Ark	8.5
50	The Lion King	8.5
51	Alien	8.5
52	The Pianist	8.5
53	Apocalypse Now	8.5
54	Psycho	8.5
55	Whiplash	8.5
56	Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb	8.5
57	The Lives of Others	8.5
58	Entourage	8.5
59	Outlander	8.5
60	Airlift	8.5
61	Children of Heaven	8.5
62	American Beauty	8.4
63	Braveheart	8.4
64	WALL-E	8.4
65	Star Wars: Episode VI - Return of the Jedi	8.4
66	Reservoir Dogs	8.4
67	The Shining	8.4
68	Requiem for a Dream	8.4
69	Amélie	8.4
70	Aliens	8.4
71	Oldboy	8.4

72	Princess Mononoke	8.4
73	Once Upon a Time in America	8.4
74	To Kill a Mockingbird	8.4
75	Lawrence of Arabia	8.4
76	Das Boot	8.4
77	A Separation	8.4
78	Rang De Basanti	8.4
79	Psych	8.4
80	Stargate SG-1	8.4
81	Baahubali: The Beginning	8.4
82	The Inbetweeners	8.4
83	Veronica Mars	8.4
84	M*A*S*H	8.4
85	Batman: The Dark Knight Returns, Part 2	8.4
86	Batman Begins	8.3
87	Inglourious Basterds	8.3
88	Eternal Sunshine of the Spotless Mind	8.3
89	Up	8.3
90	Toy Story	8.3
91	Good Will Hunting	8.3
92	Snatch	8.3
93	Toy Story 3	8.3
94	Scarface	8.3
95	Indiana Jones and the Last Crusade	8.3
96	Taxi Driver	8.3
97	2001: A Space Odyssey	8.3
98	L.A. Confidential	8.3
99	Monty Python and the Holy Grail	8.3
100	Inside Out	8.3
101	Unforgiven	8.3
102	Amadeus	8.3
103	Downfall	8.3
104	Raging Bull	8.3
105	The Sting	8.3
106	Some Like It Hot	8.3
107	The Hunt	8.3
108	The Great Escape	8.3
109	Room	8.3
110	Singin' in the Rain	8.3
111	Metropolis	8.3
112	The Apartment	8.3

113	Inside Job	8.3
114	Judgment at Nuremberg	8.3
115	Life	8.3
116	V for Vendetta	8.2
117	The Wolf of Wall Street	8.2
118	Finding Nemo	8.2
119	A Beautiful Mind	8.2
120	Die Hard	8.2
121	Gran Torino	8.2
122	The Big Lebowski	8.2
123	How to Train Your Dragon	8.2
124	Trainspotting	8.2
125	Pan's Labyrinth	8.2
126	Blade Runner	8.2
127	Into the Wild	8.2
128	Lock, Stock and Two Smoking Barrels	8.2
129	Casino	8.2
130	Warrior	8.2
131	Captain America: Civil War	8.2
132	The Thing	8.2
133	The Deer Hunter	8.2
134	Gone with the Wind	8.2
135	Howl's Moving Castle	8.2
136	The Elephant Man	8.2
137	The Bridge on the River Kwai	8.2
138	The Secret in Their Eyes	8.2
139	Buffy the Vampire Slayer	8.2
140	On the Waterfront	8.2
141	Rebecca	8.2
142	Incendies	8.2
143	Mr. Smith Goes to Washington	8.2
144	It Happened One Night	8.2
145	Lage Raho Munna Bhai	8.2
146	The Avengers	8.1
147	Pirates of the Caribbean: The Curse of the Black Pearl	8.1
148	Shutter Island	8.1
149	Kill Bill: Vol. 1	8.1
150	The Sixth Sense	8.1
151	Guardians of the Galaxy	8.1
152	The Truman Show	8.1



153	Sin City	8.1
154	Jurassic Park	8.1
155	No Country for Old Men	8.1
156	The Terminator	8.1
157	Monsters, Inc.	8.1
158	Donnie Darko	8.1
159	Gone Girl	8.1
160	Mad Max: Fury Road	8.1
161	The Bourne Ultimatum	8.1
162	Million Dollar Baby	8.1
163	Deadpool	8.1
164	The Grand Budapest Hotel	8.1
165	The Martian	8.1
166	The Imitation Game	8.1
167	12 Years a Slave	8.1
168	Groundhog Day	8.1
169	The Revenant	8.1
170	Prisoners	8.1
171	Rocky	8.1
172	There Will Be Blood	8.1
173	The Help	8.1
174	Rush	8.1
175	The Princess Bride	8.1
176	The Wizard of Oz	8.1
177	Platoon	8.1
178	Stand by Me	8.1
179	Hotel Rwanda	8.1
180	Spotlight	8.1
181	Annie Hall	8.1
182	Before Sunrise	8.1
183	Amores Perros	8.1
184	Gandhi	8.1
185	Hachi: A Dog's Tale	8.1
186	Butch Cassidy and the Sundance Kid	8.1
187	Akira	8.1
188	A Christmas Story	8.1
189	Network	8.1
190	Barry Lyndon	8.1
191	Elite Squad	8.1
192	High Noon	8.1
193	The Celebration	8.1

194	The Sea Inside	8.1
195	Solaris	8.1
196	The Man Who Shot Liberty Valance	8.1
197	The Best Years of Our Lives	8.1
198	Cat on a Hot Tin Roof	8.1
199	Tae Guk Gi: The Brotherhood of War	8.1
200	Touching the Void	8.1
201	Slumdog Millionaire	8
202	Black Swan	8
203	District 9	8
204	Catch Me If You Can	8
205	X-Men: Days of Future Past	8
206	Kill Bill: Vol. 2	8
207	Star Trek	8
208	The King's Speech	8
209	The Incredibles	8
210	Ratatouille	8
211	Casino Royale	8
212	Life of Pi	8
213	Jaws	8
214	Blood Diamond	8
215	Shaun of the Dead	8
216	Rain Man	8
217	Her	8
218	The Perks of Being a Wallflower	8
219	Big Fish	8
220	Mystic River	8
221	The Pursuit of Happyness	8
222	Dallas Buyers Club	8
223	In Bruges	8
224	The Exorcist	8
225	Dead Poets Society	8
226	Boyhood	8
227	Aladdin	8
228	Serenity	8
229	Magnolia	8
230	Mulholland Drive	8
231	The Artist	8
232	Dances with Wolves	8
233	Before Sunset	8
234	True Romance	8

235	Brazil	8
236	Cinderella Man	8
237	The Sound of Music	8
238	A Fistful of Dollars	8
239	Rosemary's Baby	8
240	The Man from Earth	8
241	The Iron Giant	8
242	Bowling for Columbine	8
243	JFK	8
244	Young Frankenstein	8
245	The Diving Bell and the Butterfly	8
246	Night of the Living Dead	8
247	Dancer in the Dark	8
248	A Streetcar Named Desire	8
249	Patton	8
250	Sling Blade	8

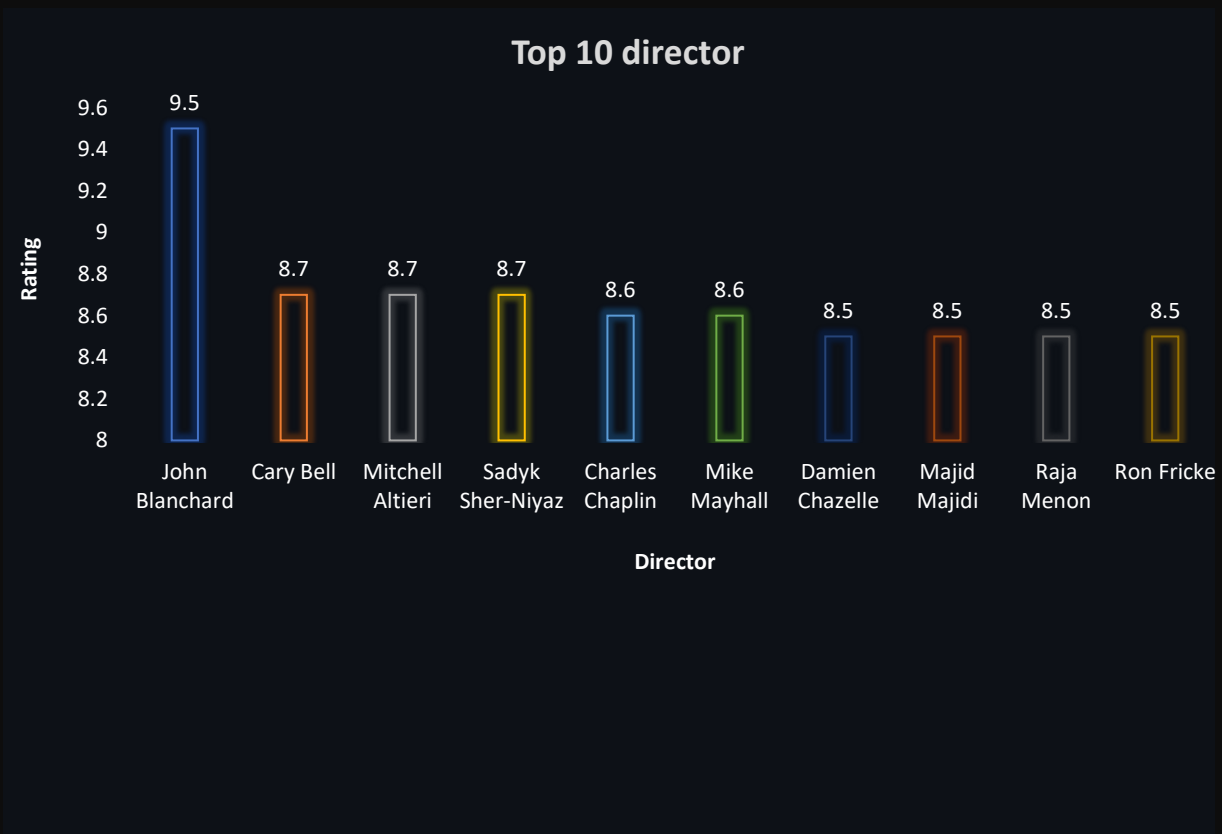
#### 4. Top foreign language movie which are not in English language.

Rank	Movie Title	IMDB Score	Language
1	The Good, the Bad and the Ugly	8.9	Italian
2	City of God	8.7	Portuguese
3	Seven Samurai	8.7	Japanese
4	Spirited Away	8.6	Japanese
5	The Lives of Others	8.5	German
6	Airlift	8.5	Hindi
7	Children of Heaven	8.5	Persian
8	Amélie	8.4	French
9	Oldboy	8.4	Korean
10	Princess Mononoke	8.4	Japanese
11	Das Boot	8.4	German
12	A Separation	8.4	Persian
13	Rang De Basanti	8.4	Hindi
14	Baahubali: The Beginning	8.4	Telugu
15	Downfall	8.3	German
16	The Hunt	8.3	Danish
17	Metropolis	8.3	German

18	Pan's Labyrinth	8.2	Spanish
19	Howl's Moving Castle	8.2	Japanese
20	The Secret in Their Eyes	8.2	Spanish
21	Incendies	8.2	French
22	Lage Raho Munna Bhai	8.2	Hindi
23	Amores Perros	8.1	Spanish
24	Akira	8.1	Japanese
25	Elite Squad	8.1	Portuguese
26	The Celebration	8.1	Danish
27	The Sea Inside	8.1	Spanish
28	Solaris	8.1	Russian
29	Tae Guk Gi: The Brotherhood of War	8.1	Korean
30	A Fistful of Dollars	8	Italian
31	The Diving Bell and the Butterfly	8	French

## 5. Best Directors:

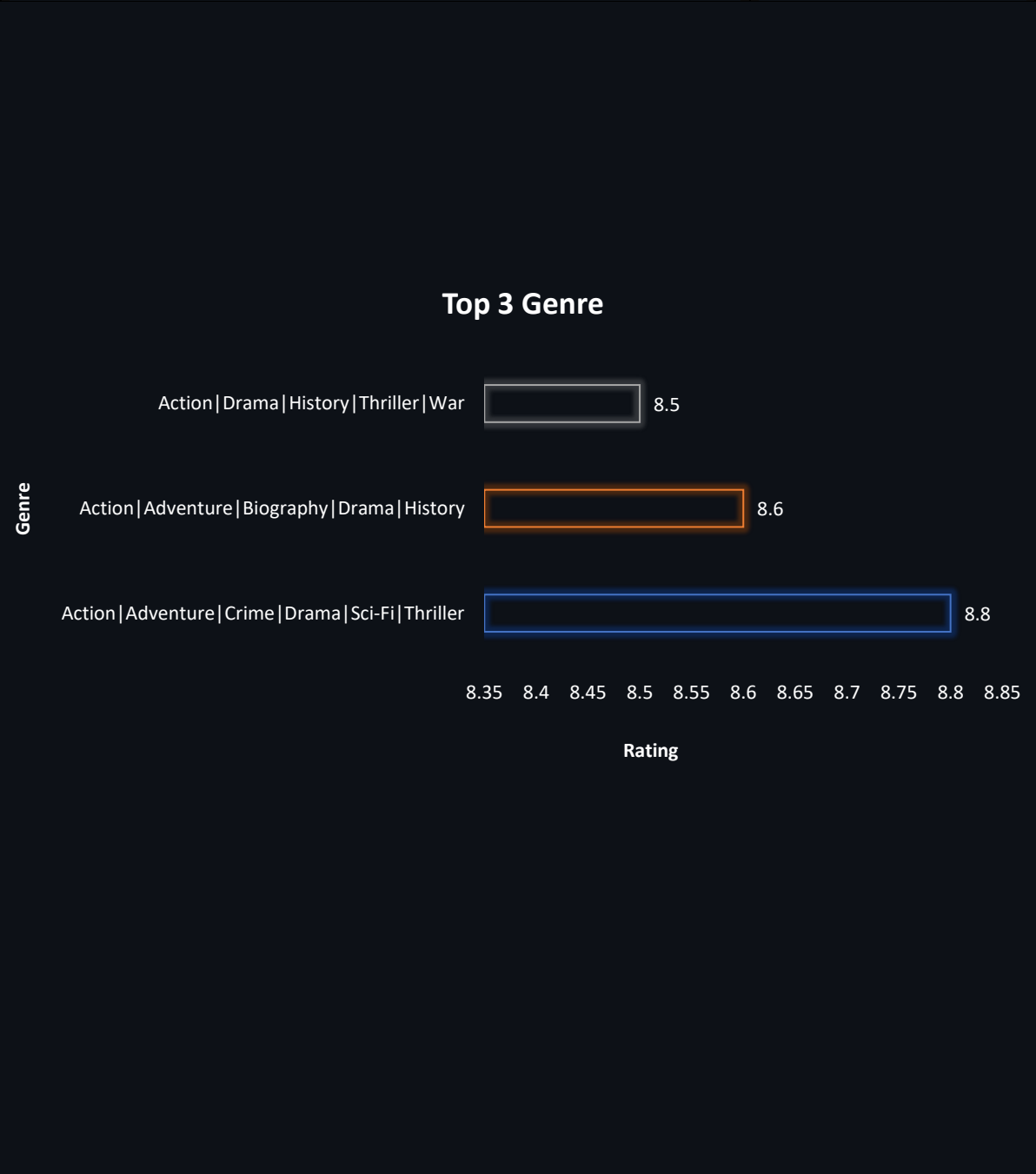
The top 10 directors for whose mean of imdb score is the highest.



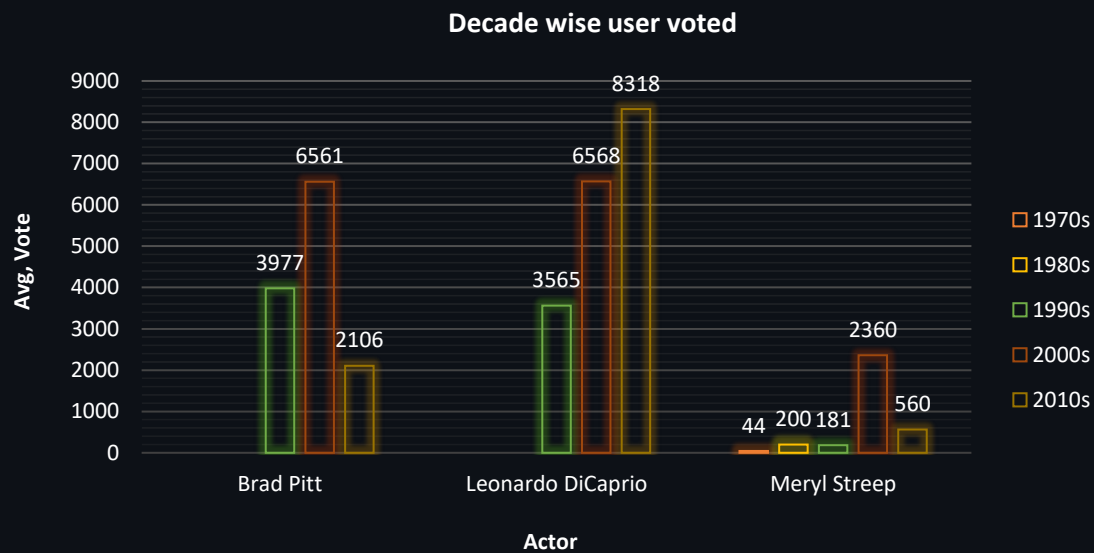
## 6. Popular Genres:

The top 3 popular genres are:

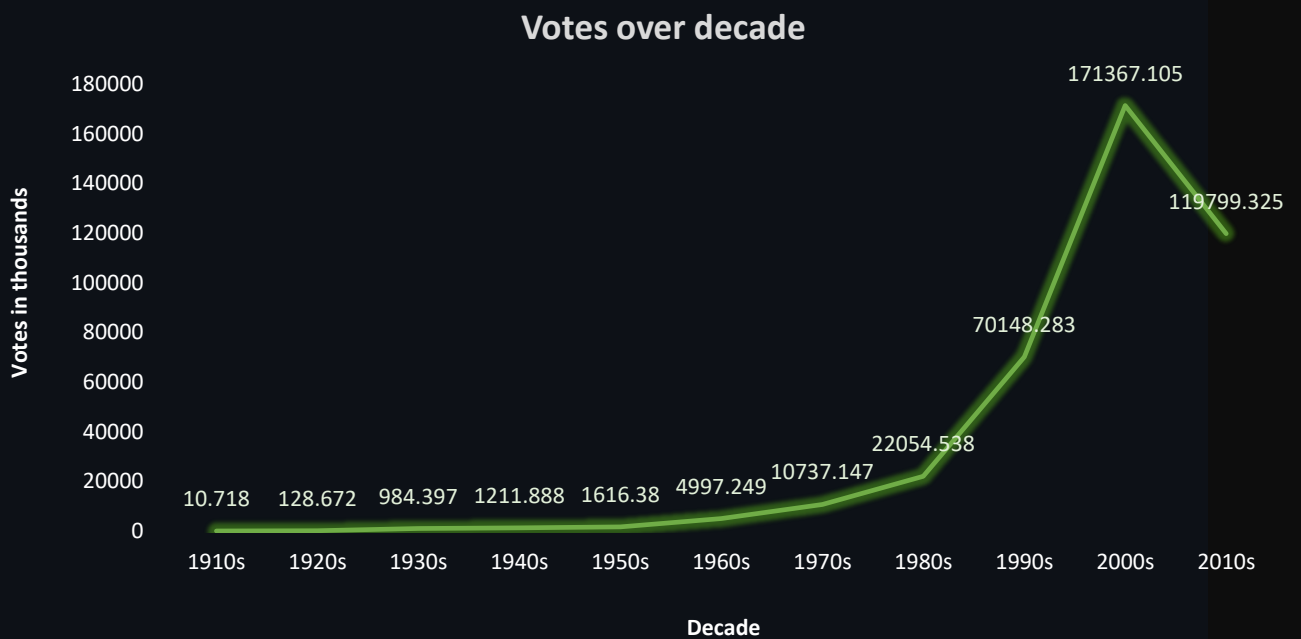
Genre	Average Rating
Action   Adventure   Crime   Drama   Sci-Fi   Thriller	8.8
Action  Adventure   Biography   Drama   History	8.6
Action   Drama   History   Thriller   War	8.5



## 6. The critic-favorite and audience-favorite actors over decade:

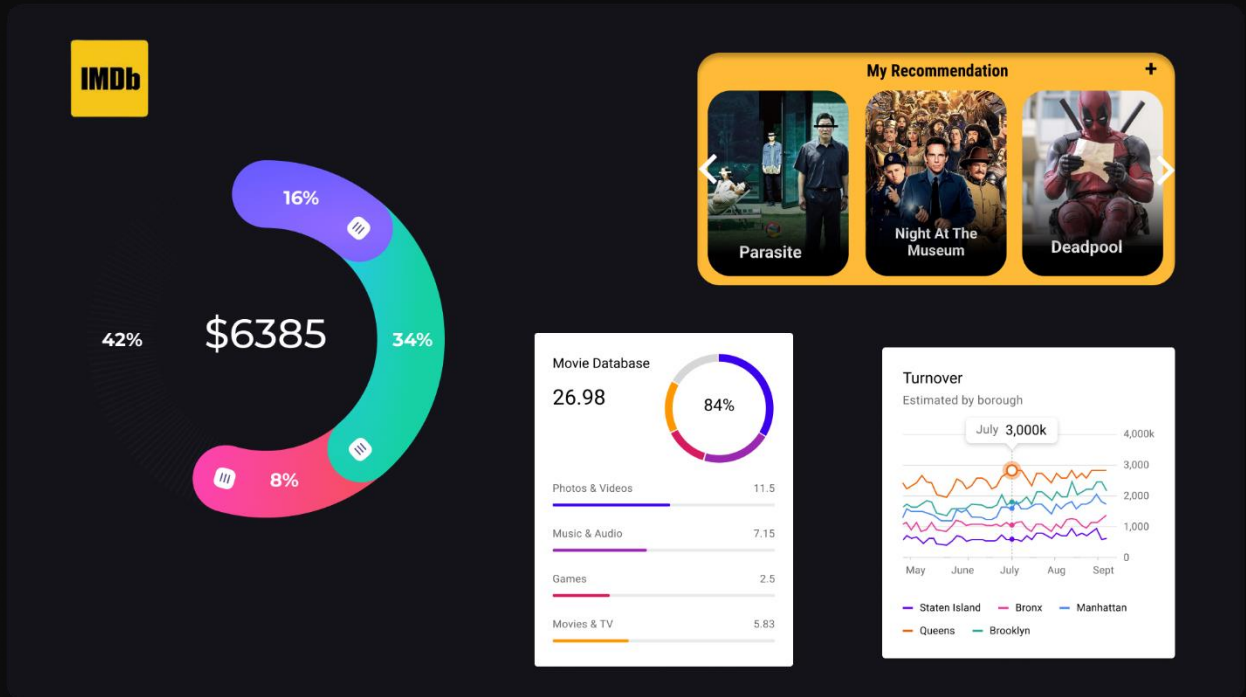


## 7. The change in number of voted users over decades:



# IMDb

## Movie Analysis



Drive Link:

[https://drive.google.com/drive/folders/1REi66KsRA\\_4LxbqhXuLFc3Ruy0H1S9uP?usp=share\\_link](https://drive.google.com/drive/folders/1REi66KsRA_4LxbqhXuLFc3Ruy0H1S9uP?usp=share_link)

Excel Sheet link:

[https://docs.google.com/spreadsheets/d/1Xl8xTo0LTRNVdc1bWTHNt1luyyINTxAH/edit?usp=share\\_link&ouid=117017262849463051494&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1Xl8xTo0LTRNVdc1bWTHNt1luyyINTxAH/edit?usp=share_link&ouid=117017262849463051494&rtpof=true&sd=true)

# Thank You