IMDB MOVIE ANALYSIS

KEERTHI NANNEPAMULA

PROJECT DESCRIPTION

- The task we are dealing with in this project is finding the popular genres of the movies among the masses, the top IMDB movies, and looking into the other details of the movies.
- All the insights are being drawn by implementing the "5 why approach", also known as the Root Cause Analysis and necessary tech stack.
- The questions answered in this process are: cleaning the data, the movies with the highest profit, the IMDB top 250 movies, the best directors and their movies, the most popular genres, and a complete chart of movies with specific lead actors,

APPROACH

IMDB movie analysis

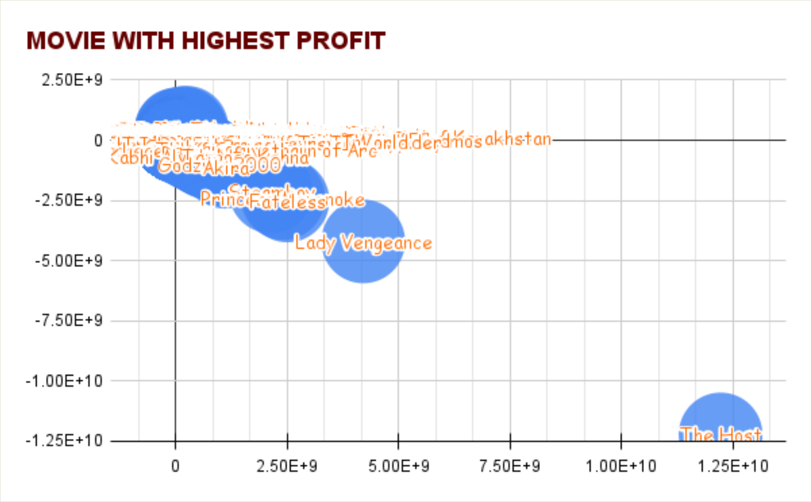
1. CLEANING THE DATA

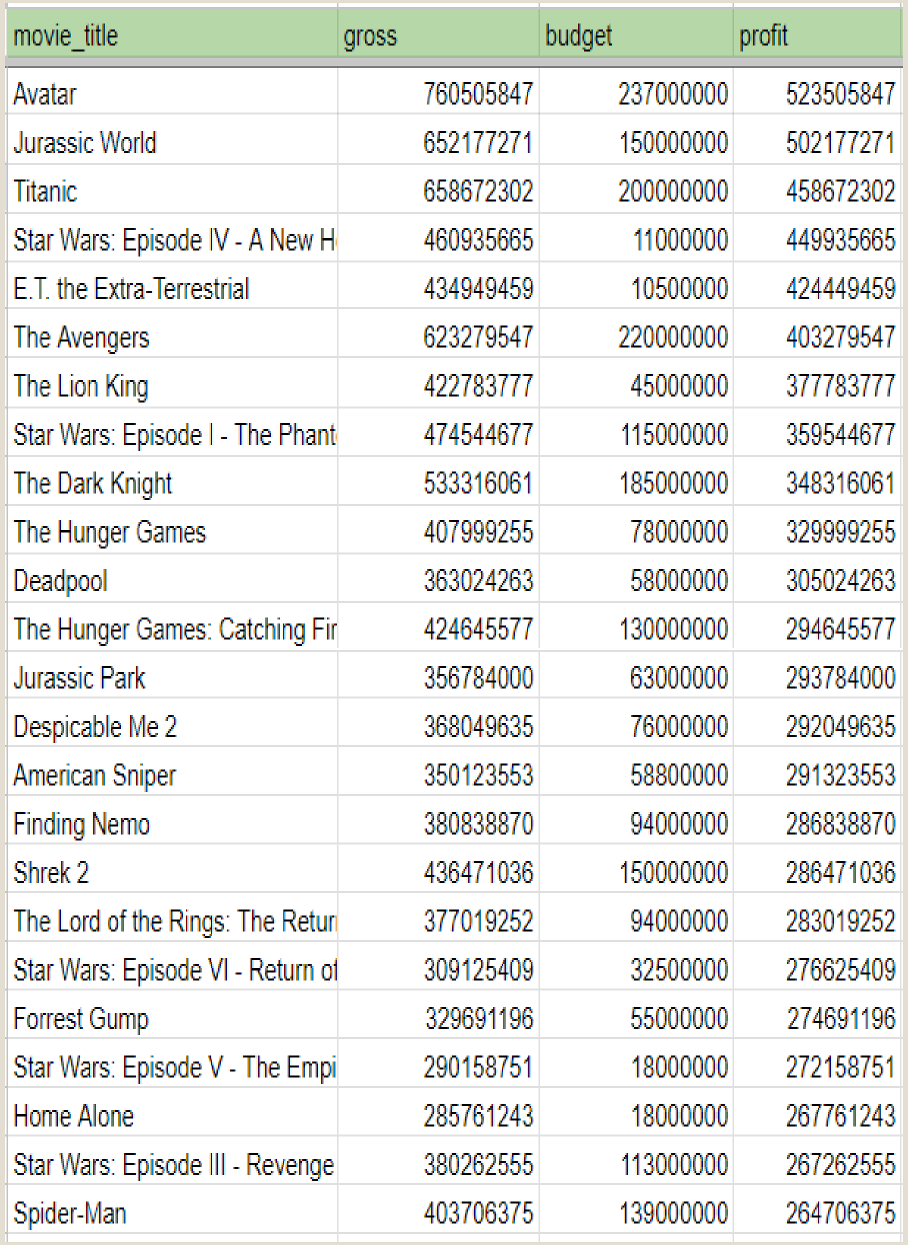
- Dropping the columns unnecessary for analysis
  1. Colour
  2. Duration
  3. Director_facebook_likes
  4. Actor_3_facebook_likes
  5. Actor_1_facbook_likes
  6. Cast_total_facebook_likes
  7. Facenumber_in_poster
  8. Plot_keywords
  9. Movie_imdb_link
  10. Actor_2_faceebook_likes
  11. Aspect_ratio
  12. Movie_facebook_likes
  13. Content_rating
- Removing all rows with null values
- Deleting duplicate rows

IMDB movie analysis

2. MOVIES WITH THE HIGHEST PROFIT

- A new column PROFIT is created from (GROSS-BUDGET)
- The PROFIT column is sorted
- A graph is created by
  - Profit-> x-axis
  - Budget-> y-axis
- The outliers are observed
- Highest profit movies- avatar, Jurassic world

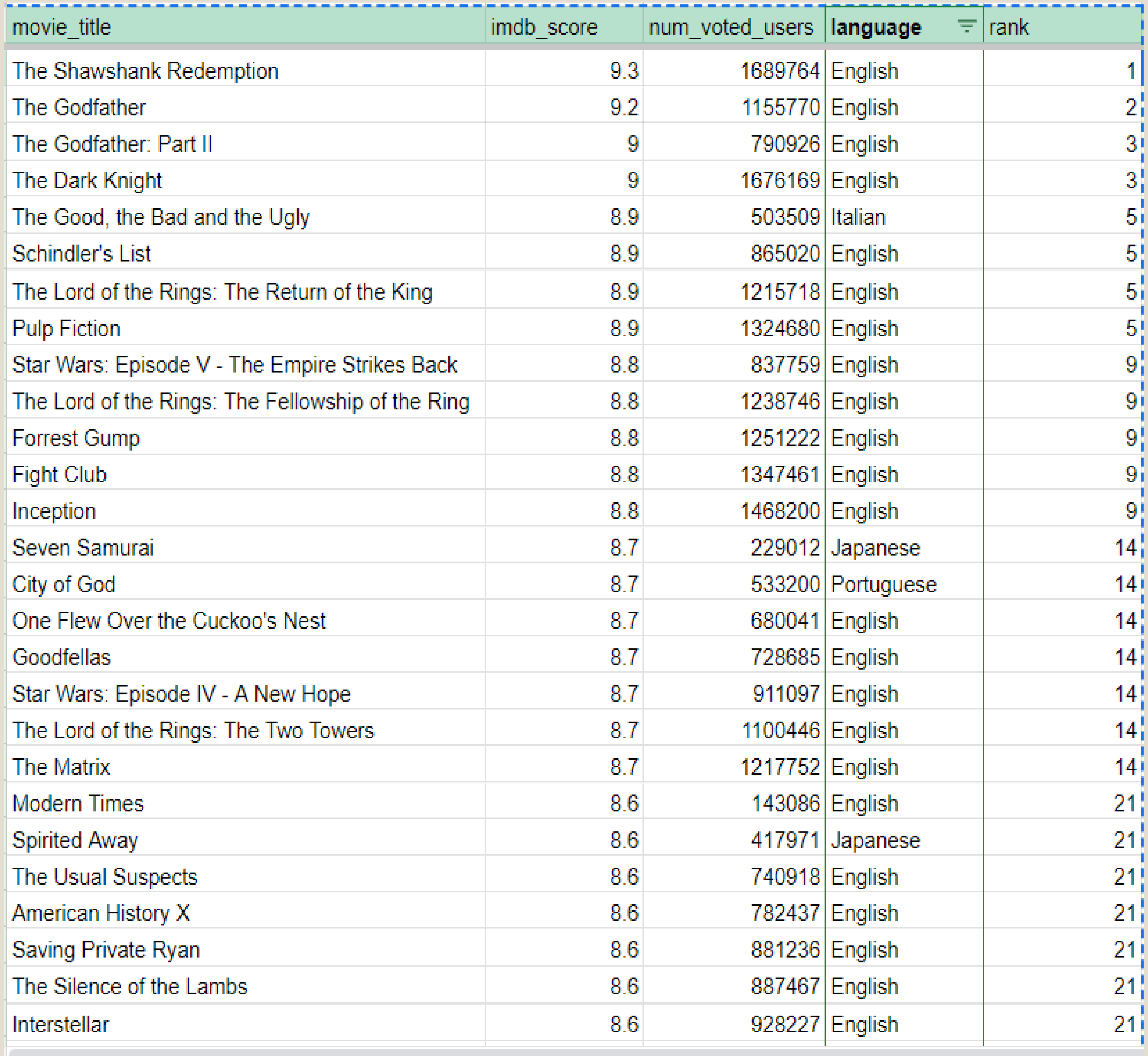




movie_title	gross	budget	profit
Avatar	760505847	237000000	523505847
Jurassic World	652177271	150000000	502177271
Titanic	658672302	200000000	458672302
Star Wars: Episode IV - A New H	460935665	11000000	449935665
E.T. the Extra-Terrestrial	434949459	10500000	424449459
The Avengers	623279547	220000000	403279547
The Lion King	422783777	45000000	377783777
Star Wars: Episode I - The Phant	474544677	115000000	359544677
The Dark Knight	533316061	185000000	348316061
The Hunger Games	407999255	78000000	329999255
Deadpool	363024263	58000000	305024263
The Hunger Games: Catching Fir	424645577	130000000	294645577
Jurassic Park	356784000	63000000	293784000
Despicable Me 2	368049635	76000000	292049635
American Sniper	350123553	58800000	291323553
Finding Nemo	380838870	94000000	286838870
Shrek 2	436471036	150000000	286471036
The Lord of the Rings: The Retur	377019252	94000000	283019252
Star Wars: Episode VI - Return of	309125409	32500000	276625409
Forrest Gump	329691196	55000000	274691196
Star Wars: Episode V - The Empi	290158751	18000000	272158751
Home Alone	285761243	18000000	267761243
Star Wars: Episode III - Revenge	380262555	113000000	267262555
Spider-Man	403706375	139000000	264706375

IMDB movie analysis

3. LIST OF THE imdb TOP 250 MOVIES

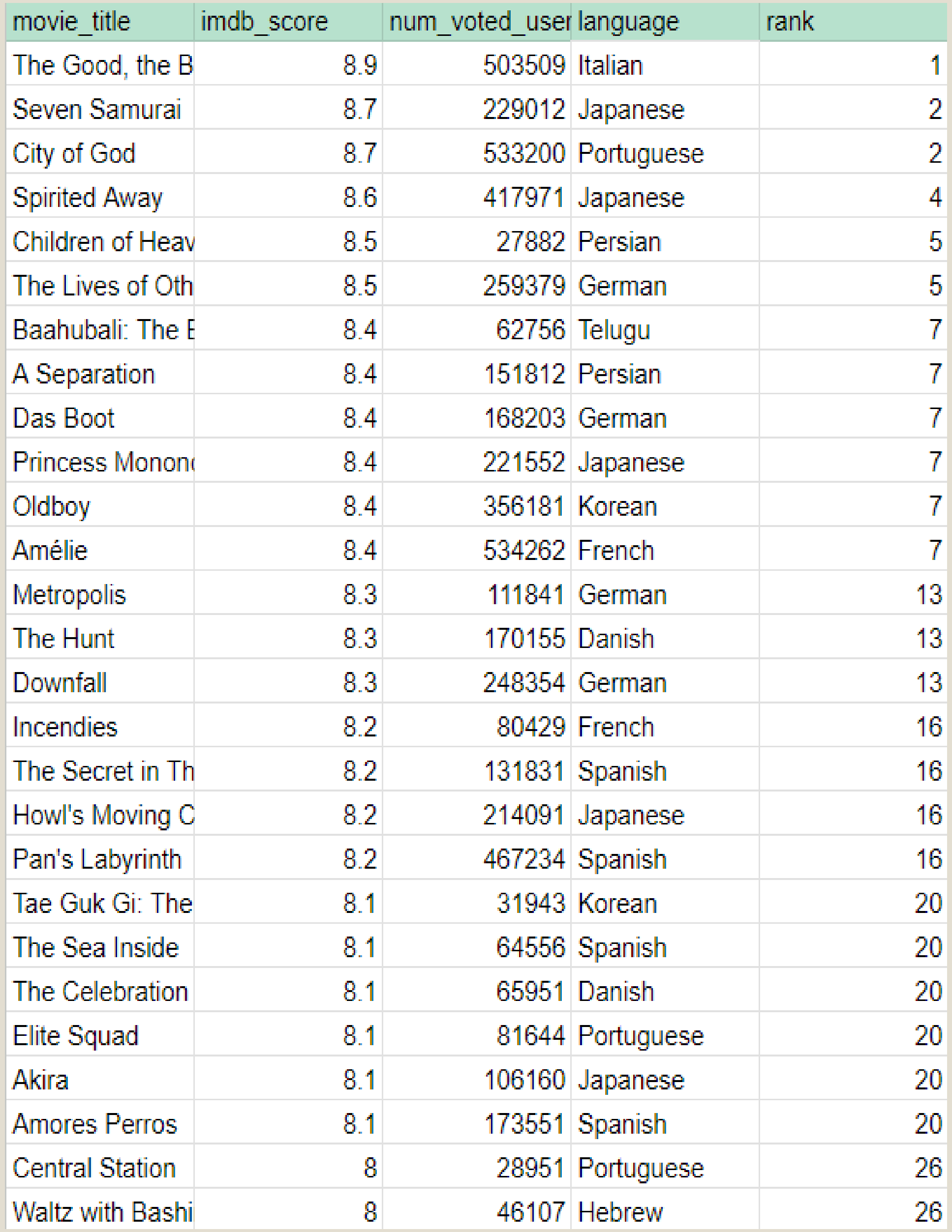


movie_title	imdb_score	num_voted_users	language	rank
The Shawshank Redemption	9.3	1689764	English	1
The Godfather	9.2	1155770	English	2
The Godfather: Part II	9	790926	English	3
The Dark Knight	9	1676169	English	3
The Good, the Bad and the Ugly	8.9	503509	Italian	5
Schindler's List	8.9	865020	English	5
The Lord of the Rings: The Return of the King	8.9	1215718	English	5
Pulp Fiction	8.9	1324680	English	5
Star Wars: Episode V - The Empire Strikes Back	8.8	837759	English	9
The Lord of the Rings: The Fellowship of the Ring	8.8	1238746	English	9
Forrest Gump	8.8	1251222	English	9
Fight Club	8.8	1347461	English	9
Inception	8.8	1468200	English	9
Seven Samurai	8.7	229012	Japanese	14
City of God	8.7	533200	Portuguese	14
One Flew Over the Cuckoo's Nest	8.7	680041	English	14
Goodfellas	8.7	728685	English	14
Star Wars: Episode IV - A New Hope	8.7	911097	English	14
The Lord of the Rings: The Two Towers	8.7	1100446	English	14
The Matrix	8.7	1217752	English	14
Modern Times	8.6	143086	English	21
Spirited Away	8.6	417971	Japanese	21
The Usual Suspects	8.6	740918	English	21
American History X	8.6	782437	English	21
Saving Private Ryan	8.6	881236	English	21
The Silence of the Lambs	8.6	887467	English	21
Interstellar	8.6	928227	English	21

- The num_voted_users column is filtered
- Created a new column that contains the list of top 250 IMDB movies of various languages by sorting the imdb_score column

IMDB movie analysis

3.2 LIST OF IMDB TOP 250 MOVIE OF FOREIGN LANGUAGE

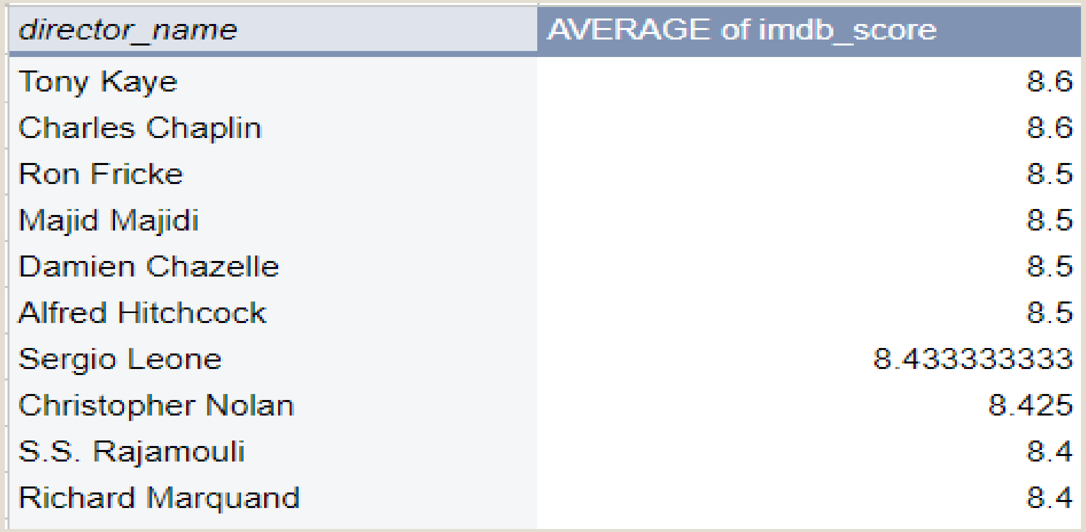


movie_title	imdb_score	num_voted_user	language	rank
The Good, the B	8.9	503509	Italian	1
Seven Samurai	8.7	229012	Japanese	2
City of God	8.7	533200	Portuguese	2
Spirited Away	8.6	417971	Japanese	4
Children of Heav	8.5	27882	Persian	5
The Lives of Oth	8.5	259379	German	5
Baahubali: The E	8.4	62756	Telugu	7
A Separation	8.4	151812	Persian	7
Das Boot	8.4	168203	German	7
Princess Monon	8.4	221552	Japanese	7
Oldboy	8.4	356181	Korean	7
Amélie	8.4	534262	French	7
Metropolis	8.3	111841	German	13
The Hunt	8.3	170155	Danish	13
Downfall	8.3	248354	German	13
Incendies	8.2	80429	French	16
The Secret in Th	8.2	131831	Spanish	16
Howl's Moving C	8.2	214091	Japanese	16
Pan's Labyrinth	8.2	467234	Spanish	16
Tae Guk Gi: The	8.1	31943	Korean	20
The Sea Inside	8.1	64556	Spanish	20
The Celebration	8.1	65951	Danish	20
Elite Squad	8.1	81644	Portuguese	20
Akira	8.1	106160	Japanese	20
Amores Perros	8.1	173551	Spanish	20
Central Station	8	28951	Portuguese	26
Waltz with Bashi	8	46107	Hebrew	26

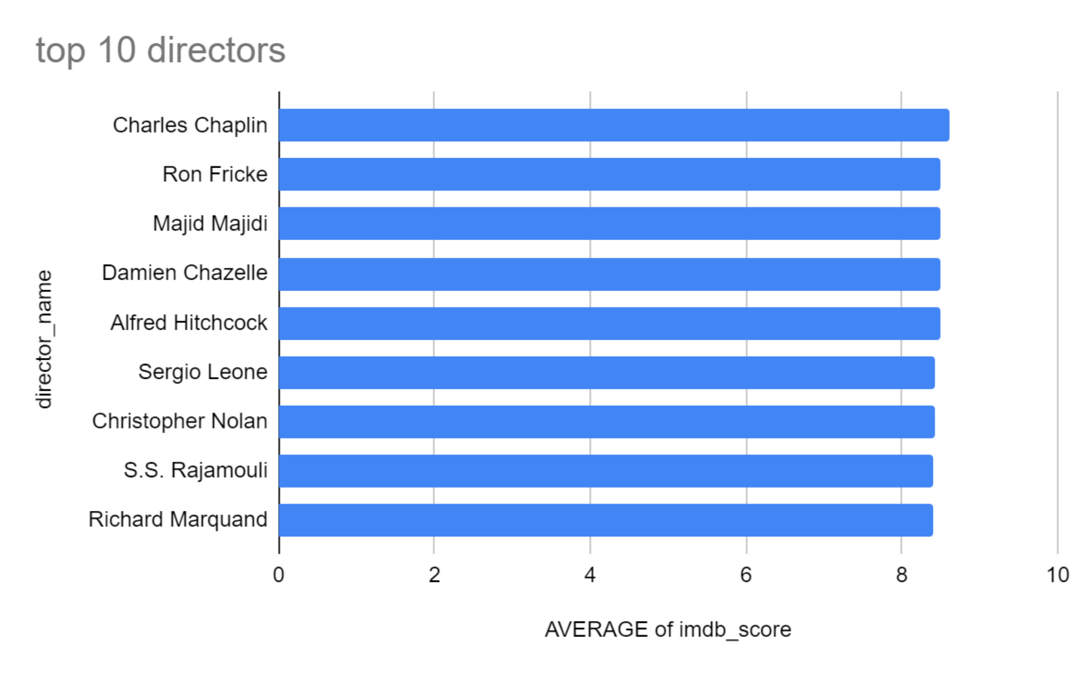
- The IMDB top 250 movie list is again filtered based on the language column.
- The movies of foreign languages are listed by applying conditional formatting of the language column.

IMDB movie analysis

4. THE BEST DIRECTORS



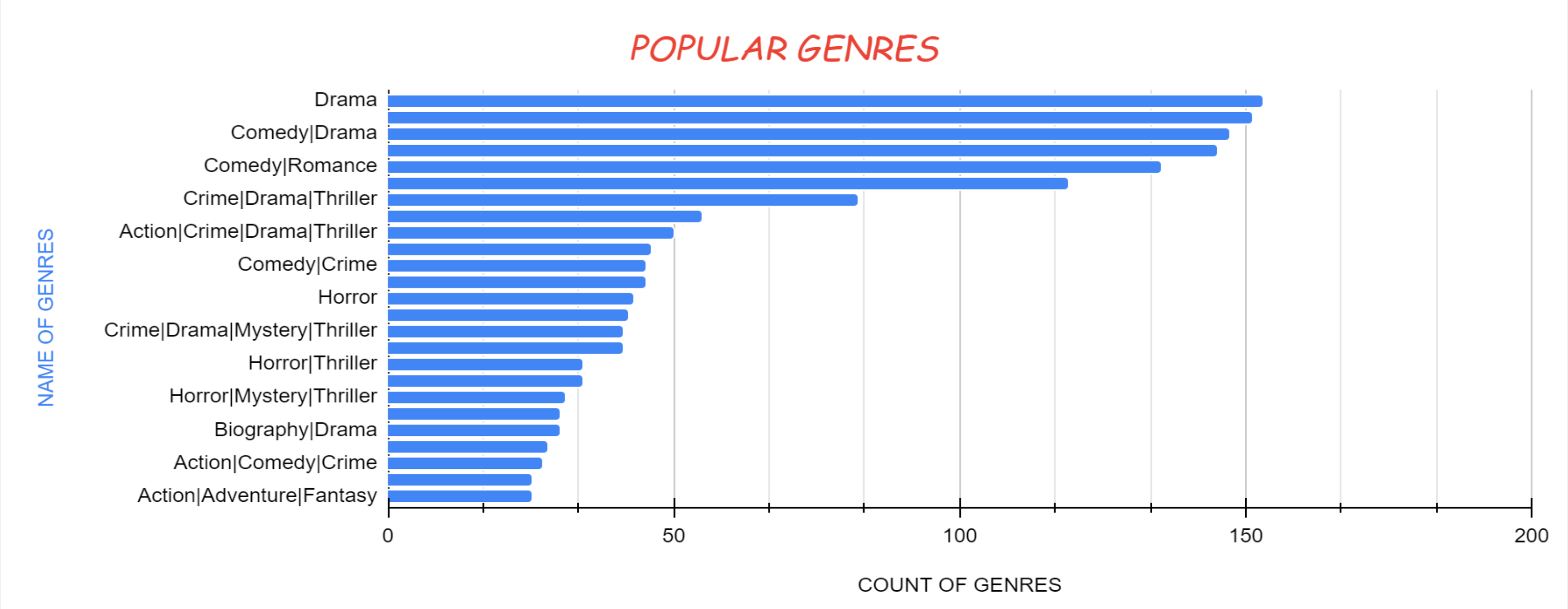
director_name	AVERAGE of imdb_score
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
S.S. Rajamouli	8.4
Richard Marquand	8.4

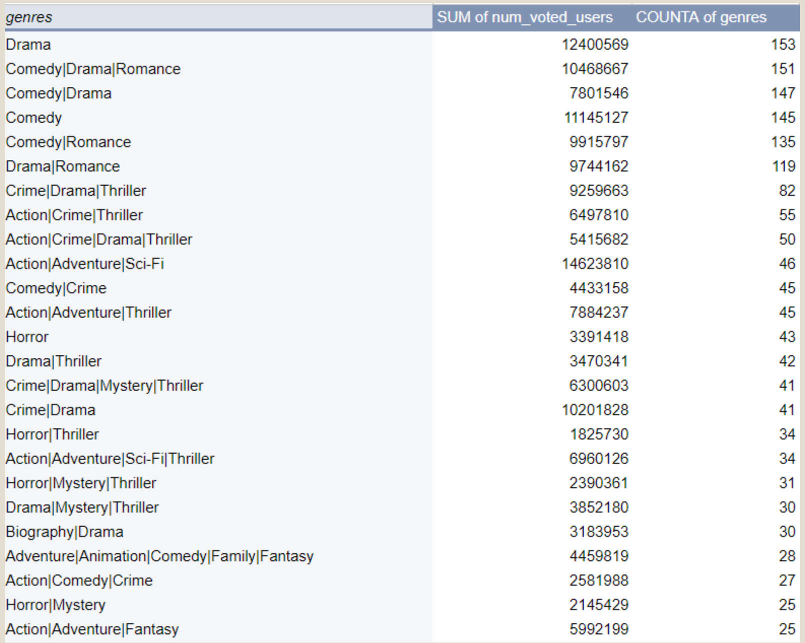


- From the cleaned data obtained in question1, we create a pivot table
- Rows: director_name
- Values: imdb_score
- The list of directors is filtered to obtain the top 10
- The same is represented as a graph

IMDB movie analysis

5. THE MOST POPULAR GENRES IN THE MOVIE INDUSTRY





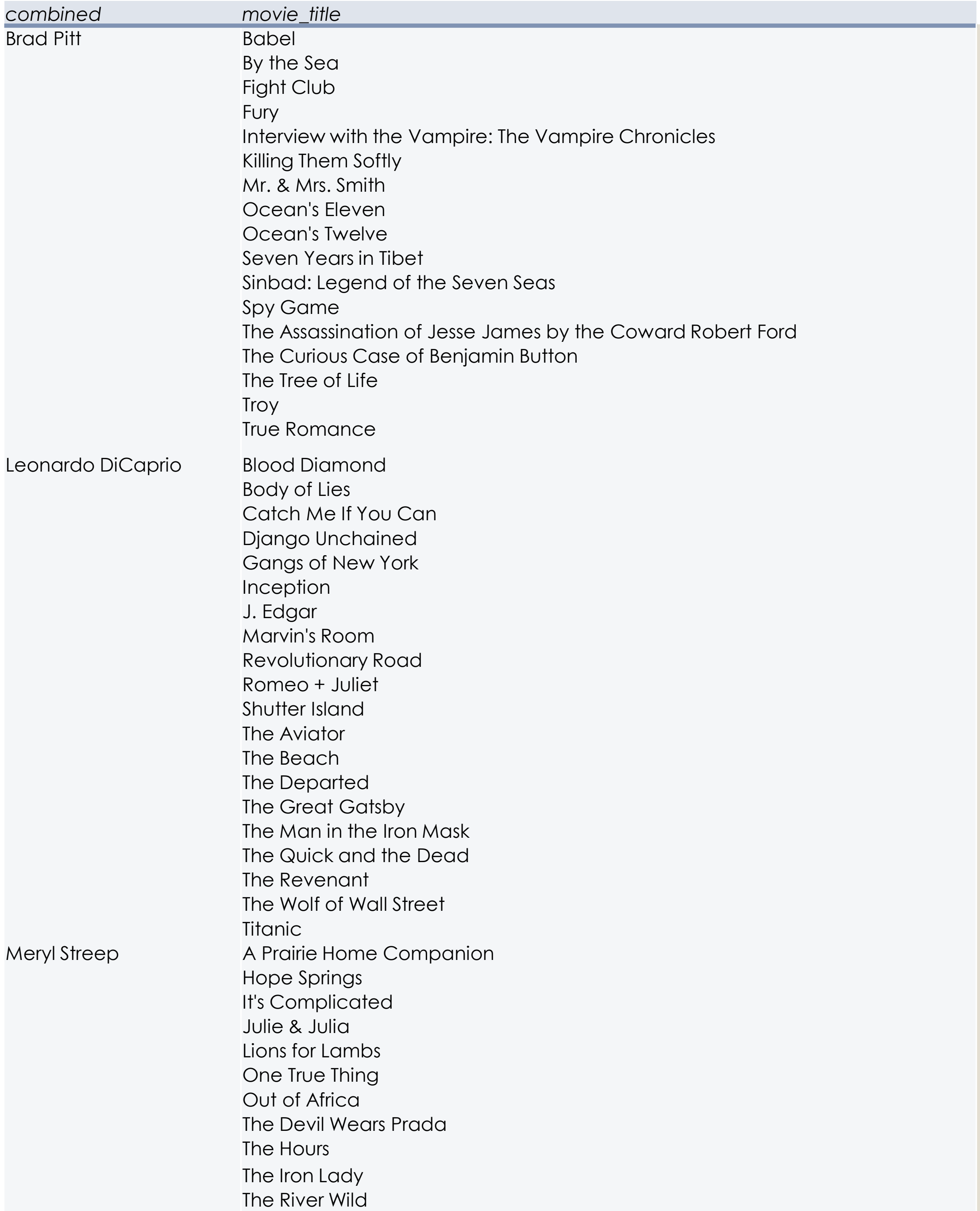
genres	SUM of num_voted_users	COUNTA of genres
Drama	12400569	153
Comedy|Drama|Romance	10468667	151
Comedy|Drama	7801546	147
Comedy	11145127	145
Comedy|Romance	9915797	135
Drama|Romance	9744162	119
Crime|Drama|Thriller	9259663	82
Action|Crime|Thriller	6497810	55
Action|Crime|Drama|Thriller	5415682	50
Action|Adventure|Sci-Fi	14623810	46
Comedy|Crime	4433158	45
Action|Adventure|Thriller	7884237	45
Horror	3391418	43
Drama|Thriller	3470341	42
Crime|Drama|Mystery|Thriller	6300603	41
Crime|Drama	10201828	41
Horror|Thriller	1825730	34
Action|Adventure|Sci-Fi|Thriller	6960126	34
Horror|Mystery|Thriller	2390361	31
Drama|Mystery|Thriller	3852180	30
Biography|Drama	3183953	30
Adventure|Animation|Comedy|Family|Fantasy	4459819	28
Action|Comedy|Crime	2581988	27
Horror|Mystery	2145429	25
Action|Adventure|Fantasy	5992199	25

- A pivot table is created from the base cleaned data
- Rows: genres
- values: num_voted_users
  count of each genre

IMDB movie analysis

6.1 THE CRITIC-FAVOURITE AND AUDIENCE-FAVOURITE ACTORS

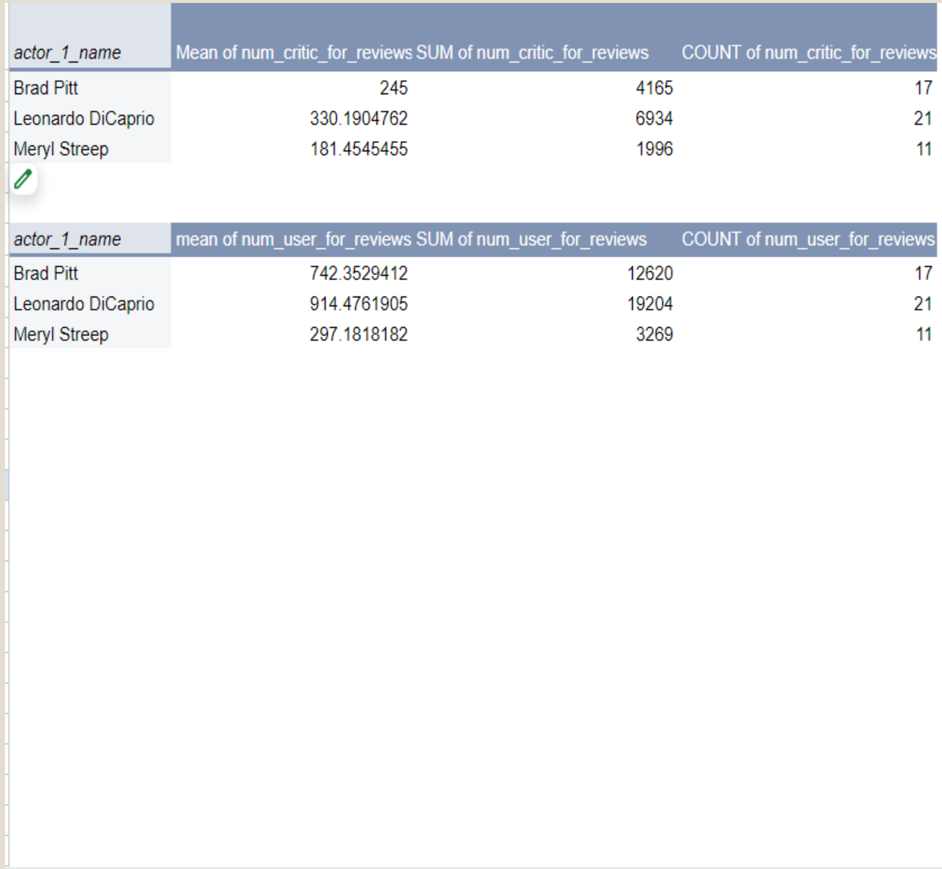
- Data extraction is performed based on the actor_1_name column
- The movies acted by the respective actors are listed using the necessary steps



combined	*movie_title*
Brad Pitt	Babel
	By the Sea
	Fight Club
	Fury
	Interview with the Vampire: The Vampire Chronicles
	Killing Them Softly
	Mr. & Mrs. Smith
	Ocean's Eleven
	Ocean's Twelve
	Seven Years in Tibet
	Sinbad: Legend of the Seven Seas
	Spy Game
	The Assassination of Jesse James by the Coward Robert Ford
	The Curious Case of Benjamin Button
	The Tree of Life
	Troy
	True Romance
Leonardo DiCaprio	Blood Diamond
	Body of Lies
	Catch Me If You Can
	Django Unchained
	Gangs of New York
	Inception
	J. Edgar
	Marvin's Room
	Revolutionary Road
	Romeo + Juliet
	Shutter Island
	The Aviator
	The Beach
	The Departed
	The Great Gatsby
	The Man in the Iron Mask
	The Quick and the Dead
	The Revenant
	The Wolf of Wall Street
	Titanic
Meryl Streep	A Prairie Home Companion
	Hope Springs
	It's Complicated
	Julie & Julia
	Lions for Lambs
	One True Thing
	Out of Africa
	The Devil Wears Prada
	The Hours
	The Iron Lady
	The River Wild

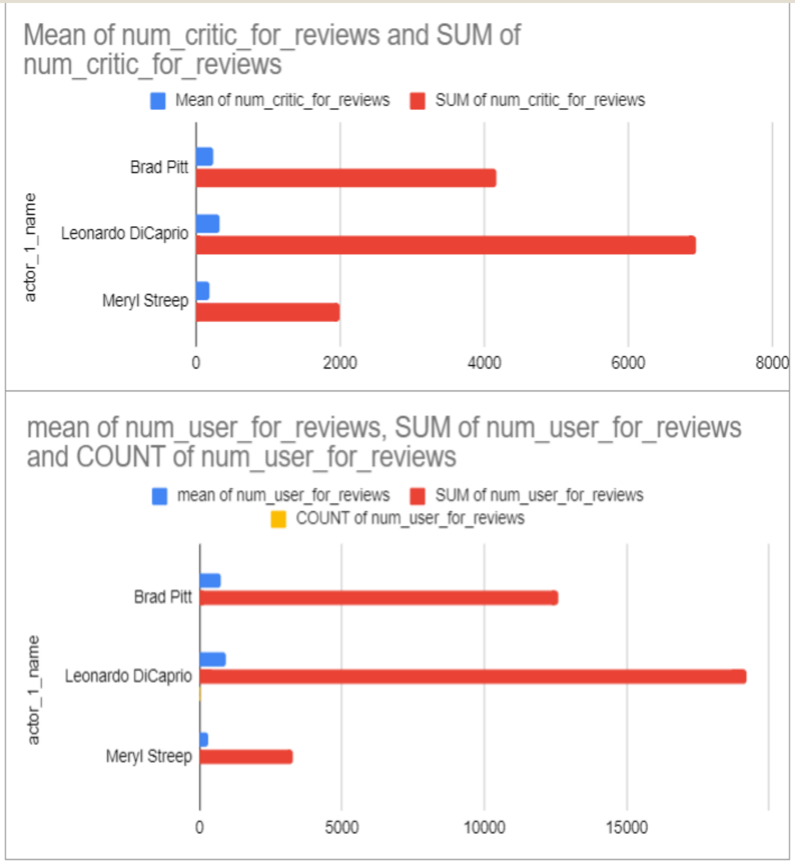
IMDB movie analysis

6.2 THE GIVEN ACTORS WITH THE HIGHEST MEAN OF CRITIC REVIEWS AND USER REVIEWS ARE OBSERVED



actor_1_name	Mean of num_critic_for_reviews	SUM of num_critic_for_reviews	COUNT of num_critic_for_reviews
Brad Pitt	245	4165	17
Leonardo DiCaprio	330.1904762	6934	21
Meryl Streep	181.4545455	1996	11

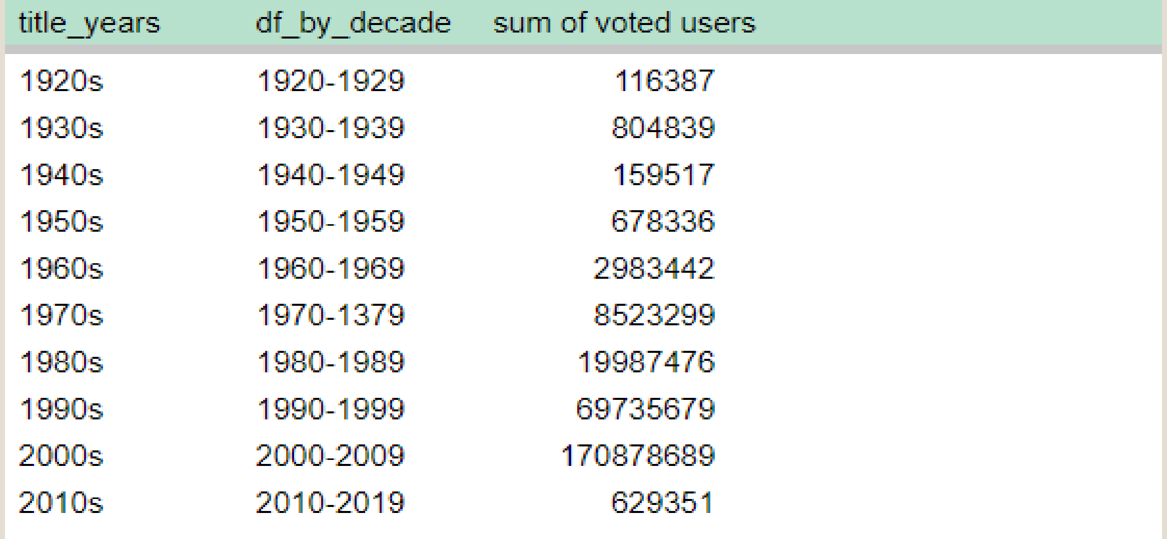
actor_1_name	mean of num_user_for_reviews	SUM of num_user_for_reviews	COUNT of num_user_for_reviews
Brad Pitt	742.3529412	12620	17
Leonardo DiCaprio	914.4761905	19204	21
Meryl Streep	297.1818182	3269	11



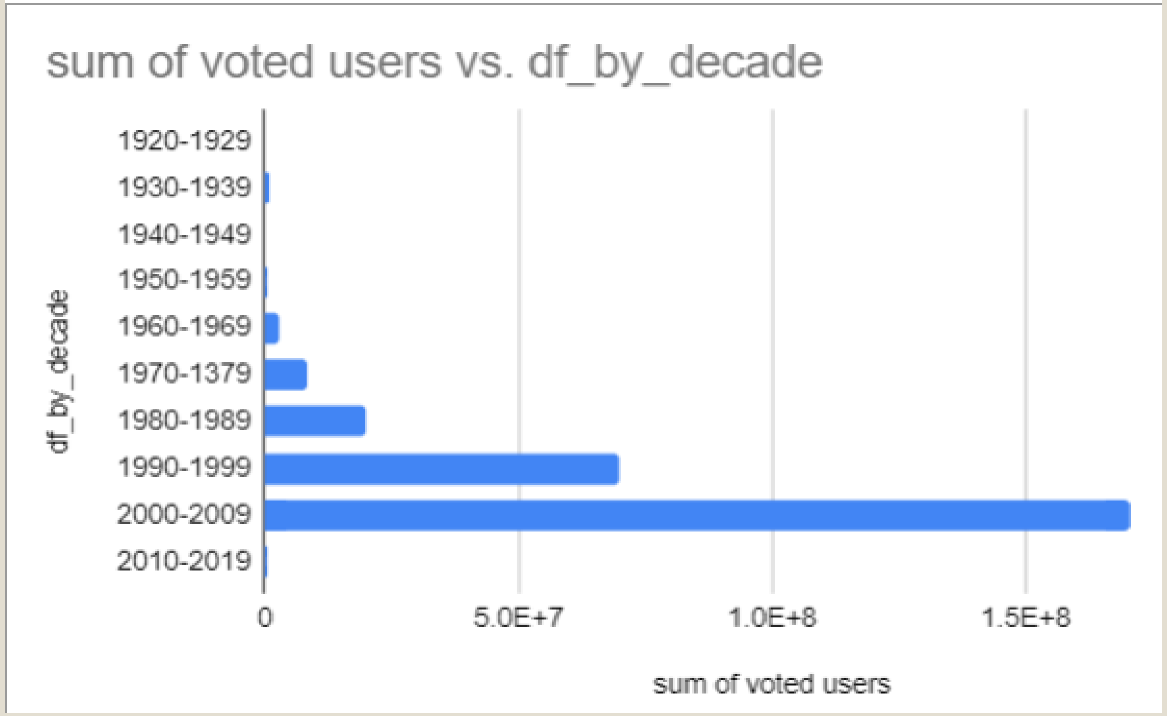
- Two pivot tables are created with the rows as actors names and values of mean of critic reviews and mean of user reviews respectively.
- Graphs of the same are also presented
- Highest critic reviews- Leonardo DiCaprio
- Highest user reviews- Leonardo Dicaprio

IMDB movie analysis

6.3 NUMBER OF VOTED VOTERS BY EACH DECADE IS PICTORIALLY REPRESENTED



title_years	df_by_decade	sum of voted users
1920s	1920-1929	116387
1930s	1930-1939	804839
1940s	1940-1949	159517
1950s	1950-1959	678336
1960s	1960-1969	2983442
1970s	1970-1379	8523299
1980s	1980-1989	19987476
1990s	1990-1999	69735679
2000s	2000-2009	170878689
2010s	2010-2019	629351



- Movies are listed with their year of release and sum of voted users.
- The number of voted users for each decade is calcuted and represented using a bar graph

TECH STACK USED

- Google sheets
  - ❑The google sheets platform has been used to check for duplicates, clean the data, apply the necessary functions and arrive at the final results
  - ❑It has also been used to represent the data pictorially using the charts feature
  - ❑The pivot table feature, conditional formatting, freeze feature, and more have been useful to complete the given task
- Microsoft PowerPoint
  - ❑Used to give a detailed report of the case study

INSIGHTS

- The given dataset is cleansed of all duplicates, null values, and other unnecessary data.
- The profit of each movie is calculated and represented in a graph
- The highest-profit movies are Avatar, Jurassic World, and so on
- The IMDB top 250 movies in English and other international languages have been recorded
- The list of top 10 directors based on their IMDB score has been deduced
- The popular genres in the movie industry have been discovered
- The data of the actors' Brad Pitt, Leonardo DiCaprio, and Meryl Streep have been extracted and ranked based on the mean of critic reviews and user reviews respectively
- Leonardo DiCaprio has the highest mean of critic reviews and user reviews.
- Finally, each decade's vote users are calculated and represented graphically.

RESULT

- I applied the knowledge of advanced Google Sheets in this project
- This project has been a suitable case study for the application of all the concepts learned
- This project has also helped me to explore the other features and functions in google sheets and enhanced my understanding of pivot tables

LINK

https://docs.google.com/spreadsheets/d/1LrwF-mdCOQyW7sQS1tg_z-lvTJYx0QaZWJVndtf3x3I/edit?usp=drive_link

APOLOGY: I was unable to be active in the internship period and keep up with the project deadlines due to internals, practicals, and semester examinations(I study in a 3rd tier college in Tamil Nadu). Hopefully, I will submit all the projects before the internship ends. I request you to kindly consider my delay in submissions. Thank you.