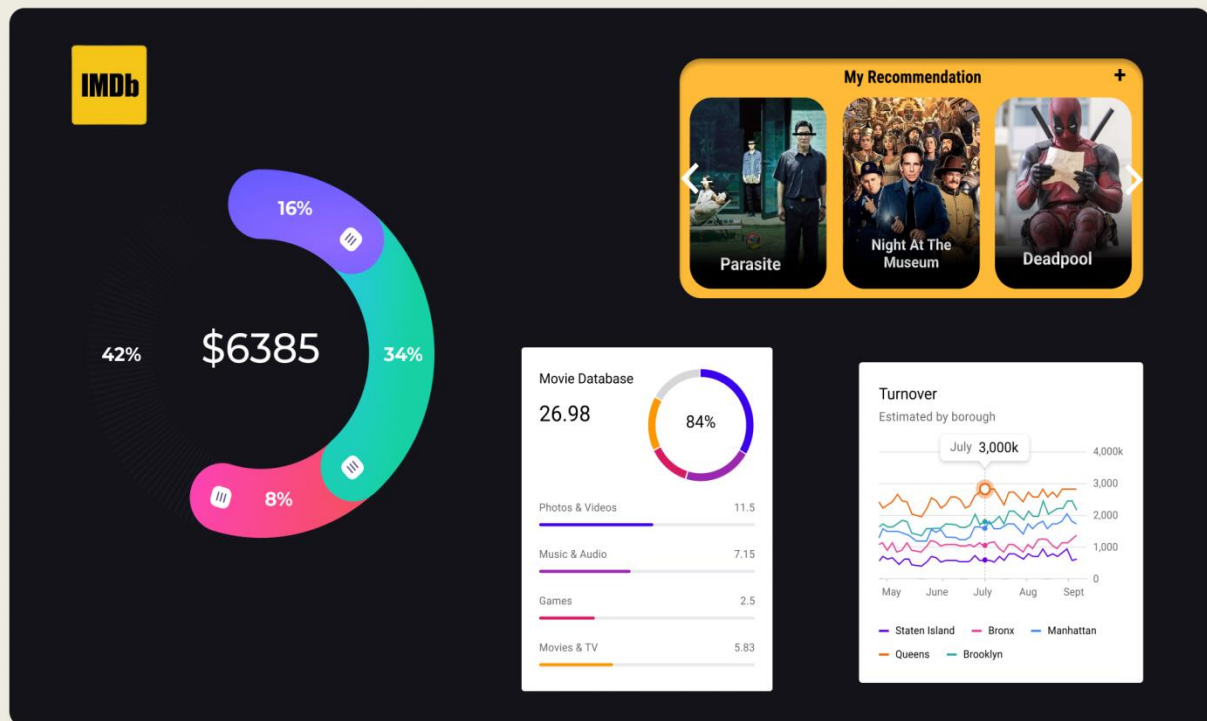


IMDB Movie Analysis

Final Project-1

(Project – 5)



Sowmiya R | Data Analytics | 20 July 2023

PROJECT DESCRIPTION

For this Final Project, they were providing me with dataset having various columns of different IMDB Movies. For this task, I will need to define a problem I want to shed some light on.

Use these questions to guide your thinking:

- What do you see happening?
- What is your hypothesis for the cause of the problem? (this will be broadly based on intuition initially)
- What is the impact of the problem on stakeholders?
- What is the impact of the problem not being solved?

Once defined a problem, clean the data as necessary, and using our Data Analysis skills to explore the data set and derive insights.

- What do you see happening?
- What are the specific symptoms of the problem?
- What is your hypothesis for the cause of the problem?

Five 'Whys' approach

Q: “Why did we go over budget on our project?”

A: It took much longer than we expected to complete.

Q: “Why did it take longer than expected to complete?”

A: We had to redesign several elements of the product.

Q: “Why did we have to redesign elements of the product?”

A: Features of the product were confusing to use.

Q: “Why were the features of the product confusing to use?”

A: We made incorrect assumptions about what users wanted.

Q: “Why did we make incorrect assumptions about what users wanted?”

A: Our user experience research team didn't ask effective questions.

I have been solving these questions by using the 5 why's approach analysis and making insights of using the excel data functions and mysql queries to get the desired results as well. Finding out the case studies as there are as follows....

CASE STUDIES

- ✓ Clean the data

(Dropping columns, removing null values, etc)

- ✓ Find the movies with the highest profit?

(observing the outliers using the appropriate chart type)

- ✓ Find IMDB Top 250

(Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film.)

- ✓ Find the best directors

(Top 10 director by highest mean of IMDB score)

- ✓ Find popular genres

(Perform this by step while performing previous steps)

- ✓ Find the critic-favorite and audience-favorite actors

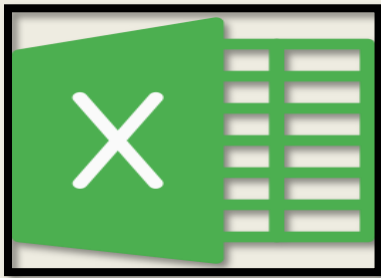
(Finding the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean) and with decade of new data frame called df_by_decade.

APPROACH

First of all, I have downloaded the provided dataset and then load the dataset in the MS Excel , performing technique of data cleaning strategy Open dataset in the form of tables and perform few process in it, with that I have been used with a MySQL workbench to query certain case studies.

TECH-STACK USED

MySQL workbench and Microsoft Excel is used to answer the required questions with the analysis and insights.



INSIGHTS AND RESULTS OF ANALYSIS

1. **Cleaning the data:** This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.

Your task: Clean the data

- ❖ First of all I have analyzed that with the provided dataset , there are 5043 rows and 28 columns and 1 header row are present there.
- ❖ After then, I looked a project description that related and required columns can only be used for the upcoming analysis so then I have been used a data cleaning strategy using MySQL workbench and excel functions to solve the data cleaning task.
- ❖ So I need only the columns and they are :

- **director_name**
- **num_critic_for_reviews**
- **gross**
- **genres**
- **actor_1_name**
- **movie_title**
- **num_voted_users**
- **num_user_for_reviews**

- language
- budget
- title_year
- imdb_score
- movie_facebook_likes

❖ After these are the only included then I left with a **13 columns** .

Then I go with a null values and null rows , this was get rid by using the excel function (press F5 – go to special – select blanks – press ok) , it will show the blank rows and then (data – remove duplicates) it will remove the duplicates and save this sheet . Then load this currently saved dataset in the MySQL workbench , queried for removing the extra null values precisely.

❖ And save the removed null values dataset and again load in the excel to see how many entries has been left .

❖ By performing these functions in the excel I found that the total number of rows is then 3789.

❖ Below this , I have attached with cleaned dataset as a picture of removing duplicates and with the blanks, required columns , and set the blank(language) column been default as 'English'.

❖ This is my attached excel file.

https://docs.google.com/spreadsheets/d/1Ha--2RrVp2mce6NS6hj5dGjqWq4jOpiw/edit?usp=drive_link

Director	num_crits_for_reviews	gross	genres	actor_name	movie_title	num_voted_users	num_user_for_reviews	average	budget	gross_year	imdb_score	movie_facebook_likes	likes
James Cameron	721	76505447	Action/Adventure/Fantasy/Sci-Fi	CCO Theatrical	Avatar	85054	210	8.5	237000000	2009	7.8	33000	13300
Gore Verbinski	302	309404152	Action/Adventure/Fantasy	Johnny Depp	Prince of the Caribbean: At World's End	471220	0	7.2	1238	2006	7.1	0	0
Sam Mendes	402	20074175	Action/Adventure/Fantasy	Christian Bale	Quantum Break	24500000	0	8.2	273680	2008	6.8	8500	0
Christopher Nolan	811	44810042	Action/Thriller	Tom Hardy	The Dark Knight Rises	1144317	2701	8.5	250000000	2012	8.5	144000	0
Sam Raimi	462	7308089	Action/Adventure/Fantasy	Daryl Sallars	John Carter	212204	718	6.8	26300000	2012	6.6	24000	0
Nathan Gross	392	33630109	Action/Adventure/Fantasy	J.L. Simmons	Spider-Man 3	189056	189056	6.7	250000000	2007	6.2	15000	0
John Wood	324	20080782	Adventure/Animation/Comedy/Family/Fantasy/Musical/Romance	Brad Pitt	Mr. & Mrs. Smith	294810	387	6.8	200000000	2005	7.8	29000	0
David Yates	431	41089109	Action/Adventure/Fantasy	Chris Hemsworth	Thor	2000000	2000000	7.5	150000000	2011	7.5	118000	0
Jack Snyder	375	3304062	Adventure/Animation/Comedy/Family/Mystery	Alan Rickman	Harry Potter and the Half-Blood Prince	312795	971	6.8	200000000	2009	7.5	10000	0
Bryan Singer	434	20060608	Action/Adventure/Fantasy	Henry Cavill	Man of Steel	2500000	2500000	7.5	220000000	2013	6.9	197000	0
Marc Forster	311	42102528	Action/Adventure	Kevin Spacey	Quantum of Solace	240396	2367	6.8	200000000	2008	6.1	0	0
Gore Verbinski	311	42102528	Action/Adventure/Fantasy	Giancarlo Giannini	Prince of the Caribbean: Dead Man's Chest	330784	1241	6.8	200000000	2006	6.7	0	0
Gore Verbinski	403	8028910	Action/Adventure/Western	Johnny Depp	Prince of the Caribbean: The Curse of the Black Pearl	522400	22200000	2006	7.3	7500	0	0	
Jack Snyder	219	20120156	Action/Adventure/Fantasy	Henry Cavill	The Lone Ranger	181792	711	6.8	215000000	2013	6.5	14800	0
Andrew Adamson	258	14164603	Action/Adventure/Family/Fantasy	Robin McKee	The Chronicles of Narnia: Prince Caspian	548573	438	6.8	250000000	2010	6.8	17000	0
John Wood	705	6212954	Action/Adventure/Sci-Fi	Robert DeNiro	The Avengers	1995415	1722	8.5	220000000	2012	8.1	12800	0
Rob Marshall	440	24106379	Action/Adventure/Fantasy	Chris Hemsworth	Prince of the Caribbean: On Stranger Tides	370704	484	6.8	200000000	2011	6.7	4000	0
Peter Jackson	451	17502054	Action/Adventure/Comedy/Family/Fantasy/Sci-Fi	Will Smith	Men in Black 3	265154	341	6.8	250000000	2012	6.8	40000	0
Marc Webb	422	25108170	Adventure/Animation/Comedy/Family/Fantasy	Adam Turner	The Hobbit: The Battle of the Five Armies	354208	802	6.8	200000000	2014	7.5	60000	0
Marc Webb	599	26200863	Action/Adventure/Fantasy	Emma Stone	The Amazing Spider-Man	451003	1225	6.8	230000000	2012	7	50000	0
Kelly Scott	520	12510139	Action/Adventure/Drama/Mystery	Mark Addy	Robin Hood	211765	546	6.8	200000000	2010	6.8	17000	0
Peter Jackson	509	25855554	Adventure/Fantasy	Adam Turner	The Hobbit: The Desolation of Smaug	483540	951	6.8	250000000	2013	7.8	80000	0
Chris Weitz	251	70083319	Adventure/Family/Fantasy	Christopher Lee	The Golden Compass	149019	664	6.8	180000000	2007	6.1	0	0
Peter Jackson	446	21401140	Action/Adventure/Drama/Romance	Naomi Watts	King Kong	316818	2018	6.8	207000000	2005	7.2	5000	0
James Cameron	315	65087262	Drama/Romance	Leonardo DiCaprio	Titanic	729870	2524	6.8	200000000	1997	7.7	26000	0
Anthony Russo	407	40718782	Action/Adventure/Sci-Fi	Robert Downey Jr.	Captain America: Civil War	278059	1021	6.8	200000000	2016	8.2	70000	0
Peter Berg	37	61373160	Action/Adventure/Sci-Fi/Thriller	Liam Neeson	Raiders	202282	751	6.8	200000000	2012	5.9	0	0
Cole Truitt	644	61177731	Action/Adventure/Sci-Fi/Thriller	Bryan Dallas Howard	Transformers: Age of Extinction	418214	1300	6.8	200000000	2015	7	10000	0
Sam Mendes	750	30496277	Action/Adventure/Thriller	Albert Finney	Slayer	523010	1494	6.8	200000000	2012	7.8	80000	0
Shane Black	373	37377893	Action/Adventure/Fantasy/Romance	J.L. Simmons	Spider-Man 2	411144	1303	6.8	200000000	2004	7.5	95000	0
Tim Burton	601	40509272	Action/Adventure/Sci-Fi	Robert Downey Jr.	Iron Man 3	553989	1187	6.8	200000000	2013	7.8	95000	0
Tin Burton	451	33485206	Adventure/Family/Fantasy	Johnny Depp	Jack & the Beanstalk	306120	736	6.8	200000000	2010	6.5	2400	0
Ben Rutter	334	24450014	Action/Adventure/Sci-Fi/Thriller	Hugh Jackman	X-Men: The Last Stand	305120	21000000	2006	6.8	10000	0	0	
Don Scamion	376	268488329	Adventure/Animation/Comedy/Family/Fantasy	Steve Buscemi	Monsters University	235025	205	6.8	200000000	2013	7.3	44000	0
Michael Bay	465	40207689	Action/Adventure/Animation/Comedy/Family/Fantasy	Michael Bay	Transformers: The Movie	2405207	1449	6.8	200000000	2007	6	6	0
Michael Bay	378	24542813	Adventure/Animation/Comedy/Family/Fantasy	Brigitte Lirio	Transformers: Age of Extinction	242420	918	6.8	210000000	2014	5.7	56000	0
Sam Raimi	525	23403076	Adventure/Family/Fantasy	Tin Holmes	On the Great and Powerful	175409	511	6.8	210000000	2013	6.4	60000	0
Marc Webb	495	20583913	Action/Adventure/Comedy/Family/Sci-Fi	Emma Stone	The Amazing Spider-Man 2	328527	1067	6.8	200000000	2015	6.8	40000	0
Joseph Kosinski	407	17025178	Adventure/Animation/Comedy/Family/Fantasy	Jeff Bridges	Thor: The Dark World	264283	665	6.8	170000000	2010	6.8	30000	0
John Lasseter	304	15450079	Adventure/Animation/Comedy/Family/Fantasy/Sport	Jon Williams	Cars 2	1051138	20000000	2011	6.8	10000	0	0	
Martin Campbell	432	11589191	Action/Adventure/Sci-Fi	Green Day	223893	510	6.8	200000000	2011	5.6	24000	0	
Lee Unkrich	431	41884847	Adventure/Animation/Comedy/Family/Fantasy	Tom Hanks	Toy Story 4	544884	731	6.8	200000000	2010	8.1	30000	0
M.O.	422	12530007	Action/Adventure/Sci-Fi	Christian Bale	Terminator: Salvation	271000	721	6.8	200000000	2009	6.4	0	0
James Wan	434	35034131	Action/Crime/Thriller	Jason Statham	Fast Five	279122	657	6.8	190000000	2011	7.2	84000	0
Mark Forster	439	20334111	Action/Adventure/Animation/Comedy/Family/Sci-Fi/Thriller	Peter Capaldi	Doctor Who	18000000	18000000	2013	7	120000	0	0	
Bryan Singer	334	23194886	Action/Adventure/Animation/Comedy/Family/Sci-Fi/Thriller	Jennifer Lawrence	X-Men: Days of Future Past	154125	751	6.8	200000000	2014	8	82000	0
J.J. Abrams	509	23876532	Action/Adventure/Animation/Comedy/Family/Sci-Fi/Thriller	Benedit Cumberbatch	Star Wars: The Force Awakens	1001619	1001619	2015	7.8	19000	0	0	
Bryan Singer	338	6313480	Adventure/Fantasy	Eddie Marston	Jack the Giant Slayer	106416	205	6.8	190000000	2013	6.3	22000	0
Bar Lufmann	440	14617776	Drama/Romance	Leonardo DiCaprio	The Great Gatsby	306212	751	6.8	190000000	2013	7.3	11000	0
Mike Newell	300	9075648	Action/Adventure/Fantasy/Romance	Jack O'Connell	Prince of Persia: The Sands of Time	221000	434	6.8	190000000	2010	6.4	2800	0
Guillermo del Toro	451	95178542	Action/Adventure/Sci-Fi	Charlie Hunnam	Pacific Rim	381248	1206	6.8	190000000	2013	7	83000	0
Michael Bay	425	3523879	Action/Adventure/Sci-Fi	Glenn Morshover	Transformers: Dark of the Moon	2000000	2000000	2011	6.8	40000	0	0	
Steven Spielberg	470	31701114	Action/Adventure/Fantasy	Harrison Ford	Indiana Jones and the Kingdom of the Crystal Skull	339347	2054	6.8	180000000	2008	6.2	5000	0
Mark Andrews	480	23782182	Action/Adventure/Animation/Comedy/Family/Fantasy	Brad Pitt	How to Train Your Dragon 2	18500000	18500000	2016	7.2	30000	0	0	
Justin Lin	322	13648626	Action/Adventure/Sci-Fi/Thriller	Sofia Boutella	Sanctuary	53607	421	6.8	180000000	2016	7.5	30000	0
Andrew Stanton	322	13280888	Adventure/Animation/Comedy/Family/Sci-Fi	John Ratzenberger	WALL-E	718817	1041	6.8	180000000	2008	8.4	16000	0
Brett Ratner	362	14008089	Action/Comedy/Crime/Thriller	Tai Ma	Knockout	121104	14000000	2012	6.8	2007	6.2	18000	0
Roland Emmerich	367	16611267	Action/Adventure/Sci-Fi	Oliver Platt	2012	283418	1055	6.8	200000000	2009	5.8	13000	0
Robert Zemeckis	240	17190006	Animation/Drama/Comedy/Family/Fantasy	Robin Wright	A Christmas Carol	71889	400	6.8	200000000	2009	6.8	6.8	0
Lana Wachowski	384	47375127	Action/Adventure/Sci-Fi	Channing Tatum	Johnny Ascending	139933	728	6.8	170000000	2015	5.4	44000	0
Robert Zemeckis	240	13405179	Action/Adventure/Animation/Comedy/Family/Fantasy	Chris Rock	The Legend of Sleepy Hollow	121077	238	6.8	170000000	2010	6.1	16000	0
Andrew Adamson	284	29150845	Adventure/Family/Fantasy	Jim Broadbent	The Chronicles of Narnia: The Lion, the Witch and the Wardrobe	286106	1461	6.8	180000000	2005	6.9	0	0
Bryan Singer	440	1508088	Action/Adventure/Sci-Fi	Jennifer Lawrence	X-Men: Apocalypse	148479	627	6.8	178000000	2016	7.3	54000	0
Christopher Nolan	445	15311661	Action/Crime/Thriller	Christian Bale	The Dark Knight	1876488	1876488	2008	190000000	2008	9	17000	0
Peter Dinklage	408	20297956	Adventure/Animation/Comedy/Family	John Ratzenberger	Up	601075	704	6.8	175000000	2009	8.3	72000	0
Rob Letterman	219	13933128	Action/Adventure/Animation/Comedy/Family/Sci-Fi	Andy Patinkin	Monsters vs. Aliens	17600000	17600000	2009	6.8	6.8	0	0	
John Favreau	480	31268180	Action/Adventure/Sci-Fi	Robert Downey Jr.	Iron Man 2	696138	1055	6.8	140000000	2008	7.9	10000	0
Martin Scorsese	682	7382008	Adventure/Drama/Family/Mystery	Chadwick Boseman	Spider-Man: Homecoming	245131	678	6.8	170000000	2017	7.5	10000	0
Bryan Singer	445	11274548	Action/Adventure/Comedy/Family/Sci-Fi	Will Smith	Men in Black III	211001	501	6.8	160000000	2012	6.8	40000	0
Rob Cohen	410	16507183	Action/Adventure/Sci-Fi	Will Smith	The Man with the Iron Fists	117927	1007	6.8	140000000	2008	5.2	0	0
David Ayer	410	16507183	Action/Adventure/Sci-Fi	Will Smith	Men in Black III	120892	1007	6.8	140000000	2012	6.8	40000	0
Tom Shadyac	186	10088900	Comedy/Family/Fantasy	Jimmy Bennett	Ever After	115099	257	6.8	175000000	2007	5.4	2000	0
Doug Liman	100	10089902	Action/Adventure/Sci-Fi	Tom Cruise	Edge of Tomorrow	27800000	27800000	2015	7.8	70000	0	0	
Kevin Reynolds	91	88462420	Action/Adventure/Sci-Fi/Thriller	Joanna Trapp	Waterworld	144317	309	6.8	170000000	1995	6.1	0	0
Stephen Sommers	250	15067830	Action/Adventure/Sci-Fi/Thriller	Joseph Gordon-Livitt	G.I. Joe: The Rise of Cobra	17476	534	6.8	170000000	2009	5.8	0	0
Peter Dinklage	354	36545487	Adventure/Animation/Comedy/Drama/Family/Fantasy	Andy Paster	Monsters vs. Aliens	17600000	17600000	2009	6.8	6.8	0	0	
John Favreau	370	36245414	Adventure/Drama/Fantasy	Scarlett Johansson	The Jungle Book	100071	398	6.8	170000000	2016	7.8	60000	0
John Favreau	313	31057433	Action/Adventure/Sci-Fi	Robert Downey Jr.	Iron Man 2	522171	1000	6.8	170000000	2008	7.9	18000	0
Rupert Sanders	451	15111415	Action/Adventure/Drama/Fantasy	Chris Hemsworth	Thor: The Dark World	228054	718	6.8	170000000	2012	6.1	53000	0
Robert Zemeckis	402	14407328	Action/Adventure/Drama/Fantasy/Romance	Angela Julia Pitt	Madagascar 3: Europe's Most Wanted	1804817	1804817	2012	190000000	2012	6.8	10000	0
Mark Reeves	521	20644795	Action/Adventure/Drama/Sci-Fi	Gary Oldman	Dawn of the Planet of the Apes	317462	628	6.8	170000000	2014	7.6	40000	0
Carl Rinsler	218	3829765	Action/Adventure/Drama/Fantasy	Keanu Reeves	At World's End	116894	324	6.8	170000000	2013	6.3	10000	0
Anthony Russo	576	24274605	Action/Adventure/Sci-Fi	Caplan Johnson	Capitan America: The Winter Soldier	138861	171	6.8	160000000	2015	6.4	0	0
Mike Mitchell	220	23817387	Adventure/Animation/Comedy/Family/Fantasy	Jon Hamm	Seven Years Ache	138861	171	6.8	160000000	2015	6.4	0	0
Brad Bird	441	9147160	Action/Adventure/Animation/Comedy/Family/Sci-Fi	Tom Hanks	Toy Story	10000000	10000000	2016	7.8	37000	0	0	
Don Hall	384	2247711	Adventure/Animation/Comedy/Drama/Family/Sci-Fi	Damon Wayans Jr.	Big Hero 6	279093	421	6.8	160000000	2014	7.9	40000	0
Robert Zemeckis	186	1802627	Adventure/Animation/Comedy/Family/Fantasy	Will Smith	Men in Black	272534	376	6.8	160000000	2002	7.3	10000	0
Robert Zemeckis	186	655426	Adventure/Animation/Comedy/Family/Fantasy	Tom Hanks	The Polar Express	130798	444	6.8	160000000	2004	6.6	10000	0
Roland Emmerich	280	10235545	Action/Adventure/Sci-Fi	Nicki Velez	Independence Day: Resurgence	58137	526	6.8	160000000	2016	5.5	67000	0
David DeSilva	280	21780797	Adventure/Animation/Comedy/Family/Fantasy	George Baker	The Last of Us: Part II	488428	411	6.8	160000000	2017	8.2	33000	0
Jonathan Mostow	280</												

2. **Movies with highest profit:** Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.

Your task: Find the movies with the highest profit?

- ❖ Now as the values in gross and budget columns are very large so I have used a excel to get rid of these large values is not readable that (go to format cells – select custom – type #,,”million” – then select ok) it will give us the readable values , i.e 274M.
- ❖ Next , we create a column called ‘profit’ as gross-budget difference in the MySQL workbench to get the desired results.
The query can be as follows:

```
SELECT director_name,  
       movie_title,  
       ROUND(budget, 0) AS budget,  
       ROUND(gross - budget, 0) AS profit  
FROM  
       movisp.cleanedm  
ORDER BY gross - budget DESC  
LIMIT 10;
```

With this query , I have found out the result of movie_title are as follows:

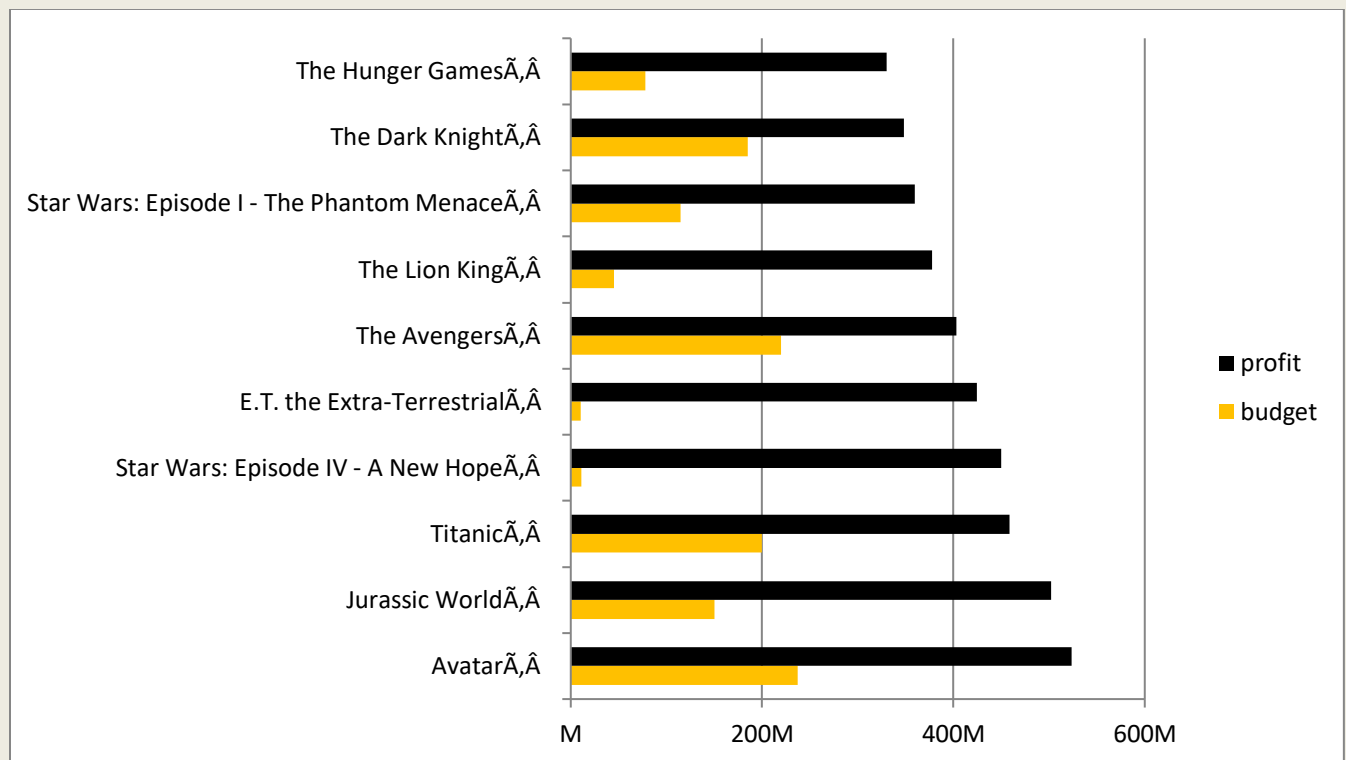
Avatar
Jurassic World
Titanic
Star Wars: Episode IV - A New Hope
E.T. the Extra-Terrestrial
The Avengers
The Lion King
Star Wars: Episode I - The Phantom Menace
The Dark Knight
The Hunger Games

director_name	movie_title	budget	profit
James Cameron	Avatar	237M	524M
Colin Trevorrow	Jurassic World	150M	502M
James Cameron	Titanic	200M	459M
George Lucas	Star Wars: Episode IV - A New Hope	11M	450M
Steven Spielberg	E.T. the Extra-Terrestrial	11M	424M
Joss Whedon	The Avengers	220M	403M
Roger Allers	The Lion King	45M	378M
George Lucas	Star Wars: Episode I - The Phantom Menace	115M	360M
Christopher Nolan	The Dark Knight	185M	348M
Gary Ross	The Hunger Games	78M	330M

This the result of above query , that we have found that Movies with Highest profit . And the insights of this table are:

This is my attached excel file:

https://docs.google.com/spreadsheets/d/1vU6T3pqE0nJI688PbSpdxmA8LBAjKr0R/edit?usp=drive_link



From the above bar chart represents that the highest profit of respected movies have been found to be analyzed and gave insights.

3. **Top 250:** Create a new column `IMDb_Top_250` and store the top 250 movies with the highest IMDb Rating (corresponding to the column: `imdb_score`). Also make sure that for all of these movies, the `num_voted_users` is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film. You can use your own imagination also!

Your task: Find IMDB Top 250

a) For the first part we should look for all the movies which has higher imdb scores user votes. So that I had solved this question by query with the user votes of higher than 30000 are as follows:

```
SELECT row_number() over(order by imdb_score DESC,
num_voted_users DESC) as ranking,
imdb_score, num_voted_users, movie_title AS IMDb_Top_250, language
FROM moviesp.cleanedm
WHERE num_voted_users > 30000
LIMIT 250;
```

IMDB Top 250 Movies			
ranking	imdb_score	IMDb_Top_250	language
1	9.3	The Shawshank Redemption	English
2	9	The Dark Knight	English
3	9	The Godfather: Part II	English
4	8.9	The Lord of the Rings: The Return of the King	English
5	8.9	Schindler's List	English
6	8.8	Inception	English
7	8.8	Fight Club	English
8	8.8	Forrest Gump	English
9	8.8	The Lord of the Rings: The Fellowship of the Ring	English

10	8.8	Star Wars: Episode V - The Empire Strikes Back	English
11	8.7	The Matrix	English
12	8.7	The Lord of the Rings: The Two Towers	English
13	8.7	Star Wars: Episode IV - A New Hope	English
14	8.7	Goodfellas	English
15	8.6	Se7en	English
16	8.6	Interstellar	English
17	8.6	The Silence of the Lambs	English
18	8.6	Saving Private Ryan	English
19	8.6	American History X	English
20	8.6	Spirited Away	Japanese
21	8.5	The Dark Knight Rises	English
22	8.5	Gladiator	English
23	8.5	Django Unchained	English
24	8.5	The Departed	English
25	8.5	The Prestige	English
26	8.5	The Green Mile	English
27	8.5	Terminator 2: Judgment Day	English
28	8.5	Back to the Future	English
29	8.5	Raiders of the Lost Ark	English
30	8.5	The Lion King	English
31	8.5	The Pianist	English
32	8.5	Apocalypse Now	English
33	8.5	Psycho	English
34	8.4	American Beauty	English
35	8.4	Braveheart	English
36	8.4	WALL-E	English
37	8.4	Star Wars: Episode VI - Return of the Jedi	English
38	8.4	Amélie	French
39	8.4	Aliens	English
40	8.4	Once Upon a Time in America	English
41	8.4	Lawrence of Arabia	English
42	8.4	Das Boot	German
43	8.4	Baahubali: The Beginning	Telugu
44	8.3	Batman Begins	English
45	8.3	Inglourious Basterds	English
46	8.3	Eternal Sunshine of the Spotless Mind	English
47	8.3	Up	English
48	8.3	Toy Story	English
49	8.3	Good Will Hunting	English
50	8.3	Snatch	English
51	8.3	Toy Story 3	English

52	8.3	ScarfaceÃ,Â	English
53	8.3	Indiana Jones and the Last CrusadeÃ,Â	English
54	8.3	2001: A Space OdysseyÃ,Â	English
55	8.3	L.A. ConfidentialÃ,Â	English
56	8.3	Inside OutÃ,Â	English
57	8.3	UnforgivenÃ,Â	English
58	8.3	AmadeusÃ,Â	English
59	8.3	DownfallÃ,Â	German
60	8.3	Raging BullÃ,Â	English
61	8.3	RoomÃ,Â	English
62	8.3	MetropolisÃ,Â	German
63	8.2	V for VendettaÃ,Â	English
64	8.2	The Wolf of Wall StreetÃ,Â	English
65	8.2	Finding NemoÃ,Â	English
66	8.2	A Beautiful MindÃ,Â	English
67	8.2	Die HardÃ,Â	English
68	8.2	Gran TorinoÃ,Â	English
69	8.2	The Big LebowskiÃ,Â	English
70	8.2	How to Train Your DragonÃ,Â	English
71	8.2	Pan's LabyrinthÃ,Â	Spanish
72	8.2	Blade RunnerÃ,Â	English
73	8.2	Into the WildÃ,Â	English
74	8.2	CasinoÃ,Â	English
75	8.2	WarriorÃ,Â	English
76	8.2	Captain America: Civil WarÃ,Â	English
77	8.2	The ThingÃ,Â	English
78	8.2	Howl's Moving CastleÃ,Â	Japanese
79	8.1	The AvengersÃ,Â	English
80	8.1	Pirates of the Caribbean: The Curse of the Black PearlÃ,Â	English
81	8.1	Shutter IslandÃ,Â	English
82	8.1	Kill Bill: Vol. 1Ã,Â	English
83	8.1	The Sixth SenseÃ,Â	English
84	8.1	Guardians of the GalaxyÃ,Â	English
85	8.1	The Truman ShowÃ,Â	English
86	8.1	Sin CityÃ,Â	English
87	8.1	Jurassic ParkÃ,Â	English
88	8.1	No Country for Old MenÃ,Â	English
89	8.1	Monsters, Inc.Ã,Â	English
90	8.1	Gone GirlÃ,Â	English
91	8.1	Mad Max: Fury RoadÃ,Â	English
92	8.1	The Bourne UltimatumÃ,Â	English

93	8.1	Million Dollar BabyÃ,Â	English
94	8.1	DeadpoolÃ,Â	English
95	8.1	The Grand Budapest HotelÃ,Â	English
96	8.1	The MartianÃ,Â	English
97	8.1	The Imitation GameÃ,Â	English
98	8.1	12 Years a SlaveÃ,Â	English
99	8.1	Groundhog DayÃ,Â	English
100	8.1	The RevenantÃ,Â	English
101	8.1	PrisonersÃ,Â	English
102	8.1	There Will Be BloodÃ,Â	English
103	8.1	The HelpÃ,Â	English
104	8.1	RushÃ,Â	English
105	8.1	The Princess BrideÃ,Â	English
106	8.1	Hotel RwandaÃ,Â	English
107	8.1	SpotlightÃ,Â	English
108	8.1	The Sea InsideÃ,Â	Spanish
109	8.1	Tae Guk Gi: The Brotherhood of WarÃ,Â	Korean
110	8	Slumdog MillionaireÃ,Â	English
111	8	Black SwanÃ,Â	English
112	8	District 9Ã,Â	English
113	8	Catch Me If You CanÃ,Â	English
114	8	X-Men: Days of Future PastÃ,Â	English
115	8	Kill Bill: Vol. 2Ã,Â	English
116	8	Star TrekÃ,Â	English
117	8	The King's SpeechÃ,Â	English
118	8	The IncrediblesÃ,Â	English
119	8	RatatouilleÃ,Â	English
120	8	Casino RoyaleÃ,Â	English
121	8	Casino RoyaleÃ,Â	English
122	8	Life of PiÃ,Â	English
123	8	JawsÃ,Â	English
124	8	Blood DiamondÃ,Â	English
125	8	Rain ManÃ,Â	English
126	8	HerÃ,Â	English
127	8	The Perks of Being a WallflowerÃ,Â	English
128	8	Big FishÃ,Â	English
129	8	Mystic RiverÃ,Â	English
130	8	The Pursuit of HappynessÃ,Â	English
131	8	In BrugesÃ,Â	English
132	8	The ExorcistÃ,Â	English
133	8	Dead Poets SocietyÃ,Â	English
134	8	AladdinÃ,Â	English

135	8	SerenityÃ,Â	English
136	8	MagnoliaÃ,Â	English
137	8	Mulholland DriveÃ,Â	English
138	8	The ArtistÃ,Â	English
139	8	Dances with WolvesÃ,Â	English
140	8	True RomanceÃ,Â	English
141	8	BrazilÃ,Â	English
142	8	Cinderella ManÃ,Â	English
143	8	The Iron GiantÃ,Â	English
144	8	JFKÃ,Â	English
145	8	Dancer in the DarkÃ,Â	English
146	8	Doctor ZhivagoÃ,Â	English
147	7.9	AvatarÃ,Â	English
148	7.9	Iron ManÃ,Â	English
149	7.9	The Hobbit: An Unexpected JourneyÃ,Â	English
150	7.9	TakenÃ,Â	English
151	7.9	The Hobbit: The Desolation of SmaugÃ,Â	English
152	7.9	ShrekÃ,Â	English
153	7.9	Edge of TomorrowÃ,Â	English
154	7.9	The Bourne IdentityÃ,Â	English
155	7.9	The NotebookÃ,Â	English
156	7.9	Toy Story 2Ã,Â	English
157	7.9	Children of MenÃ,Â	English
158	7.9	Edward ScissorhandsÃ,Â	English
159	7.9	Hot FuzzÃ,Â	English
160	7.9	Captain PhillipsÃ,Â	English
161	7.9	E.T. the Extra-TerrestrialÃ,Â	English
162	7.9	Big Hero 6Ã,Â	English
163	7.9	The FighterÃ,Â	English
164	7.9	The Hateful EightÃ,Â	English
165	7.9	How to Train Your Dragon 2Ã,Â	English
166	7.9	The UntouchablesÃ,Â	English
167	7.9	Crouching Tiger, Hidden DragonÃ,Â	Mandarin
168	7.9	Almost FamousÃ,Â	English
169	7.9	Boogie NightsÃ,Â	English
170	7.9	Walk the LineÃ,Â	English
171	7.9	HalloweenÃ,Â	English
172	7.9	HeroÃ,Â	Mandarin
173	7.9	The Blues BrothersÃ,Â	English
174	7.9	Ed WoodÃ,Â	English
175	7.9	The InsiderÃ,Â	English
176	7.9	Letters from Iwo JimaÃ,Â	Japanese

177	7.9	Straight Outta ComptonÃ,Â	English
178	7.9	GloryÃ,Â	English
179	7.9	GloryÃ,Â	English
180	7.9	My Fair LadyÃ,Â	English
181	7.9	The Remains of the DayÃ,Â	English
182	7.9	The Right StuffÃ,Â	English
183	7.9	The World's Fastest IndianÃ,Â	English
184	7.8	The HangoverÃ,Â	English
185	7.8	GravityÃ,Â	English
186	7.8	Silver Linings PlaybookÃ,Â	English
187	7.8	SkyfallÃ,Â	English
188	7.8	X-Men: First ClassÃ,Â	English
189	7.8	Captain America: The Winter SoldierÃ,Â	English
190	7.8	The Curious Case of Benjamin ButtonÃ,Â	English
191	7.8	DriveÃ,Â	English
192	7.8	Ocean's ElevenÃ,Â	English
193	7.8	Star Trek Into DarknessÃ,Â	English
194	7.8	Birdman or (The Unexpected Virtue of Ignorance)Ã,Â	English
195	7.8	Harry Potter and the Prisoner of AzkabanÃ,Â	English
196	7.8	The Bourne SupremacyÃ,Â	English
197	7.8	Back to the Future Part IIÃ,Â	English
198	7.8	The Girl with the Dragon TattooÃ,Â	English
199	7.8	American GangsterÃ,Â	English
200	7.8	TangledÃ,Â	English
201	7.8	PredatorÃ,Â	English
202	7.8	Wreck-It RalphÃ,Â	English
203	7.8	Lucky Number SlevinÃ,Â	English
204	7.8	The GameÃ,Â	English
205	7.8	Being John MalkovichÃ,Â	English
206	7.8	The Fault in Our StarsÃ,Â	English
207	7.8	The Lego MovieÃ,Â	English
208	7.8	3:10 to YumaÃ,Â	English
209	7.8	Moonrise KingdomÃ,Â	English
210	7.8	ApocalyptoÃ,Â	Maya
211	7.8	Donnie BrascoÃ,Â	English
212	7.8	O Brother, Where Art Thou?Ã,Â	English
213	7.8	GattacaÃ,Â	English
214	7.8	The FugitiveÃ,Â	English
215	7.8	About TimeÃ,Â	English
216	7.8	Office SpaceÃ,Â	English
217	7.8	ChangelingÃ,Â	English
218	7.8	Pride & PrejudiceÃ,Â	English

219	7.8	AtonementÃ,Â	English
220	7.8	The Big ShortÃ,Â	English
221	7.8	Finding NeverlandÃ,Â	English
222	7.8	What's Eating Gilbert GrapeÃ,Â	English
223	7.8	South Park: Bigger Longer & UncutÃ,Â	English
224	7.8	Remember the TitansÃ,Â	English
225	7.8	Fantastic Mr. FoxÃ,Â	English
226	7.8	The Boy in the Striped PajamasÃ,Â	English
227	7.8	The Last of the MohicansÃ,Â	English
228	7.8	The Jungle BookÃ,Â	English
229	7.8	The Jungle BookÃ,Â	English
230	7.8	Kung Fu HustleÃ,Â	Cantonese
231	7.8	TombstoneÃ,Â	English
232	7.8	NebraskaÃ,Â	English
233	7.8	Glengarry Glen RossÃ,Â	English
234	7.8	The Last EmperorÃ,Â	English
235	7.8	The Best OfferÃ,Â	English
236	7.8	The Conjuring 2Ã,Â	English
237	7.8	The Color PurpleÃ,Â	English
238	7.8	Black BookÃ,Â	Dutch
239	7.8	The White RibbonÃ,Â	German
240	7.8	HamletÃ,Â	English
241	7.8	The Little PrinceÃ,Â	English
242	7.8	The VerdictÃ,Â	English
243	7.7	TitanicÃ,Â	English
244	7.7	300Ã,Â	English
245	7.7	The Social NetworkÃ,Â	English
246	7.7	ArgoÃ,Â	English
247	7.7	Kick-AssÃ,Â	English
248	7.7	Minority ReportÃ,Â	English
249	7.7	Cast AwayÃ,Â	English
250	7.7	WatchmenÃ,Â	English

From the above query , I got this **IMDb_Top_250** , I had attached as proof of this.

This is my attached excel file:

https://docs.google.com/spreadsheets/d/16NnCBQPpx46wDCiuNatK76Pn8Ra47QAzb/edit?usp=drive_link

b)For the next part as , out of these top250 movies with a highest user votes , we search for all the **Top_Foreign_Lang_Film** using the below query :

```
SELECT row_number() over(order by imdb_score DESC,
num_voted_users DESC) as ranking,
      imdb_score,
      num_voted_users,
      movie_title AS Top_non_english,
      language
FROM
      moviesp.cleanedm
WHERE
      num_voted_users > 30000 and language <> 'English'
LIMIT 36;
```

ranking	imdb_score	num_voted_users	Top_non_english	language
1	8.6	417971	Spirited Away	Japanese
2	8.4	534262	Amélie	French
3	8.4	168203	Das Boot	German
4	8.4	62756	Baahubali: The Beginning	Telugu
5	8.3	248354	Downfall	German
6	8.3	111841	Metropolis	German
7	8.2	467234	Pan's Labyrinth	Spanish
8	8.2	214091	Howl's Moving Castle	Japanese
9	8.1	64556	The Sea Inside	Spanish

10	8.1	31943	Tae Guk Gi: The Brotherhood of WarÃ,Â	Korean
11	7.9	217740	Crouching Tiger, Hidden DragonÃ,Â	Mandarin
12	7.9	149414	HeroÃ,Â	Mandarin
13	7.9	132149	Letters from Iwo JimaÃ,Â	Japanese
14	7.8	236000	ApocalyptoÃ,Â	Maya
15	7.8	99353	Kung Fu HustleÃ,Â	Cantonese
16	7.8	59507	Black BookÃ,Â	Dutch
17	7.8	52958	The White RibbonÃ,Â	German
18	7.7	85589	PonyoÃ,Â	Japanese
19	7.7	62607	A Very Long EngagementÃ,Â	French
20	7.7	55516	The Great BeautyÃ,Â	Italian
21	7.6	92295	House of Flying DaggersÃ,Â	Mandarin
22	7.6	68119	The Kite RunnerÃ,Â	Dari
23	7.6	38690	The Flowers of WarÃ,Â	Mandarin
24	7.5	77656	The Painted VeilÃ,Â	Mandarin
25	7.5	41496	2046Ã,Â	Cantonese
26	7.4	36894	Red CliffÃ,Â	Mandarin
27	7.3	63084	Paris, je t'aimeÃ,Â	French
28	7.3	37635	Mongol: The Rise of Genghis KhanÃ,Â	Mongolia n
29	7.2	55928	District B13Ã,Â	French
30	7.1	179235	The Passion of the ChristÃ,Â	Aramaic
31	7	36455	Curse of the Golden FlowerÃ,Â	Mandarin
32	6.7	32003	Coco Before ChanelÃ,Â	French
33	6.4	86152	The InterpreterÃ,Â	Aboriginal
34	5.9	71574	The Legend of ZorroÃ,Â	Spanish
35	4.3	31414	In the Land of Blood and HoneyÃ,Â	Bosnian

- ❖ From the above query , I have found **35 movies list** with respect to the **num_voted_users** for non English movies with the top 250 lists.
- ❖ Then with this results , have been queried with respect to the number of voters for the most favorable movies with ratings of imdb_score, that is non_english that is foreign language movies that have been listed using the query. So in this question Top

250 and Top_Foreign_Lang_Film(35 movies) , MySQL workbench had been used to get the desired results and with the results , I had transferred the results to excel to make table to be readable.

❖ This is my file attached here:

https://drive.google.com/file/d/1sDk5gBzVoLVdENnGPDrv77P4AksNeHbW/view?usp=drive_link

4. **Best Directors:** Group the column using the director_name column.

Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.

Your task: Find the best directors

For this question ,want to find out the top 10 directors whom with the imdb_score of highest one . This will be found out by using the below query:

```
SELECT
```

```
director_name AS Top_10_Director,
```

```
ROUND(AVG(imdb_score), 2) AS avg_rating
```

FROM

moviesp.cleanedm

GROUP BY director_name

ORDER BY avg_rating DESC , director_name

LIMIT 10;

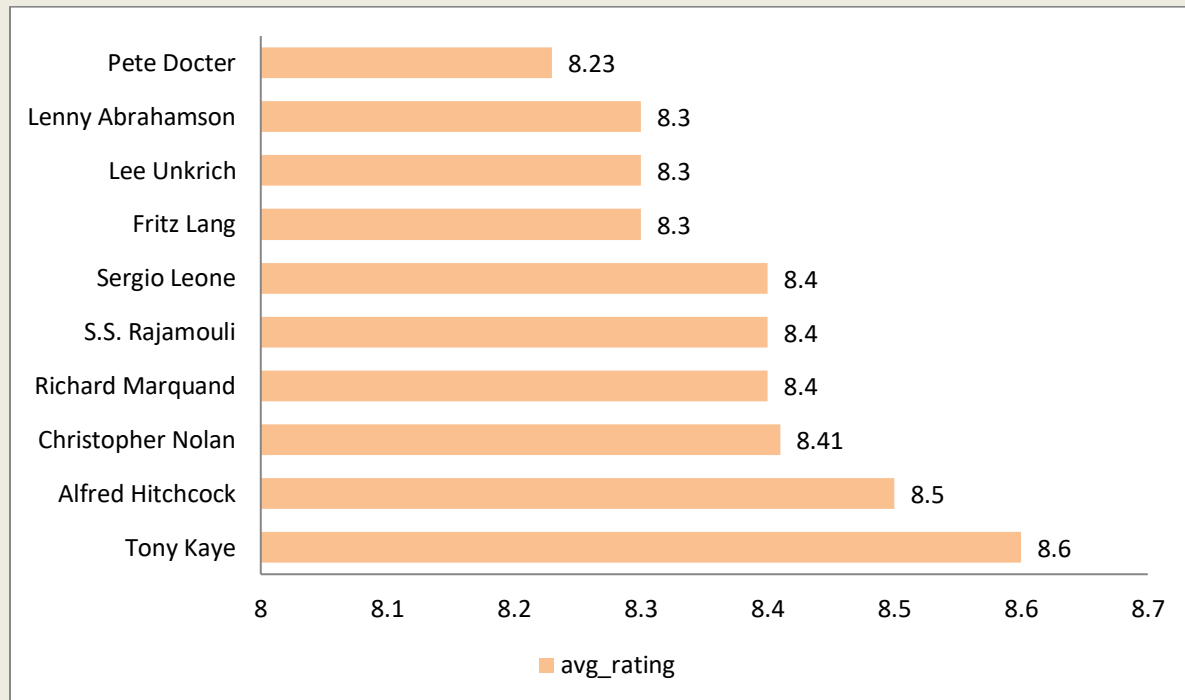
With this query , I have got the results.....

Top 10 Directors	
Top_10_Director	avg_rating
Tony Kaye	8.6
Alfred Hitchcock	8.5
Christopher Nolan	8.41
Richard Marquand	8.4
S.S. Rajamouli	8.4
Sergio Leone	8.4
Fritz Lang	8.3
Lee Unkrich	8.3
Lenny Abrahamson	8.3
Pete Docter	8.23

From the above results , I have inferred that the top 10 directors have been listed by using the sql query results with the below insights.

This is my excel file attached:

https://docs.google.com/spreadsheets/d/1kDmzbNlehAevghQAK8UPAuYi1rA3jlXj/edit?usp=drive_link



5. **Popular Genres:** Perform this step using the knowledge gained while performing previous steps.

Your task: Find popular genres

- ❖ For this question, we first split the genres of each movie into individual genres into 8 separate levels of genres. Then, we make a list of all genres by uniting all levels of genres and then count the popularity of each genre.

- ❖ I have been using sql query to solve this description of separating 8 levels of genres with list of all genres the with mean of gross and the number of movies.

With d8 as

```
(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_1 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_2 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_3 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_4 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_5 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_6 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_7 THEN NULL
ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) END) AS
genre_8
FROM
(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_1 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_2 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_3 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_4 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_5 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_6 THEN NULL
ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) END) AS
genre_7
FROM
(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_1 THEN NULL
```



```

WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_2
THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_3
THEN NULL
    WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_4
THEN NULL
      WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_5
THEN NULL
        ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) END) AS
genre_6
FROM
  (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|',
-1) = genre_1 THEN NULL
    WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_2
THEN NULL
      WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_3
THEN NULL
        WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_4
THEN NULL
          ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) END) AS
genre_5
FROM(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|',
4), '|', -1) = genre_1 THEN NULL
  WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) = genre_2
THEN NULL
    WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) = genre_3
THEN NULL
      ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) END) AS
genre_4
FROM
  (SELECT *,(CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|',
-1) = genre_1 THEN NULL
    WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|', -1) = genre_2
THEN NULL
      ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|', -1) END) AS
genre_3
FROM
  (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 2), '|',
-1) = genre_1 THEN NULL

```

```

ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 2), '|', -1) END) AS
genre_2
FROM
(SELECT
SUBSTRING(genres, 2) as genres,
SUBSTRING_INDEX(SUBSTRING_INDEX(SUBSTRING(genres, 2), '|', 1), '|', - 1) AS
genre_1
FROM moviesp.cleanedm) d1) d2) d3) d4) d5) d6) d7),
list as
(SELECT DISTINCT genre_1 AS genre_list
FROM
(
(SELECT genre_1 FROM d8) UNION ALL
(SELECT genre_2 FROM d8) UNION ALL
(SELECT genre_3 FROM d8) UNION ALL
(SELECT genre_4 FROM d8) UNION ALL
(SELECT genre_5 FROM d8) UNION ALL
(SELECT genre_6 FROM d8) UNION ALL
(SELECT genre_7 FROM d8) UNION ALL
(SELECT genre_8 FROM d8)) g12
WHERE genre_1 IS NOT NULL ORDER BY genre_1)
SELECT genre_list,round(AVG(DISTINCT CASE WHEN INSTR(genres, genre_list)
THEN gross ELSE NULL END), 2) AS avg_grossing,
COUNT(DISTINCT CASE WHEN INSTR(genres, genre_list) THEN genres ELSE
NULL END) AS no_of_movies
FROM cleanedm JOIN list
GROUP BY genre_list
ORDER BY avg_grossing DESC;

SELECT actor_1_name, COUNT(movie_title) AS no_of_movies,
ROUND(AVG(num_user_for_reviews), 2) AS user_reviews,
ROUND(AVG(num_critic_for_reviews), 2) AS critic_reviews
FROM
(
(SELECT actor_1_name, movie_title,num_user_for_reviews, num_critic_for_reviews
FROM moviesp.cleanedm
WHERE actor_1_name = ' CCH Pounder')
UNION ALL
(SELECT actor_1_name, movie_title,num_user_for_reviews, num_critic_for_reviews
FROM moviesp.cleanedm

```

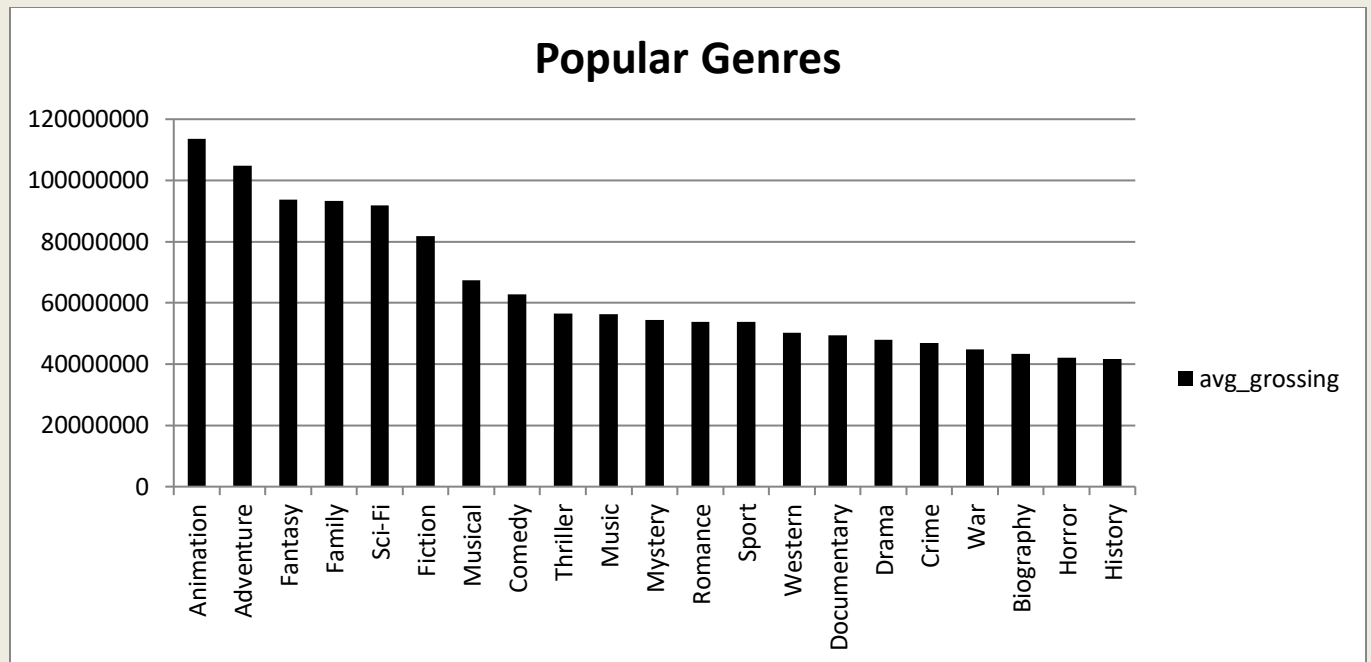
```

WHERE actor_1_name = ' J.K. Simmons')
UNION ALL
(SELECT actor_1_name, movie_title, num_user_for_reviews,
num_critic_for_reviews
FROM moviesp.cleanedm
WHERE actor_1_name = ' Jennifer Lawrence')) c
GROUP BY actor_1_name
ORDER BY user_reviews DESC , critic_reviews DESC;

```

Popular Genres		
genre_list	avg_grossing	no_of_movies
Animation	113547233.2	84
Adventure	104744835.7	238
Fantasy	93680359.4	170
Family	93247050.7	178
Sci-Fi	91891509.05	134
Fiction	81796457.68	235
Musical	67450846.54	53
Comedy	62876377.93	263
Thriller	56544554.87	179
Music	56385988.68	101
Mystery	54357927.28	97
Romance	53797743.97	171
Sport	53731204.96	39
Western	50326576.6	33
Documentary	49478940.43	4
Drama	47978352.15	310
Crime	46975720.56	121
War	44756370.57	60
Biography	43329712.5	63
Horror	42187374.12	70
History	41782106.2	51

From the above query results , There are **23 Genres** got to inferred that the splitted genres with the mean of gross and number of movies match lists. And got the analysis of insights of column chart are as follows:



This is my attached excel file:

https://docs.google.com/spreadsheets/d/13gHz7H3oDULFNXcZJIRAB2g6VOB-UgVZ/edit?usp=drive_link

- Charts:** Create three new columns namely, **Meryl Streep**, **Leo Caprio**, and **Brad Pitt** which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined.

Group the combined column using the **actor_1_name** column. Find the mean of the **num_critic_for_reviews** and **num_users_for_review** and identify the actors which have the highest mean.

Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called **df_by_decade**.

Your task: Find the critic-favorite and audience-favorite actors

a) For this critic-favourite , the assigned favourite movies has been extracted which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors.

❖ So I have queried for extracting the provided favourite actors with the number of movies , user reviews (audience-favourite) and critic-reviews. The query for finding this as follows:

```
SELECT actor_1_name, COUNT(movie_title) AS no_of_movies,
```

```

ROUND(AVG(num_user_for_reviews), 2) AS user_reviews,
ROUND(AVG(num_critic_for_reviews), 2) AS critic_reviews
FROM
(
(SELECT actor_1_name, movie_title,num_user_for_reviews,
num_critic_for_reviews
FROM moviesp.cleanedm
WHERE actor_1_name = 'Meryl Streep')
UNION ALL
(SELECT actor_1_name, movie_title,num_user_for_reviews,
num_critic_for_reviews
FROM moviesp.cleanedm
WHERE actor_1_name = 'Leonardo DiCaprio')
UNION ALL
(SELECT actor_1_name, movie_title, num_user_for_reviews,
num_critic_for_reviews
FROM moviesp.cleanedm
WHERE actor_1_name = 'Brad Pitt')) c
GROUP BY actor_1_name
ORDER BY user_reviews DESC , critic_reviews DESC;

```

Favourite actors

actor_1_name	num_of_movies	user_reviews	critic_reviews
--------------	---------------	--------------	----------------

Leonardo DiCaprio	20	922.55	322.2
Brad Pitt	17	742.35	245
Meryl Streep	11	297.18	181.45

- ❖ From the above query , I got this results which inferred of extracting from the actor name 1 with he help of user reviews and critic reviews.
- ❖ And this query helped me to find out that **Leonardo DiCaprio** is the **most favorite actor** for both the **audience and the critics**.
- ❖ This is the file attached here:

https://docs.google.com/spreadsheets/d/1046zg_F9vOST467Ewu6HOAwdWET-C-Bn/edit?usp=drive_link

- b) Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called **df_by_decade**.

- ❖ For this decades per number user votes , we will bw finding out extracting the years from MySQL query workbench are as follows:

```
SELECT
    CONCAT(CONVERT( FLOOR(title_year / 10) * 10 ,
    CHAR),
        's') AS decade,
```

```

SUM(num_voted_users) AS total_votes
FROM
    moviewsp.cleanedm
GROUP BY decade
ORDER BY decade;

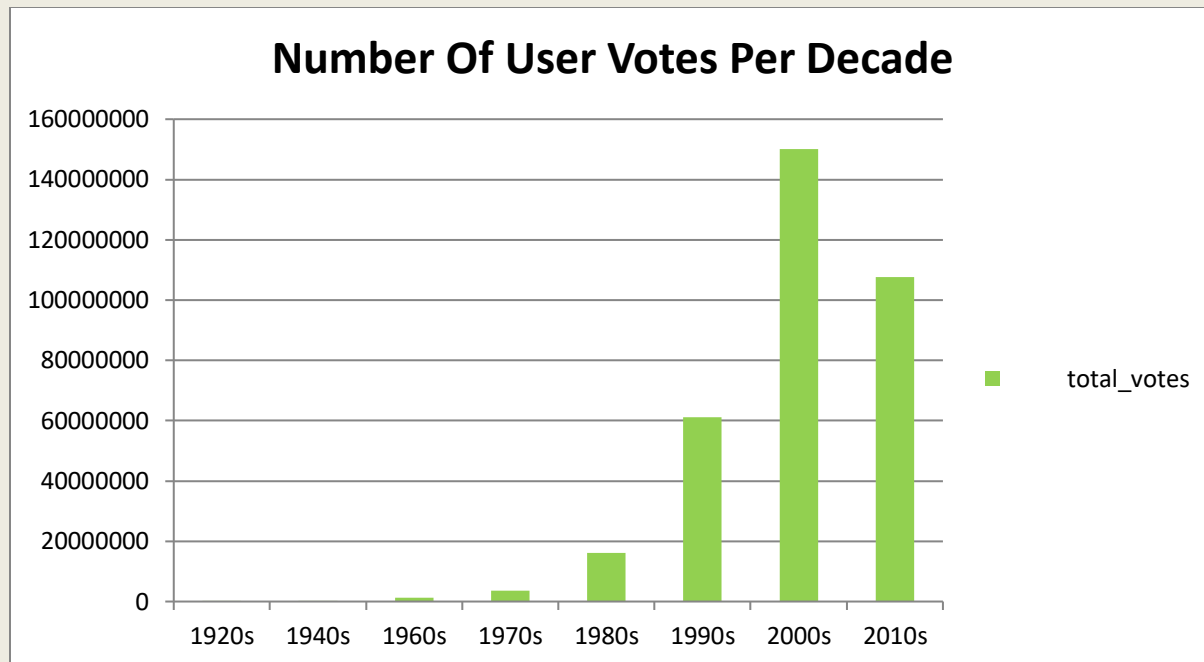
```

This is excel attached here:

https://docs.google.com/spreadsheets/d/1B3LzOFwL59h8d2jBRb6YJ73Y3tmHTAiR/edit?usp=drive_link

Number Of User Votes Per Decade	
decade	total_votes
1920s	111841
1940s	90360
1960s	1218680
1970s	3679621
1980s	16136813
1990s	61213572
2000s	150036196
2010s	107714560

From the above query , I have inferred that the results of decades per number of user votes with the voting that are extracted using the query using aggregation function .



From the above column chart , I have analyzed a insights of decades over the number of user votes.

RESULT

- So in this IMDB movie analysis project , I have analyzed with the useful insights that are making good impression with a anyone can readable format. In this project , MySQL workbench and MS Excel helped me throughout the one , which helps me getting and extracting the more and more useful strategy.
- Then I have learned how to remove outliers , data cleaning strategy , Advanced SQL query , exporting the results of sql query to the excel to make visual insights and so on.

DRIVE LINK (for whole excel files)

https://drive.google.com/drive/folders/1L1Lmf10XD2d6N2DZx08B4saOG2RrdUwi?usp=drive_link