IMDB Movie Analysis

Final Project-1

PROJECT DESCRIPTION

We can start by framing a problem with a series of questions such "What do you see happening?" and "What is the problem's impact on stakeholders?" This will help you in defining the issue and collecting the relevant information to address it. Once the issue has been identified, apply data analysis to investigate the information and provide conclusions. In addition, by continuously asking "Why?" and digging deeper into each response, apply the Five Whys Analysis to identify the problem's underlying causes.

To identify the root cause of a problem, apply the Five Whys technique, sometimes referred to as Root Cause Analysis. When employing this method, begin by asking "Why?" and keep doing so until you identify the fundamental source of the issue. To get to the root of the problem, each response should be carefully considered. You can discover that there are many responses to the same question as you inquire why.

APPROACH

First we have to open the Excel workbook and insert the dataset.

The dataset that is given contains information on applicants for a certain position in one of this organisation's departments. You must apply your statistical expertise, make use of various Excel formulas, and derive the appropriate business-related conclusions.

I have used all the rows to work on the problems after analysing these things having knowledge of data columns, examining missing data, combining columns that have several categories, looking for outliers, taking out null values, drawing data overview.

TECH-STACK USED

I have used Microsoft Excel to download the dataset and work on it and I have used One drive to save the file. I have Microsoft word to create the pdf file. I have used MySQL command line client to work on these problems too.

INSIGHTS

I have used required excel formulas and functions to solve the problems which I learnt from the learning session in Trainity Dashboard.

A. **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

These are required steps to clean the data

1. Remove any columns that are not required for the analysis: Remove any columns that are not necessary. For instance, if the analysis only needs the title, director, genre, year of release, and box office receipts, you can omit columns for the names of the actors in the film, the production company, etc.
2. Delete any rows that have null or missing values: Find and delete any rows that have null or missing values. These might be deceptive and compromise the analysis's accuracy.
3. Eliminate duplicates: Inspect the dataset for any duplicate rows, and delete them if necessary. The analysis's conclusions may be skewed by duplicate rows.
4. Ensure that data is consistent and formatted correctly by standardising its format. For instance, you can standardise the format if the release year column contains both two- and four-digit years.

	A	B	C	D
1	color	director_n	num_critic	duration
2	Color	James Can	723	1₹
3	Color	Gore Verbi	302	16
4	Color	Sam Mend	602	14
5	Color	Christophe	813	16
6		Doug Walker		
7	Color	Andrew St.	462	13
8	Color	Sam Raimi	392	15
9	Color	Nathan Gr	324	10
10	Color			14
11	Color			15
12	Color			18
13	Color			16
14	Color			10
15	Color			15
16	Color			15
17	Color			14
18	Color			15
19	Color	Joss Whed	703	1₹
20	Color	Rob Marsh	448	13
21	Color	Barry Sonr	451	10
22	Color	Peter Jack:	422	16
23	Color	Marc Web	599	15
24	Color	Ridley Sco	343	15
25	Color	Peter Jack:	509	18
26	Color	Chris Weit	251	11
27	Color	Peter Jack:	446	20
28	Color	James Can	315	1₹

Delete ? ✕

Delete

○ Shift cells left
○ Shift cells up
○ Entire row
● Entire column

OK   Cancel

In MySQL we can use this method

ALTER TABLE tab
DROP COLUMN color;

```
mysql> alter table tab
    -> drop column color;
Query OK, 0 rows affected (0.10 sec)
Records: 0  Duplicates: 0  Warnings: 0
```

DELETE FROM t1 WHERE director_name IS NULL OR
num_critic_for_reviews IS NULL OR duration IS NULL OR
director_facebook_likes IS NULL OR actor_3_facebook_likes IS NULL OR
actor_2_name IS NULL OR actor_1_facebook_likes IS NULL OR gross IS
NULL OR genres IS NULL OR actor_1_name IS NULL OR movie_title IS
NULL OR num_voted_users IS NULL OR cast_total_facebook_likes IS
NULL OR actor_3_name IS NULL OR facenumber_in_poster IS NULL OR
plot_keywords IS NULL OR movie_imdb_link IS NULL OR
num_user_for_reviews IS NULL OR language IS NULL OR country IS NULL
OR content_rating IS NULL OR budget IS NULL OR title_year IS NULL OR
actor_2_facebook_likes IS NULL OR imdb_score IS NULL OR aspect_ratio
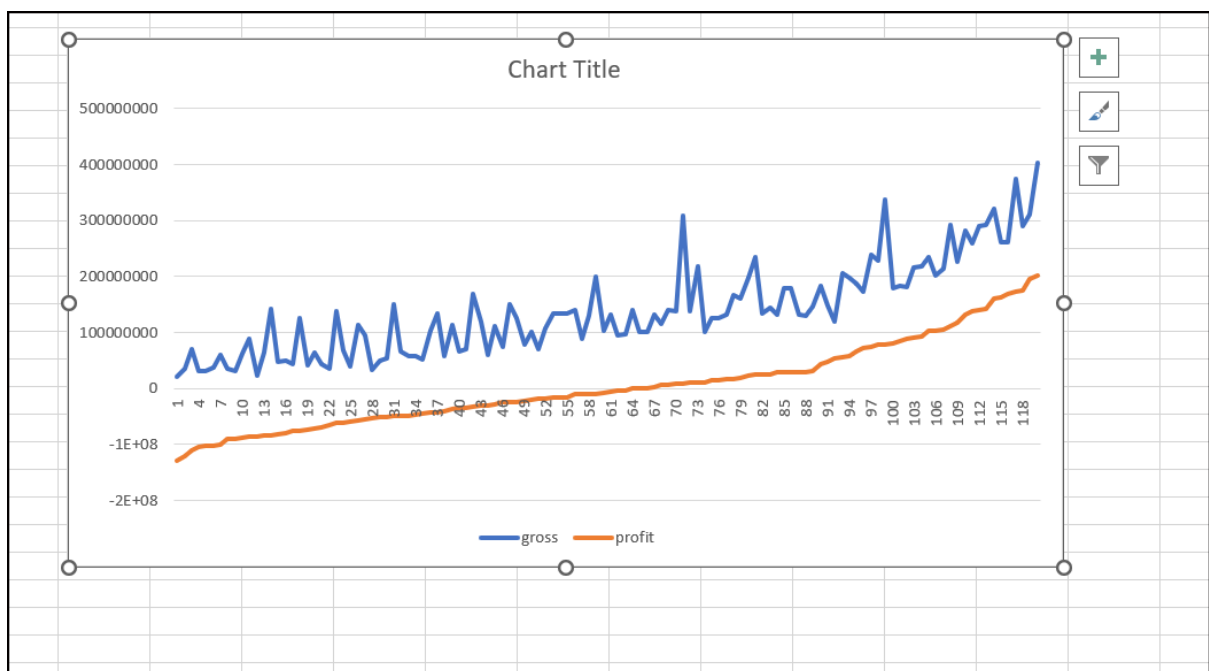IS NULL OR movie_facebook_likes is null;

```
mysql> delete from tab where duration is null or director_facebook_likes is null or actor_3_facebook_likes is null or actor_2_name is null or actor_1_facebook_likes is null or gross is null or genres is null or
actor_1_name is null or movie_title is null or num_voted_users is null or cast_total_facebook_likes is null or actor_3_name is null or facenumber_in_poster is null or plot_keywords is null or movie_imdb_link is
null or num_user_for_reviews is null or language is null or country is null or content_rating is null or budget is null or title_year is null or actor_2_facebook_likes is null or imdb_score is null or aspect_rat
io is null or movie_facebook_likes;
Query OK, 219 rows affected (0.09 sec)
```

B. **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?

We can find the movies with profit by profit column . These are the movies
with highest profit.

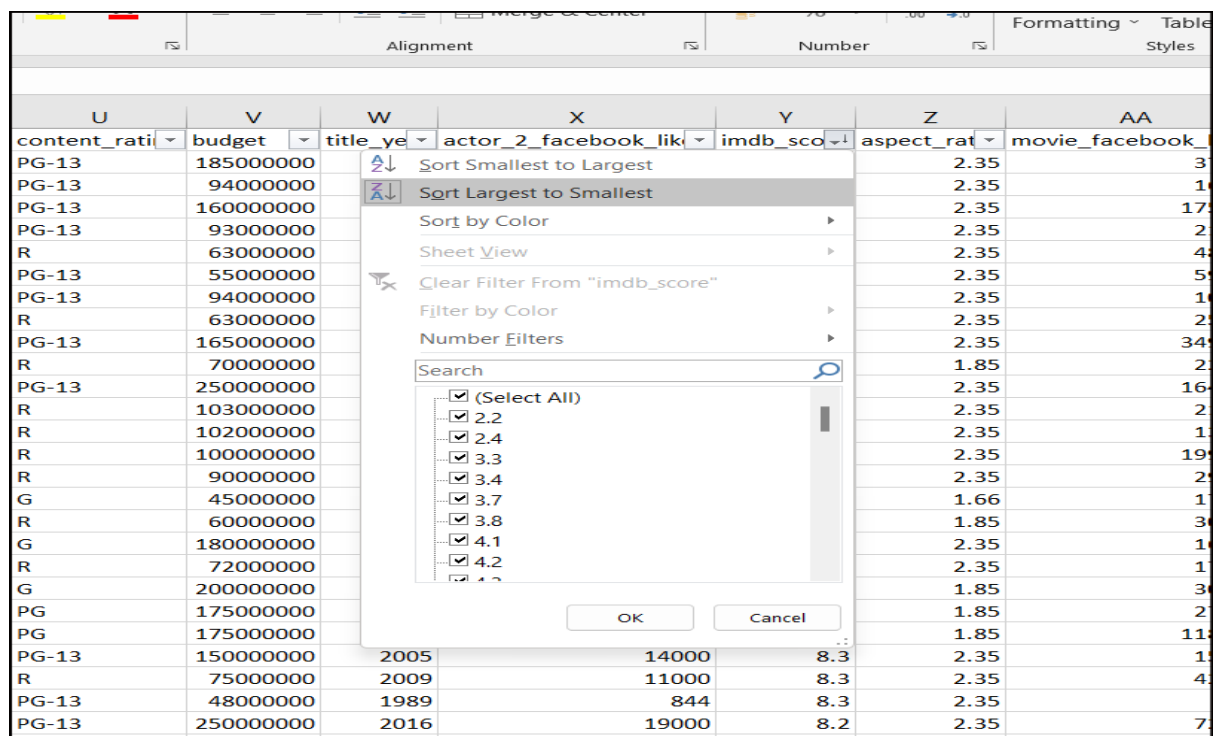| K movie_title	L num_vo	M cast_tot	N ac	O fac	P plot	Q movie	R num_us	S langu	T cou	U cont	V bu	W title	X actor	Y im	Z aspe	AA movie_facebook_lik	AB profit
AvatarÃ‚Â	886204	4834	Wes	0	avatar	http://ww	3054	English	USA	PG-13	####	2009	936	7.9	1.78	33000	523505847
Jurassic WorldÃ‚Â	418214	8458	Oma	0	dinosau	http://ww	1290	English	USA	PG-13	####	2015	2000	7	2	150000	502177271
TitanicÃ‚Â	793059	45223	Glori	0	artist|lc	http://ww	2528	English	USA	PG-13	####	1997	14000	7.7	2.35	26000	458672302
The AvengersÃ‚Â	995415	87697	Scarl	3	alien inv	http://ww	1722	English	USA	PG-13	####	2012	21000	8.1	1.85	123000	403279547
The AvengersÃ‚Â	995415	87697	Scarl	3	alien inv	http://ww	1722	English	USA	PG-13	####	2012	21000	8.1	1.85	123000	403279547
The Lion KingÃ‚Â	644348	6458	Niket	0	king|pri	http://ww	656	English	USA	G	####	1994	886	8.5	1.66	17000	377783777
Star Wars: Episode I - The Phantom MenaceÃ‚Â	534658	37723	Ian N	1	alien|ch	http://ww	3597	English	USA	PG	####	1999	14000	6.5	2.35	13000	359544677
The Dark KnightÃ‚Â	1676169	57802	Morg	0	based o	http://ww	4667	English	USA	PG-13	####	2008	13000	9	2.35	37000	348316061
The Hunger GamesÃ‚Â	701607	49942	Anth	0	fight to	http://ww	1959	English	USA	PG-13	####	2012	14000	7.3	2.35	140000	329999255
DeadpoolÃ‚Â	479047	18239	Stefa	0	based o	http://ww	1058	English	USA	R	####	2016	805	8.1	2.35	117000	305024263
The Hunger Games: Catching FireÃ‚Â	498397	49355	Sand	1	preside	http://ww	706	English	USA	PG-13	####	2013	14000	7.6	2.35	82000	294645577
Jurassic ParkÃ‚Â	613473	2129	Bob	0	dinosau	http://ww	895	English	USA	PG-13	####	1993	610	8.1	1.85	19000	293784000
Despicable Me 2Ã‚Â	286877	11905	Steve	0	dating|	http://ww	284	English	USA	PG	####	2013	2000	7.5	1.85	56000	292049635
American SniperÃ‚Â	325264	16277	Keir (	0	assassir	http://ww	916	English	USA	R	####	2014	962	7.3	2.35	112000	291323553
Finding NemoÃ‚Â	692482	5641	Brad	0	great ba	http://ww	866	English	USA	G	####	2003	939	8.2	1.85	11000	286838870
Shrek 2Ã‚Â	314630	1148	Conr	1	drag qu	http://ww	483	English	USA	PG	####	2004	309	7.2	1.85	0	286471036
The Lord of the Rings: The Return of the KingÃ‚Â	1215718	6434	Bern	2	battle|c	http://ww	3189	English	USA	PG-13	####	2003	857	8.9	2.35	16000	283019252
Forrest GumpÃ‚Â	1251222	15700	Sam	0	ampute	http://ww	1398	English	USA	PG-13	####	1994	294	8.8	2.35	59000	274691196
Star Wars: Episode III - Revenge of the SithÃ‚Â	520104	44042	Hayd	4	elongat	http://ww	3286	English	USA	PG-13	####	2005	16000	7.6	2.35	10000	267262555
Spider-ManÃ‚Â	544665	40484	Kirst	0	evil|gol	http://ww	2012	English	USA	PG-13	####	2002	11000	7.3	1.85	5000	264706375
MinionsÃ‚Â	142403	2582	Steve	0	antarcti	http://ww	275	English	USA	PG	####	2015	4000	6.4	1.85	70000	262029560
FrozenÃ‚Â	421658	2582	Livvy	0	magic|s	http://ww	904	English	USA	PG	####	2013	523	7.6	2.24	58000	250736600
The Secret Life of PetsÃ‚Â	24407	4782	Alber	0	animal		http://ww	155	English	Japan	PG	####	2016	904	6.8	1.85	36000	248505540
The Twilight Saga: New MoonÃ‚Â	220758	64040	Taylo	2	forks w	http://ww	919	English	USA	PG-13	####	2009	17000	4.6	2.35	13000	246623634
The Lord of the Rings: The Two TowersÃ‚Â	1100446	23052	Billy	1	epic|ev	http://ww	2417	English	USA	PG-13	####	2002	5000	8.7	2.35	10000	246478898
The Twilight Saga: EclipseÃ‚Â	184637	52547	Anna	3	friendsh	http://ww	498	English	USA	PG-13	####	2010	17000	4.9	2.35	29000	232523113
Independence DayÃ‚Â	430055	16785	Mary	0	alien|ar	http://ww	856	English	USA	PG-13	####	1996	2000	6.9	2.35	16000	231124059

This is the plot of profit (y-axis) vs budget (x- axis)

Chart Title

C. **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

Step 1: Sort the imdb_score from largest to smallest

Step 2: Custom filter of num_voted_users column to be above

K	L	M	N	O	P
	num_voted_use ▼	cast_total_facebook_lik ▼	actor_3_name ▼	facenumber_in_post ▼	plot_keywords
	1676169	57802 Morgan Freeman		0 based on comic book|dc comics|psychopath|star died before release|urban setti
gs: The Return of the KingÃ‚Â	1215718	6434 Bernard Hill		2 battle|epic|king|orc|ring
	1468200	81115 Joseph Gordon-Levitt		0 ambiguous ending|corporate espionage|dream|subconscious|thief
gs: The Fellowship of the RingÃ‚Â	1238746			dle earth|quest|ring
	1347461			ent|dark humor|fighting|multiple personality disorder|surprise end
	1251222			vietnam|vietnam war|war hero
gs: The Two TowersÃ‚Â	1100446			l|middle earth|ring|wizard
	1217752			|computer|matrix|questioning reality|simulated reality
	928227			er daughter relationship|saving the world|space travel|wormhole
ÃÂ	881236			killed in action|normandy|soldier
esÃ‚Â	1144337			isonment|lawlessness|police officer|terrorist plot
	982637			od|combat|gladiator|roman empire
ment DayÃÂ	744891			etal|multiple cameos|sexy woman|time travel
ÃÂ	955174			rically inaccurate|ku klux klan|n word|slavery
	873649			olice|undercover|undercover cop
	644348	6458 Niketa Calame		0 king|prince|scar|uncle|unnecessary guilt
	782610	17716 Michael Jeter		0 1930s|death row inmate|healing|prison guard|southern u.s.
	718837	2975 Jeff Garlin		0 earth|obesity|plant|robot|soil
	736638	1976 James Robinson		1 14th century|legend|revolt|scotland|tyranny
	544884	19085 Don Rickles		3 college|day care|escape|teddy bear|toy
	665575	2635 Jess Harnell		1 balloon|house|promise|retirement|skeleton
	345198	2944 Phyllis Smith		1 anger|joy|memory|running away|sadness
	980946	59558 Morgan Freeman		0 armored car|gangster|gotham|mixed martial arts|social decay
lsÃ‚Â	885175	36741 Christoph Waltz		1 france|german|nazis|revenge|scalping
he Last CrusadeÃ‚Â	515306	12884 Alison Doody		5 castle|diary|holy grail|map|nazi
ivil WarÃ‚Â	272670	64798 Chris Evans		0 based on comic book|knife|marvel cinematic universe|returning character killed c
DragonÃ‚Â	485430	20453 Craig Ferguson		0 dragon|island|training|viking|village
treetÃ‚Â	780588	46057 Jon Favreau		3 based on true story|fellatio|female rear nudity|sex in an airplane|stockbroker

Custom Autofilter dialog:
Show rows where:
num_voted_users
is greater than | 25000
● And ○ Or

Use ? to represent any single character
Use * to represent any series of characters

OK   Cancel

Step 3: Observe the top 250 directors

	A	B	C	D	E	F	G	H	I
1	director_name ▼	num_critic_for_revie ▼	durati ▼	director_facebook_lik ▼	actor_3_facebook_lik ▼	actor_2_name ▼	actor_1_facebook_lik ▼	gross ▼	genres
2	Christopher Nolan	645	152	22000	11000 Heath Ledger		23000	533316061	Action|Crime|Drama|Thriller
3	Peter Jackson	328	192	0	416 Billy Boyd		5000	377019252	Action|Adventure|Drama|Fantasy
4	Christopher Nolan	642	148	22000	23000 Tom Hardy		29000	292568851	Action|Adventure|Sci-Fi|Thriller
5	Peter Jackson	297	171	0	857 Orlando Bloom		16000	313837577	Action|Adventure|Drama|Fantasy
6	David Fincher	315	151	21000	637 Meat Loaf		11000	37023395	Drama
7	Robert Zemeckis	149	142	0	194 Siobhan Fallon Hogan		15000	329691196	Comedy|Drama
8	Peter Jackson	294	172	0	857 Orlando Bloom		16000	340478898	Action|Adventure|Drama|Fantasy
9	Lana Wachowski	313	136	0	99 Marcus Chong		18000	171383253	Action|Sci-Fi
10	Christopher Nolan	712	169	22000	6000 Anne Hathaway		11000	187991439	Adventure|Drama|Sci-Fi
11	Steven Spielberg	219	169	14000	13000 Vin Diesel		15000	216119491	Action|Drama|War
12	Christopher Nolan	813	164	22000	23000 Christian Bale		27000	448130642	Action|Thriller
13	Ridley Scott	265	171	0	695 Connie Nielsen		3000	187670866	Action|Drama|Romance
14	James Cameron	210	153	0	539 Jenette Goldstein		780	204843350	Action|Sci-Fi
15	Quentin Tarantino	765	165	16000	265 Christoph Waltz		29000	162804648	Drama|Western
16	Martin Scorsese	352	151	17000	1000 Matt Damon		29000	132373442	Crime|Drama|Thriller
17	Roger Allers	186	73	28	847 Nathan Lane		2000	422783777	Adventure|Animation|Drama|Family|Musical
18	Frank Darabont	186	189	0	693 Jeffrey DeMunn		15000	136801374	Crime|Drama|Fantasy|Mystery
19	Andrew Stanton	421	98	475	522 Fred Willard		1000	223806889	Adventure|Animation|Family|Sci-Fi
20	Mel Gibson	132	178	0	403 Patrick McGoohan		906	75600000	Biography|Drama|History|War
21	Lee Unkrich	453	103	125	721 John Ratzenberger		15000	414984497	Adventure|Animation|Comedy|Family|Fantasy
22	Pete Docter	408	96	0	262 Delroy Lindo		1000	292979556	Adventure|Animation|Comedy|Family
23	Pete Docter	536	95	0	384 Mindy Kaling		1000	356454367	Adventure|Animation|Comedy|Drama|Family|Fantasy
24	Christopher Nolan	478	128	22000	11000 Liam Neeson		23000	205343774	Action|Adventure
25	Quentin Tarantino	486	153	16000	11000 Brad Pitt		13000	120523073	Adventure|Drama|War
26	Steven Spielberg	149	127	14000	440 Julian Glover		11000	197171806	Action|Adventure|Fantasy
27	Anthony Russo	516	147	94	11000 Scarlett Johansson		21000	407197282	Action|Adventure|Sci-Fi
28	Dean DeBlois	288	98	255	759 America Ferrera		18000	217387997	Adventure|Animation|Family|Fantasy
29	Martin Scorsese	606	240	17000	4000 Matthew McConaughey		29000	116866727	Biography|Comedy|Crime|Drama

Step 4: Select all languages except English which lies in top 250 rows after applying all the filters

Step 5: these are the movies which are in top 250 other than English language.

In MySQL I tried to create a new column and got this output using these queries.

1. ALTER TABLE tab

ADD COLUMN imdb_Top_250 int(100);

2. ALTER TABLE tab

ADD COLUMN movie_rank int(100);

3. UPDATE tab SET IMDb_Top_250 = 1
   WHERE num_voted_users > 25000
   order by imdb_score;

4. SET @row_number = 0;
   UPDATE tab SET movie_rank = (@row_number:=@row_number+1)
    WHERE IMDb_Top_250 = 1;

5. SELECT IMDb_Top_250 FROM tab;

6. SELECT movie_rank FROM tab;

```
ERROR 1064 (42000): You have an error in your SQL syntax; check the
mysql> select imdb_top_250 from tab limit 250;
+--------------+
| imdb_top_250 |
+--------------+
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
|            1 |
```

```
mysql> select movie_rank from tab;
+------------+
| movie_rank |
+------------+
|          1 |
|          2 |
|          3 |
|          4 |
|          5 |
|          6 |
|          7 |
|          8 |
|          9 |
|         10 |
|         11 |
|         12 |
|         13 |
|         14 |
|         15 |
|         16 |
|         17 |
|         18 |
|         19 |
|         20 |
|         21 |
|         22 |
|         23 |
|         24 |
|         25 |
|         26 |
|         27 |
|         28 |
|         29 |
|         30 |
|         31 |
|         32 |
|         33 |
|         34 |
|         35 |
|         36 |
|         37 |
|         38 |
|         39 |
|         40 |
```

D. **Best Directors:** TGroup 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

In excel I have used filter to sort the table using imdb_score in descending order which gives the highest imdb_score on the top along with the director name. These are the top directors name.

	director_name	num_critic_for_review	d
1	director_name	num_critic_for_review	d
2	Christopher Nolan	645	
3	Peter Jackson	328	
4	Christopher Nolan	642	
5	David Fincher	315	
6	Peter Jackson	297	
7	Robert Zemeckis	149	
8	Lana Wachowski	313	
9	Peter Jackson	294	
10	Christopher Nolan	712	
11	Steven Spielberg	219	
12	Christopher Nolan	813	
13	Frank Darabont	186	
14	James Cameron	210	
15	Martin Scorsese	352	
16	Quentin Tarantino	765	
17	Ridley Scott	265	
18	Roger Allers	186	
19	Andrew Stanton	421	
20	Mel Gibson	132	
21	Christopher Nolan	478	
22	Lee Unkrich	453	
23	Pete Docter	408	
24	Pete Docter	536	
25	Quentin Tarantino	486	
26	Steven Spielberg	149	

These are the top directors name in MySQL using this command.

SELECT director_name, imdb_score FROM t1 ORDER BY imdb_score DESC LIMIT 10;

```
10 rows in set (0.00 sec)

mysql> select director_name, imdb_score from t1 order by imdb_score desc limit 10;
+-------------------+------------+
| director_name     | imdb_score |
+-------------------+------------+
| Christopher Nolan |          9 |
| Peter Jackson     |        8.9 |
| Christopher Nolan |        8.8 |
| Peter Jackson     |        8.8 |
| David Fincher     |        8.8 |
| Robert Zemeckis   |        8.8 |
| Peter Jackson     |        8.7 |
| Lana Wachowski    |        8.7 |
| Steven Spielberg  |        8.6 |
| Christopher Nolan |        8.6 |
+-------------------+------------+
10 rows in set (0.00 sec)
```

E. **Popular Genres:** Perform this step using the knowledge gained while performing previous steps.
Your task: Find popular genres

In excel I have used filter to sort the table using imdb_score in descending order which gives the highest imdb_score on the top along with the genres. These are the top genres.

	G	H	I	J
	actor_1_facebook_lik...	gross	genres	actor_1_name
	23000	533316061	Action|Crime|Drama|Thriller	Christian Bale
	5000	377019252	Action|Adventure|Drama|Fantasy	Orlando Bloom
	29000	292568851	Action|Adventure|Sci-Fi|Thriller	Leonardo DiCaprio
	11000	37023395	Drama	Brad Pitt
	16000	313837577	Action|Adventure|Drama|Fantasy	Christopher Lee
	15000	329691196	Comedy|Drama	Tom Hanks
	18000	171383253	Action|Sci-Fi	Keanu Reeves
	16000	340478898	Action|Adventure|Drama|Fantasy	Christopher Lee
	11000	187991439	Adventure|Drama|Sci-Fi	Matthew McConaughey
	15000	216119491	Action|Drama|War	Tom Hanks
	27000	448130642	Action|Thriller	Tom Hardy
	15000	136801374	Crime|Drama|Fantasy|Mystery	Tom Hanks
	780	204843350	Action|Sci-Fi	Joe Morton
	29000	132373442	Crime|Drama|Thriller	Leonardo DiCaprio
	29000	162804648	Drama|Western	Leonardo DiCaprio
	3000	187670866	Action|Drama|Romance	Djimon Hounsou
	2000	422783777	Adventure|Animation|Drama|Family|Musical	Matthew Broderick
	1000	223806889	Adventure|Animation|Family|Sci-Fi	John Ratzenberger
	906	75600000	Biography|Drama|History|War	Mhairi Calvey
	23000	205343774	Action|Adventure	Christian Bale
	15000	414984497	Adventure|Animation|Comedy|Family|Fantasy	Tom Hanks
	1000	292979556	Adventure|Animation|Comedy|Family	John Ratzenberger
	1000	356454367	Adventure|Animation|Comedy|Drama|Family|Fantasy	Amy Poehler
	13000	120523073	Adventure|Drama|War	Michael Fassbender
	11000	197171806	Action|Adventure|Fantasy	Harrison Ford
	1000	380838870	Adventure|Animation|Comedy|Family	Alexander Gould
	21000	407197282	Action|Adventure|Sci-Fi	Robert Downey Jr.
	18000	217387997	Adventure|Animation|Family|Fantasy	Gerard Butler

These are the top genres in MySQL using this command.

SELECT director_name, genres, imdb_score FROM t1 ORDER BY imdb_score DESC LIMIT 10;

```
mysql> select director_name,genres, imdb_score from t1 order by imdb_score desc limit 10;
+------------------+------------------------------------+------------+
| director_name    | genres                             | imdb_score |
+------------------+------------------------------------+------------+
| Christopher Nolan | Action|Crime|Drama|Thriller       |          9 |
| Peter Jackson    | Action|Adventure|Drama|Fantasy     |        8.9 |
| Christopher Nolan | Action|Adventure|Sci-Fi|Thriller  |        8.8 |
| Peter Jackson    | Action|Adventure|Drama|Fantasy     |        8.8 |
| David Fincher    | Drama                              |        8.8 |
| Robert Zemeckis  | Comedy|Drama                       |        8.8 |
| Peter Jackson    | Action|Adventure|Drama|Fantasy     |        8.7 |
| Lana Wachowski   | Action|Sci-Fi                      |        8.7 |
| Steven Spielberg | Action|Drama|War                   |        8.6 |
| Christopher Nolan | Adventure|Drama|Sci-Fi            |        8.6 |
+------------------+------------------------------------+------------+
10 rows in set (0.00 sec)
```

F. **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

These are the movies in Meryl Streep is the lead actor.

I	J	K	L
	actor_1_name	movie_title	num_voted_us
	Meryl Streep	It's ComplicatedÃ‚Â	69

These are the movies in Leonardo DiCaprio is the lead actor.

actor_1_name	movie_title
Leonardo DiCaprio	InceptionÃ,Â
Leonardo DiCaprio	The DepartedÃ,Â
Leonardo DiCaprio	Django UnchainedÃ,Â
Leonardo DiCaprio	The Wolf of Wall StreetÃ,Â
Leonardo DiCaprio	The RevenantÃ,Â
Leonardo DiCaprio	Shutter IslandÃ,Â
Leonardo DiCaprio	Blood DiamondÃ,Â
Leonardo DiCaprio	Catch Me If You CanÃ,Â
Leonardo DiCaprio	TitanicÃ,Â
Leonardo DiCaprio	The AviatorÃ,Â
Leonardo DiCaprio	Gangs of New YorkÃ,Â
Leonardo DiCaprio	The Great GatsbyÃ,Â
Leonardo DiCaprio	Body of LiesÃ,Â
Leonardo DiCaprio	The BeachÃ,Â

These are the movies in Brad Pitt is the lead actor.

actor_1_name	movie_title
Brad Pitt	Fight ClubÃ,Â
Brad Pitt	The Curious Case of Benjamin ButtonÃ,Â
Brad Pitt	Ocean's ElevenÃ,Â
Brad Pitt	FuryÃ,Â
Brad Pitt	Interview with the Vampire: The Vampire ChroniclesÃ,Â
Brad Pitt	TroyÃ,Â
Brad Pitt	Seven Years in TibetÃ,Â
Brad Pitt	Spy GameÃ,Â
Brad Pitt	Sinbad: Legend of the Seven SeasÃ,Â
Brad Pitt	Mr. & Mrs. SmithÃ,Â
Brad Pitt	Ocean's TwelveÃ,Â

These are the combined movies in which Meryl Streep, Leonardo DiCaprio and Brad Pitt are the lead actors in the movies

actor_1_name	movie_title
Meryl Streep	It's ComplicatedÃ‚Â
Leonardo DiCaprio	InceptionÃ‚Â
Leonardo DiCaprio	The DepartedÃ‚Â
Leonardo DiCaprio	Django UnchainedÃ‚Â
Leonardo DiCaprio	The Wolf of Wall StreetÃ‚Â
Leonardo DiCaprio	The RevenantÃ‚Â
Leonardo DiCaprio	Shutter IslandÃ‚Â
Leonardo DiCaprio	Blood DiamondÃ‚Â
Leonardo DiCaprio	Catch Me If You CanÃ‚Â
Leonardo DiCaprio	TitanicÃ‚Â
Leonardo DiCaprio	The AviatorÃ‚Â
Leonardo DiCaprio	Gangs of New YorkÃ‚Â
Leonardo DiCaprio	The Great GatsbyÃ‚Â
Leonardo DiCaprio	Body of LiesÃ‚Â
Leonardo DiCaprio	The BeachÃ‚Â
Brad Pitt	Fight ClubÃ‚Â
Brad Pitt	The Curious Case of Benjamin ButtonÃ‚Â
Brad Pitt	Ocean's ElevenÃ‚Â
Brad Pitt	FuryÃ‚Â
Brad Pitt	Interview with the Vampire: The Vampire ChroniclesÃ‚Â
Brad Pitt	TroyÃ‚Â
Brad Pitt	Seven Years in TibetÃ‚Â
Brad Pitt	Spy GameÃ‚Â
Brad Pitt	Sinbad: Legend of the Seven SeasÃ‚Â
Brad Pitt	Mr. & Mrs. SmithÃ‚Â
Brad Pitt	Ocean's TwelveÃ‚Â

After applying filter for num_critic_for_reviews column in descending order.

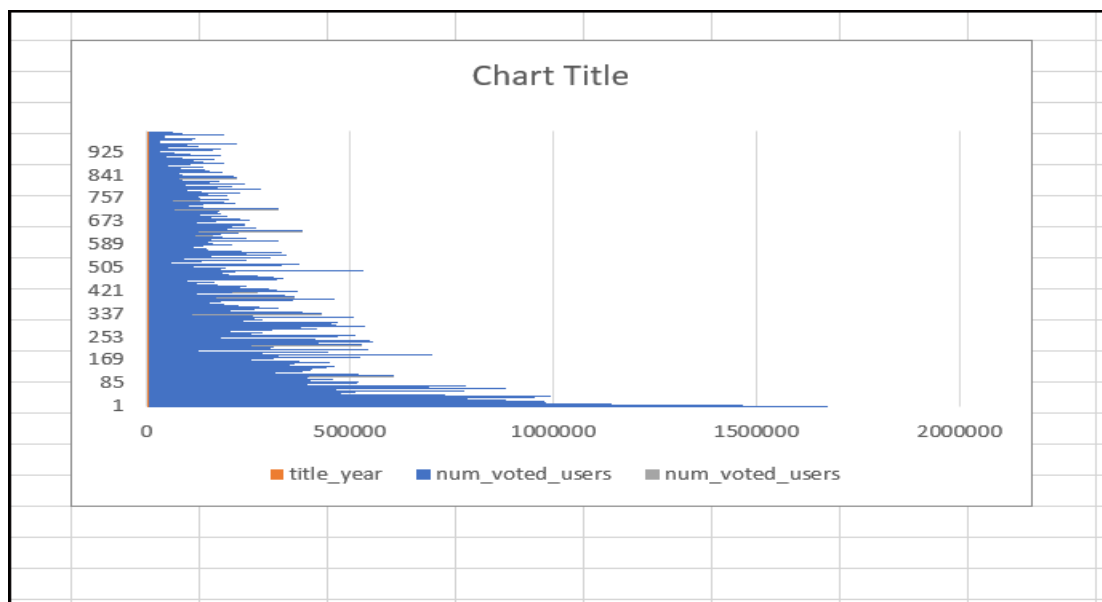These are the actors in lead who are critic-favorite actors. In top Leonardo DiCaprio is there.

num_critic_for_review	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	genres	actor_1_name
765	165	16000	265	Christoph Waltz	29000	162804648	Drama|Western	Leonardo DiCaprio
642	148	22000	23000	Tom Hardy	29000	292568851	Action|Adventure|Sci-Fi|Thriller	Leonardo DiCaprio
606	240	17000	4000	Matthew McConaughey	29000	116866727	Biography|Comedy|Crime|Drama	Leonardo DiCaprio
556	156	0	733	Tom Hardy	29000	183635922	Adventure|Drama|Thriller|Western	Leonardo DiCaprio
490	138	17000	163	Joseph Sikora	29000	127968405	Mystery|Thriller	Leonardo DiCaprio
490	143	1000	77	Elizabeth Debicki	29000	144812796	Drama|Romance	Leonardo DiCaprio
406	134	452	697	Logan Lerman	11000	85707116	Action|Drama|War	Brad Pitt
362	166	21000	919	Jason Flemyng	11000	127490802	Drama|Fantasy|Romance	Brad Pitt
352	151	17000	1000	Matt Damon	29000	132373442	Crime|Drama|Thriller	Leonardo DiCaprio
315	151	21000	637	Meat Loaf	11000	37023395	Drama	Brad Pitt
315	194	0	794	Kate Winslet	29000	658672302	Drama|Romance	Leonardo DiCaprio
267	170	17000	827	Adam Scott	29000	102608827	Biography|Drama	Leonardo DiCaprio
238	128	0	135	Simon McBurney	29000	39380442	Action|Drama|Thriller	Leonardo DiCaprio
233	216	17000	1000	Liam Neeson	29000	77679638	Crime|Drama	Leonardo DiCaprio
233	126	218	322	Angelina Jolie Pitt	11000	186336103	Action|Comedy|Crime|Romance|Thriller	Brad Pitt
220	196	249	844	Orlando Bloom	11000	133228348	Adventure	Brad Pitt
198	125	0	350	Julia Roberts	11000	125531634	Crime|Thriller	Brad Pitt
194	141	14000	3000	Tom Hanks	29000	164435221	Biography|Crime|Drama	Leonardo DiCaprio
187	120	278	963	Hunter Parrish	11000	112703470	Comedy|Drama|Romance	Meryl Streep
186	116	0	471	Bernie Mac	11000	183405771	Crime|Thriller	Brad Pitt
166	143	380	452	Djimon Hounsou	29000	57366262	Adventure|Drama|Thriller	Leonardo DiCaprio
142	114	12000	307	Stephen Dillane	11000	26871	Action|Crime|Thriller	Brad Pitt
120	123	277	4000	Tom Cruise	11000	105264608	Drama|Fantasy|Horror	Brad Pitt
118	119	0	39	Virginie Ledoyen	29000	39778599	Adventure|Drama|Thriller	Leonardo DiCaprio
98	85	0	82	Adriano Giannini	11000	26288320	Adventure|Animation|Comedy|Drama|Family|Fantasy|Romance	Brad Pitt
76	136	218	400	Mako	11000	37901509	Adventure|Biography|Drama|History|War	Brad Pitt

After applying filter for num_users_for_review column in descending order.

These are the actors in lead who are audience-favorite actors. In top Brad Pitt is there.

actor_1_name	movie_title	num	cast_to	acto	facenur	plot_keywor	movie	num_user_for_review	la
Brad Pitt	Fight ClubÃ,Â	1E+06	13209	Eugenie	2	anti establishm	http://wv	2968	En
Leonardo DiCaprio	InceptionÃ,Â	1E+06	81115	Joseph	0	ambiguous endi	http://wv	2803	En
Leonardo DiCaprio	TitanicÃ,Â	793059	45223	Gloria S	0	artist|love|ship	http://wv	2528	En
Leonardo DiCaprio	The DepartedÃ,Â	873649	45648	Ray Wir	0	boston|mole|p	http://wv	2054	En
Brad Pitt	TroyÃ,Â	381672	17944	Julian G	2	greek|mytholog	http://wv	1694	En
Leonardo DiCaprio	Django UnchainedÃ,Â	955174	40978	Ato Essa	1	dynamite|histo	http://wv	1193	En
Leonardo DiCaprio	The RevenantÃ,Â	406020	57108	Lukas H	0	bear attack|cau	http://wv	1188	En
Leonardo DiCaprio	Gangs of New YorkÃ,Â	314033	47657	Jim Bro	3	butcher|civil wa	http://wv	1166	En
Leonardo DiCaprio	The Wolf of Wall StreetÃ,Â	780588	46057	Jon Fav	3	based on true s	http://wv	1138	En
Leonardo DiCaprio	Shutter IslandÃ,Â	786092	29585	Nellie S	0	female prisoner	http://wv	964	En
Brad Pitt	Ocean's ElevenÃ,Â	402645	13028	Elliott G	4	card dealer|cas	http://wv	845	En
Brad Pitt	The Curious Case of Benjamin ButtonÃ,Â	459346	13333	Julia Or	2	deformed baby	http://wv	822	En
Leonardo DiCaprio	The AviatorÃ,Â	264318	34582	Frances	0	1920s|aviation	http://wv	799	En
Brad Pitt	Mr. & Mrs. SmithÃ,Â	348861	22722	Stephan	2	assassin|marria	http://wv	798	En
Leonardo DiCaprio	The Great GatsbyÃ,Â	362912	29770	Steve B	4	ingratitude|ma	http://wv	753	En
Brad Pitt	FuryÃ,Â	303185	20411	Jim Parr	4	battle|battlefie	http://wv	701	En
Leonardo DiCaprio	Catch Me If You CanÃ,Â	525801	48153	Jennifer	0	attorney|cat ar	http://wv	667	En
Leonardo DiCaprio	Blood DiamondÃ,Â	400292	33160	Stepher	0	diamond|eleph	http://wv	657	En
Brad Pitt	Ocean's TwelveÃ,Â	284852	19359	Mini An	0	father daughter	http://wv	627	En
Leonardo DiCaprio	The BeachÃ,Â	176169	29461	Peter Yo	0	love|map|para	http://wv	548	En
Brad Pitt	Interview with the Vampire: The Vampire ChroniclesÃ,Â	239752	25697	Kirsten	0	betrayal|dange	http://wv	406	En
Brad Pitt	Spy GameÃ,Â	121259	12499	Catherii	2	china|cia|cold	http://wv	361	En
Leonardo DiCaprio	Body of LiesÃ,Â	174248	29824	Michae	1	cia|jordan|mid	http://wv	263	En
Meryl Streep	It's ComplicatedÃ,Â	69860	15226	Zoe Kaz	2	affair|divorce|	http://wv	214	En

This the bar chart used to observe the change in number of voted users over decades.

This is based on the column decade, which I have grouped it by decade and counted the users votes and found the sum of users voted in each decade.

decade	count of voted_users	sum of voted_users
1970s	1	126357
1980s	7	1016161
1990s	185	28704737
2000s	474	84247831
2010s	333	12087154

RESULT

This project helped me to learn the methods on how to solve the problems in Excel. I have learnt some new ways to solve the problems. It was fun to do this project, I totally enjoyed it.

THANK YOU

DONE BY

JEEVIKA K