

IMDB Movie Analytics

PROJECT DESCRIPTION

Here I am trying to build a system through which I will be able to track the highest profit movies or top movies. Through this IMDB movie analytics I will measure certain parameters which in turn will help us grow as a company and also most certainly help in understanding our flaws. By this way the MNCs get to know about the actors or directors whom most people like. These analytics are the foundation pillar of the success of any organisation. Trends such as- imdb rating, top directors, most popular genres, top foreign language movie etc. are important for a company to analyse before producing a movie.

I have been given a dataset of a company various columns of different IMDB Movies is given. Knowledge in statistics and different formulas in excel are used to draw necessary conclusions about the company.

APPROACH

I have tried to understand the dataset before trying to execute any of the requirements. I related each given data with what exactly I require to derive e.g. If I have been given the dataset and asked to find out the most relevant columns, so I segregated the significant columns from the total number of given columns by first listing all the columns then visualising the necessity of a column in deriving a conclusion and then at last deleting the non significant column such as Color column . So over in all my approach was quite simple I just kept on connecting the dots to build these graphs and charts.

TECH-STACK

I had used MS-Excel provide by Microsoft. I have used the office home and student version of 2019.

The reason for using it is that it has very user friendly interface and it is also hassle free with all the provided services such as creating visual illustrations, administering it, modifying it etc. I have particularly used it to create several required charts and graphs to perfectly understand the data then I have used multiple pivot tables to derive the outcome I required out of the given dataset.

RESULT

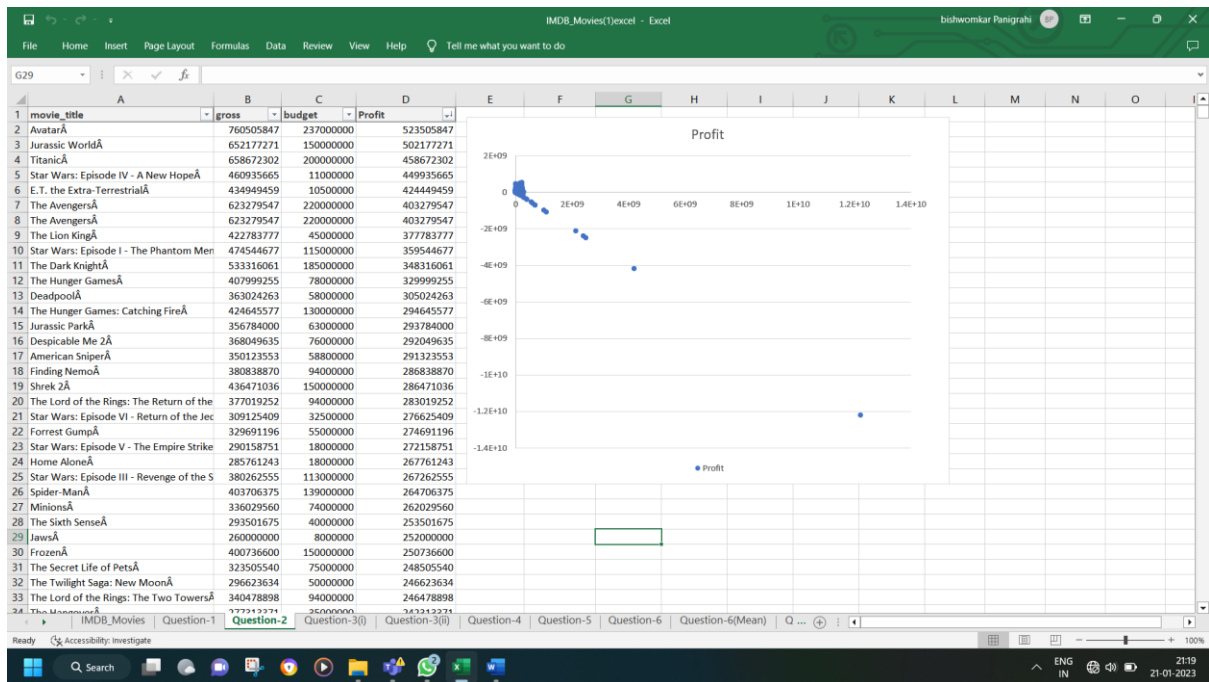
A.) Clean the data

	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_reviews	language	country	budget	title_year	imdb_score		
1	83	1127331	Drama	Marina Golbahari	Osama	7559	77	Dari	Afghanistan	46000	2003	7.4		
2	106	304124	Drama	MÃ-a Maestro	The Holy Girl	2720	37	Spanish	Argentina	1400000	2004	6.7		
3	114	1221261	Crime	Dram Ricardo DarÃ-n	Nine Queens	38215	125	Spanish	Argentina	1500000	2000	7.9		
4	129	20167424	Drama	Mys Ricardo DarÃ-n	The Secret in Their Eyes	131831	231	Spanish	Argentina	2000000	2009	8.2		
5	91	10076136	Action	Com Paul Sorvino	Knock Off	11512	141	English	Aruba	35000000	1998	4.8		
6	76	212285	Comedy	Roi Radha Mitchell	Love and Other Catastrophes	1727	21	English	Australia	250000	1996	6.4		
7	84	17411331	Horror	Mys Bruce Greenwood	Disturbing Behavior	17328	237	English	Australia	15000000	1998	5.5		
8	85	265107	Comedy	Ho Lauren Ambrose	Psycho Beach Party	4617	93	English	Australia	1500000	2000	6.3		
9	86	13751	Drama	Fam Jacqueline McKenzie	Opal Dream	813	6	English	Australia	11400000	2006	6.6		
10	87	9003011	Action	Adve Vernon Wells	Mad Max 2: The Road Warrior	133966	294	English	Australia	2000000	1981	7.6		
11	88	110029	Comedy	Dr Deborah Mailman	Bran Nue Dae	1388	31	English	Australia	6500000	2009	6.3		
12	90	882290	Action	Sci-F Delroy Lindo	The Blood of Heroes	4792	43	English	Australia	10000000	1989	6.5		
13	90	28399192	Action	Adve Steve Irwin	The Crocodile Hunter: Collision Course	5663	106	English	Australia	13000000	2002	5.4		
14	91	66600000	Comedy	Dr Miriam Margolyes	Babe	94435	182	English	Australia	30000000	1995	6.8		
15	92	25590119	Adventure	(Jere Burns	Crocodile Dundee in Los Angeles	19699	119	English	Australia	21150000	2001	4.8		
16	92	19059018	Action	Fant Caitlin Stasey	I, Frankenstein	65709	194	English	Australia	65000000	2014	5.1		
17	93	174635000	Adventure	(Paul Hogan	Crocodile Dundee	74743	93	English	Australia	88000000	1986	6.5		
18	94	6165429	Adventure	(Roy Billing	Rabbit-Proof Fence	23486	249	Aboriginal	Australia	60000000	2002	7.5		
19	95	6739141	Action	Crim Christopher Lambert	Fortress	23383	68	English	Australia	80000000	1992	5.9		
20	97	18318000	Adventure	(Adam Goldberg	Babe: Pig in the City	24868	216	English	Australia	80000000	1998	5.8		
21	98	10544143	Comedy	Roi Matthew Perry	Three to Tango	16372	121	English	Australia	20000000	1999	6.1		
22	98	29975979	Action	Hor Jay Laga'aia	Daybreakers	105797	281	English	Australia	20000000	2009	6.5		
23	100	17300889	Action	Adve Cary-Hiroyuki Tagawa	The Phantom	27648	148	English	Australia	45000000	1996	4.9		
24	100	63992328	Animation	(Robin Williams	Happy Feet 2	32399	79	English	Australia	135000000	2011	5.9		
25	100	257756197	Action	Adve Morgan Freeman	The Lego Movie	246698	471	English	Australia	60000000	2014	7.8		
26	104	1900725	Crime	Dram Ray Winstone	The Proposition	43205	232	English	Australia	20000000	2005	7.4		
27	105	35811509	Biography	(Noah Taylor	Shine	43013	104	English	Australia	5500000	1996	7.7		
28	107	36200000	Action	Adve Tina Turner	Mad Max Beyond Thunderdome	101840	193	English	Australia	12305523	1985	6.2		
29	107	141853	Comedy	Dr Kyra Sedgwick	Critical Care	1427	12	English	Australia	12000000	1997	6		
30	108	109306210	Action	Adve Paul Hogan	Crocodile Dundee II	44096	62	English	Australia	15800000	1988	5.5		
31	108	505295	Adventure	(Tshewang Dendup	Travelers and Magicians	1747	28	Dzongkha	Australia	1800000	2003	7.5		
32	108	32048809	Horror	Robert R'chard	House of Wax	94456	576	English	Australia	30000000	2005	5.3		
33	108	23048809	Horror	Robert R'chard	House of Wax	94456	576	English	Australia	30000000	2005	5.3		

INSIGHT:-

1. This is the first and very important step before moving forward.
2. First, I dropped the columns which have no use for the analysis.
3. Second, I dropped the columns which are blank/null.
4. Third, I removed the duplicate row values.

B.) Find the movies with the highest profit



INSIGHT:-

1. First, I created a column 'Profit' by subtracting budget column from gross column.
2. Second, I sorted the columns using the profit column from small to large.
3. Then, I plotted an XY Scatter chart to find the outliers between budget and profit.
4. There are 5 outliers in profit column.
5. The highest profited movie is 'Avatar'.

C.) Find IMDB Top 250

IMDb_Movies1.xlsx - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

A2 fx The Shawshank Redemption

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	IMDb_Top_250	num_voted_users	language	imdb_score	Rank														
2	The Shawshank Redemption	1689764	English	9.3	1														
3	The Godfather	1155770	English	9.2	2														
4	The Dark Knight	1676169	English	9	3														
5	The Godfather: Part II	790926	English	9	4														
6	Pulp Fiction	1324680	English	8.9	5														
7	Schindler's List	865020	English	8.9	6														
8	The Good, the Bad and the Ugly	503509	Italian	8.9	7														
9	The Lord of the Rings: The Return of the King	1215718	English	8.9	8														
10	Fight Club	1347461	English	8.8	9														
11	Forrest Gump	1251222	English	8.8	10														
12	Inception	1468200	English	8.8	11														
13	Star Wars: Episode V - The Empire Strikes Back	837759	English	8.8	12														
14	The Lord of the Rings: The Fellowship of the Ring	1238746	English	8.8	13														
15	City of God	533200	Portuguese	8.7	14														
16	Goodfellas	728685	English	8.7	15														
17	One Flew Over the Cuckoo's Nest	680041	English	8.7	16														
18	Seven Samurai	229012	Japanese	8.7	17														
19	Star Wars: Episode IV - A New Hope	911097	English	8.7	18														
20	The Lord of the Rings: The Two Towers	1100446	English	8.7	19														
21	The Matrix	1217752	English	8.7	20														
22	American History X	782437	English	8.6	21														
23	Interstellar	928227	English	8.6	22														
24	Modern Times	143086	English	8.6	23														
25	Saving Private Ryan	881236	English	8.6	24														
26	Se7en	1023511	English	8.6	25														
27	Spirited Away	413971	Japanese	8.6	26														
28	The Silence of the Lambs	887467	English	8.6	27														
29	The Usual Suspects	740918	English	8.6	28														
30	Alien	563827	English	8.5	29														
31	Apocalypse Now	450676	English	8.5	30														
32	Back to the Future	732212	English	8.5	31														
33	Children of Heaven	27882	Persian	8.5	32														
34	Das Leben der Jojo	105717	German	8.5	33														

IMDb_Movies Question-1 Question-2 Question-3(1) Question-3(1) Question-4 Question-5 Question-6 Question-6(Mean) Q ...

Ready Accessibility: Investigate

ENG 22:03 21-01-2023

IMDb_Movies(T)excel - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

A2 X ✓ fx The Good, the Bad and the Ugly

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Top_Foreign_Lang_Film	num_voted_users	imdb_score	language														
2	The Good, the Bad and the Ugly	503509	8.9	Italian														
3	City of God	533200	8.7	Portuguese														
4	Seven Samurai	229012	8.7	Japanese														
5	Spirited Away	417971	8.6	Japanese														
6	Children of Heaven	27882	8.5	Persian														
7	The Lives of Others	259379	8.5	German														
8	A Separation	151812	8.4	Persian														
9	Amélie	534262	8.4	French														
10	Baahubali: The Beginning	62756	8.4	Telugu														
11	Das Boot	168203	8.4	German														
12	Oldboy	356181	8.4	Korean														
13	Princess Mononoke	221552	8.4	Japanese														
14	Metropolis	111841	8.3	German														
15	The Hunt	170155	8.3	Danish														
16	Unforgiven	248354	8.3	German														
17	Pan's Labyrinth	80429	8.2	French														
18	The Bridge on the River Kwai	131831	8.2	Spanish														
19	The Thing	467234	8.2	Spanish														
20	Warrior	214091	8.2	Japanese														
21	Annie Hall	81644	8.1	Portuguese														
22	In the Shadow of the Moon	65951	8.1	Danish														
23	Sling Blade	106160	8.1	Japanese														
24	Tae Guk Gi: The Brotherhood of War	64556	8.1	Spanish														
25	The Best Years of Our Lives	173551	8.1	Spanish														
26	The Imitation Game	31943	8.1	Korean														
27	Bowling for Columbine	28951	8	Portuguese														
28	Jaws	70194	8	French														
29	True Romance	69759	8	Hindi														
30	Waltz with Bashir	147566	8	Italian														
31	Young Frankenstein	46107	8	Hebrew														
32	Children of Men	149414	7.9	Mandarin														
33	E.T. the Extra-Terrestrial	34449	7.9	Hindi														
34	Mon... Mon...	44151	7.9	French														

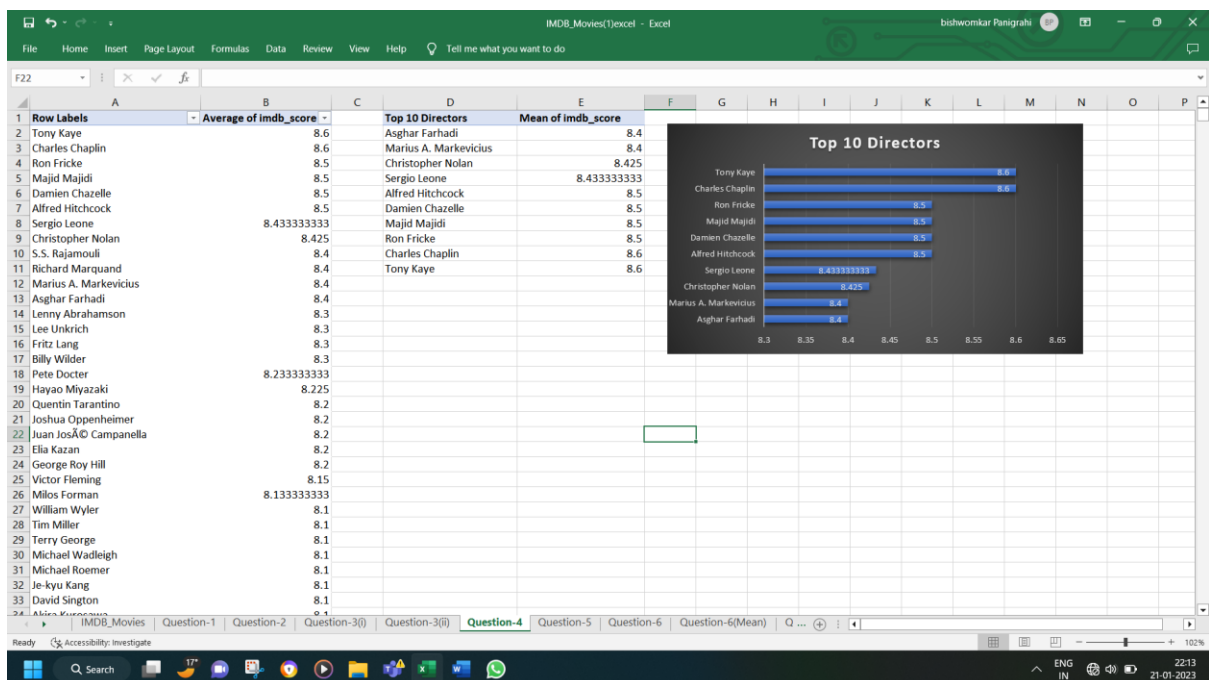
IMDb_Movies Question-1 Question-2 Question-3(1) Question-4 Question-5 Question-6 Question-6(Mean) Q ... + : |

Ready Accessibility: Investigate ENG IN 22:04 21-01-2023

INSIGHT:-

1. At first, I kept only those movies which have 'num_voted_users' greater than 25,000.
2. Made a column named 'IMDb_Top_250' and kept the top movies with highest 'imdb_score' to list 'imdb_score'.
3. Then, added a column 'Rank' containing the values 1 to 250 using the RANK() function + COUNTIFS() function.
4. After that, I extracted all the movies in the IMDb_Top_250 column by filtering the 'language' column (unselecting English language) and stored them in a new column named 'Top_Foreign_Lang_Film'.

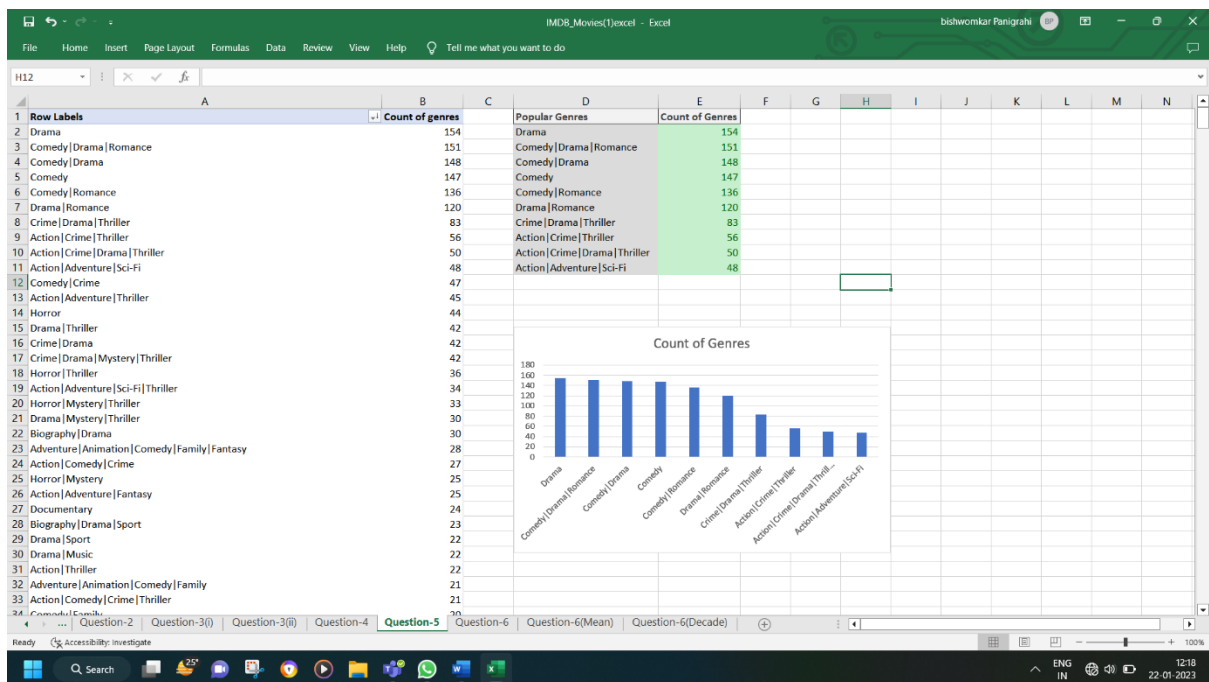
D.) Find the best directors



INSIGHT:-

1. First, I created a pivot table done in Question 1.
2. Then, I fed 'director_name' into the Rows and took average of 'imdb_score' in the Values section then sorted it from largest to smallest.
3. After that I selected top 10 directors and their mean of imdb_score in other columns, then I made a bar chart.

E.) Find popular genres

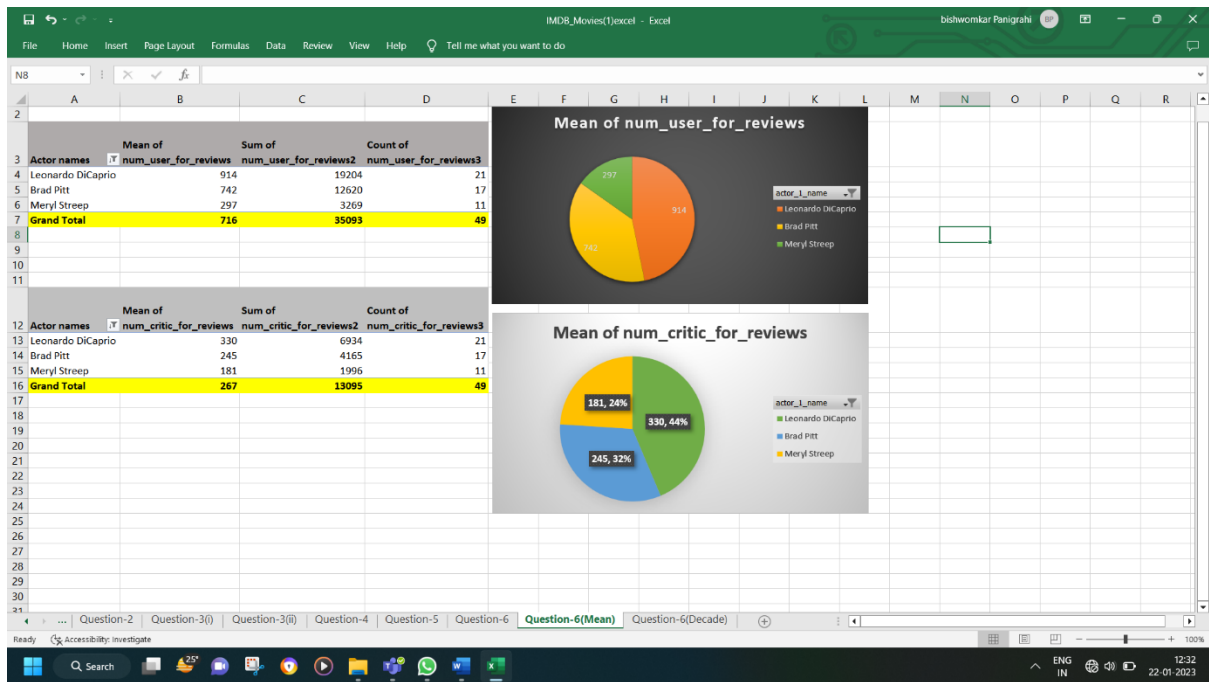


INSIGHT:-

1. First, I selected the 'genres' column from the cleaned dataset done in Question 1 and created a pivot table.
2. Then I fed the 'genres' into the Rows and took count of 'genres' in the Values section.
3. After that, I sorted the 'Count of genres' in descending order.
4. Then I copied the top 10 genres and their count and pasted it in the other columns.
5. Next, I made a clustered column chart of the top 10 genres for the better insights.

F.) Find the critic-favourite and audience-favourite Actors

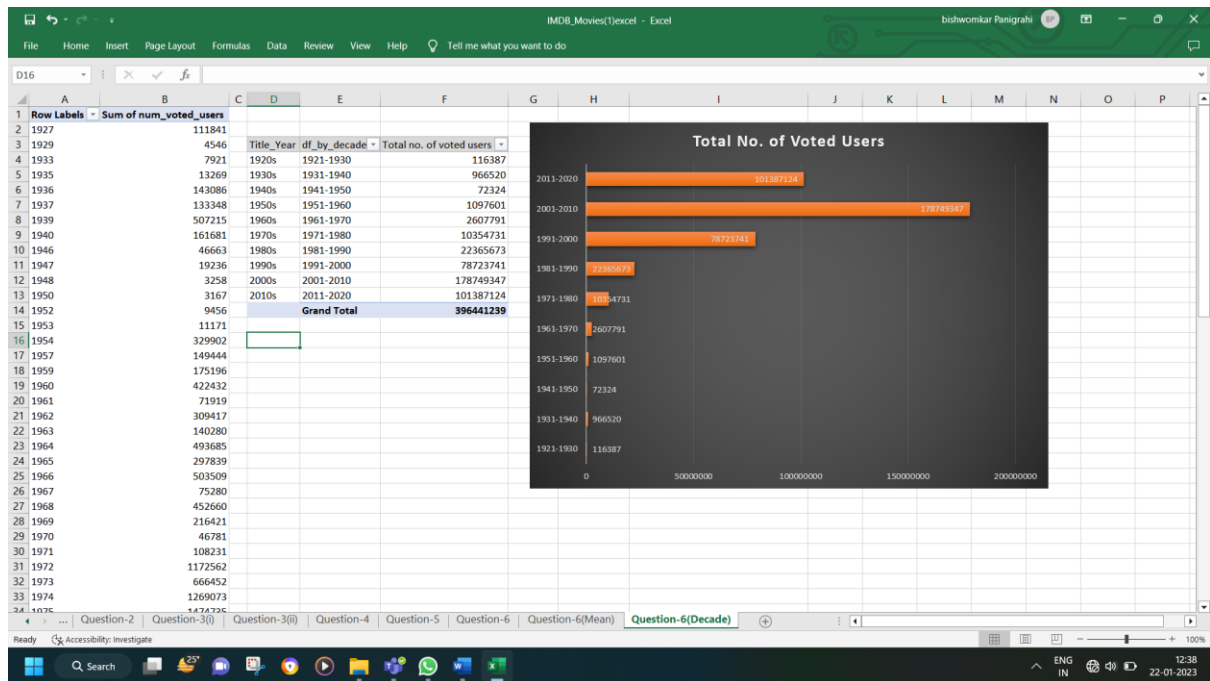
	D	E	F	G	H	I	J	K	L	M	N	O	P
1	num_voted_users	movie_title	title_year	Meryl Streep	Leo Caprio	Brad Pitt	Combined	Group By					
2	1148137	The Dark Knight Rises	2012	Lucy	Django Unchained	Divergent	Lucy	Meryl Streep					
3	456260	Soul Food	2012	Maleficent	Inception	Shanghai Knights	Maleficent	Meryl Streep					
4	951174	Django Unchained	2012	Transcendence	How to Train Your Dragon	The Circle	Transcendence	Meryl Streep					
5	522048	This Is England	2012	Down and Out with the Dolls	Pirates of the Caribbean: The Curse of the Black Pearl	42nd Street	Down and Out with the Dolls	Meryl Streep					
6	520380	Saw	2012	Machine Gun Preacher	The Sessions	Fight Club	Machine Gun Preacher	Meryl Streep					
7	352503	Guardians of the Galaxy	2015	Like Crazy	Frozen River	Scott Pilgrim vs. the World	Like Crazy	Meryl Streep					
8	582917	Weekend	2013	Brown Sugar	The Martian	Mulan	Brown Sugar	Meryl Streep					
9	548573	Green Room	2015	Bridge of Spies	Lethal Weapon 4	Leap Year	Bridge of Spies	Meryl Streep					
10	886294	Mauro	2009	True Lies	The Departed	The Kids Are All Right	True Lies	Meryl Streep					
11	928227	Interstellar	2014	Hellboy II: The Golden Army	The Hunger Games	Connie and Carla	Hellboy II: The Golden Army	Meryl Streep					
12	995415	A Fistful of Dollars	2012	Gerry	Animal House	Star Trek Into Darkness	Gerry	Meryl Streep					
13	995415	Akira	2012		The Squid and the Whale	Sunshine State	Django Unchained	Leonardo DiCaprio					
14	245333	A Room with a View	2011		Dream with the Fishes	Zookeeper	Inception	Leonardo DiCaprio					
15	431578	Black Book	2011		Transamerica	Black Swan	How to Train Your Dragon	Leonardo DiCaprio					
16	371639	Sliding Doors	2016		The Iron Giant	Kung Fu Panda	Pirates of the Caribbean: The Curse of the Black Pearl	Leonardo DiCaprio					
17	701607	Tarnation	2012		Casino Royale	The Puffy Chair	The Sessions	Leonardo DiCaprio					
18	551363	In Bruges	2010		Hurricane Streets	It Follows	Frozen River	Leonardo DiCaprio					
19	70336	The Queen	2012		The Devil Wears Prada		The Martian	Leonardo DiCaprio					
20	452465	Cast Away	2012		Ocean's Twelve		Lethal Weapon 4	Leonardo DiCaprio					
21	465019	Sunshine Cleaning	2012		The Longest Yard		The Departed	Leonardo DiCaprio					
22	682153	The Avengers	2014		The Family Man		The Hunger Games	Leonardo DiCaprio					
23	637246	Gangster's Paradise: Jerusalem	2012				Animal House	Leonardo DiCaprio					
24	1676169	The Dark Knight	2008				The Squid and the Whale	Leonardo DiCaprio					
25	418214	Wild Target	2015				Dream with the Fishes	Leonardo DiCaprio					
26	1465200	Inception	2010				Transamerica	Leonardo DiCaprio					
27	462669	Lady Cake	2015				The Iron Giant	Leonardo DiCaprio					
28	277172	Rio	2012				Casino Royale	Leonardo DiCaprio					
29	557489	Secretary	2013				Hurricane Streets	Leonardo DiCaprio					
30	780588	How to Train Your Dragon	2013				The Devil Wears Prada	Leonardo DiCaprio					
31	275868	Machete	2015				Ocean's Twelve	Leonardo DiCaprio					
32	451803	The Rose	2012				The Longest Yard	Leonardo DiCaprio					
33	439176	There Will Be Blood	2013				The Family Man	Leonardo DiCaprio					
34	272893	Iron Man	2015				Divergent	Brad Pitt					
35	395573	You Can Count on Me	2013				Shanghai Knights	Brad Pitt					
36	428916	I Served the King of England	2012				The Circle	Brad Pitt					
37	395087	O Brother, Where Art Thou?	2014				42nd Street	Brad Pitt					
38	431620	Murderball	2014				Fight Club	Brad Pitt					
39	136367	Divergent	2011				Scott Pilgrim vs. the World	Brad Pitt					
40	479047	The Truman Show	2016				Mulan	Brad Pitt					
41	496749	Dead Man's Shoes	2014				Leap Year	Brad Pitt					
42	190038	Slumdog Millionaire	2011				The Kids Are All Right	Brad Pitt					
43	381148	Chadw of the Vamora	2013				Connie and Carla	Brad Pitt					



INSIGHT:-

1. I created columns named Meryl_Streep, Leo_Caprio, and Brad_Pitt which have the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors using the 'actor_1_name' column.
2. Then, added the rows of all these columns and stored them in a new column named 'Combined'.
3. I stacked the columns by the actor's name: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt'.
4. Then, I selected the cleaned dataset done in Question 1 and created a pivot table.
5. Then, I put the 'actor_1_name' into the Rows and took mean/average of 'num_users_for_review' in the Values section.
6. Sorted the column from largest to smallest by mean of 'num_users_for_review'.
7. Then, I made a pie chart of the mean of 'num_users_for_review'.
8. I did all the same process for the mean of 'num_critic_for_review'.

Second part (change in number of voted users over decades):



INSIGHT:-

1. First, I selected the cleaned dataset done in Question 1 and created a pivot table.
2. Second, I put the 'title_year' into the Rows and took the sum of 'num_voted_users'.
3. Third, I grouped the title_year by decade and stored in df_by_decade column.
4. Finally, I plotted the total no. of voted users against the decade in a bar chart.

- ✚ There are total 5 outliers in the profit columns.
- ✚ The movie with the highest profit is 'Avatar' .
- ✚ Shawshank Redemption is having the highest IMDB rating.
- ✚ The Good, the Bad and the Ugly (Italian) is the top foreign language movie.
- ✚ Charles Chaplin is the top director .
- ✚ The most popular genres is Drama.
- ✚ 'Leonardo DiCaprio' is both the critic-favourite and audience-favourite actor.
- ✚ Most of the users voted in the decade 2000s and the least in decade 1940s.

In this project, I applied the basic and advance Excel concepts. The concepts related to statistics and EDA have been implemented here by using MS Excel. In this task, the concepts regarding the sort, filter, pivot table, charts, different functions like rank, etc have been implemented. I learned to implement the learning of Excel in the real-time project. I learned how to frame the problem by asking 'what' looking at the dataset. It helped me in learning the '5 Why Analysis' to determine the root cause of the problem. I learned how a data analyst think deeper and deeper to generate the valuable insights.

It was a great learning experience while doing this project and it was challenging too while asking the different questions and finding their answers.

Here is the excel sheet link https://1drv.ms/x/s!Arxa-xC1P_LYgQBI8Dz53PcZpo4S .



**Thank
You!!!**

www.dreamstime.com

Submitted by :- **Bishwomkar Panigrahi**