

# IMDB DATA ANALYSIS

## DESCRIPTION:

I have used the 5 whys approach to better understand the problem. This project is more about deep analysis of IMDB Movie data. It's also called the Root Cause Analysis, developed by Sakichi Toyoda, founder of Toyota Industries. Once I have defined a problem, cleaned the necessary data, and used Data Analysis skills to explore the data set and derive insights. This analysis will help the company to get insights about what users like the most so it would be easy for them to improve the product.

## APPROACH :

First I tried my best to understand the problem statement and gather the requirements of what details the project asked me to fetch using Excel. And then I planned how to use the logic to get desired results. And then I tried to implement it. I have used my knowledge of Excel here. All the tasks asked in this project were performed on Microsoft Excel. Excel functions have been used here

## TECH-STACK USED :

I have used Microsoft Excel from Microsoft 365

## INSIGHTS :

1. I first imported the file from google sheets to excel.
2. I have used many excel functions like SUMIF, AVERAGEIF, etc in this project.
3. I came to know about many excel shortcuts while doing this project.

## Clean the data:

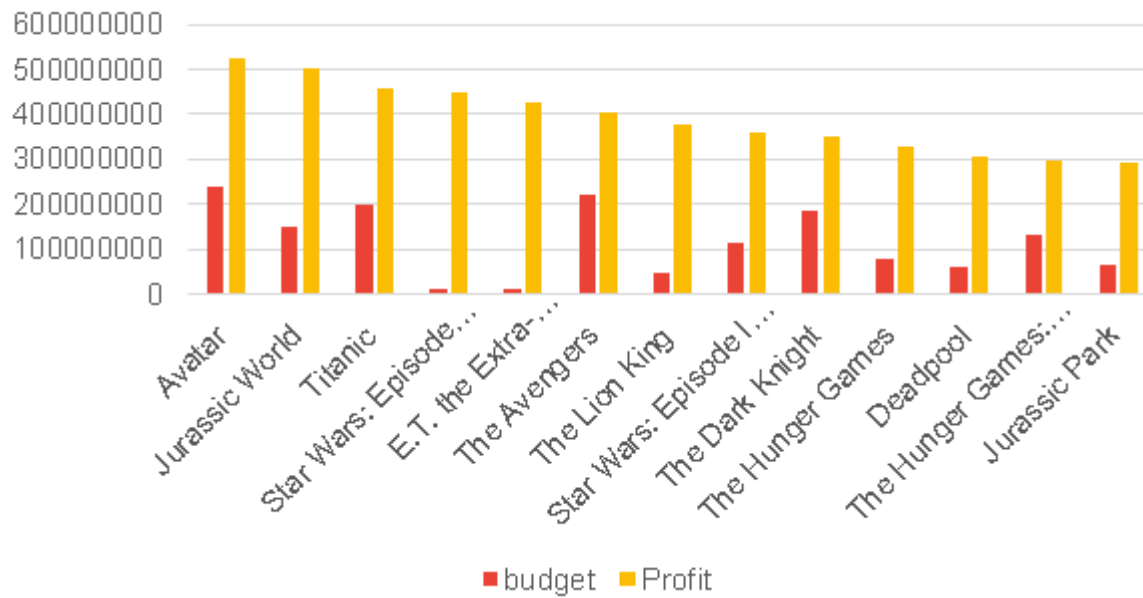
I have first import the .csv file in excel and then performed a clean-up on data like the **Removal of duplicate data, Removal of blank cells, etc.** cleaning the data is the most important task before starting the analysis of data so that it would provide accurate results.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	color	director_name	num_critic_for_revie	duration	director_facet	actor_3_faceb	actor_2_name	actor_1_faceb	gross	genres	actor_1_nai	movie_title	num_voted_U
2	Color	James Cameron	723	178	0	855	Joel David Moore	1000	760505847	Action Adventure	CCH Pounder	Avatar	8
3	Color	Colin Trevorrow	644	124	365	1000	Judy Greer	3000	652177271	Action Adventure	Bryce Dallas Howard	Jurassic World	4
4	Color	James Cameron	315	194	0	794	Kate Winslet	29000	658672302	Drama	Roman Leonardo DiCaprio	Titanic	7
5	Color	George Lucas	282	125	0	504	Peter Cushing	11000	460935665	Action Adventure	Harrison Ford	Star Wars: Episode I - The Phantom Menace	9
6	Color	Steven Spielberg	215	120	14000	548	Dee Wallace	861	434949459	Family Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	2
7	Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	26000	623279547	Action Adventure	Chris Evans	The Avengers	9
8	Color	Roger Allers	186	73	28	847	Nathan Lane	2000	422783777	Adventure Animation	Matthew Broderick	The Lion King	6
9	Color	George Lucas	320	136	0	1000	Liam Neeson	20000	474544677	Action Adventure	Natalie Portman	Star Wars: Episode II - Attack of the Clones	8
10	Color	Christopher Nolan	645	152	22000	11000	Heath Ledger	23000	533316061	Action Crime	Christian Bale	The Dark Knight	16
11	Color	Gary Ross	673	142	378	575	Josh Hutcherson	34000	407999255	Adventure Drama	Jennifer Lawrence	The Hunger Games	7
12	Color	Tim Miller	579	108	84	361	Ed Skrein	16000	363024263	Action Adventure	Ryan Reynolds	Deadpool	4
13	Color	Francis Lawrence	502	146	508	523	Josh Hutcherson	34000	424645577	Adventure Sci-Fi	Jennifer Lawrence	The Hunger Games	4
14	Color	Steven Spielberg	308	127	14000	191	Ariana Richards	967	356784000	Adventure Sci-Fi	Wayne Knight	Jurassic Park	6
15	Color	Pierre Coffin	306	98	275	1000	Miranda Cosgrove	7000	368049635	Animation Comedy	Steve Carell	Despicable Me	2
16	Color	Clint Eastwood	490	133	16000	318	Leonard Roberts	14000	350123553	Action Biography	Bradley Cooper	American Sniper	3
17	Color	Andrew Stanton	301	100	475	799	Stephen Root	1000	380838870	Adventure Animation	Alexander Gould	Finding Nemo	6
18	Color	Andrew Adamson	205	93	80	48	Jennifer Saunders	692	436471036	Adventure Animation	Rupert Everett	Shrek 2	3
19	Color	Peter Jackson	328	192	0	416	Billy Boyd	5000	377019252	Action Adventure	Orlando Bloom	The Lord of the Rings: The Two Towers	12
20	Color	Richard Marquand	197	134	37	504	Ian McDiarmid	11000	309125409	Action Adventure	Harrison Ford	Star Wars: Episode IV - A New Hope	6
21	Black and White	Robert Zemeckis	149	142	0	194	Siobhan Fallon	15000	329691196	Comedy Drama	Tom Hanks	Forrest Gump	12
22	Color	Irvin Kershner	223	127	883	441	Kenny Baker	11000	290158751	Action Adventure	Harrison Ford	Star Wars: Episode V - The Empire Strikes Back	8
23	Color	Chris Columbus	102	103	0	925	Devin Ratray	3000	285761243	Comedy Family	Macaulay Culkin	Home Alone	3
24	Color	George Lucas	359	140	0	4000	Christopher Lee	20000	380262555	Action Adventure	Natalie Portman	Star Wars: Episode III - Revenge of the Sith	6
25	Color	Sam Raimi	291	121	0	4000	James Franco	24000	403706375	Action Adventure	J.K. Simmons	Spider-Man	6
26	Color	Kyle Balda	308	91	22	1000	Jon Hamm	7000	336029560	Action Animation	Steve Carell	Minions	1
27	Color	M. Night Shyamalan	234	107	0	766	Haley Joel Osment	13000	293501675	Drama Mystery	Bruce Willis	The Sixth Sense	7
28	Color	Steven Spielberg	385	130	14000	366	Robert Shaw	813	260000000	Adventure Drama	Roy Scheider	Jaws	4
29	Color	Chris Buck	406	102	69	490	Maurice LaMarche	1000	400736600	Adventure Animation	Josh Gad	Frozen	4
30	Color	Yarrow Cheney	165	87	11	745	Eric Stonestreet	1000	323505540	Animation Comedy	Steve Coogan	The Secret Life of Pets	1
31	Color	Chris Weitz	299	130	129	12000	Kristen Steward	21000	296623634	Adventure Drama	Robert Pattinson	The Twilight Saga: Eclipse	2
32	Color	Peter Jackson	294	172	0	857	Orlando Bloom	16000	340478898	Action Adventure	Christopher Lee	The Lord of the Rings: The Fellowship Ring	12

## Find the movies with the highest profit.

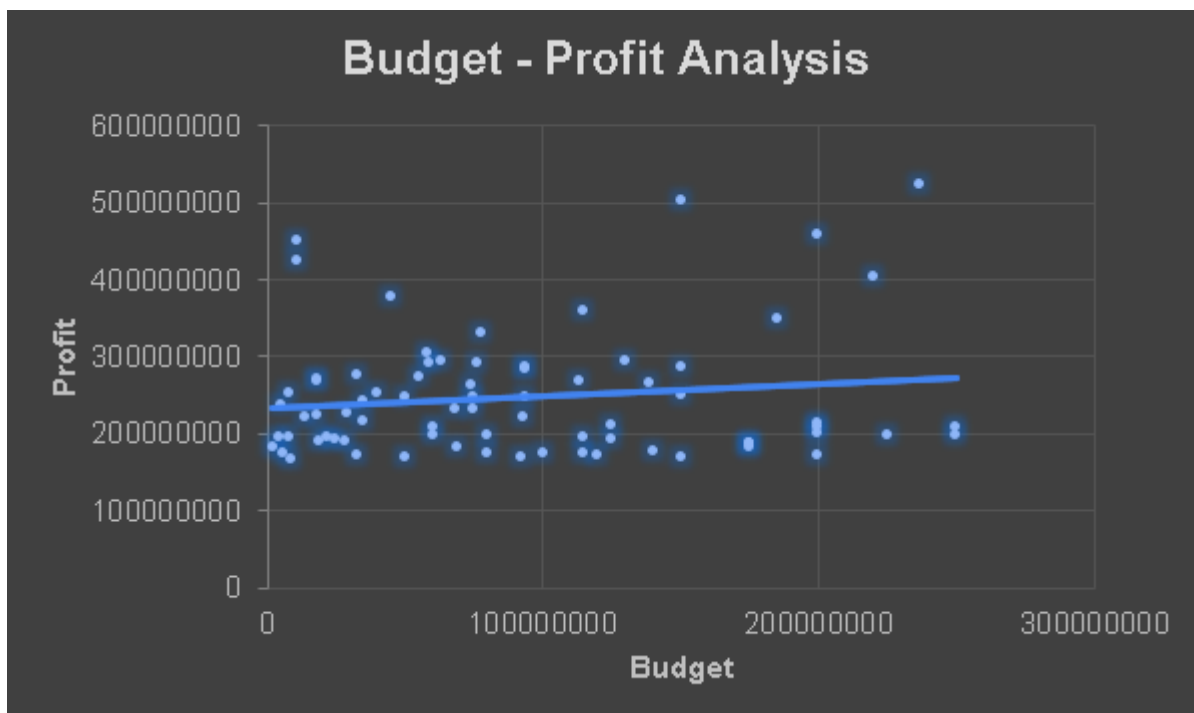
I have created a new column called profit which contains the difference between the two columns: gross and budget. and Sort the column using the profit column as a reference. And then Plotted profit (y-axis) vs budget (x-axis) using the appropriate chart type. I observed some outliers, there are some extreme rises in profit compared to budget that cause outliers

I have shown the highest-profit movies using the chart.



I have plotted the **trend line** which shows the trend of profit and is also used to predict future up-downs of profit.

## BUDGET VS PROFIT CHART



## Find IMDB Top 250

I have created a new column IMDb\_Top\_250 and stored the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb\_score and num\_voted\_users is greater than 25,000. With Rank representing the rank of movies based on IMDb rating

1. MOVIES WITH THE HIGHEST IMDB SCORE, NUM\_VOTED\_USERS GREATER THAN 20000, IN ALL LANGUAGES EXCEPT ENGLISH.
2. MOVIES WITH THE HIGHEST SCORE, NUM\_VOTED\_USERS GREATER THAN 20000, IN ENGLISH LANGUAGE ONLY.

I have used advanced filters to get the results.

L	M	N	O	P	Q	R	S	T	U	V	W
num_voted_users	Rank	actor_1_name	movie_title	language	imdb_score	num_voted_users					
>25000		1 Clint Eastwood	The Good, the Bad and	Italian	8.9	503509					
		2 Alice Braga	City of God	Portuguese	8.7	533200					
		2 Takashi Shimura	Seven Samurai	Japanese	8.7	229012					
		3 Bunta Sugawara	Spirited Away	Japanese	8.6	417971					
		4 Sebastian Koch	The Lives of Others	German	8.5	259379					
		4 Bahare Seddigi	Children of Heaven	Persian	8.5	27882					
		5 Shahab Hosseini	A Separation	Persian	8.4	151812					
		5 Min-sik Choi	Oldboy	Korean	8.4	356181					
		5 Jürgen Prochnow	Das Boot	German	8.4	168203					
		5 Mathieu Kassovitz	Amélie	French	8.4	534262					
		5 Minnie Driver	Princess Mononoke	Japanese	8.4	221552					
		6 Thomas Bo Larsen	The Hunt	Danish	8.3	170155					
		6 Brigitte Helm	Metropolis	German	8.3	111841					
		6 Thomas Kretschma	Downfall	German	8.3	248354					
		7 Ivana Baquero	Pan's Labyrinth	Spanish	8.2	467234					
		7 Ricardo Darin	The Secret in Their Eyes	Spanish	8.2	131831					
		7 Lubna Azabal	Incendies	French	8.2	80429					
		7 Christian Bale	Howl's Moving Castle	Japanese	8.2	214091					
		8 Adriana Barraza	Amores Perros	Spanish	8.1	173551					
		8 Ulrich Thomsen	The Celebration	Danish	8.1	65951					
		8 Wagner Moura	Elite Squad	Portuguese	8.1	81644					
		8 Belén Rueda	The Sea Inside	Spanish	8.1	64556					
		8 Min-sik Choi	Tae Guk Gi: The Brothe	Korean	8.1	31943					
		8 Mitsuo Iwata	Akira	Japanese	8.1	106160					
		9 Clint Eastwood	A Fistful of Dollars	Italian	8	147566					
		9 Fernanda Montenegro	Central Station	Portuguese	8	28951					
		9 Ari Folman	Waltz with Bashir	Hebrew	8	46107					
		9 Catherine Deneuve	Persepolis	French	8	70194					
		9 Shah Rukh Khan	My Name Is Khan	Hindi	8	69759					
		10 Chen Chang	Crouching Tiger, Hidden	Mandarin	7.9	217740					
		10 Anamaria Marinca	4 Months, 3 Weeks and	Romanian	7.9	44763					
		10 Ricardo Darin	Nine Queens	Spanish	7.9	38215					
		10 Jean-Baptiste Maur	The Chorus	French	7.9	44151					
		10 Yuki Matsuzaki	Letters from Iwo Jima	Japanese	7.9	132149					
		10 Isabelle Huppert	Amour	French	7.9	70382					
		10 Jet Li	Hero	Mandarin	7.9	149414					

MOVIES BASED ON  
HIGHEST IMDB  
SCORE IN ALL  
LANGUAGES EXCEPT  
"ENGLISH" WITH  
NUM\_VOTED\_USERS  
GREATER THAN 20000  
WITH RANK

A	B	C	D	E	F	G	H	I
Rank	actor_1_name	movie_title	language	imdb_score	num_voted_users			
1	Morgan Freeman	The Shawshank Redemption	English	9.3	1689764			
2	Al Pacino	The Godfather	English	9.2	1155770			
3	Christian Bale	The Dark Knight	English	9	1676169			
4	Robert De Niro	The Godfather: Part II	English	9	790926			
4	Orlando Bloom	The Lord of the Rings: The Return of the King	English	8.9	1215718			
5	Bruce Willis	Pulp Fiction	English	8.9	1324680			
5	Liam Neeson	Schindler's List	English	8.9	865020			
5	Tom Hanks	Forrest Gump	English	8.8	1251222			
6	Harrison Ford	Star Wars: Episode V - The Empire Strikes Back	English	8.8	837759			
6	Christopher Lee	The Lord of the Rings: The Fellowship Ring	English	8.8	1238746			
6	Leonardo DiCaprio	Inception	English	8.8	1468200			
6	Brad Pitt	Fight Club	English	8.8	1347461			
6	Harrison Ford	Star Wars: Episode IV - A New Hope	English	8.7	911097			
7	Christopher Lee	The Lord of the Rings: The Two Towers	English	8.7	1100446			
7	Keanu Reeves	The Matrix	English	8.7	1217752			
7	Scatman Crothers	One Flew Over the Cuckoo's Nest	English	8.7	680041			
7	Robert De Niro	Goodfellas	English	8.7	728685			
7	Tom Hanks	Saving Private Ryan	English	8.6	881236			
8	Anthony Hopkins	The Silence of the Lambs	English	8.6	887467			
8	Morgan Freeman	Se7en	English	8.6	1023511			
8	Matthew McConaughey	Interstellar	English	8.6	928227			
8	Kevin Spacey	The Usual Suspects	English	8.6	740918			
8	Ethan Suplee	American History X	English	8.6	782437			
8	Paulette Goddard	Modern Times	English	8.6	143086			
8	Matthew Broderick	The Lion King	English	8.5	644348			
9	Harrison Ford	Raiders of the Lost Ark	English	8.5	661017			
9	Tom Hardy	The Dark Knight Rises	English	8.5	1144337			
9	Lea Thompson	Back to the Future	English	8.5	732212			
9	Joe Morton	Terminator 2: Judgment Day	English	8.5	744891			
9	Djimon Hounsou	Gladiator	English	8.5	982637			
9	Tom Hanks	The Green Mile	English	8.5	782610			
9	Tom Skerritt	Alien	English	8.5	563827			
9	Leonardo DiCaprio	Django Unchained	English	8.5	955174			
9	Harrison Ford	Apocalypse Now	English	8.5	450676			
9	Leonardo DiCaprio	The Departed	English	8.5	873649			
9	Janet Leigh	Psycho	English	8.5	422432			

I first made a dummy table and specifies the condition that the table should follow to produce desired results. and used an advanced filter to get the results.

## Find TOP10 best directors

I have shown the top 10 directors for whom the mean of imdb\_score is the highest.

I have used AVERAGEIF(RANGE, CRITERIA, AVERAGE RANGE) function to get the results.

	A	B	C	D
1	SR NO.	Director_name	IMDB_SCORE	
2	1	Christopher Nolan	☆ 8.425	
3	2	Richard Marquand	☆ 8.4	
4	3	Alfred Hitchcock	☆ 8.5	
5	4	Damien Chazelle	☆ 8.5	
6	5	Asghar Farhadi	☆ 8.4	
7	6	Sergio Leone	☆ 8.433333333	
8	7	Majid Majidi	☆ 8.5	
9	8	Tony Kaye	☆ 8.6	
10	9	Charles Chaplin	☆ 8.6	
11	10	Ron Fricke	☆ 8.5	
12	11	Akira Kurosawa	☆ 8.7	

## Find popular genres

Firstly I split the genre information into different columns using the “text to columns” option in excel and then I used the “COUNTIF” function to calculate the most popular genre.

	I	J	K	L
	<b>genres</b>	<b>Count</b>		<b>Popular genre:</b>
	Action	951		Drama
	Drama	1876		
	Family	440		
	Adventure	771		
	Animation	196		
	Comedy	1455		
	Biography	238		
	Horror	386		
	Musical	93		
	Crime	703		
	Documentary	45		
	Mystery	377		
	Fantasy	500		
	Sci-Fi	475		
	Western	57		
	Romance	843		
	Thriller	1095		
	War	148		

**Find the mean of the num\_critic\_for\_reviews,num\_users\_for\_review  
And Identify the actor with the Highest mean.**

I have used ROUNDUP(AVERAGEIF(RANGE, CRITERIA, AVERAGE RANGE),0) Function to get the results. Here Roundup function is used to round the digits to the whole number. AVERAGEIF is used here to get an average based on certain conditions

D	E	F	G	H	I	J	K	L
		actor_1_name	Average num_critic	Average num_user_for_reviews				
		CCH Pounder	260	966				
		Bryce Dallas Howard	333	1104				
		Leonardo DiCaprio	331	915				
		Harrison Ford	182	529				
		Henry Thomas	117	272				
		Chris Hemsworth	391	630				
		Matthew Broderick	143	307				
		Natalie Portman	250	1032				
		Christian Bale	258	800				
		Jennifer Lawrence	419	584				
		Ryan Reynolds	232	330				
		Wayne Knight	308	895				
		Steve Carell	237	315				
		Bradley Cooper	278	380				
		Alexander Gould	191	569				
		Rupert Everett	205	483				
		Orlando Bloom	259	1842				
		Tom Hanks	237	678				
		Macaulay Culkin	59	189				
		J.K. Simmons	238	504				
		Bruce Willis	174	443				
		Roy Scheider	230	614				
		Josh Gad	326	593				
		Steve Coogan	159	127				
		Robert Pattinson	283	768				
		Christopher Lee	155	1238				
		Nia Vardalos	146	312				
		Will Smith	223	564				
		Catherine Dyer	261	420				
		Judge Reinhold	80	134				
		Michael Gough	167	746				
		Robert Downey Jr.	246	384				
		Kathleen Freeman	212	945				
		Glenn Morshower	315	935				
		Robert De Niro	142	298				
		Tom Hardy	342	745				

CRITIC FAVOURITE :  
ALBERT FINNEY  
AUDIENCE FAVOURITE :  
HEATHER DONHUE

Sheet1 | Sheet8 | Sheet9 | Sheet2 | Sheet3 | Sheet4 | **Sheet5** | Sheet6 | ... + : ◀

## I HAVE EXTRACTED MOVIES DONE BY MERYL STREEP, BRAD PITT, AND LEONARDO DI CAPRIO.

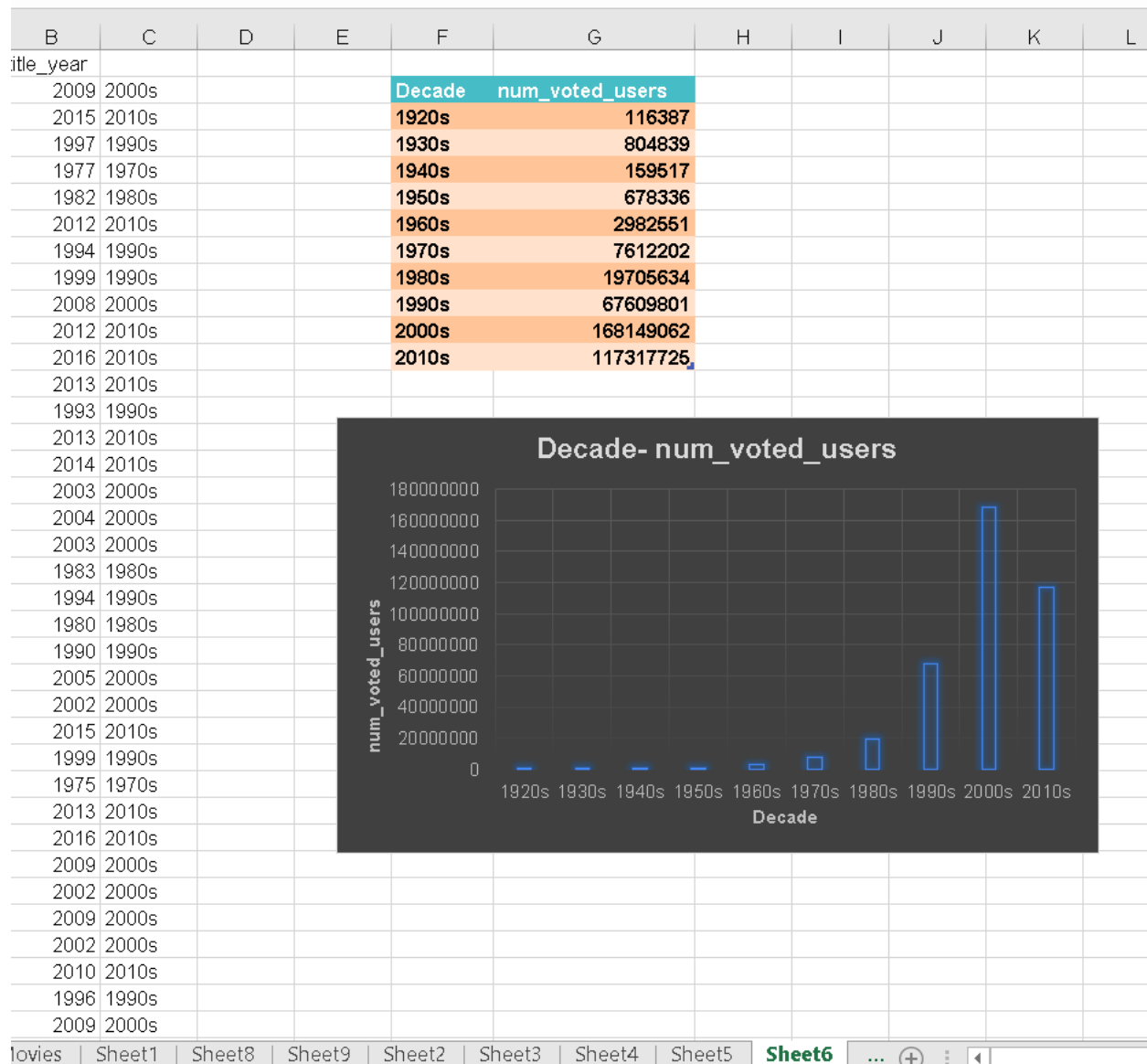
	A	B	C
1	MERYL STREEP	BRAD PITT	LEONARDO DI CAPRIO
2	The Devil Wears Prada	Ocean's Eleven	Titanic
3	Out of Africa	Mr. & Mrs. Smith	Inception
4	Julie & Julia	Interview with the Vampire: The Vampire Chronicles	Catch Me If You Can
5	Hope Springs	Fury	Django Unchained
6	It's Complicated	Ocean's Twelve	The Revenant
7	The Iron Lady	Babel	Shutter Island
8	The Hours	Killing Them Softly	The Departed
9	A Prairie Home Companion	True Romance	The Great Gatsby
10	The River Wild	By the Sea	The Great Gatsby
11	One True Thing	The Tree of Life	Romeo + Juliet
12	Lions for Lambs	The Curious Case of Benjamin Button	The Man in the Iron Mask
13		Fight Club	The Wolf of Wall Street
14		The Assassination of Jesse James by the Coward Robert For J. Edgar	
15		Seven Years in Tibet	The Aviator
16		Sinbad: Legend of the Seven Seas	Marvin's Room
17		Troy	The Beach
18		Spy Game	Revolutionary Road
19			The Quick and the Dead
20			Gangs of New York
21			Body of Lies
22			Blood Diamond

## Observe the change in the number of voted users over decades using a bar chart.

I have created a column called decade which represents the decade to which every movie belongs. For example, the title\_year year 1923, 1925 should be stored as the 1920s.

I have used SUMIF function to get results.





## RESULTS:

I have learned a lot while working on this project. I have gained more in-depth knowledge. This project has improved my ability to write efficient and optimized Excel functions and my problem-solving skills. I can help organizations extract the information they need, by understanding their intent and providing a proper answer to their question.