IMDB Movie Analysis

Project Description:

The dataset provided is having various columns of different IMDB Movies. i have to Frame the problem. For this task, i will need to define a problem that i want to shed some light on.

We can do this by asking 'What?' This is where you frame the problem i.e. What is the problem?

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 I have defined a problem, I should clean the data as necessary, and use my Data Analysis skills to explore the data set and derive insights.

Approach:

Step1: To download and analyse the Dataset provided to work on.

Step2: Checking for missing data

Step3: Clubbing columns with multiple categories

Step4: Checking and Removing the outliers

Step5: Drawing Data Summary

Tech Stack Used:

I used

- **Microsoft® Excel® 2019 MSO (Version 2304 Build 16.0.16327.20200) 64-bit** for the IMDB Movie Analysis.
- **Ms Word 2019** for the preparation of the document to be presented.

Insights:

Based on the achieved results I can conclude the following things:

1. There are Outliers in the dataset which can be observed from the XY Scatter chart.
2. There are duplicates in the dataset which are found and removed.
3. There are blanks/Null values in the dataset which are removed.
4. The number of voted users are increasing over decades.
5. People often like the genre Comedy|Drama|Romance.
6. The IMDB score average of English movies are higher than that of foreign movies.

Results: The detailed answers to the questions are below:

A. **Cleaning the data:** Clean the data provided

Result:

1. First, I removed all the columns that are unwanted for the analysis.
2. Second, I removed the rows that are blank/null.
3. Third, I removed all the duplicate values found

	A	B	C	D	E	F	G	H	I	J	K	L
1	director_name	genres	actor_1_name	movie_title	num_voted_users	num_user_for_reviews	num_critic_for_reviews	language	gross	budget	title_year	imdb_score
2	James Cameron	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	AvatarÂ	886204	3054	723	English	760505847	237000000	2009	7.9
3	Gore Verbinski	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: At World's EndÂ	471220	1238	302	English	309404152	300000000	2007	7.1
4	Sam Mendes	Action|Adventure|Thriller	Christoph Waltz	SpectreÂ	275868	994	602	English	200074175	245000000	2015	6.8
5	Christopher Nolan	Action|Thriller	Tom Hardy	The Dark Knight RisesÂ	1144337	2701	813	English	448130642	250000000	2012	8.5
6	Andrew Stanton	Action|Adventure|Sci-Fi	Daryl Sabara	John CarterÂ	212204	738	462	English	73058679	263700000	2012	6.6
7	Sam Raimi	Action|Adventure|Romance	J.K. Simmons	Spider-Man 3Â	383056	1902	392	English	336530303	258000000	2007	6.2
8	Nathan Greno	Adventure|Animation|Comedy|Fa	Brad Garrett	TangledÂ	294810	387	324	English	200807262	260000000	2010	7.8
9	Joss Whedon	Action|Adventure|Sci-Fi	Chris Hemsworth	Avengers: Age of UltronÂ	462669	1117	635	English	458991599	250000000	2015	7.5
10	David Yates	Adventure|Family|Fantasy|Myste	Alan Rickman	Harry Potter and the Half-Blood PrinceÂ	321795	973	375	English	301956980	250000000	2009	7.5
11	Zack Snyder	Action|Adventure|Sci-Fi	Henry Cavill	Batman v Superman: Dawn of JusticeÂ	371639	3018	673	English	330249062	250000000	2016	6.9
12	Bryan Singer	Action|Adventure|Sci-Fi	Kevin Spacey	Superman ReturnsÂ	240396	2367	434	English	200069408	209000000	2006	6.1
13	Marc Forster	Action|Adventure	Giancarlo Giannini	Quantum of SolaceÂ	330784	1243	403	English	168368427	200000000	2008	6.7
14	Gore Verbinski	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: Dead Man's ChestÂ	522040	1832	313	English	423032628	225000000	2006	7.3
15	Gore Verbinski	Action|Adventure|Western	Johnny Depp	The Lone RangerÂ	181792	711	450	English	89289910	215000000	2013	6.5
16	Zack Snyder	Action|Adventure|Fantasy|Sci-Fi	Henry Cavill	Man of SteelÂ	548573	2536	733	English	291021565	225000000	2013	7.2
17	Andrew Adamson	Action|Adventure|Family|Fantasy	Peter Dinklage	The Chronicles of Narnia: Prince CaspianÂ	149922	438	258	English	141614023	225000000	2008	6.6
18	Joss Whedon	Action|Adventure|Sci-Fi	Chris Hemsworth	The AvengersÂ	995415	1722	703	English	623279547	220000000	2012	8.1
19	Rob Marshall	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: On Stranger TidesÂ	370704	484	448	English	241063875	250000000	2011	6.7
20	Barry Sonnenfeld	Action|Adventure|Comedy|Famil	Will Smith	Men in Black 3Â	268154	341	451	English	179020854	225000000	2012	6.8
21	Peter Jackson	Adventure|Fantasy	Aidan Turner	The Hobbit: The Battle of the Five ArmiesÂ	354228	802	422	English	255108370	250000000	2014	7.5
22	Marc Webb	Action|Adventure|Fantasy	Emma Stone	The Amazing Spider-ManÂ	451803	1225	599	English	262030663	230000000	2012	7
23	Ridley Scott	Action|Adventure|Drama|History	Mark Addy	Robin HoodÂ	211765	546	343	English	105219735	200000000	2010	6.7
24	Peter Jackson	Adventure|Fantasy	Aidan Turner	The Hobbit: The Desolation of SmaugÂ	483540	951	509	English	258355354	225000000	2013	7.9
25	Chris Weitz	Adventure|Family|Fantasy	Christopher Lee	The Golden CompassÂ	149019	666	251	English	70083519	180000000	2007	6.1
26	Peter Jackson	Action|Adventure|Drama|Roman	Naomi Watts	King KongÂ	316018	2618	446	English	218051260	207000000	2005	7.2
27	James Cameron	Drama|Romance	Leonardo DiCaprio	TitanicÂ	793059	2528	315	English	658672302	200000000	1997	7.7
28	Anthony Russo	Action|Adventure|Sci-Fi	Robert Downey Jr.	Captain America: Civil WarÂ	272670	1022	516	English	407197282	250000000	2016	8.2
29	Peter Berg	Action|Adventure|Sci-Fi|Thriller	Liam Neeson	BattleshipÂ	202382	751	377	English	65173160	209000000	2012	5.9

A B C1 C2 D E F1 F2 F3

B. Movies with highest profit: Find the movies with the highest profit?

Result:

1. I created a new column ‘Profit’ which is the difference between the Gross and the Budget of the movie
2. Sorted the columns using the ‘Profit’ column as Reference from largest to smallest.
3. Plotted the XY Scatter Chart for observing the Outliers.

4. Outliers: -12,21,55,00,000

   -4,20,00,00,000

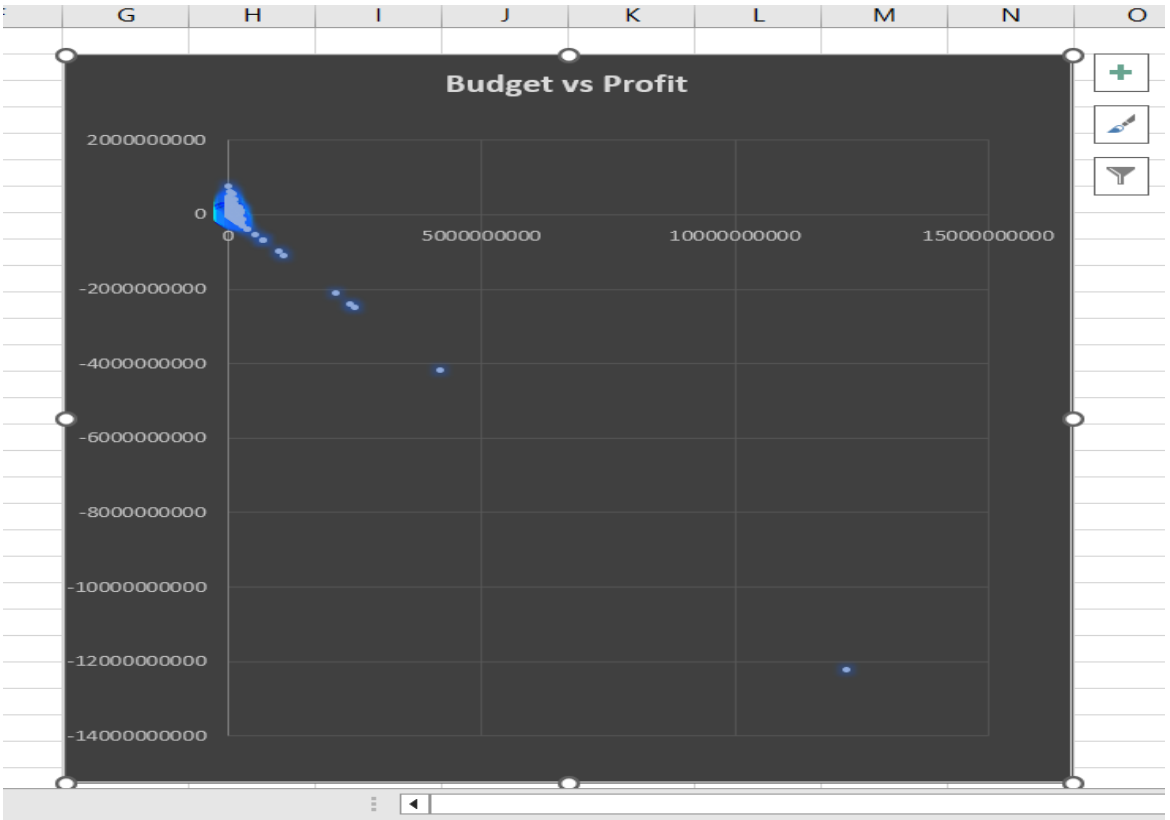
   -2,50,00,00,000

   -2,40,00,00,000

   -2,12,75,19,898

5. The movies with highest profit are Avatar,
   Jurassic World ,
   Titanic,
   Star Wars: Episode I - ThePhantomMenace
   and so on.

	A	B	C	D
1	movie_title	gross	budget	profit
2	AvatarÂ	760505847	237000000	523505847
3	Jurassic WorldÂ	652177271	150000000	502177271
4	TitanicÂ	658672302	200000000	458672302
5	Star Wars: Episode IV - A New HopeÂ	460935665	11000000	449935665
6	E.T. the Extra-TerrestrialÂ	434949459	10500000	424449459
7	The AvengersÂ	623279547	220000000	403279547
8	The Lion KingÂ	422783777	45000000	377783777
9	Star Wars: Episode I - The Phantom MenaceÂ	474544677	115000000	359544677
10	The Dark KnightÂ	533316061	185000000	348316061
11	The Hunger GamesÂ	407999255	78000000	329999255
12	DeadpoolÂ	363024263	58000000	305024263
13	The Hunger Games: Catching FireÂ	424645577	130000000	294645577
14	Jurassic ParkÂ	356784000	63000000	293784000
15	Despicable Me 2Â	368049635	76000000	292049635
16	American SniperÂ	350123553	58800000	291323553
17	Finding NemoÂ	380838870	94000000	286838870
18	Shrek 2Â	436471036	150000000	286471036
19	The Lord of the Rings: The Return of the KingÂ	377019252	94000000	283019252
20	Star Wars: Episode VI - Return of the JediÂ	309125409	32500000	276625409
21	Forrest GumpÂ	329691196	55000000	274691196
22	Star Wars: Episode V - The Empire Strikes Back	290158751	18000000	272158751
23	Home AloneÂ	285761243	18000000	267761243
24	Star Wars: Episode III - Revenge of the SithÂ	380262555	113000000	267262555
25	Spider-ManÂ	403706375	139000000	264706375
26	MinionsÂ	336029560	74000000	262029560
27	The Sixth SenseÂ	293501675	40000000	253501675
28	JawsÂ	260000000	8000000	252000000
29	FrozenÂ	400736600	150000000	250736600



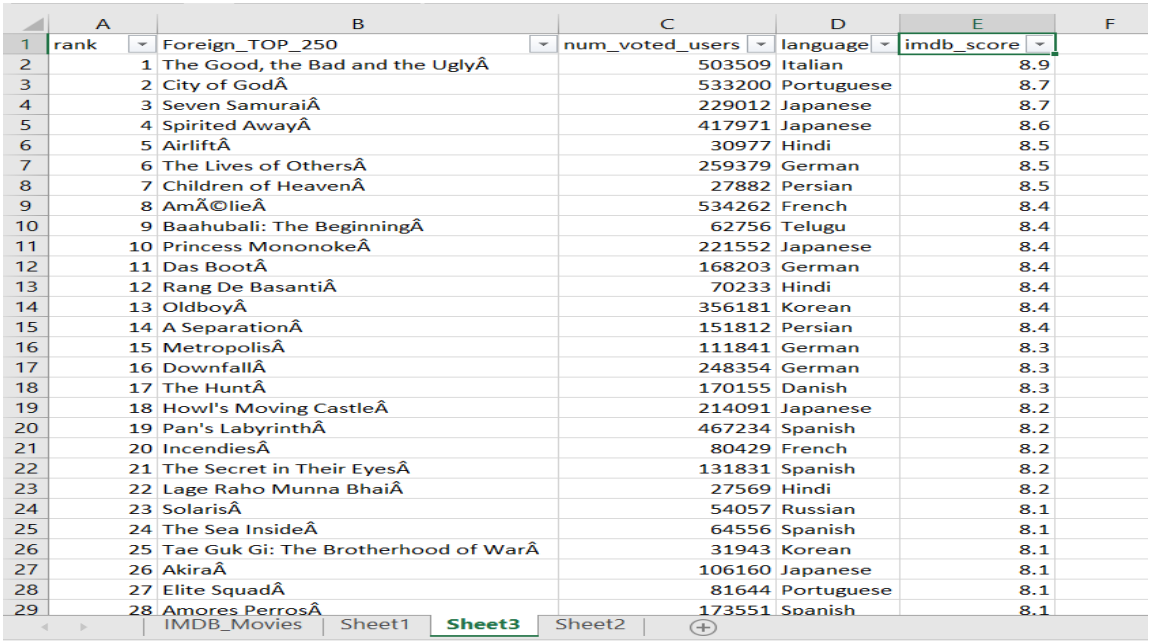
C. Top 250: Find IMDB Top 250

Result:

1. Filtered the num_voted_users greater than the value ’25,000’
2. Sorted the IMDB_score from Largest to Smallest.
3. Added a Column Rank using
   =RANK(E2,$E$2:$E$2735,0)+COUNTIF($E$2:E2,E2)-1
4. Filtered the movies of english language and pasted top 250 movies them in a new sheet
5. Similarly filtered movies of Foreign languages and pasted top 250 in a new sheet.

rank	IMDB_TOP_250	num_voted_users	language	imdb_score
1	The Shawshank RedemptionÂ	1689764	English	9.3
2	The GodfatherÂ	1155770	English	9.2
3	The Dark KnightÂ	1676169	English	9
4	The Godfather: Part IIÂ	790926	English	9
5	FargoÂ	170055	English	9
6	The Lord of the Rings: The Return of the KingÂ	1215718	English	8.9
7	Schindler's ListÂ	865020	English	8.9
8	Pulp FictionÂ	1324680	English	8.9
9	12 Angry MenÂ	447785	English	8.9
10	InceptionÂ	1468200	English	8.8
11	The Lord of the Rings: The Fellowship of the RingÂ	1238746	English	8.8
12	DaredevilÂ	213483	English	8.8
13	Fight ClubÂ	1347461	English	8.8
14	Forrest GumpÂ	1251222	English	8.8
15	It's Always Sunny in PhiladelphiaÂ	133415	English	8.8
16	Star Wars: Episode V - The Empire Strikes BackÂ	837759	English	8.8
17	The Lord of the Rings: The Two TowersÂ	1100446	English	8.7
18	The MatrixÂ	1217752	English	8.7
19	Friday Night LightsÂ	42746	English	8.7
20	GoodfellasÂ	728685	English	8.7
21	Star Wars: Episode IV - A New HopeÂ	911097	English	8.7
22	One Flew Over the Cuckoo's NestÂ	680041	English	8.7
23	InterstellarÂ	928227	English	8.6
24	HannibalÂ	159910	English	8.6
25	Saving Private RyanÂ	881236	English	8.6
26	LutherÂ	70568	English	8.6
27	Se7enÂ	1023511	English	8.6
28	The Silence of the LambsÂ	887467	English	8.6

IMDB_Movies | Sheet1 | Sheet3 | **Sheet2**



	A	B	C	D	E	F
1	rank	Foreign_TOP_250	num_voted_users	language	imdb_score	
2	1	The Good, the Bad and the UglyÂ	503509	Italian	8.9	
3	2	City of GodÂ	533200	Portuguese	8.7	
4	3	Seven SamuraiÂ	229012	Japanese	8.7	
5	4	Spirited AwayÂ	417971	Japanese	8.6	
6	5	AirliftÂ	30977	Hindi	8.5	
7	6	The Lives of OthersÂ	259379	German	8.5	
8	7	Children of HeavenÂ	27882	Persian	8.5	
9	8	AmÃ©lieÂ	534262	French	8.4	
10	9	Baahubali: The BeginningÂ	62756	Telugu	8.4	
11	10	Princess MononokeÂ	221552	Japanese	8.4	
12	11	Das BootÂ	168203	German	8.4	
13	12	Rang De BasantiÂ	70233	Hindi	8.4	
14	13	OldboyÂ	356181	Korean	8.4	
15	14	A SeparationÂ	151812	Persian	8.4	
16	15	MetropolisÂ	111841	German	8.3	
17	16	DownfallÂ	248354	German	8.3	
18	17	The HuntÂ	170155	Danish	8.3	
19	18	Howl's Moving CastleÂ	214091	Japanese	8.2	
20	19	Pan's LabyrinthÂ	467234	Spanish	8.2	
21	20	IncendiesÂ	80429	French	8.2	
22	21	The Secret in Their EyesÂ	131831	Spanish	8.2	
23	22	Lage Raho Munna BhaiÂ	27569	Hindi	8.2	
24	23	SolarisÂ	54057	Russian	8.1	
25	24	The Sea InsideÂ	64556	Spanish	8.1	
26	25	Tae Guk Gi: The Brotherhood of WarÂ	31943	Korean	8.1	
27	26	AkiraÂ	106160	Japanese	8.1	
28	27	Elite SquadÂ	81644	Portuguese	8.1	
29	28	Amores PerrosÂ	173551	Spanish	8.1	

IMDB_Movies | Sheet1 | **Sheet3** | Sheet2

D. Best Directors: Find the best directors?

Result:

1. Created a Pivot table from the Cleaned dataset
2. Took the director's name into rows section and took average of imdb score into values section.
3. Sorted the director's name in ascending order(A to Z)
4. Sorted the Average of imdb score from Largest to Smallest
5. Selected the Top 10 directors and pasted in a new sheet

	A	B
1	Task4: Top 10 directors	
2	Top 10 Directors	Average of imdb_score
3	Frank Darabont	9.3
4	Francis Ford Coppola	9.1
5	Christopher Nolan	8.9
6	Quentin Tarantino	8.9
7	Sergio Leone	8.9
8	Steven Spielberg	8.9
9	Peter Jackson	8.85
10	David Fincher	8.8
11	Irvin Kershner	8.8
12	Robert Zemeckis	8.8

E. **Popular Genres:** Find popular genres?

Result:

1. Selected the genres column from the cleaned dataset .
2. Created a Pivot Table.
3. Took Genres into rows section and took count of genre in values section.
4. Sorted the count of genres in Descending order
5. Sorted the genres in ascending order
6. Selected the Top 10 Genres and pasted in a new sheet.

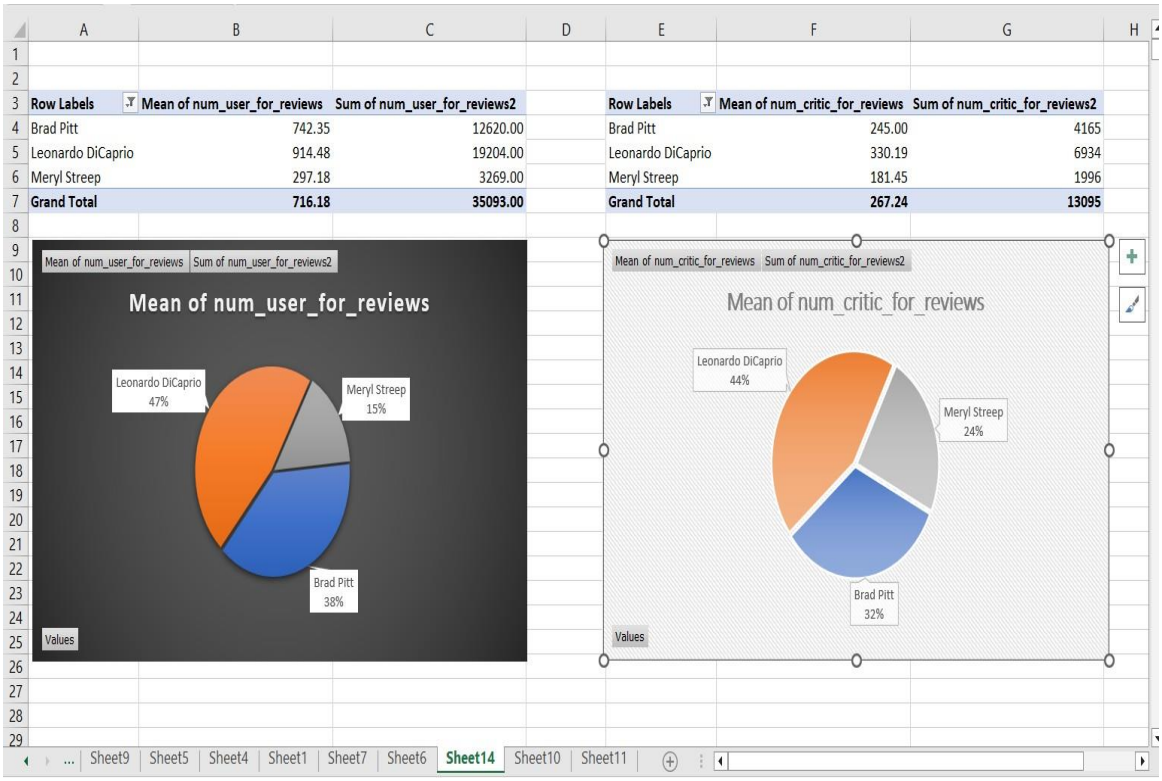
	A	B
1	GENRES	Count of genres
2	Comedy|Drama|Romance	147
3	Comedy	143
4	Drama	143
5	Comedy|Drama	138
6	Comedy|Romance	132
7	Drama|Romance	116
8	Crime|Drama|Thriller	81
9	Action|Crime|Thriller	56
10	Action|Crime|Drama|Thriller	50
11	Action|Adventure|Sci-Fi	46
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		

IMDB_Movies | Sheet1 | Sheet5 | Sheet4 | Sheet7 | **Sheet6**

F. Charts: Find the critic-favorite and audience-favorite actors?

Result:

1. Created 3 new columns namely: Meryl streep, Leonardo di caprio, Brad pitt which contains the movies they are the lead actors which is taken from the column actor_1_name.
2. Then, Appended all these movies in a new column named ‘Combined’ which are grouped by the actor names.
3. Created a Pivot table from the dataset that was cleaned earlier in task1
4. Took actor_1_name into rows section and mean of num_user_for_reviews into values section and plotted a pie chart for better insights.
5. Similarly, Took actor_1_name into rows section and mean of num_critic_for_reviews into values section and plotted a pie chart for better insights.

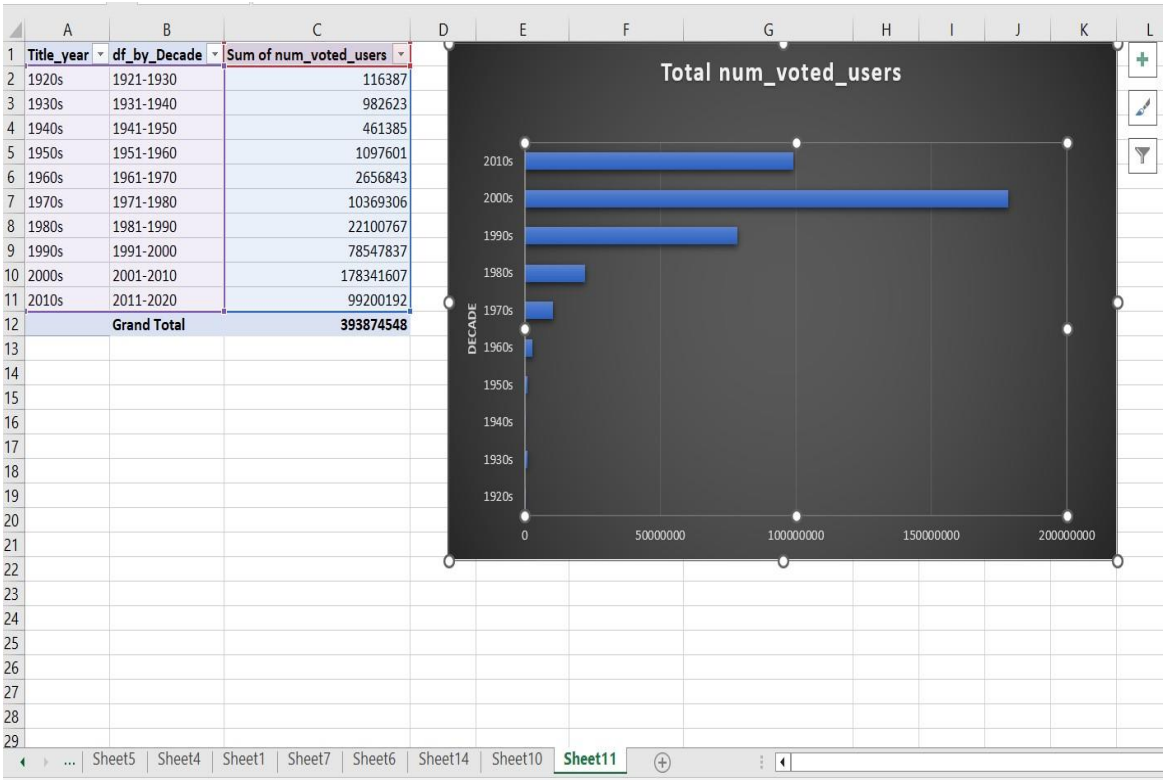


Row Labels	Mean of num_user_for_reviews	Sum of num_user_for_reviews2
Brad Pitt	742.35	12620.00
Leonardo DiCaprio	914.48	19204.00
Meryl Streep	297.18	3269.00
Grand Total	**716.18**	**35093.00**

Row Labels	Mean of num_critic_for_reviews	Sum of num_critic_for_reviews2
Brad Pitt	245.00	4165
Leonardo DiCaprio	330.19	6934
Meryl Streep	181.45	1996
Grand Total	**267.24**	**13095**

Change in num_of_voted_users over decades:

1. Selected the Required data from the cleaned dataset and created a pivot table.
2. Took the Title_year into the rows and sum of num_of_voted_users into values section.
3. Grouped the title_year by decades and inserted the data into df_by_decade column
4. Plotted a Bar chart showing number of voted users over different decades for better insights.



Title_year	df_by_Decade	Sum of num_voted_users
1920s	1921-1930	116387
1930s	1931-1940	982623
1940s	1941-1950	461385
1950s	1951-1960	1097601
1960s	1961-1970	2656843
1970s	1971-1980	10369306
1980s	1981-1990	22100767
1990s	1991-2000	78547837
2000s	2001-2010	178341607
2010s	2011-2020	99200192
	Grand Total	393874548

MY EXCEL FILE:
https://docs.google.com/spreadsheets/d/11_MjffhjrE78tgqgvCJoKlI5cdMjZR33/edit?usp=share_link&ouid=104755012826368900391&rtpof=true&sd=true

Thank You