

IMDB Movie Analysis

Project Description:

In this project we were provided with a CSV dataset having various columns of different IMDB movies. We are required to Frame the problem. For this task, we need to define a problem we want to shed some light on like top 250 movie, top director, highest grossing movies, top actors, etc. Once we have defined a problem, clean the data as necessary, and use Data Analysis skills to explore the data set and derive insights.

Approach:

My approach was first to study the dataset to better understand it then I cleaned the data set by removing blank spaces after that I used my excel knowledge which I learned in course video to draw out insights from the dataset to answer the questions.

Tech-Stack Used:

I have used MS Excel 2016 for data analysis and MS word to make the report.

Excel Dataset:

<https://docs.google.com/spreadsheets/d/1tGN12pveejGIFPs5zj8aDHxHICtjQMm/edit?usp=sharing&ouid=115977276322304781744&rtpof=true&sd=true>

Insights:

In this project I learned several new excel function and I also learned how to clean the data using whereas shortcut techniques i.e. Find & Select option. I also learned how to create table using charts to analyze the something which in my point very useful thing and save a lot of time.

Results:

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.)

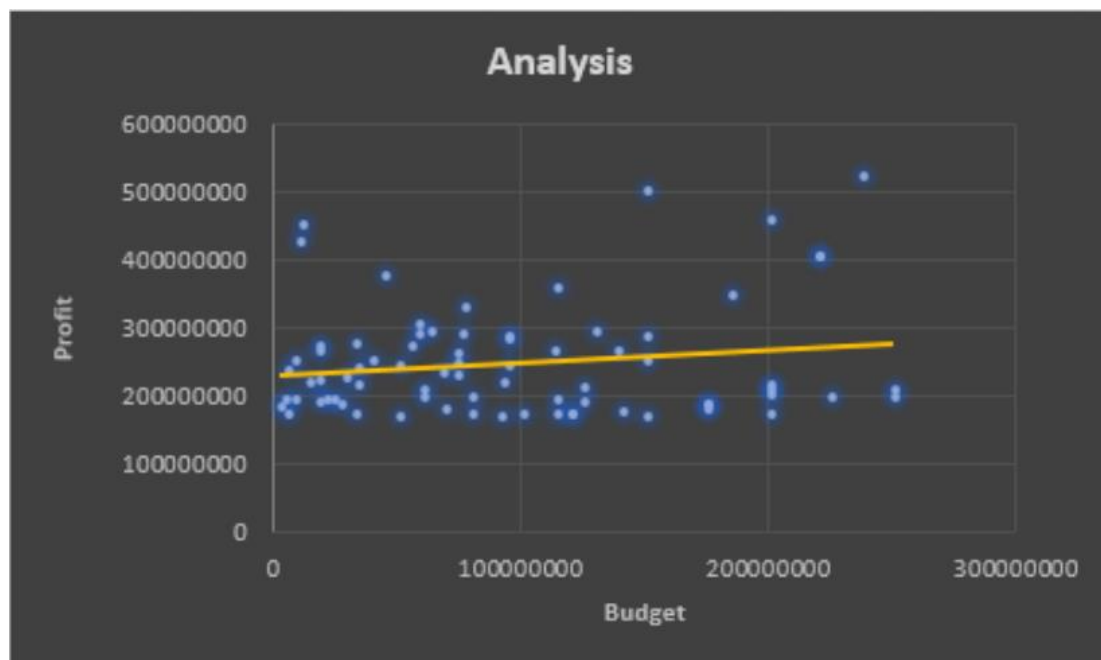
Insights: I clean the dataset by delete all the blank space using Find & Select Option, it reduce my dataset from 5044 rows to 3757 rows.

	A	B	C	D	E	F	G	H	I	J	K	
1	color	director_name	num_critc_for_rev	duration	director_facebook	actor_3_facebook	actor_2_name	actor_1_face	gross	genres	actor_1_name	movie_title
2	Color	James Cameron	723	178	0	855	Joel David Moore	1000	760505847	Action/Adventure/Fantasy/Sci-Fi	CCH Pounder	Avatar
3	Color	Colin Trevorrow	644	124	365	1000	Judy Greer	3000	652177271	Action/Adventure/Sci-Fi/Thriller	Bryce Dallas Howard	Jurassic World
4	Color	James Cameron	315	194	0	794	Kate Winslet	29000	658672302	Drama/Romance	Leonardo DiCaprio	Titanic
5	Color	George Lucas	282	125	0	504	Peter Cushing	11000	460935665	Action/Adventure/Fantasy/Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope
6	Color	Steven Spielberg	215	120	14000	548	Dee Wallace	861	434949459	Family/Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial
7	Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	26000	623279547	Action/Adventure/Sci-Fi	Chris Hemsworth	The Avengers
8	Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	26000	623279547	Action/Adventure/Sci-Fi	Chris Hemsworth	The Avengers
9	Color	Roger Allers	186	73	28	847	Nathan Lane	2000	422783777	Adventure/Animation/Drama/Family/Musical	Matthew Broderick	The Lion King
10	Color	George Lucas	320	136	0	1000	Liam Neeson	20000	474544677	Action/Adventure/Fantasy/Sci-Fi	Natalie Portman	Star Wars: Episode I - The Phantom Menace
11	Color	Christopher Nolan	645	152	22000	11000	Heath Ledger	23000	533316561	Action/Crime/Drama/Thriller	Christian Bale	The Dark Knight
12	Color	Gary Ross	673	142	378	575	Josh Hutcherson	34000	407992255	Adventure/Drama/Sci-Fi/Thriller	Jennifer Lawrence	The Hunger Games
13	Color	Tim Miller	579	108	84	361	Ed Skrein	16000	363024263	Action/Adventure/Comedy/Romance/Sci-Fi	Ryan Reynolds	Deadpool
14	Color	Francis Lawrence	502	146	508	523	Josh Hutcherson	34000	424645577	Adventure/Sci-Fi/Thriller	Jennifer Lawrence	The Hunger Games: Catching Fire
15	Color	Steven Spielberg	308	127	14000	191	Ariana Richards	967	356784000	Adventure/Sci-Fi/Thriller	Wayne Knight	Jurassic Park
16	Color	Pierre Coffin	306	98	275	1000	Miranda Cosgrove	7000	368049635	Animation/Comedy/Family/Sci-Fi	Steve Carell	Despicable Me 2
17	Color	Clint Eastwood	490	133	16000	318	Leonard Roberts	14000	350123553	Action/Biography/Drama/History/Thriller/War	Bradley Cooper	American Sniper
18	Color	Andrew Stanton	301	100	475	799	Stephen Root	1000	380838870	Adventure/Animation/Comedy/Family	Alexander Gould	Finding Nemo
19	Color	Andrew Adamson	205	93	80	48	Jennifer Saunders	692	436471036	Adventure/Animation/Comedy/Family/Fantasy/Romance	Rupert Everett	Shrek 2
20	Color	Peter Jackson	328	192	0	416	Billy Boyd	5000	377019252	Action/Adventure/Drama/Fantasy	Orlando Bloom	The Lord of the Rings: The Two Towers
21	Color	Richard Marquand	197	134	37	504	Ian McDiarmid	11000	309125409	Action/Adventure/Fantasy/Sci-Fi	Harrison Ford	Star Wars: Episode VI - The Return of the Jedi
22	Black and White	Robert Zemeckis	149	142	0	194	Siobhan Fallon Hogan	15000	329691196	Comedy/Drama	Tom Hanks	Forrest Gump
23	Color	Irvin Kershner	223	127	883	441	Kenny Baker	11000	290158751	Action/Adventure/Fantasy/Sci-Fi	Harrison Ford	Star Wars: Episode V - The Empire Strikes Back
24	Color	Chris Columbus	102	103	0	925	Devin Ratray	3000	285761243	Comedy/Family	Macaulay Culkin	Home Alone
25	Color	George Lucas	359	140	0	4000	Christopher Lee	20000	380262555	Action/Adventure/Fantasy/Sci-Fi	Natalie Portman	Star Wars: Episode III - Revenge of the Sith
26	Color	Sam Raimi	291	121	0	4000	James Franco	24000	403706375	Action/Adventure/Fantasy/Romance	J.K. Simmons	Spider-Man
27	Color	Kyle Balda	308	91	22	1000	Jon Hamm	7000	386029560	Action/Animation/Comedy/Family/Sci-Fi	Steve Carell	Minions
28	Color	M. Night Shyamalan	234	107	0	766	Haley Joel Osment	13000	293501675	Drama/Mystery/Thriller	Bruce Willis	The Sixth Sense
29	Color	Steven Spielberg	385	130	14000	366	Robert Shaw	813	260000000	Adventure/Drama/Thriller	Roy Scheider	Jaws
30	Color	Chris Buck	406	102	69	490	Maurice LaMarche	1000	400736600	Adventure/Animation/Comedy/Family/Fantasy/Musical	Josh Gad	Frozen
31	Color	Yarrow Cheney	165	87	11	745	Eric Stonestreet	1000	323505550	Animation/Comedy/Family	Steve Coogan	The Secret Life of Pets
32	Color	Chris Weitz	299	130	129	12000	Kristen Stewart	21000	296623634	Adventure/Drama/Fantasy/Romance	Robert Pattinson	The Twilight Saga: New Moon
33	Color	Peter Jackson	294	172	0	857	Orlando Bloom	16000	340478898	Action/Adventure/Drama/Fantasy	Christopher Lee	The Lord of the Rings: The Fellowship Ring
34	Color	Todd Phillips	334	108	480	706	Rob Riggle	14000	277313371	Comedy	Bradley Cooper	The Hangover
35	Color	Joel Zwick	168	95	38	249	Louis Mandylor	567	241437427	Comedy/Family/Romance	Nia Vardalos	My Big Fat Greek Wedding
36	Color	David Slade	293	124	171	10000	Kristen Stewart	21000	300523113	Adventure/Drama/Fantasy/Romance	Robert Pattinson	The Twilight Saga: Eclipse
37	Color	Roland Emmerich	185	154	776	933	Adam Baldwin	10000	306124059	Action/Adventure/Sci-Fi	Will Smith	Independence Day
38	Color	John Lee Hancock	261	129	102	624	Quinton Aaron	768	255950375	Biography/Drama/Sport	Catherine Dyer	The Blind Side
39	Color	Steven Spielberg	234	115	14000	488	Karen Allen	11000	242374454	Action/Adventure	Harrison Ford	Raiders of the Lost Ark

Fig: Cleaned Excel Data

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.

Insights:



	A	B	C	D	E
1	Rank	movie_title	gross	budget	Profit
2	1	Avatar	760505847	237000000	523505847
3	2	Jurassic World	652177271	150000000	502177271
4	3	Titanic	658672302	200000000	458672302
5	4	Star Wars: Episode IV - A New Hope	460935665	110000000	449935665
6	5	E.T. the Extra-Terrestrial	434949459	105000000	424449459
7	6	The Avengers	623279547	220000000	403279547
8	7	The Avengers	623279547	220000000	403279547
9	8	The Lion King	422783777	450000000	377783777
10	9	Star Wars: Episode I - The Phantom Menace	474544677	115000000	359544677
11	10	The Dark Knight	533316061	185000000	348316061
12	11	The Hunger Games	407999255	780000000	329999255
13	12	Deadpool	363024263	580000000	305024263
14	13	The Hunger Games: Catching Fire	424645577	130000000	294645577
15	14	Jurassic Park	356784000	630000000	293784000
16	15	Despicable Me 2	368049635	760000000	292049635
17	16	American Sniper	350123553	588000000	291323553
18	17	Finding Nemo	380838870	940000000	286838870
19	18	Shrek 2	436471036	150000000	286471036
20	19	The Lord of the Rings: The Return of the King	377019252	940000000	283019252
21	20	Star Wars: Episode VI - Return of the Jedi	309125409	325000000	276625409
22	21	Forrest Gump	329691196	550000000	274691196
23	22	Star Wars: Episode V - The Empire Strikes Back	290158751	180000000	272158751
24	23	Home Alone	285761243	180000000	267761243
25	24	Star Wars: Episode III - Revenge of the Sith	380262555	113000000	267262555
26	25	Spider-Man	403706375	139000000	264706375
27	26	Minions	336029560	740000000	262029560
28	27	The Sixth Sense	293501675	400000000	253501675
29	28	Jaws	260000000	80000000	252000000
30	29	Frozen	400736600	150000000	250736600
31	30	The Secret Life of Pets	323505540	750000000	248505540
32	31	The Twilight Saga: New Moon	296623634	500000000	246623634

Fig: Movies with Most Profit

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!

Insights:

	A	B	C	D
1	Ranking	movie_title	num_voted_users	imdb_score
2	1	The Shawshank Redemption	1689764	9.3
3	2	The Godfather	1155770	9.2
4	4	The Dark Knight	1676169	9
5	3	The Godfather: Part II	790926	9
6	8	Pulp Fiction	1324680	8.9
7	7	The Lord of the Rings: The Return of the King	1215718	8.9
8	6	Schindler's List	865020	8.9
9	5	The Good, the Bad and the Ugly	503509	8.9
10	13	Inception	1468200	8.8
11	12	Fight Club	1347461	8.8
12	11	Forrest Gump	1251222	8.8
13	10	The Lord of the Rings: The Fellowship of the Ring	1238746	8.8
14	9	Star Wars: Episode V - The Empire Strikes Back	837759	8.8
15	20	The Matrix	1217752	8.7
16	19	The Lord of the Rings: The Two Towers	1100446	8.7
17	18	Star Wars: Episode IV - A New Hope	911097	8.7
18	17	Goodfellas	728685	8.7
19	16	One Flew Over the Cuckoo's Nest	680041	8.7
20	15	City of God	533200	8.7
21	14	Seven Samurai	229012	8.7
22	28	Se7en	1023511	8.6
23	27	Interstellar	928227	8.6
24	26	The Silence of the Lambs	887467	8.6
25	25	Saving Private Ryan	881236	8.6
26	24	American History X	782437	8.6
27	23	The Usual Suspects	740918	8.6
28	22	Spirited Away	417971	8.6
29	21	Modern Times	143086	8.6
30	46	The Dark Knight Rises	1144337	8.5
31	45	Gladiator	982637	8.5
32	44	Django Unchained	955174	8.5

Fig: Top 250 movies based on IMDB Score

Top_Foreign_Lang_film	num_voted_users	imdb_score	language
The Good, the Bad and the Ugly	503509	8.9	Italian
Seven Samurai	229012	8.7	Japanese
City of God	533200	8.7	Portuguese
Spirited Away	417971	8.6	Japanese
Children of Heaven	27882	8.5	Persian
The Lives of Others	259379	8.5	German
Samsara	22457	8.5	None
A Separation	151812	8.4	Persian
Das Boot	168203	8.4	German
Princess Mononoke	221552	8.4	Japanese
Oldboy	356181	8.4	Korean
Amélie	534262	8.4	French
Metropolis	111841	8.3	German
The Hunt	170155	8.3	Danish
Downfall	248354	8.3	German
Incendies	80429	8.2	French
The Secret in Their Eyes	131831	8.2	Spanish
Howl's Moving Castle	214091	8.2	Japanese
Pan's Labyrinth	467234	8.2	Spanish
The Act of Killing	23836	8.2	Indonesian
Tae Guk Gi: The Brotherhood of War	31943	8.1	Korean
The Sea Inside	64556	8.1	Spanish
The Celebration	65951	8.1	Danish
Elite Squad	81644	8.1	Portuguese
Akira	106160	8.1	Japanese
Amores Perros	173551	8.1	Spanish
Central Station	28951	8	Portuguese
Waltz with Bashir	46107	8	Hebrew
My Name Is Khan	69759	8	Hindi
Persepolis	70194	8	French
A Fistful of Dollars	147566	8	Italian

Fig: Top Foreign Language Movie

D. 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.

Insights:

1	Director	Mean imdb_score
2	Akira Kurosawa	8.7
3	Charles Chaplin	8.6
4	Michael Curtiz	8.6
5	Tony Kaye	8.6
6	Damien Chazelle	8.5
7	Majid Majidi	8.5
8	Ron Fricke	8.5
9	Sergio Leone	8.433333333
10	Christopher Nolan	8.425
11	Asghar Farhadi	8.4

Fig: Top 10 Director Based on Mean IMDB Score

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

Insights:

1	Genre	count
2	Drama	1893
3	Comedy	1461
4	Thriller	1115
5	Action	959
6	Romance	858
7	Adventure	781
8	Crime	709
9	Fantasy	507
10	Sci-Fi	496
11	Family	442
12	Horror	392
13	Mystery	384
14	Biography	239
15	Animation	196
16	War	152
17	Music	151
18	History	149
19	Sport	148
20	Musical	96
21	Western	59
22	Documentary	45
23	Film-Noir	1

Fig: Top Genre Based on their count

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.

Insights:

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

Fig: Movie Done by Meryl Streep, Leonardo DiCaprio, Brad pitt

	A	B	C	D	E	F
1	Actor Name	Average of num_critic_for_review	Average of num_user_for_reviews			
2	Albert Finney	750	1498			
3	Phaldut Sharma	738	1885			
4	Peter Capaldi	654	995			
5	Craig Stark	596	1018			
6	Bérénice Bejo	576	583			
7	Suraj Sharma	552	755			
8	Ellar Coltrane	548	836		Critic Favourite	Heather Donahue
9	Mike Howard	546	405		Audience Favorite	Albert Finney
10	Lou Taylor Pucci	543	789			
11	Joel Courtney	539	849			
12	Maika Monroe	533	631			
13	Tim Holmes	525	511			
14	Elina Alminas	489	611			
15	Kurt Fuller	487	509			
16	Iko Uwais	481	316			
17	Quvenzhané Wallis	478.6666667	392.3333333			
18	Edgar Arreola	478	461			
19	Sharlto Copley	472	1262			
20	Cory Hardrict	452	326			
21	Matt Frewer	451	1229			
22	Elizabeth McGovern	447	801			
23	Isabelle Huppert	447	190			
24	Michael Fassbender	434.3333333	837.5			
25	Wood Harris	432	588			
26	Jennifer Lawrence	418.9285714	583.2142857			
27	Anil Kapoor	418	1017			
28	Jessica Barden	417	438			
29	Sally Hawkins	415	384			
30	Chris Hemsworth	411.7333333	702.3333333			
31	Danielle Kitch	411	524			
32	Micah Sloat	409	1189			

Fig: Critic Favorite and Audience Favorite

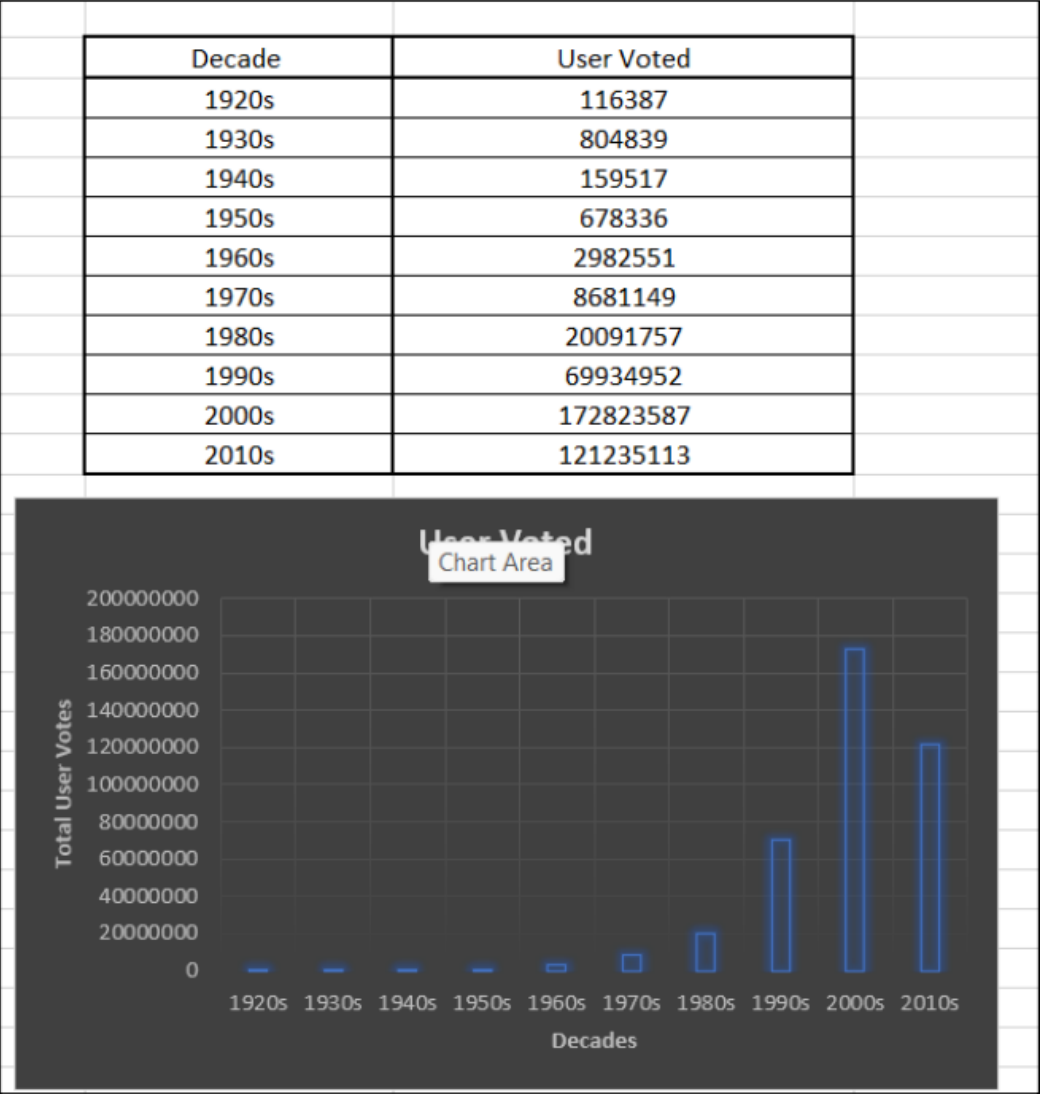


Fig: Number of voted User over Decade