IMDB Movie Analysis

Project Description:

Here we are given a dataset from IMDB, which is a movies dataset. In which we can find various details for a movie i.e title,is the movie color or black & white,director's name, duration,budget and gross collection of the movie. We also get information about the genres,plots,lead actors name and their facebook like counts, the language of the movie,the country in which the movie was released,the year in which the movie was released,IMDB rating of the movie,aspect ratio of the movie and many more details.

Our task is to analyse the given data and make business insights from it.

Approach:

Downloaded the excel sheet and analysed the data using excel. Cleaned the data. Then used various functions and pivot tables to solve the given problems. Analysed the data using the 5 whys approach. Made graphs and tables to represent the insights as needed.

Tech-Stack Used:

Microsoft Excel

Insights:

Below insights we could make like :

1. **Cleaning the data:** There were 5044 rows and 28 columns in the sheet. Many of the records were duplicated and in important rows null values were present. **I have dropped all the duplicate records** and also **dropped all he records where budget , gross values were not present** cause those fields are most important ones and if they are not having datas there is no meaning of keeping those.

   The Title of the movie was also having a special character in the end. I have cleared the special character from the end as it was of no use. i.e AvatarÂ → Avatar

   Replaced nulls with 0 in the all fields where null and 0 meant the same in our case like *num_critic_for_reviews, director_facebook_likes* **and etc.**

   At last I was left with 3857 records. I have not dropped any columns intentionally as I was not sure those will be needed in future or not.

2. **Movies with highest profit:** Made a new column as Profit by calculating gross-budget for each movies and then made a list of movies with highest profit.
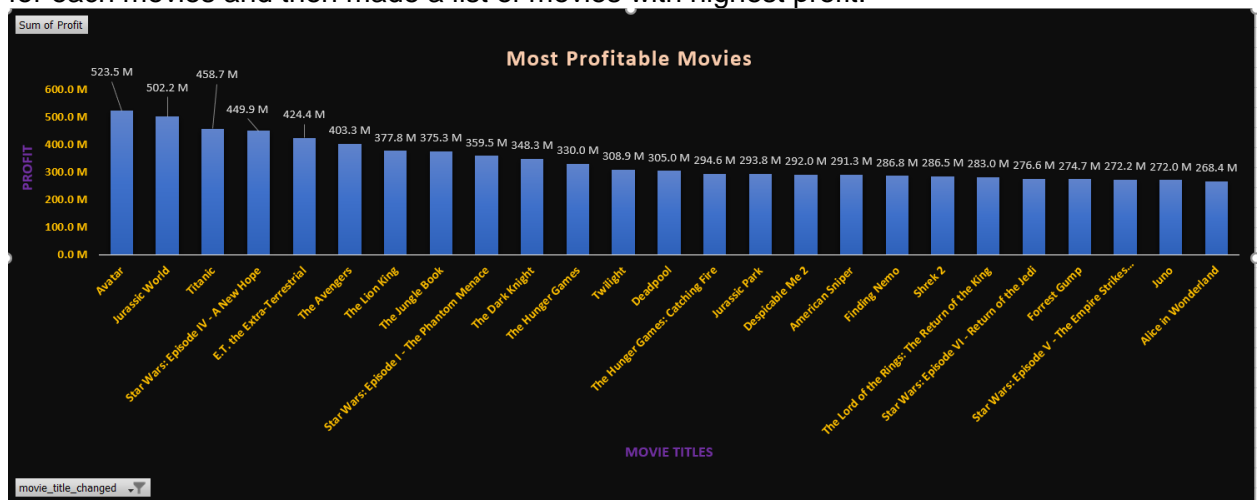
3. **Top 250 movies as per IMDB Ratings:** Sorted the dataset according to the IMDB Rating from largest to smallest and selected top 250. Also ranked them accordingly.

| movie_title_changed	num_voted_users	language	budget	gross	Profit	imdb_score	Rank
The Shawshank Redemption	1689764	English	25000000	28341469	3341469	9.3	1
The Godfather	1155770	English	6000000	134821952	128821952	9.2	2
The Dark Knight	1676169	English	185000000	533316061	348316061	9	3
The Godfather: Part II	790926	English	13000000	57300000	44300000	9	3
The Lord of the Rings: The Return of the King	1215718	English	94000000	377019252	283019252	8.9	5
Schindler's List	865020	English	22000000	96067179	74067179	8.9	5
Pulp Fiction	1324680	English	8000000	107930000	99930000	8.9	5
The Good, the Bad and the Ugly	503509	Italian	1200000	6100000	4900000	8.9	5
Inception	1468200	English	160000000	292568851	132568851	8.8	9
The Lord of the Rings: The Fellowship of the Ring	1238746	English	93000000	313837577	220837577	8.8	9
Fight Club	1347461	English	63000000	37023395	-25976605	8.8	9
Star Wars: Episode V - The Empire Strikes Back	837759	English	18000000	290158751	272158751	8.8	9
Forrest Gump	1251222	English	55000000	329691196	274691196	8.8	9
Star Wars: Episode IV - A New Hope	911097	English	11000000	460935665	449935665	8.7	14
The Lord of the Rings: The Two Towers	1100446	English	94000000	340478898	246478898	8.7	14
The Matrix	1217752	English	63000000	171383253	108383253	8.7	14
Goodfellas	728685	English	25000000	46836394	21836394	8.7	14
One Flew Over the Cuckoo's Nest	680041	English	4400000	112000000	107600000	8.7	14
City of God	533200	Portuguese	3300000	7563397	4263397	8.7	14
Seven Samurai	229012	Japanese	2000000	269061	-1730939	8.7	14
Interstellar	928227	English	165000000	187991439	22991439	8.6	21
Saving Private Ryan	881236	English	70000000	216119491	146119491	8.6	21
Se7en	1023511	English	33000000	100125340	67125340	8.6	21
The Silence of the Lambs	887467	English	19000000	130727000	111727000	8.6	21
Spirited Away	417971	Japanese	19000000	10049886	-8950114	8.6	21
American History X	782437	English	7500000	6712241	-787759	8.6	21
The Usual Suspects	740918	English	6000000	23272306	17272306	8.6	21
Modern Times	143086	English	1500000	163245	-1336755	8.6	21

Top_250_Movies.xlsx

Created a new table from the above table for foreign language films in which English is not there.

| movie_title_changed	num_voted_users	language	budget	gross	Profit	imdb_score	Rank
The Good, the Bad and the Ugly	503509	Italian	1200000	6100000	4900000	8.9	1
City of God	533200	Portuguese	3300000	7563397	4263397	8.7	2
Seven Samurai	229012	Japanese	2000000	269061	-1730939	8.7	2
Spirited Away	417971	Japanese	19000000	10049886	-8950114	8.6	4
The Lives of Others	259379	German	2000000	11284657	9284657	8.5	5
Children of Heaven	27882	Persian	180000	925402	745402	8.5	5
AmÃ©lie	534262	French	77000000	33201661	-43798339	8.4	7
Baahubali: The Beginning	62756	Telugu	18026148	6498000	-11528148	8.4	7
Princess Mononoke	221552	Japanese	2400000000	2298191	-2397701809	8.4	7
Das Boot	168203	German	14000000	11433134	-2566866	8.4	7
Oldboy	356181	Korean	3000000	2181290	-818710	8.4	7
A Separation	151812	Persian	500000	7098492	6598492	8.4	7
Metropolis	111841	German	6000000	26435	-5973565	8.3	13
Downfall	248354	German	13500000	5501940	-7998060	8.3	13
The Hunt	170155	Danish	3800000	610968	-3189032	8.3	13
Howl's Moving Castle	214091	Japanese	24000000	4710455	-19289545	8.2	16
Pan's Labyrinth	467234	Spanish	13500000	37623143	24123143	8.2	16
Incendies	80429	French	6800000	6857096	57096	8.2	16
The Secret in Their Eyes	131831	Spanish	2000000	20167424	18167424	8.2	16
The Sea Inside	64556	Spanish	10000000	2086345	-7913655	8.1	20
Tae Guk Gi: The Brotherhood of War	31943	Korean	12800000	1110186	-11689814	8.1	20
Akira	106160	Japanese	1100000000	439162	-1099560838	8.1	20
Elite Squad	81644	Portuguese	4000000	8060	-3991940	8.1	20
Amores Perros	173551	Spanish	2000000	5383834	3383834	8.1	20
The Celebration	65951	Danish	1300000	1647780	347780	8.1	20
My Name Is Khan	69759	Hindi	12000000	4018695	-7981305	8	26
Persepolis	70194	French	7300000	4443403	-2856597	8	26
Central Station	28951	Portuguese	2900000	5595428	2695428	8	26

Top_Foreign_Langua
ge_Films.xlsx

4. **Best Directors:** Using pivot table grouped all the movies directed by the director and found the mean of it. Then sorted them from largest to smallest and selected top 10. Sorted them again according to the directors name alphabetically if the mean value is same.
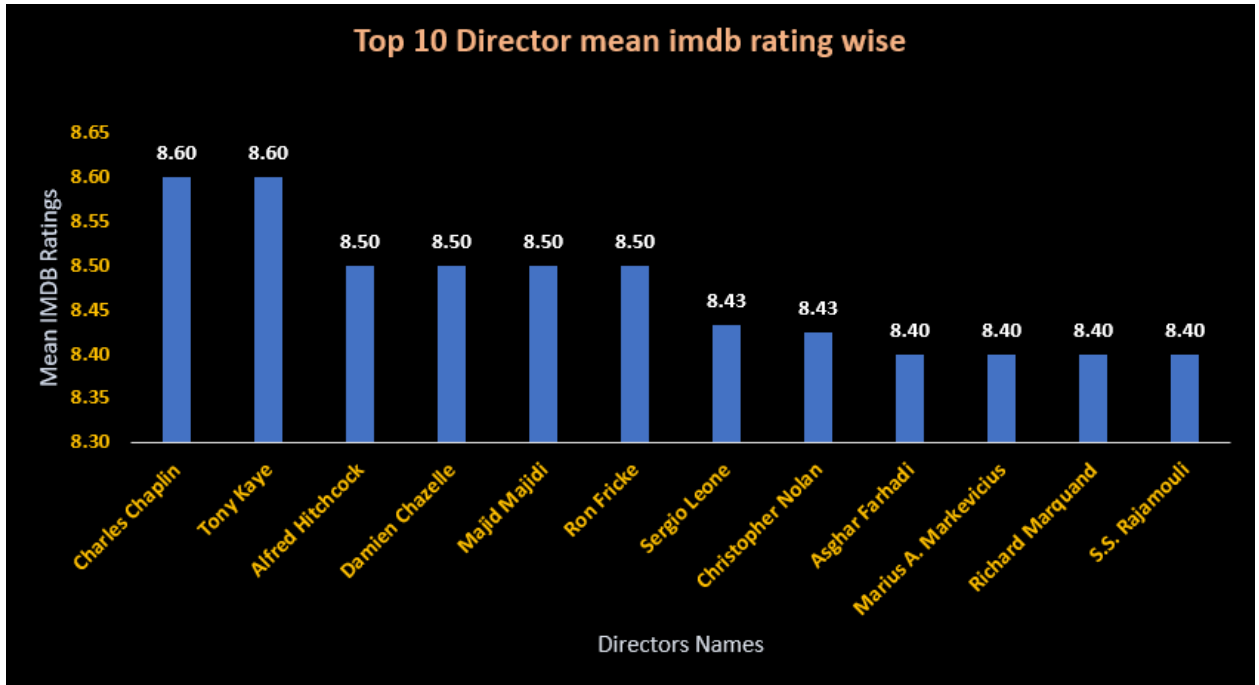
5. **Popular Genres:** Using pivot table found the top 25 favourite genres. Using the filter in the image we can find as many a we need like top 10/20 or any number. I have selected Top 25 here.
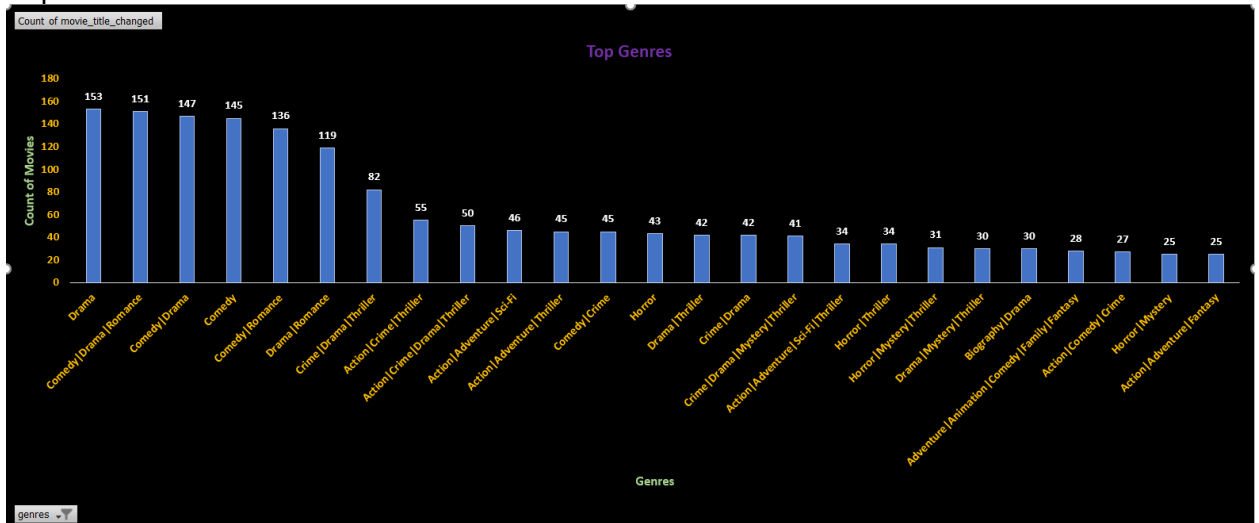
6. **Number of Movies in which Maryl Streep, Leonardo DiCaprio, Brad Pitt acted in the lead role i.e Actor 1.**
Inserted 3 new columns as per actors name. Also Plotted a graph counting the movies

gross	genres	actor_1_name	Meryl Streep	Leonardo DiCap	Brad Pi	movie_title_changed	num_voted_us
658672302	Drama|Romance	Leonardo DiCaprio	No	Yes	No	Titanic	793059
144812796	Drama|Romance	Leonardo DiCaprio	No	Yes	No	The Great Gatsby	362912
292568851	Action|Adventure|Sci-Fi|Thriller	Leonardo DiCaprio	No	Yes	No	Inception	1468200
102608827	Biography|Drama	Leonardo DiCaprio	No	Yes	No	The Aviator	264318
162804648	Drama|Western	Leonardo DiCaprio	No	Yes	No	Django Unchained	955174
132373442	Crime|Drama|Thriller	Leonardo DiCaprio	No	Yes	No	The Departed	873649
77679638	Crime|Drama	Leonardo DiCaprio	No	Yes	No	Gangs of New York	314033
39380442	Action|Drama|Thriller	Leonardo DiCaprio	No	Yes	No	Body of Lies	174248
39778599	Adventure|Drama|Thriller	Leonardo DiCaprio	No	Yes	No	The Beach	176169
22877808	Drama|Romance	Leonardo DiCaprio	No	Yes	No	Revolutionary Road	152591
56876365	Action|Adventure	Leonardo DiCaprio	No	Yes	No	The Man in the Iron Mask	125219
37304950	Biography|Crime|Drama	Leonardo DiCaprio	No	Yes	No	J. Edgar	102728
18636537	Action|Thriller|Western	Leonardo DiCaprio	No	Yes	No	The Quick and the Dead	69197
116866727	Biography|Comedy|Crime|Drama	Leonardo DiCaprio	No	Yes	No	The Wolf of Wall Street	780588
183635922	Adventure|Drama|Thriller|Western	Leonardo DiCaprio	No	Yes	No	The Revenant	406020
127968405	Mystery|Thriller	Leonardo DiCaprio	No	Yes	No	Shutter Island	786092
12782508	Drama	Leonardo DiCaprio	No	Yes	No	Marvin's Room	20163
46338728	Drama|Romance	Leonardo DiCaprio	No	Yes	No	Romeo + Juliet	167750
57366262	Adventure|Drama|Thriller	Leonardo DiCaprio	No	Yes	No	Blood Diamond	400292
164435221	Biography|Crime|Drama	Leonardo DiCaprio	No	Yes	No	Catch Me If You Can	525801
144812796	Drama|Romance	Leonardo DiCaprio	No	Yes	No	The Great Gatsby	362933

gross	genres	actor_1_name	Meryl Streep	Leonardo DiCap	Brad Pi	movie_title_changed	num_voted_us
37023395	Drama	Brad Pitt	No	No	Yes	Fight Club	1347461
127490802	Drama|Fantasy|Romance	Brad Pitt	No	No	Yes	The Curious Case of Benjamin Button	459346
133228348	Adventure	Brad Pitt	No	No	Yes	Troy	381672
125531634	Crime|Thriller	Brad Pitt	No	No	Yes	Ocean's Twelve	284852
186336103	Action|Comedy|Crime|Romance|Thriller	Brad Pitt	No	No	Yes	Mr. & Mrs. Smith	348861
26871	Action|Crime|Thriller	Brad Pitt	No	No	Yes	Spy Game	121259
183405771	Crime|Thriller	Brad Pitt	No	No	Yes	Ocean's Eleven	402645
85707116	Action|Drama|War	Brad Pitt	No	No	Yes	Fury	303185
37901509	Adventure|Biography|Drama|History|War	Brad Pitt	No	No	Yes	Seven Years in Tibet	96385
26288320	Adventure|Animation|Comedy|Drama|Family|Fantasy|Romance	Brad Pitt	No	No	Yes	Sinbad: Legend of the Seven Seas	36144
105264608	Drama|Fantasy|Horror	Brad Pitt	No	No	Yes	Interview with the Vampire: The Vampire Chronicles	239752
13303319	Drama|Fantasy	Brad Pitt	No	No	Yes	The Tree of Life	136367
3904982	Biography|Crime|Drama|History|Western	Brad Pitt	No	No	Yes	The Assassination of Jesse James by the Coward Robert Ford	136104
34300771	Drama	Brad Pitt	No	No	Yes	Babel	243799
531009	Drama|Romance	Brad Pitt	No	No	Yes	By the Sea	7976
14938570	Crime|Thriller	Brad Pitt	No	No	Yes	Killing Them Softly	111625
12281500	Action|Crime|Drama|Romance|Thriller	Brad Pitt	No	No	Yes	True Romance	163492

gross	genres	actor_1_name	Meryl Streep	Leonardo DiCap	Brad Pi	movie_title_changed	num_voted_us
112703470	Comedy|Drama|Romance	Meryl Streep	Yes	No	No	It's Complicated	69860
46815748	Action|Adventure|Crime|Thriller	Meryl Streep	Yes	No	No	The River Wild	32544
94125426	Biography|Drama|Romance	Meryl Streep	Yes	No	No	Julie & Julia	79264
124732962	Comedy|Drama|Romance	Meryl Streep	Yes	No	No	The Devil Wears Prada	286178
14998070	Drama|Thriller|War	Meryl Streep	Yes	No	No	Lions for Lambs	41170
87100000	Biography|Drama|Romance	Meryl Streep	Yes	No	No	Out of Africa	52339
63536011	Comedy|Drama|Romance	Meryl Streep	Yes	No	No	Hope Springs	34258
23209440	Drama	Meryl Streep	Yes	No	No	One True Thing	9283
41597830	Drama|Romance	Meryl Streep	Yes	No	No	The Hours	102123
29959436	Biography|Drama|History	Meryl Streep	Yes	No	No	The Iron Lady	82327
20338609	Comedy|Drama|Music	Meryl Streep	Yes	No	No	A Prairie Home Companion	19655

| Actor's Name	Count
Meryl Streep	11
Leonardo DiCaprio	21
Brad Pitt	17
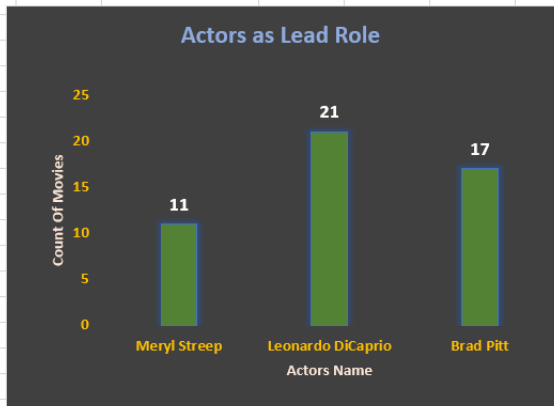
7. Top Actors by Critics Vote:

Using the pivot table plotted for actor 1 and average of number of vote by critics. I am giving the result for top 25 actors. We can modify the number as per requirement using the filter. i.e we can easily find Top 25,Top 50
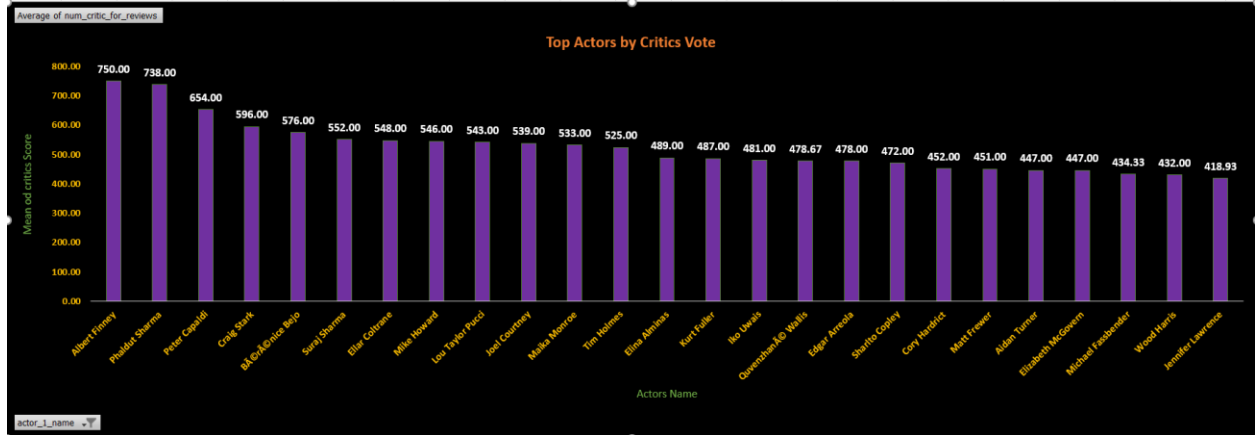

8. Top Actors by Users Vote:

Using the pivot table plotted for actor 1 and average of number of vote by users. I am giving the result for top 25 actors. We can modify the number as per requirement using the filter. i.e we can easily find Top 25,Top 50
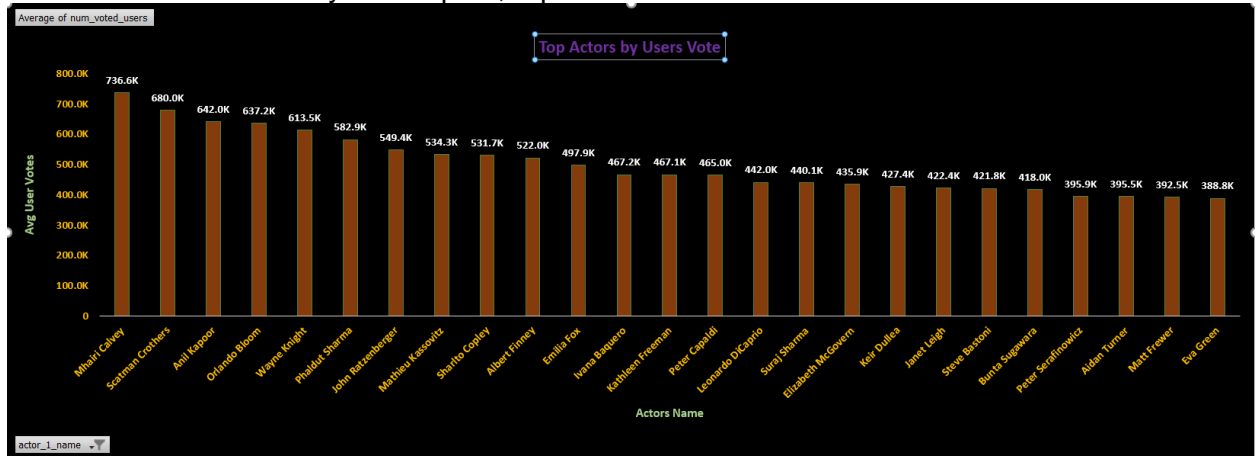
9. **Change in number of voted users over decades:** Using pivot table grouped the year of release decade wise and made a table summing up the number of voted users and made a plot.
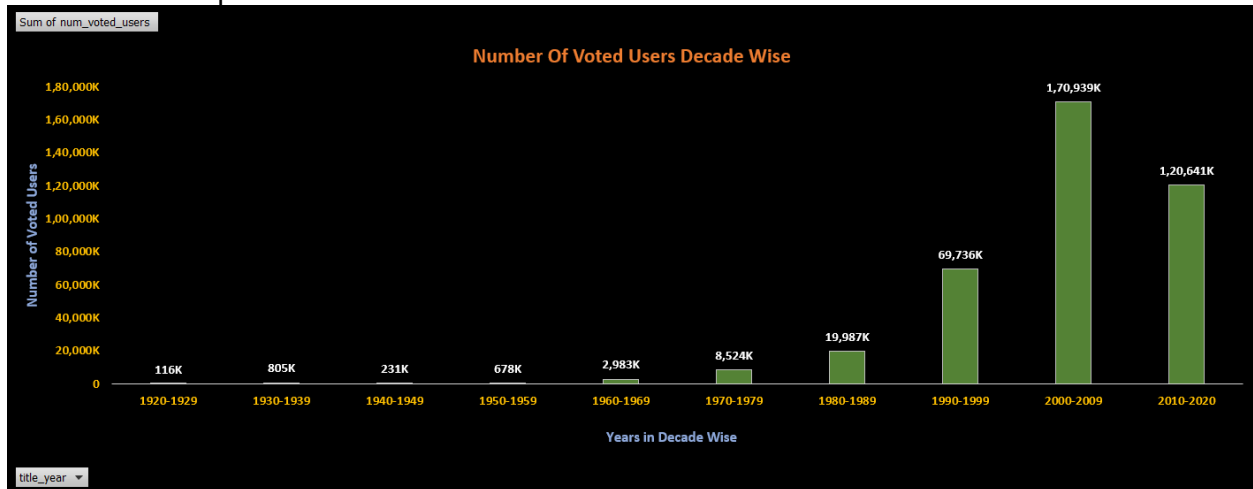
Result:

I was able to create meaningful insights from the dataset which can be crucial for business. I was able to sharpen my Excel skills from here. Successfully completing this project boosted my confidence