

IMDB Movie Analysis



Final Project – 1, By
Shubham Kumar

Project Description

In this final project, we'll understand the whole dataset and perform Data Cleaning firstly by defining the problem for the given IMDB movie dataset, followed by asking the 5 Why's:

- What do we see happening?
- What is our hypothesis for the cause of the problem?
- What is the impact of the problem on stakeholders?
- What is the impact of the problem not being solved?

Tech stack Used



Excel



Canva

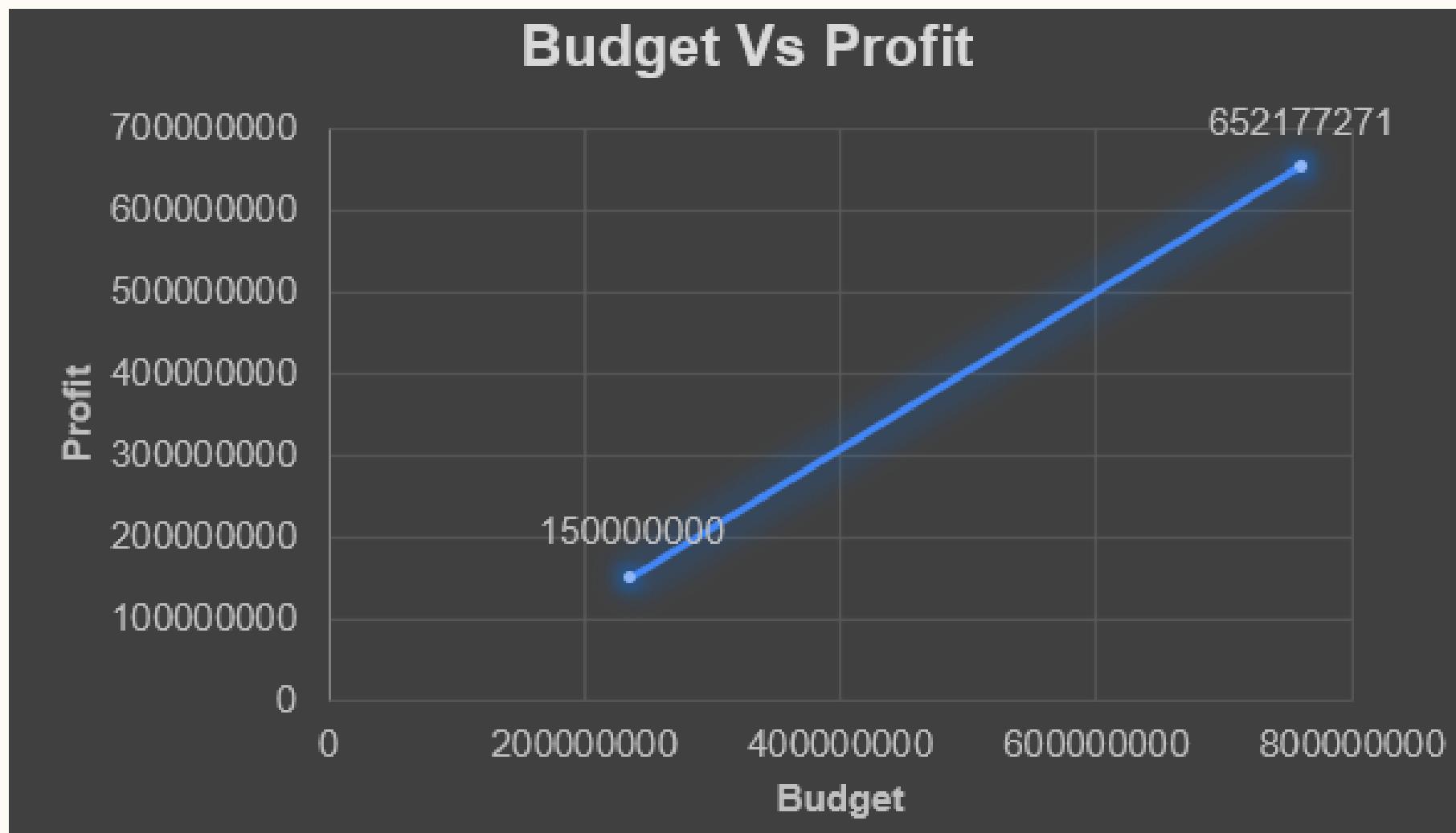
Insights & Results

A. Data Cleaning

- Using Excel to first extract the data and read to explore NULL values, Dropping Columns, etc.
- The Dataset contains 5043 rows, 28 columns, and 1 header row.
- Dropping the following 15 unnecessary columns for best results: color, director_facebook_likes, actor_1_facebook_likes, actor_2_facebook_likes, actor_3_facebook_likes, actor_2_name, cast_total_facebook_likes, actor_3_name, Duration, facenumber_in_poster, content_rating, country, movie_imdb_link, aspect_ratio, plot_keywords.
- Removing Blank cells/ NULL rows.

B. Movies With the Highest Profit

- Created profit column containing difference of Gross and Budget



TOP 10 Movies with high profit			
movie_title	highest profit	budget	Profit
Avatar		237000000	523505847
Jurassic World		150000000	502177271
Titanic		200000000	458672302
Star Wars: Episode IV - A New Hope		11000000	449935665
E.T. the Extra-Terrestrial		10500000	424449459
The Avengers		220000000	403279547
The Lion King		45000000	377783777
Star Wars: Episode I - The Phantom Me		115000000	359544677
The Dark Knight		185000000	348316061
The Hunger Games		78000000	329999255

- Avatar is the top highest profitable movie

C.Top 250

- To find the Top 250: Create a new column name "IMDb_Top_250" and store the top 250 movies with the highest IMDb Rating and num_voted_users greater than 25,000
- To find the Top 250 we filtered by sorting all the movies with the highest IMDb scores and num_voted_users >25000 & sorting the imdb_score in descending order

Top - 250 movies with Voted_users > 25000 and imdb_score largest to smallest			
movie_title	num_voted_users	imdb_score	language
The Shawshank Redemption	1689764	9.3	English
The Godfather	1155770	9.2	English
The Dark Knight	1676169	9	English
The Godfather: Part II	790926	9	English
The Lord of the Rings: The Return of the King	1215718	8.9	English
Pulp Fiction	1324680	8.9	English
Schindler's List	865020	8.9	English
The Good, the Bad and the Ugly	503509	8.9	Italian
Forrest Gump	1251222	8.8	English
Star Wars: Episode V - The Empire Strikes Back	837759	8.8	English
The Lord of the Rings: The Fellowship of the Ring	1238746	8.8	English
Inception	1468200	8.8	English
Fight Club	1347461	8.8	English
Star Wars: Episode IV - A New Hope	911097	8.7	English
The Lord of the Rings: The Two Towers	1100446	8.7	English
The Matrix	1217752	8.7	English
One Flew Over the Cuckoo's Nest	680041	8.7	English
Goodfellas	728685	8.7	English
City of God	533200	8.7	Portuguese
Seven Samurai	229012	8.7	Japanese
Saving Private Ryan	881236	8.6	English
The Silence of the Lambs	887467	8.6	English
Se7en	1023511	8.6	English
Interstellar	928227	8.6	English
The Usual Suspects	740918	8.6	English
American History X	782437	8.6	English
Modern Times	143086	8.6	English
Spirited Away	417971	8.6	Japanese
The Lion King	644348	8.5	English
Raiders of the Lost Ark	661017	8.5	English
The Dark Knight Rises	1144337	8.5	English
Back to the Future	732212	8.5	English
Terminator 2: Judgment Day	744891	8.5	English
Gladiator	982637	8.5	English
The Green Mile	782610	8.5	English

Alien	563827	8.5	English
Django Unchained	955174	8.5	English
Apocalypse Now	450676	8.5	English
The Departed	873649	8.5	English
Psycho	422432	8.5	English
Memento	845580	8.5	English
The Prestige	844052	8.5	English
Whiplash	399138	8.5	English
The Lives of Others	259379	8.5	German
Children of Heaven	27882	8.5	Persian
The Pianist	497946	8.5	English
Star Wars: Episode VI - Return of the Jedi	681857	8.4	English
American Beauty	822500	8.4	English
Aliens	488537	8.4	English
WALL·E	718837	8.4	English
A Separation	151812	8.4	Persian
Braveheart	736638	8.4	English
Reservoir Dogs	664719	8.4	English
Oldboy	356181	8.4	Korean
Requiem for a Dream	573541	8.4	English
Das Boot	168203	8.4	German
Lawrence of Arabia	192775	8.4	English
Once Upon a Time in America	221000	8.4	English
Amélie	534262	8.4	French
Princess Mononoke	221552	8.4	Japanese
Toy Story 3	544884	8.3	English
Inside Out	345198	8.3	English
Toy Story	623757	8.3	English
The Sting	175607	8.3	English
Indiana Jones and the Last Crusade	515306	8.3	English
Good Will Hunting	604904	8.3	English
Up	665575	8.3	English
Unforgiven	277505	8.3	English
Batman Begins	980946	8.3	English
Inglourious Basterds	885175	8.3	English
2001: A Space Odyssey	427357	8.3	English
Amadeus	270790	8.3	English
L.A. Confidential	414219	8.3	English
Snatch	600996	8.3	English

The Help	318955	8.1	English
Platoon	291603	8.1	English
The Martian	472488	8.1	English
The Bourne Ultimatum	491077	8.1	English
Rocky	375240	8.1	English
Gone Girl	569841	8.1	English
Butch Cassidy and the Sundance Kid	152089	8.1	English
The Imitation Game	467613	8.1	English
Million Dollar Baby	482064	8.1	English
The Truman Show	667983	8.1	English
Groundhog Day	437418	8.1	English
No Country for Old Men	612060	8.1	English
The Revenant	406020	8.1	English
Shutter Island	786092	8.1	English
Stand by Me	271794	8.1	English
Kill Bill: Vol. 1	735784	8.1	English
12 Years a Slave	439176	8.1	English
Annie Hall	192940	8.1	English
Sin City	656640	8.1	English
The Grand Budapest Hotel	475518	8.1	English
The Terminator	600266	8.1	English
Spotlight	195333	8.1	English
The Best Years of Our Lives	40359	8.1	English
The Wizard of Oz	291875	8.1	English
There Will Be Blood	372990	8.1	English
Prisoners	383591	8.1	English
The Princess Bride	294163	8.1	English
Hotel Rwanda	264533	8.1	English
Mad Max: Fury Road	552503	8.1	English
Amores Perros	173551	8.1	Spanish
Before Sunrise	183288	8.1	English
The Celebration	65951	8.1	Danish
Donnie Darko	580999	8.1	English
Elite Squad	81644	8.1	Portuguese
The Sea Inside	64556	8.1	Spanish
Rush	312629	8.1	English
Tae Guk Gi: The Brotherhood of War	31943	8.1	Korean
Akira	106160	8.1	Japanese
Jaws	412454	8	English

Some Like It Hot	175196	8.3	English
Scarface	537442	8.3	English
Eternal Sunshine of the Spotless Mind	666937	8.3	English
Room	161288	8.3	English
Monty Python and the Holy Grail	382240	8.3	English
The Hunt	170155	8.3	Danish
Metropolis	111841	8.3	German
Downfall	248354	8.3	German
Raging Bull	235133	8.3	English
Finding Nemo	692482	8.2	English
Gone with the Wind	215340	8.2	English
Captain America: Civil War	272670	8.2	English
Gran Torino	561773	8.2	English
A Beautiful Mind	610568	8.2	English
Die Hard	592582	8.2	English
How to Train Your Dragon	485430	8.2	English
The Bridge on the River Kwai	149444	8.2	English
Pan's Labyrinth	467234	8.2	Spanish
The Secret in Their Eyes	131831	8.2	Spanish
The Wolf of Wall Street	780588	8.2	English
V for Vendetta	791783	8.2	English
Trainspotting	469561	8.2	English
On the Waterfront	100890	8.2	English
Into the Wild	426359	8.2	English
Lock, Stock and Two Smoking Barrels	414976	8.2	English
The Big Lebowski	537419	8.2	English
Incendies	80429	8.2	French
Blade Runner	461609	8.2	English
The Thing	258078	8.2	English
Casino	333542	8.2	English
Warrior	332276	8.2	English
Howl's Moving Castle	214091	8.2	Japanese
The Avengers	995415	8.1	English
Deadpool	479047	8.1	English
Jurassic Park	613473	8.1	English
The Sixth Sense	704766	8.1	English
Monsters, Inc.	585659	8.1	English
Pirates of the Caribbean: The Curse of the Black Pearl	809474	8.1	English
Guardians of the Galaxy	682155	8.1	English

Shaun of the Dead	395921	8	English
Life of Pi	440084	8	English
The Perks of Being a Wallflower	351274	8	English
A Fistful of Dollars	147566	8	Italian
Before Sunset	168398	8	English
Central Station	28951	8	Portuguese
Her	355126	8	English
Waltz with Bashir	46107	8	Hebrew
True Romance	163492	8	English
Persepolis	70194	8	French
Big Fish	350698	8	English
The Straight Story	63733	8	English
Brazil	152306	8	English
In Bruges	307639	8	English
Mulholland Drive	235992	8	English
My Name Is Khan	69759	8	Hindi
Dancer in the Dark	79330	8	English
Magnolia	241030	8	English
Serenity	242599	8	English
Cinderella Man	148238	8	English
Blood Diamond	400292	8	English
The Iron Giant	128455	8	English
Avatar	886204	7.9	English
E.T. the Extra-Terrestrial	281842	7.9	English
Shrek	467113	7.9	English
Iron Man	696338	7.9	English
Toy Story 2	385871	7.9	English
Straight Outta Compton	119928	7.9	English
Taken	483756	7.9	English
Crouching Tiger, Hidden Dragon	217740	7.9	Mandarin
Walk the Line	188637	7.9	English
The Fighter	275869	7.9	English
The Bourne Identity	407601	7.9	English
Big Hero 6	279093	7.9	English
My Fair Lady	66959	7.9	English
Captain Phillips	323353	7.9	English
Little Miss Sunshine	355810	7.9	English
The Untouchables	219008	7.9	English
Crash	361169	7.9	English

Elite Squad	81644	8.1	Portuguese
The Sea Inside	64556	8.1	Spanish
Rush	312629	8.1	English
Tae Guk Gi: The Brotherhood of War	31943	8.1	Korean
Akira	106160	8.1	Japanese
Jaws	412454	8	English
The Exorcist	284252	8	English
Aladdin	260939	8	English
The Incredibles	479166	8	English
Dances with Wolves	186485	8	English
The Sound of Music	148172	8	English
Rain Man	383784	8	English
Slumdog Millionaire	641997	8	English
The King's Speech	503631	8	English
Catch Me If You Can	525801	8	English
Star Trek	504419	8	English
The Pursuit of Happyness	338383	8	English
Doctor Zhivago	55816	8	English
Black Swan	551363	8	English
District 9	531737	8	English
Young Frankenstein	112671	8	English
Dead Poets Society	277451	8	English
Mystic River	338415	8	English
Ratatouille	473887	8	English
Fiddler on the Roof	29839	8	English
Kill Bill: Vol. 2	512749	8	English
X-Men: Days of Future Past	514125	8	English
JFK	113472	8	English
The Artist	190030	8	English
Sling Blade	72443	8	English
Dallas Buyers Club	326494	8	English
Boyhood	266020	8	English
Bowling for Columbine	123090	8	English
Casino Royale	470483	8	English
Casino Royale	470501	8	English
Sicko	66610	8	English
Shaun of the Dead	395921	8	English
Life of Pi	440084	8	English
The Perks of Being a Wallflower	351274	8	English

Walk the Line	188637	7.9	English
The Fighter	275869	7.9	English
The Bourne Identity	407601	7.9	English
Big Hero 6	279093	7.9	English
My Fair Lady	66959	7.9	English
Captain Phillips	323353	7.9	English
Little Miss Sunshine	355810	7.9	English
The Untouchables	219008	7.9	English
Crash	361169	7.9	English
Halloween	157857	7.9	English
Halloween	157863	7.9	English
Edward Scissorhands	357581	7.9	English
The Hobbit: The Desolation of Smaug	483540	7.9	English
How to Train Your Dragon 2	221128	7.9	English
The Blues Brothers	142448	7.9	English
Nightcrawler	293304	7.9	English
Do the Right Thing	59524	7.9	English
The Wrestler	251349	7.9	English
Hot Fuzz	352695	7.9	English
The Remains of the Day	45703	7.9	English
Boogie Nights	189032	7.9	English
The Hateful Eight	272839	7.9	English
Once	90827	7.9	English
Glory	101888	7.9	English
Glory	101889	7.9	English
Before Midnight	95362	7.9	English
4 Months, 3 Weeks and 2 Days	44763	7.9	Romanian
Moon	260607	7.9	English
Nine Queens	38215	7.9	Spanish
The Chorus	44151	7.9	French
Letters from Iwo Jima	132149	7.9	Japanese
The Right Stuff	45271	7.9	English
Amour	70382	7.9	French
Ed Wood	142416	7.9	English
The World's Fastest Indian	44198	7.9	English
Almost Famous	207287	7.9	English
The Notebook	396396	7.9	English

Movies in Foreign Language

Task: To Extract all the movies in the IMDb_Top_250 column which are not in the English language and storing them in a new column named Top_Foreign_Lang_Film.

Extract all the movies in the IMDb_Top_250 column which are not in the English			
movie_title	num_voted_user	imdb_score	language
The Good, the Bad and the Ugly	503509	8.9	Italian
City of God	533200	8.7	Portuguese
Seven Samurai	229012	8.7	Japanese
Spirited Away	417971	8.6	Japanese
The Lives of Others	259379	8.5	German
Children of Heaven	27882	8.5	Persian
A Separation	151812	8.4	Persian
Oldboy	356181	8.4	Korean
Das Boot	168203	8.4	German
Amélie	534262	8.4	French
Princess Mononoke	221552	8.4	Japanese
The Hunt	170155	8.3	Danish
Metropolis	111841	8.3	German
Downfall	248354	8.3	German
Pan's Labyrinth	467234	8.2	Spanish
The Secret in Their Eyes	131831	8.2	Spanish
Incendies	80429	8.2	French
Howl's Moving Castle	214091	8.2	Japanese
Amores Perros	173551	8.1	Spanish
The Celebration	65951	8.1	Danish
Elite Squad	81644	8.1	Portuguese
The Sea Inside	64556	8.1	Spanish
Tae Guk Gi: The Brotherhood of War	31943	8.1	Korean
Akira	106160	8.1	Japanese
A Fistful of Dollars	147566	8	Italian
Central Station	28951	8	Portuguese
Waltz with Bashir	46107	8	Hebrew
Persepolis	70194	8	French
My Name Is Khan	69759	8	Hindi
Crouching Tiger, Hidden Dragon	217740	7.9	Mandarin
4 Months, 3 Weeks and 2 Days	44763	7.9	Romanian
Nine Queens	38215	7.9	Spanish
The Chorus	44151	7.9	French
Letters from Iwo Jima	132149	7.9	Japanese
Amour	70382	7.9	French

D. Best Directors

- Task: To Find the best directors.
- Using the pivot table to calculate the Average by keeping Director_name in row and imdb_score in values. Finally sorting directors in descending order.

1	2	TOP 10 Directors
3	Director Name	Average of imdb_score
4	Akira Kurosawa	8.7
5	Charles Chaplin	8.6
6	Tony Kaye	8.6
7	Alfred Hitchcock	8.5
8	Damien Chazelle	8.5
9	Majid Majidi	8.5
10	Ron Fricke	8.5
11	Sergio Leone	8.4
12	Christopher Nolan	8.4
13	Asghar Farhadi	8.4
14	Richard Marquand	8.4

E. Popular Genres

- Task: To Find popular genres Using the Data > Text to Column function on the ribbon to list out the genres.

Most Watched Genre

Genre	Most genre
Drama	1874
Comedy	1451
Thriller	1105
Action	951
Romance	849
Adventure	773
Crime	702
Fantasy	504
Sci-Fi	491
Family	439
Horror	386
Mystery	378
Biography	238
Animation	196
Musical	95
Western	57
Documentary	44

F. Charts

- Task: To Find the critic-favorite and audience-favorite actors
- By using the filter created 3 columns of Meryl_Streep, Leo_Caprio, Brad_Pitt and containing movie_title.

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

Bar Charts

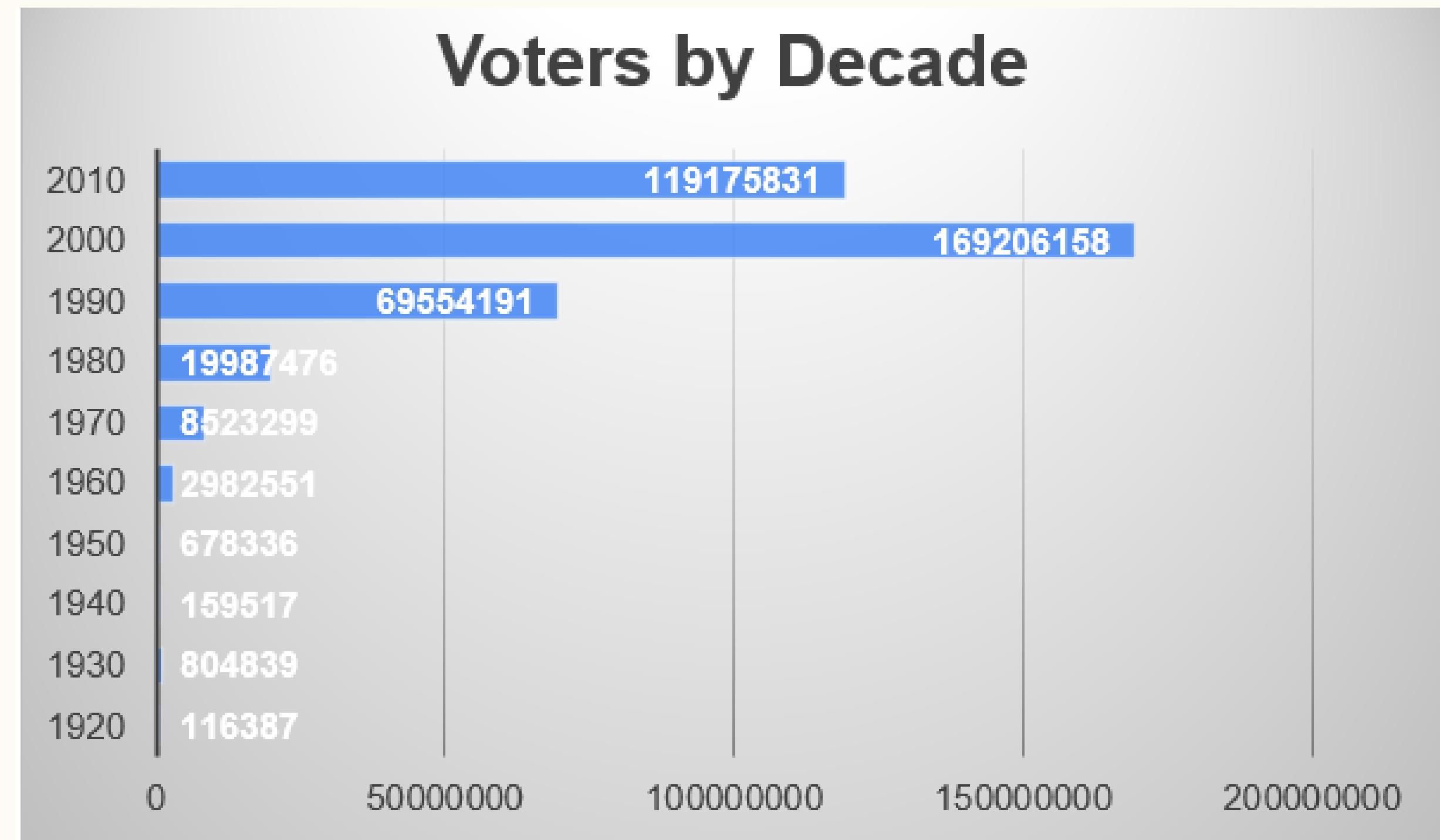
- Task: To observe the change in the number of voted users over decades.
- Using the pivot table to calculate the total sum, keeping the year in rows and no. of voted users as values.
- Taking the Min and Max value into consideration and using the following SUMIFS aggregate function to sort the decade column:

=SUMIFS(\$H\$2:\$H\$75,\$G\$2:\$G\$75, ">= 1920", \$G\$2:\$G\$75,"< 1930") to

=SUMIFS(\$H\$2:\$H\$75,\$G\$2:\$G\$75, ">= 2010", \$G\$2:\$G\$75,"< 2020")

Voters by Decade

Decade	sum_of_decade_vote
1920	116387
1930	804839
1940	159517
1950	678336
1960	2982551
1970	8523299
1980	19987476
1990	69554191
2000	169206158
2010	119175831



Thank You!
