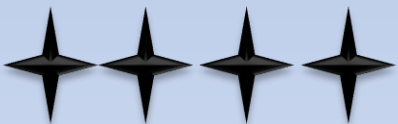# IMDB MOVIE ANALYSIS

Submitted By Harshita Mundhe

PROJECT DESCRIPTION

In this project we have given a dataset of having various columns of different IMDB movies. First we have to framed the problem by asking 'What is the problem?'. By giving answers we can easily defined the problem and can easily find out right data to solve it. After defining the problem we have to use 5 Whys technique to determine the root cause of the problem. IMDB data is widely used to show data in relation to video games, movies, directors, actors, and so on. We can access this data to take look at the information about movies, videos games, actors, and so on.

For analyzing and predicting the best movies, director from the IMDB movie dataset I have used excel as it allow us to ask questions about our data through natural language queries instead of writing complex formulas.

APPROACH

For implementing this project first we have to download the given dataset and open it using Google Sheets or Excel. After that I have cleaned the data by finding null and duplicate values.
I used different excel formulas to get the results like min, max, avg and so on. I also used pivot table and pivot chart concepts for getting output. And to present the output in more effectively I used graphs, charts and so on as they help us to quickly analyze data and see relationships.
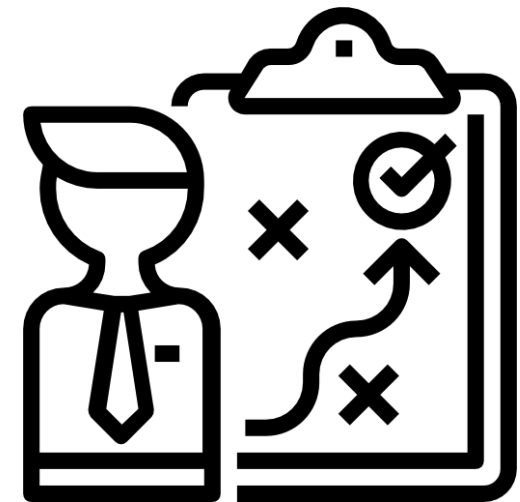
TECH-STACK USED

- **Microsoft Excel:** I have used Microsoft Excel 2019 MSO (Version 2212 Build 16.0.15928.20196) for getting the meaningful insights from given dataset as it provides high-level visual summaries, trends and it helps us to understand the data through natural language queries that allow us to ask questions about the data without having to write complicated formulas.

- **PowerPoint Presentation:** I have used Microsoft PowerPoint 2019 MSO (Version 2212 Build 16.0.15928.20196) 64-bit to create a report as it allow us to present the complex ideas, facts, or figures into easily digestible visuals.

INSIGHTS

- **Cleaning the data :-**
   For defining the problem first we have to clean the data to fetch the exact data. For cleaning the data I removed null values, duplicate values by using various option available in Excel's data tab and used filter option, formatting etc. And then I count the number of rows and columns.

- **Result:**

Row lables	Before Cleaning	After Cleaning
Rows	5044	3740
Columns	28	28
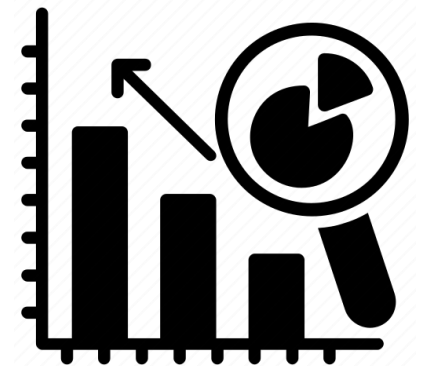
INSIGHTS

- **Movies with highest profit:-**

  For fetching movies with highest profit I have created pivot table. Then I have created a column profit which is a difference of Gross and Budget column with the help of calculated item option. And then I filtered first 20 records. I have used pivot chart to represent the results.

- **Result:**

Movie	Profit
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459
The Avengers	403279547
The Lion King	377783777
The Jungle Book	375290282
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061
The Hunger Games	329999255
Twilight	308898950
Deadpool	305024263
The Hunger Games: Catching Fire	294645577
Jurassic Park	293784000
Despicable Me 2	292049635
American Sniper	291323553
Finding Nemo	286838870
Shrek 2	286471036
The Lord of the Rings: The Return of the King	283019252
Grand Total	**7195009279**

INSIGHTS

- **IMDB Top 250:-**
  For fetching top 250 movies I have created pivot table. I have used various filter and sort functions to get the desired output. Then I have created a column rank which is containing rank of movies from 1 to 250.
- **Result:**

Movie	num_voted_users	imdb_score	Rank
The Shawshank Redemption	1689764	9.3	1
The Godfather	1155770	9.2	2
The Godfather: Part II	790926	9	3
The Dark Knight	1676169	9	4
The Lord of the Rings: The Return of the King	1215718	8.9	5
The Good, the Bad and the Ugly	503509	8.9	6
Schindler's List	865020	8.9	7
Pulp Fiction	1324680	8.9	8
The Lord of the Rings: The Fellowship of the Ring	1238746	8.8	9
Star Wars: Episode V - The Empire Strikes Back	837759	8.8	10
Forrest Gump	1251222	8.8	11
Inception	1468200	8.8	12
Fight Club	1347461	8.8	13
The Lord of the Rings: The Two Towers	1100446	8.7	14
The Matrix	1217752	8.7	15
Star Wars: Episode IV - A New Hope	911097	8.7	16
One Flew Over the Cuckoo's Nest	680041	8.7	17
Seven Samurai	229012	8.7	18
Goodfellas	728685	8.7	19
City of God	533200	8.7	20
The Silence of the Lambs	887467	8.6	21
The Usual Suspects	740918	8.6	22
Spirited Away	417971	8.6	23
Se7en	1023511	8.6	24
Saving Private Ryan	881236	8.6	25
Modern Times	143086	8.6	26
Interstellar	928227	8.6	27
American History X	782437	8.6	28
Whiplash	399138	8.5	29
The Lives of Others	259379	8.5	30
The Pianist	497946	8.5	31
The Green Mile	782610	8.5	32
The Prestige	844052	8.5	33
The Lion King	644348	8.5	34
The Departed	873649	8.5	35
Terminator 2: Judgment Day	744891	8.5	36

	Movie	num_voted_users	imdb_score	Rank
40	The Dark Knight Rises	1144337	8.5	37
41	Raiders of the Lost Ark	661017	8.5	38
42	Psycho	422432	8.5	39
43	Memento	845580	8.5	40
44	Gladiator	982637	8.5	41
45	Django Unchained	955174	8.5	42
46	Children of Heaven	27882	8.5	43
47	Back to the Future	732212	8.5	44
48	Apocalypse Now	450676	8.5	45
49	Alien	563827	8.5	46
50	WALL·E	718837	8.4	47
51	Star Wars: Episode VI - Return of the Jedi	681857	8.4	48
52	Requiem for a Dream	573541	8.4	49
53	Once Upon a Time in America	221000	8.4	50
54	Reservoir Dogs	664719	8.4	51
55	Oldboy	356181	8.4	52
56	Princess Mononoke	221552	8.4	53
57	Lawrence of Arabia	192775	8.4	54
58	Braveheart	736638	8.4	55
59	Das Boot	168203	8.4	56
60	A Separation	151812	8.4	57
61	American Beauty	822500	8.4	58
62	Amélie	534262	8.4	59
63	Aliens	488537	8.4	60
64	Unforgiven	277505	8.3	61
65	Up	665575	8.3	62
66	The Sting	175607	8.3	63
67	Toy Story	623757	8.3	64
68	Toy Story 3	544884	8.3	65
69	The Hunt	170155	8.3	66
70	Snatch	600996	8.3	67
71	Some Like It Hot	175196	8.3	68
72	Room	161288	8.3	69
73	Scarface	537442	8.3	70
74	Raging Bull	235133	8.3	71
75	Metropolis	111843	8.3	72
76	Monty Python and the Holy Grail	382240	8.3	73
77	L.A. Confidential	414219	8.3	74
78	Indiana Jones and the Last Crusade	515306	8.3	75

INSIGHTS

79	Inside Out	345198	8.3	76
80	Inglourious Basterds	885175	8.3	77
81	Good Will Hunting	604904	8.3	78
82	Downfall	248354	8.3	79
83	Eternal Sunshine of the Spotless Mind	666937	8.3	80
84	2001: A Space Odyssey	427357	8.3	81
85	Batman Begins	980946	8.3	82
86	Amadeus	270790	8.3	83
87	Trainspotting	469561	8.2	84
88	The Wolf of Wall Street	780588	8.2	85
89	V for Vendetta	791783	8.2	86
90	The Thing	258078	8.2	87
91	Warrior	332276	8.2	88
92	The Secret in Their Eyes	131831	8.2	89
93	The Bridge on the River Kwai	149444	8.2	90
94	The Big Lebowski	537419	8.2	91
95	On the Waterfront	100890	8.2	92
96	Pan's Labyrinth	467234	8.2	93
97	Lock, Stock and Two Smoking Barrels	414976	8.2	94
98	Into the Wild	426359	8.2	95
99	Incendies	80429	8.2	96
100	Gone with the Wind	215340	8.2	97
101	Finding Nemo	692482	8.2	98
102	How to Train Your Dragon	485430	8.2	99
103	Howl's Moving Castle	214091	8.2	100
104	Gran Torino	561773	8.2	101
105	Casino	333542	8.2	102
106	Captain America: Civil War	272670	8.2	103
107	Die Hard	592582	8.2	104
108	A Beautiful Mind	610568	8.2	105
109	Blade Runner	461609	8.2	106
110	The Truman Show	667983	8.1	107
111	There Will Be Blood	372990	8.1	108
112	The Sixth Sense	704766	8.1	109
113	The Terminator	600266	8.1	110
114	The Wizard of Oz	291875	8.1	111
115	The Imitation Game	467613	8.1	112
116	The Princess Bride	294163	8.1	113
117	The Revenant	406020	8.1	114

118	The Martian	472488	8.1	115
119	The Help	318955	8.1	116
120	The Grand Budapest Hotel	475518	8.1	117
121	The Sea Inside	64556	8.1	118
122	Sin City	656640	8.1	119
123	Shutter Island	786092	8.1	120
124	The Bourne Ultimatum	491077	8.1	121
125	Spotlight	195333	8.1	122
126	The Celebration	65951	8.1	123
127	Stand by Me	271794	8.1	124
128	The Best Years of Our Lives	40359	8.1	125
129	The Avengers	995415	8.1	126
130	Tae Guk Gi: The Brotherhood of War	31943	8.1	127
131	No Country for Old Men	612060	8.1	128
132	Prisoners	383591	8.1	129
133	Rocky	375240	8.1	130
134	Rush	312629	8.1	131
135	Pirates of the Caribbean: The Curse of the Black Pearl	809474	8.1	132
136	Platoon	291603	8.1	133
137	Monsters, Inc.	585659	8.1	134
138	Jurassic Park	613473	8.1	135
139	Kill Bill: Vol. 1	735784	8.1	136
140	Million Dollar Baby	482064	8.1	137
141	Mad Max: Fury Road	552503	8.1	138
142	Hotel Rwanda	264533	8.1	139
143	Guardians of the Galaxy	682155	8.1	140
144	Gone Girl	569841	8.1	141
145	Groundhog Day	437418	8.1	142
146	Deadpool	479047	8.1	143
147	Butch Cassidy and the Sundance Kid	152089	8.1	144
148	Elite Squad	81644	8.1	145
149	Donnie Darko	580999	8.1	146
150	Before Sunrise	183288	8.1	147
151	Annie Hall	192940	8.1	148
152	12 Years a Slave	439176	8.1	149
153	Akira	106160	8.1	150
154	Amores Perros	173551	8.1	151
155	The Sound of Music	148172	8	152
156	True Romance	163492	8	153

INSIGHTS

157	Waltz with Bashir	46107	8	154
158	The Straight Story	63733	8	155
159	Young Frankenstein	112671	8	156
160	X-Men: Days of Future Past	514125	8	157
161	The Exorcist	284252	8	158
162	The Incredibles	479166	8	159
163	The Perks of Being a Wallflower	351274	8	160
164	The King's Speech	503631	8	161
165	The Pursuit of Happyness	338383	8	162
166	The Iron Giant	128455	8	163
167	Sicko	66610	8	164
168	Slumdog Millionaire	641997	8	165
169	Sling Blade	72443	8	166
170	Shaun of the Dead	395921	8	167
171	Star Trek	504419	8	168
172	The Artist	190030	8	169
173	Serenity	242599	8	170
174	Ratatouille	473887	8	171
175	Persepolis	70194	8	172
176	Rain Man	383784	8	173
177	Mulholland Drive	235992	8	174
178	Jaws	412454	8	175
179	Kill Bill: Vol. 2	512749	8	176
180	Life of Pi	440084	8	177
181	JFK	113472	8	178
182	My Name Is Khan	69759	8	179
183	Mystic River	338457	8	180
184	Magnolia	241030	8	181
185	In Bruges	307639	8	182
186	Fiddler on the Roof	29839	8	183
187	Her	355126	8	184
188	Casino Royale	470483	8	185
189	Dancer in the Dark	79330	8	186
190	Brazil	152306	8	187
191	Cinderella Man	148238	8	188
192	District 9	531737	8	189
193	Central Station	28951	8	190
194	Doctor Zhivago	55816	8	191
195	Dead Poets Society	277451	8	192

	A	B	C	D
95	Dead Poets Society	277451	8	192
96	Dances with Wolves	186485	8	193
97	Catch Me If You Can	525801	8	194
98	Boyhood	266020	8	195
99	Dallas Buyers Club	326494	8	196
00	Blood Diamond	400292	8	197
01	Big Fish	350698	8	198
02	Before Sunset	168398	8	199
03	A Fistful of Dollars	147566	8	200
04	Black Swan	551363	8	201
05	Aladdin	260939	8	202
06	Bowling for Columbine	123090	8	203
07	The World's Fastest Indian	44198	7.9	204
08	The Untouchables	219008	7.9	205
09	Walk the Line	188637	7.9	206
10	The Wrestler	251349	7.9	207
11	Toy Story 2	385871	7.9	208
12	The Fighter	275869	7.9	209
13	The Insider	133526	7.9	210
14	The Hateful Eight	272839	7.9	211
15	The Notebook	396396	7.9	212
16	The Remains of the Day	45703	7.9	213
17	The Hobbit: The Desolation of Smaug	483540	7.9	214
18	The Right Stuff	45271	7.9	215
19	Taken	483756	7.9	216
20	The Chorus	44151	7.9	217
21	The Blues Brothers	142448	7.9	218
22	Shrek	467113	7.9	219
23	Straight Outta Compton	119928	7.9	220
24	The Bourne Identity	407601	7.9	221
25	Nightcrawler	293304	7.9	222
26	Once	90827	7.9	223
27	Nine Queens	38215	7.9	224
28	Moon	260607	7.9	225
29	Letters from Iwo Jima	132149	7.9	226
30	My Fair Lady	66959	7.9	227
31	Iron Man	696338	7.9	228
32	Little Miss Sunshine	355810	7.9	229
33	How to Train Your Dragon 2	221128	7.9	230

234	Halloween	157857	7.9	231
235	Glory	101888	7.9	232
236	Hero	149414	7.9	233
237	Hot Fuzz	352695	7.9	234
238	Ed Wood	142416	7.9	235
239	Children of Men	361767	7.9	236
240	Edward Scissorhands	357581	7.9	237
241	Crash	361169	7.9	238
242	E.T. the Extra-Terrestrial	281842	7.9	239
243	Crouching Tiger, Hidden Dragon	217740	7.9	240
244	Edge of Tomorrow	431620	7.9	241
245	Captain Phillips	323353	7.9	242
246	Do the Right Thing	59524	7.9	243
247	Amour	70382	7.9	244
248	Almost Famous	207287	7.9	245
249	Big Hero 6	279093	7.9	246
250	Before Midnight	95362	7.9	247
251	Avatar	886204	7.9	248
252	Boogie Nights	189032	7.9	249
253	4 Months, 3 Weeks and 2 Days	44763	7.9	250

INSIGHTS

- **Best Directors:-**
  For fetching best directors I have created pivot table. After that I have calculated mean of imdb_score. And sorted the same from largest to smallest. I have used pivot chart to represent the results.
- **Result:**

| Director	Mean of imdb_score
Akira Kurosawa	8.7
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Asghar Farhadi	8.4
Richard Marquand	8.4
Grand Total	**8.47**

Best Directors

INSIGHTS

- **Popular genres:-**
   For fetching popular genres I have created pivot table. After that I took count and sorted from highest to smallest value. I have used pivot chart to represent the results.
- **Result:**

Genres	Count of gross
Comedy|Drama|Romance	147
Drama	144
Comedy|Drama	138
Comedy	137
Comedy|Romance	131
Drama|Romance	114
Crime|Drama|Thriller	81
Action|Crime|Thriller	55
Action|Crime|Drama|Thriller	50
Action|Adventure|Sci-Fi	46
Comedy|Crime	45
Action|Adventure|Thriller	45
Horror	43
Drama|Thriller	42
Crime|Drama	41
Crime|Drama|Mystery|Thriller	41
Action|Adventure|Sci-Fi|Thriller	34
Horror|Thriller	33
Horror|Mystery|Thriller	31
Drama|Mystery|Thriller	30
Grand Total	**1428**

Popular Genres
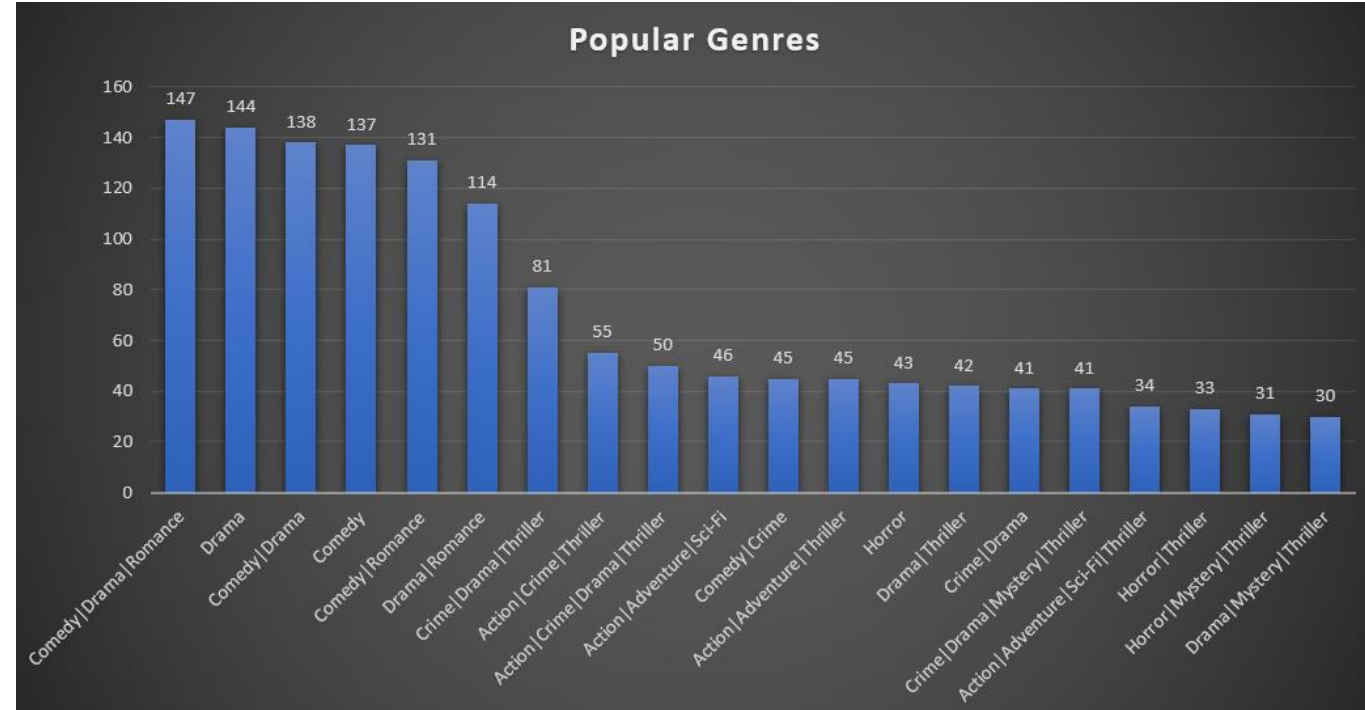
INSIGHTS

- **Critic-favorite and audience-favorite actor:-**
  For this first I created a pivot table and added actor and their movies in rows. After this I grouped them by actor name and calculated mean of num_critic_for_reviews and num_user_for_reviews. And as a result Leonardo DiCaprio has the highest mean

- **Result:**

| Actor	mean of num_critic_for_reviews	mean of num_user_for_reviews
⊞ Leonardo DiCaprio	330.1904762	914.4761905
⊞ Brad Pitt	245	742.3529412
⊞ Meryl Streep	181.4545455	297.1818182
Grand Total	267.244898	716.1836735

INSIGHTS

- **Number Of Voted User The Decades:-**
  For this first I created a pivot table and added year and I grouped them by decades and calculated sum of num_voted_user. And I used pivot chart to represent the data.

- **Result:**

Row Labels	Sum of num_voted_users
⊞ 1920s	116387
⊞ 1930s	804839
⊞ 1940s	159517
⊞ 1950s	678336
⊞ 1960s	2982551
⊞ 1970s	8523299
⊞ 1980s	19987476
⊞ 1990s	69581866
⊞ 2000s	170714771
⊞ 2010s	119464590
Grand Total	**393013632**

Number Of Voted User Over The Decade
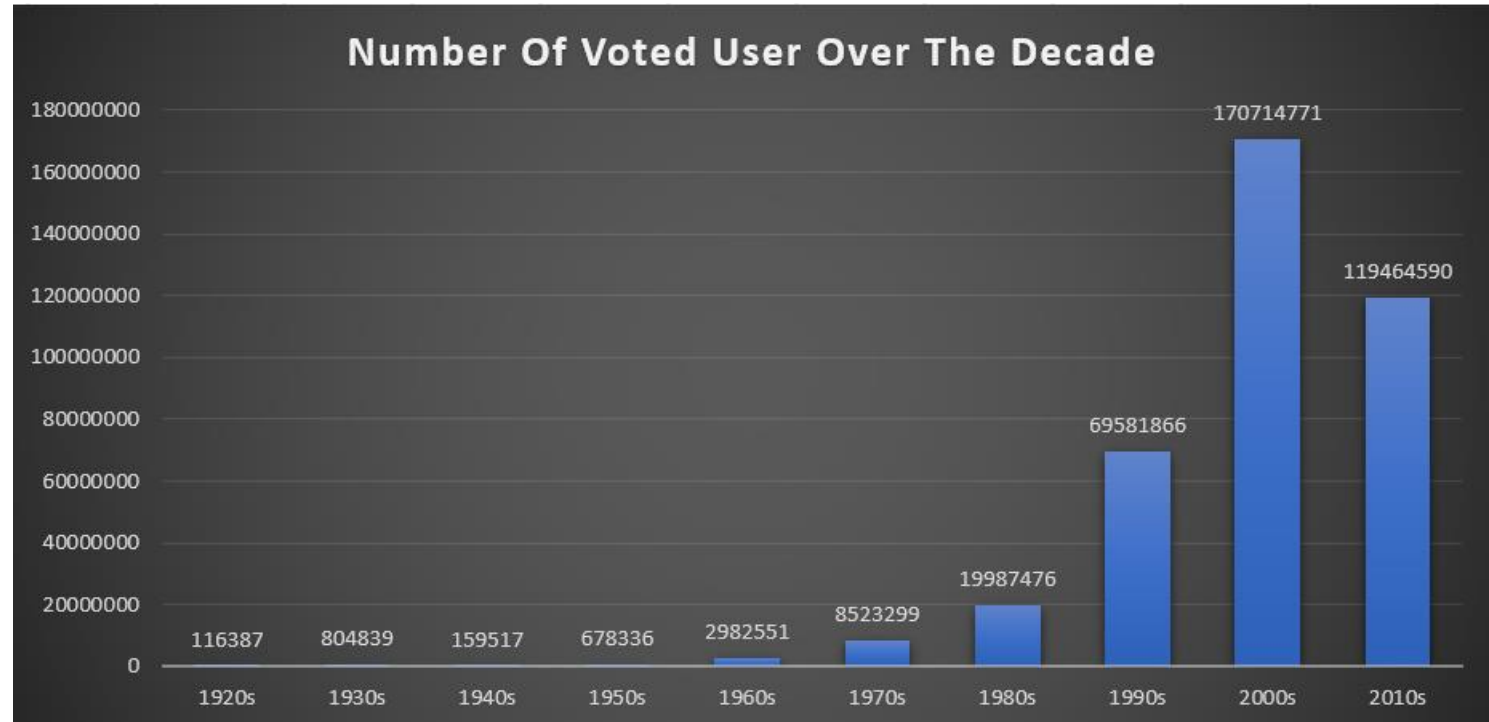
RESULTS

From these project I understood how to clean the data and extract the exact data which is required to find out the root cause of defined problem. I also understood how to use excel formulas to get the desired output. I got to learn various excel formulas like mean,min, max, avg, count and so on. I got to learn what is pivot table, how to create it and importance of the same. I also got to learn how to represent the data using graphs/charts to make the data / insights quickly and easily understandable.

Drive link:

https://docs.google.com/spreadsheets/d/1i4ub44ZiphJjIrQ8QIeF3tYF8ZqkC4tj/edit?usp=sharing&ouid=10772485983792814 6273&rtpof=true&sd=true

THANK YOU