



# Movie Analysis

Debarshi mahato

# Project description

---

- ❑ The dataset includes information from IMDB about various films, including ratings.
- ❑ Our goal is to formulate inquiries and come up with answers in order to draw information from the dataset. We begin by cleaning the dataset.
- ❑ We'll use Excel's many features and functions to get results, which will allow us to study the dataset and gain insightful knowledge.

# Insights

## Task 1

### Cleaning the data:

- This is one of the most important step to perform before moving forward with the analysis.



Cleaning / Removing  
the blanks and null  
rows



Removing all the  
duplicates



Removing all the  
unnecessary  
columns

To carry out the above Data analysis project.

The following steps are taken:

- ❑ Initially to carry out the first task, the datasets were cleaned using transform option while importing the data set into excel and some columns which are not required are removed.
- ❑ After dropping the irrelevant columns, now I have removed the rows from the dataset having any of its column values as blank/NULL.
- ❑ I get rid of the duplicate values in the dataset which can be achieved by using the 'Remove Duplicate Values/Cells' available in the 'Data' tab

**Drive link:** [https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr\\_B04Wz6mJiTtYFloZST/edit?usp=sharing&oid=105000439991683267564&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr_B04Wz6mJiTtYFloZST/edit?usp=sharing&oid=105000439991683267564&rtpof=true&sd=true)

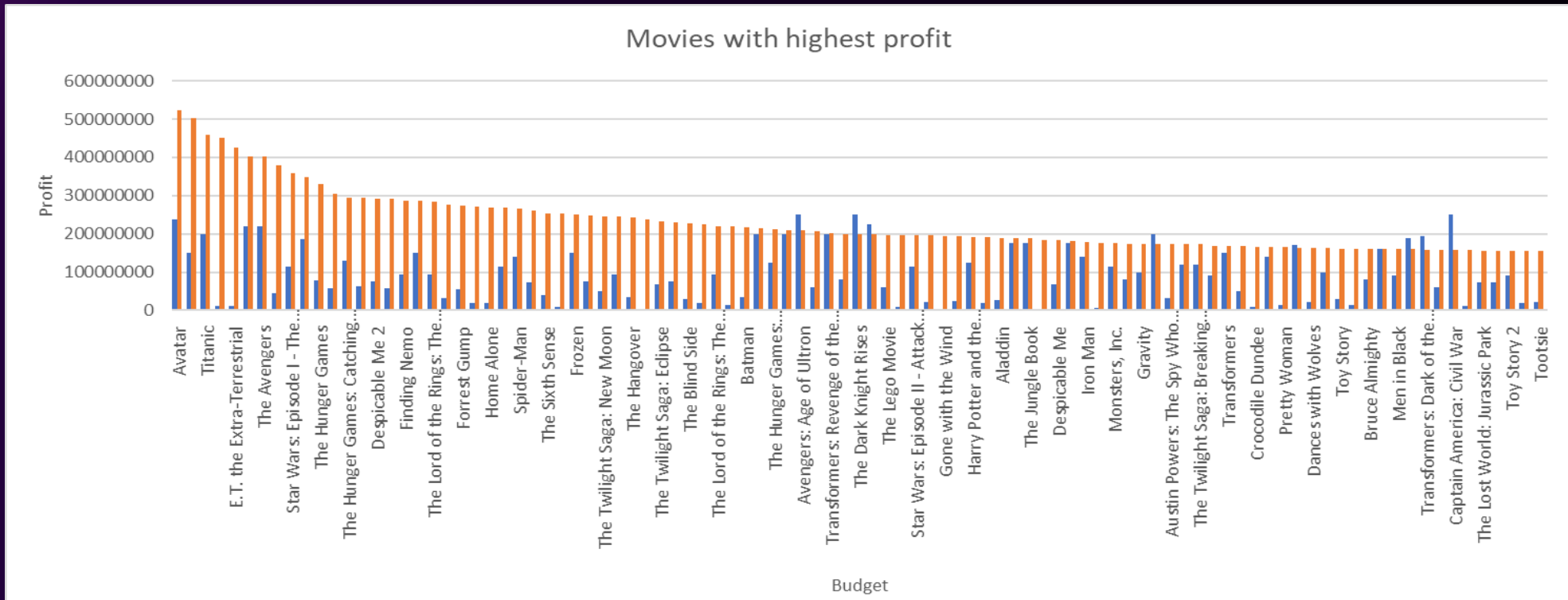


# Task 2

## Movies with highest profit

To find the movies with the highest profit: -

- ❑ First I have subtract the budget value from the gross value to get the profit.
- ❑ Then, by using the line chart option, I plot the values of profit(y\_axis) and budget(x\_axis).



**Drive link:** [https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr\\_B04Wz6mJiTtYfLoZST/edit?usp=sharing&oid=105000439991683267564&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr_B04Wz6mJiTtYfLoZST/edit?usp=sharing&oid=105000439991683267564&rtpof=true&sd=true)

# TASK 3

## Top 250

### To find the IMDB Top 250:-

- Initially we will filter out those rows whose num\_voted\_users > 25000 using the sort and filter option followed by arranging the dataset on the basis of imdb\_score in descending order and selecting only the top 250 rows for the further analysis

### Drive link:

[https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr\\_B04Wz6mJiTtYFloZST/edit?usp=sharing&oid=105000439991683267564&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr_B04Wz6mJiTtYFloZST/edit?usp=sharing&oid=105000439991683267564&rtpof=true&sd=true)

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

# TASK 4

## Best Directors

To carryout this task:

- Initially after selecting the imdb\_score column of the cleaned dataset. click on pivot table to add director\_name and average imdb\_score into the values section of the pivot table to get desire output

### Drive link:

[https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr\\_B04Wz6mJiTiyFloZST/edit?usp=sharing&ouid=105000439991683267564&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr_B04Wz6mJiTiyFloZST/edit?usp=sharing&ouid=105000439991683267564&rtpof=true&sd=true)

		alphabetically	
Directors	Mean of imdb_score	Directors	Mean of imdb_score
Charles Chaplin	8.6	Alfred Hit	8.5
Tony Kaye	8.6	Asghar Far	8.4
Alfred Hitchcock	8.5	Billy Wilder	8.3
Damien Chazelle	8.5	Charles Ch	8.6
Majid Majidi	8.5	Christoph	8.425
Ron Fricke	8.5	Damien Cl	8.5
Sergio Leone	8.433333333	Elia Kazan	8.2
Christopher Nolan	8.425	Fritz Lang	8.3
Asghar Farhadi	8.4	George Ro	8.2
Marius A. Markevicius	8.4	Hayao Miy	8.225
Richard Marquand	8.4	Joshua Op	8.2
S.S. Rajamouli	8.4	Juan José	8.2
Billy Wilder	8.3	Lee Unkric	8.3
Fritz Lang	8.3	Lenny Abr	8.3
Lee Unkrich	8.3	Majid Maj	8.5
Lenny Abrahamson	8.3	Marius A.	8.4
Pete Docter	8.233333333	Pete Doct	8.233333
Hayao Miyazaki	8.225	Quentin T	8.2
Quentin Tarantino	8.2	Richard M	8.4
Elia Kazan	8.2	Ron Fricke	8.5
George Roy Hill	8.2	S.S. Rajam	8.4
Joshua Oppenheimer	8.2		
Juan José Campanella	8.2		

# TASK 5

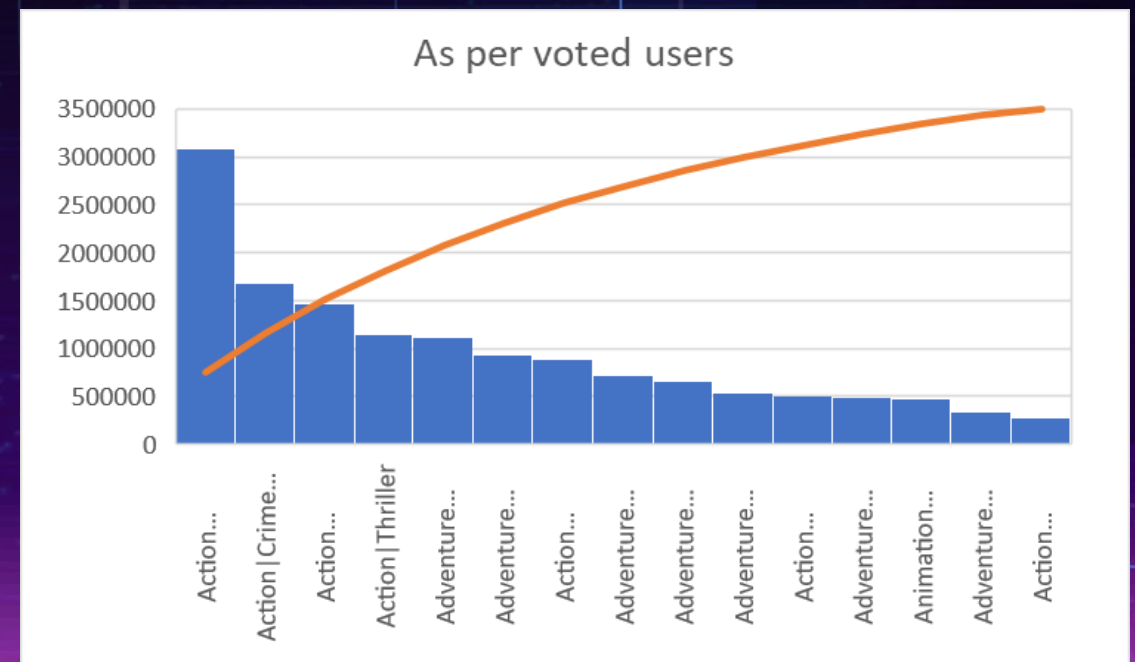
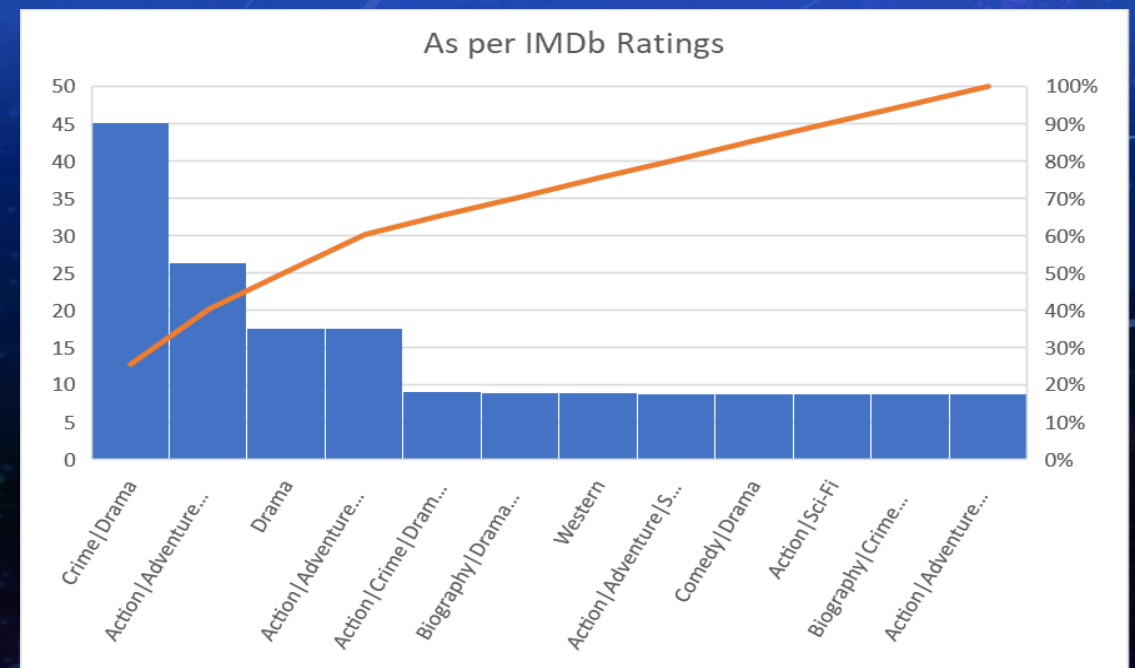
## Popular Genres

To carry out this task;

- ❑ Similar method was followed from the previous task but in this case we inserted genres against imdb ratings and voted users

### Drive link:

[https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr\\_B04Wz6mJiTiyFloZST/edit?usp=sharing&ouid=105000439991683267564&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr_B04Wz6mJiTiyFloZST/edit?usp=sharing&ouid=105000439991683267564&rtpof=true&sd=true)



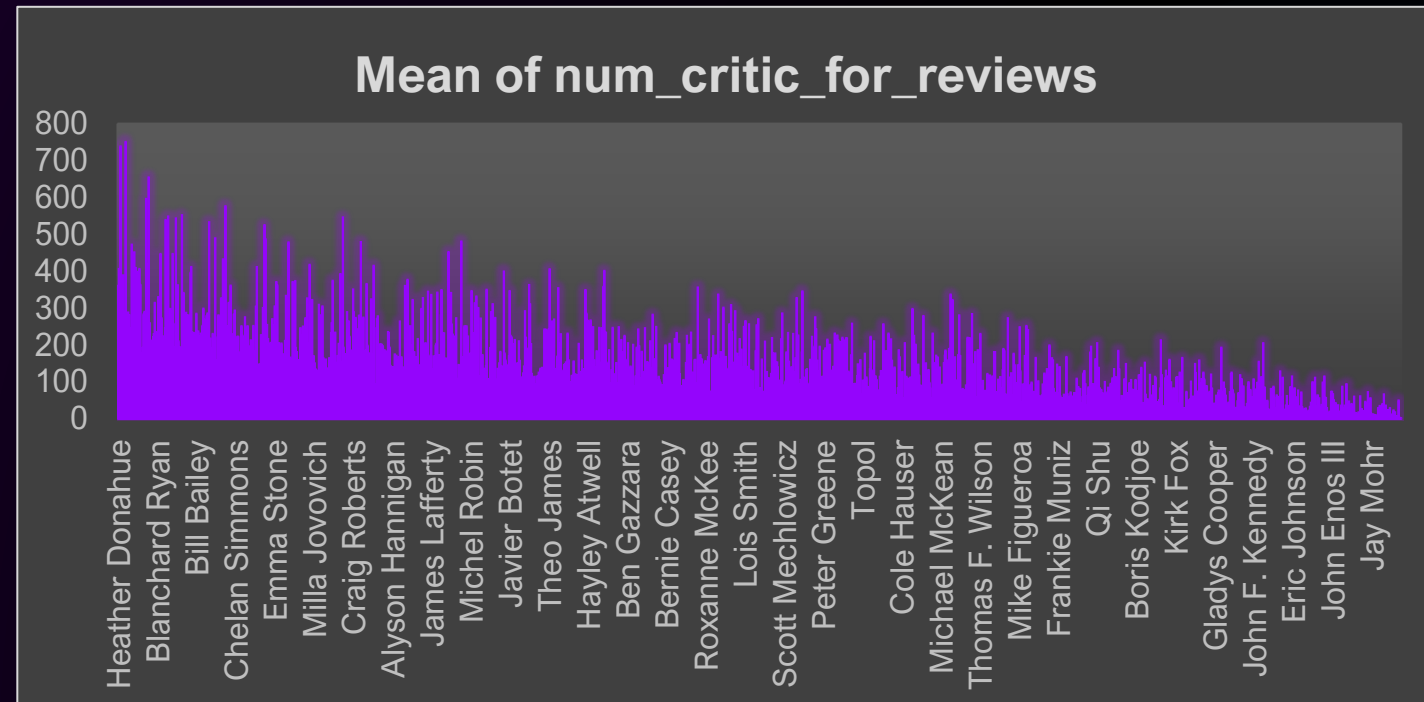
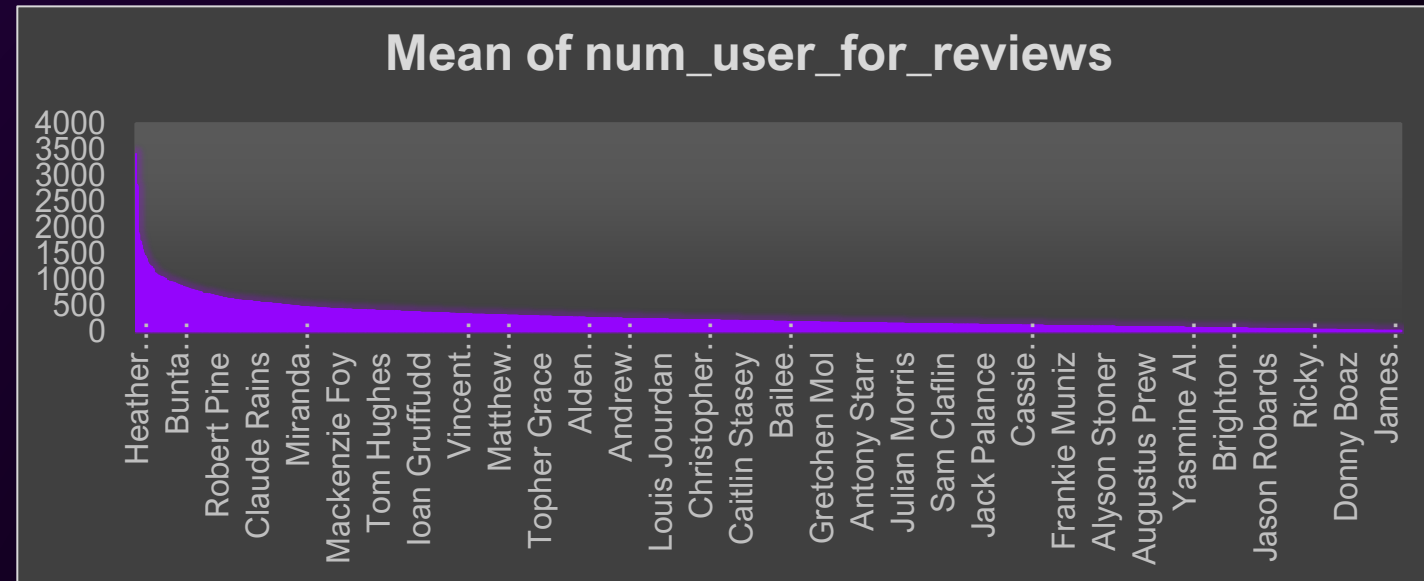
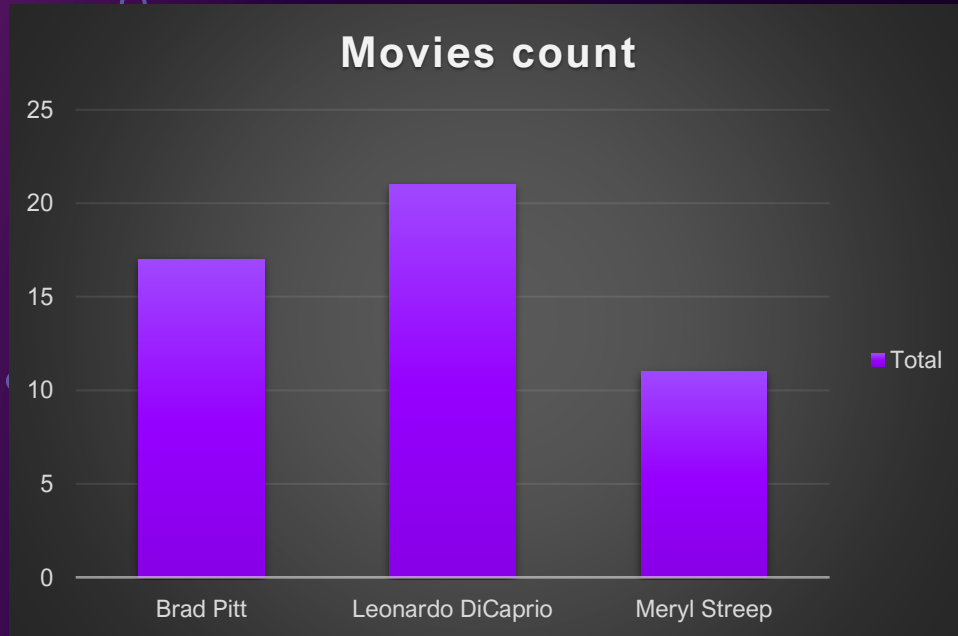


# TASK 6

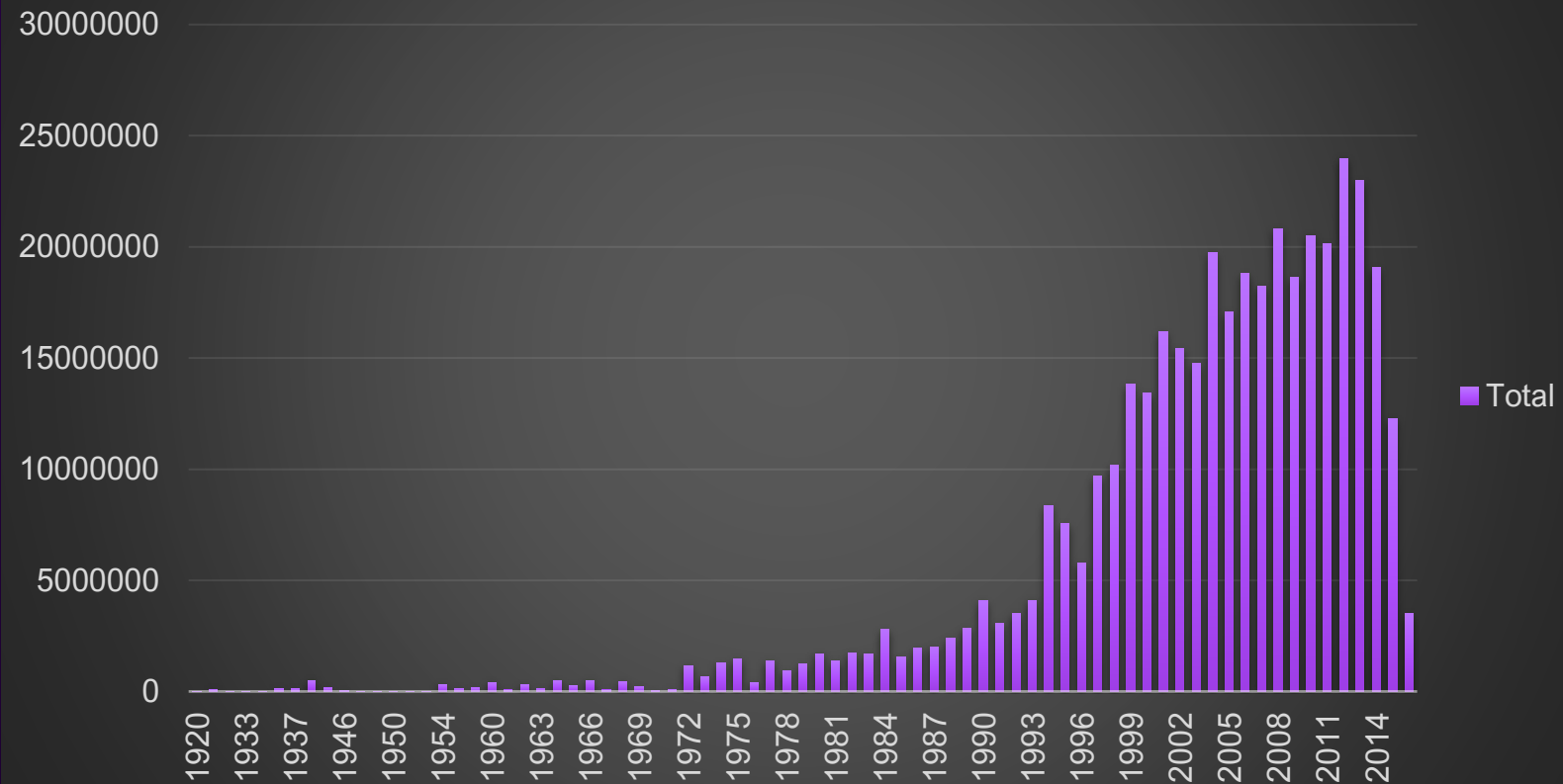
## Charts

To carry out this task;

- ❑ Similar method was followed from the previous task using the pivot table option from the excel.



## Voted users over the years



**Drive link:** [https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr\\_B04Wz6mJiTtYFloZST/edit?usp=sharing&ouid=105000439991683267564&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1tIS5xhCFw-0Hr_B04Wz6mJiTtYFloZST/edit?usp=sharing&ouid=105000439991683267564&rtpof=true&sd=true)



Thank you