

Drive Link for the project-

[https://drive.google.com/drive/u/0/folders/1z6azyOrtJhOmBiMGLIH\\_SBJNiug1vEQF](https://drive.google.com/drive/u/0/folders/1z6azyOrtJhOmBiMGLIH_SBJNiug1vEQF)

Presentation Link for the project-

<https://www.loom.com/share/1bafdcfd8e3944ab96887586103ef812>

Project by –

Mohd Faiz

faiz.mohd340@gmail.com

# IMDB Movie Analysis

## Project Description

This project will go through to find some of the insights on IMDB movies like most popular actor, most popular director etc.

## Project Approach

In order to find the insight, Excel was used. Using Excel, we performed Exploratory Analysis first like understanding data columns, checking for missing data, checking and removing outliers, etc. After that, we performed various operations like selecting, sorting, applying different formulas, etc. to get the insights we needed.

## Tech Stack Used

Microsoft Excel 2021, Microsoft Word 2021, and Google Drive.

## Project Insight

1. **Cleaning the data:** At first, removed all the rows which were empty. Removed & from all movie titles. Removed Duplicate values as well. Hide the columns which were not going to be used for finding insights.
2. **Movies with the highest profit:** Find the movies with the highest profit.

### Excel Formula-

Created Profit column by subtracting gross with budget using formula-

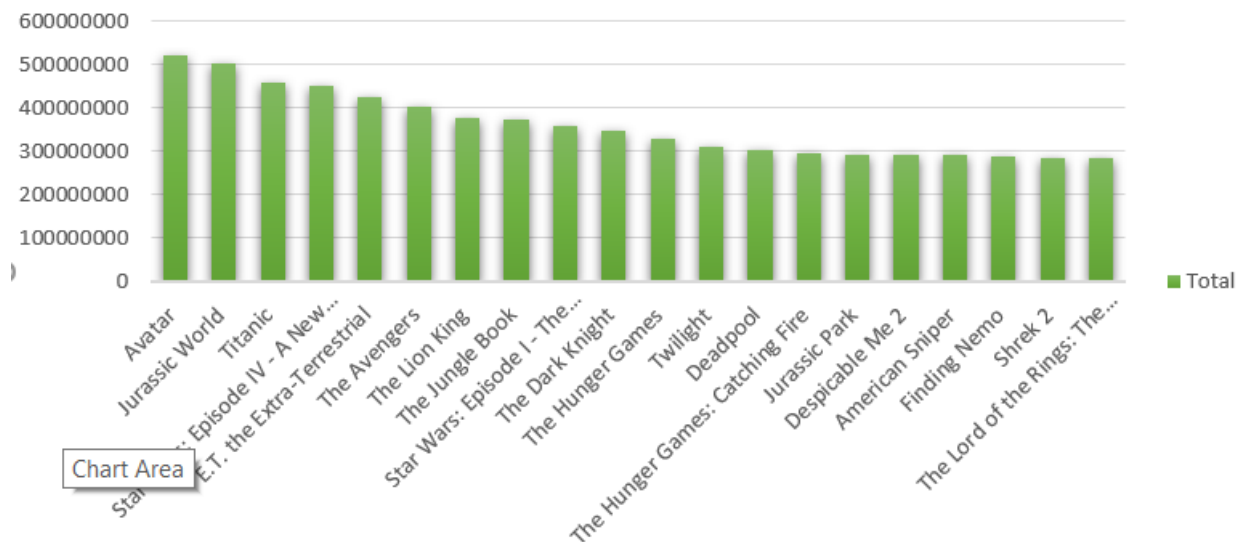
=I2-W2

Created Pivot Table and found out the top 20 movies. Created Chart for the bet

### Result-

Row Labels	Sum of 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

**Top 20 highest profitable movies**



movie\_title

### 3. Top 250: Find IMDB top 250.

#### Excel Formula-

Applied filters based on the requirement and extracted columns which needed in a different sheet.

#### Result-

A new sheet has been created to show IMDB's top 250 movies with votes>25000 and IMDB's Top Foreign language film

A	B	C	D	E	F	H	I	J	K	L
rank	movie_title	num_voted_users>25000	imdb_score	language		rank	top_foreign_language_film	num_vote	imdb_scor	language
1	The Shawshank Redemption	1689764	9.3	English		1	The Good, the Bad and the Ugly	503509	8.9	Italian
2	The Godfather	1155770	9.2	English		2	City of God	533200	8.7	Portuguese
3	The Dark Knight	1676169	9	English		3	Seven Samurai	229012	8.7	Japanese
4	The Godfather: Part II	790926	9	English		4	Spirited Away	417971	8.6	Japanese
5	The Lord of the Rings: The Return of the King	1215718	8.9	English		5	The Lives of Others	259379	8.5	German
6	The Good, the Bad and the Ugly	503509	8.9	Italian		6	Children of Heaven	27882	8.5	Persian
7	Pulp Fiction	1324680	8.9	English		7	A Separation	151812	8.4	Persian
8	Schindler's List	865020	8.9	English		8	Oldboy	356181	8.4	Korean
9	Forrest Gump	1251222	8.8	English		9	Das Boot	168203	8.4	German
10	Star Wars: Episode V - The Empire Strikes Back	837759	8.8	English		10	Amélie	534262	8.4	French
11	The Lord of the Rings: The Fellowship of the Ring	1238746	8.8	English		11	Princess Mononoke	221552	8.4	Japanese
12	Inception	1468200	8.8	English		12	The Hunt	170155	8.3	Danish
13	Fight Club	1347461	8.8	English		13	Metropolis	111841	8.3	German
14	Star Wars: Episode IV - A New Hope	911097	8.7	English		14	Downfall	248354	8.3	German
15	The Lord of the Rings: The Two Towers	1100446	8.7	English		15	Pan's Labyrinth	467234	8.2	Spanish
16	The Matrix	1217752	8.7	English		16	The Secret in Their Eyes	131831	8.2	Spanish
17	One Flew Over the Cuckoo's Nest	680041	8.7	English		17	Incendies	80429	8.2	French
18	City of God	533200	8.7	Portuguese		18	Howl's Moving Castle	214091	8.2	Japanese
19	Seven Samurai	229012	8.7	Japanese		19	Amores Perros	173551	8.1	Spanish
20	Goodfellas	728685	8.7	English		20	The Celebration	65951	8.1	Danish
21	Saving Private Ryan	881236	8.6	English		21	Elite Squad	81644	8.1	Portuguese
22	The Silence of the Lambs	887467	8.6	English		22	The Sea Inside	64556	8.1	Spanish
23	Se7en	1023511	8.6	English		23	Tae Guk Gi: The Brotherhood of War	31943	8.1	Korean
24	Spirited Away	417971	8.6	Japanese		24	Akira	106160	8.1	Japanese
25	Interstellar	928227	8.6	English		25	A Fistful of Dollars	147566	8	Italian
26	The Usual Suspects	740918	8.6	English		26	Central Station	28951	8	Portuguese
27	American History X	782437	8.6	English		27	Waltz with Bashir	46107	8	Hebrew
28	Modern Times	143086	8.6	English		28	Persepolis	70194	8	French
29	The Lion King	241220	8.6	English		29	My Neighbor Totoro	63750	8.1	Japanese

4. **Best Directors:** The top 20 directors for whom the mean of imdb\_score is the highest. The best directors list is shown.

#### Excel Formula-

Done using Pivot table. The grouping feature was used to find the mean of imdb score.

#### Result-

3	Row Labels	Average of imdb_score
4	Akira Kurosawa	8.7
5	Tony Kaye	8.6
6	Charles Chaplin	8.6
7	Alfred Hitchcock	8.5
8	Majid Majidi	8.5
9	Ron Fricke	8.5
10	Damien Chazelle	8.5
11	Sergio Leone	8.433333333
12	Christopher Nolan	8.425
13	Asghar Farhadi	8.4
14	Richard Marquand	8.4
15	Lenny Abrahamson	8.3
16	Lee Unkrich	8.3
17	Billy Wilder	8.3
18	Fritz Lang	8.3



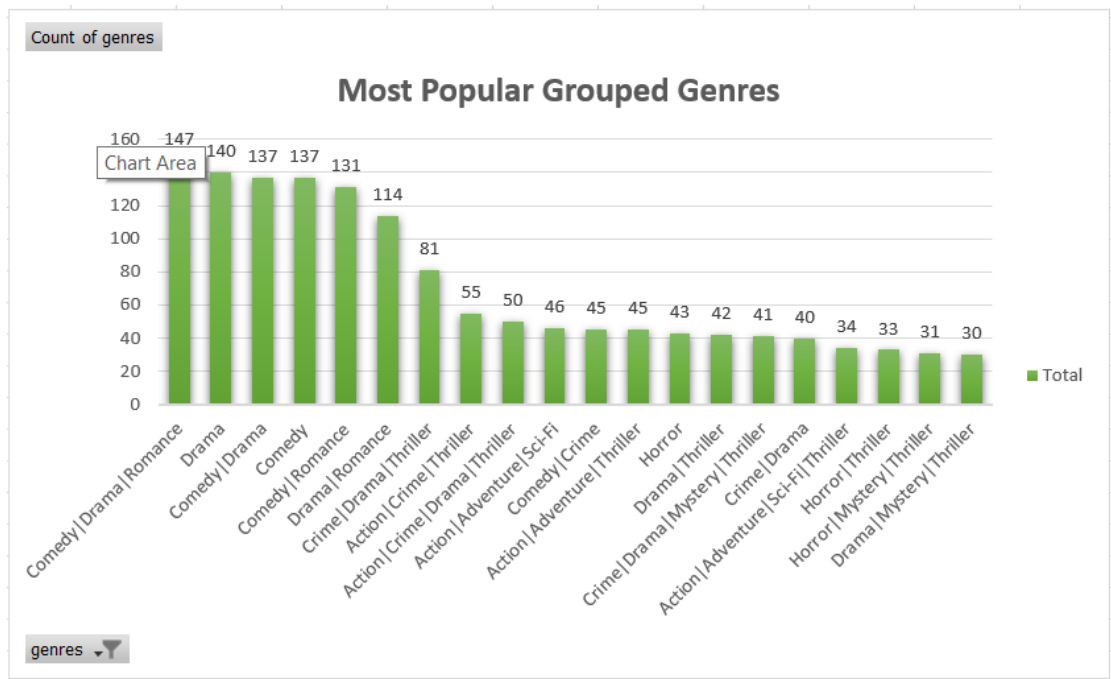
## 5. Popular Genres: Find the most popular genres.

### Excel Formula-

Done using Pivot table. The grouping feature was used to find the count of grouped genres.

### Result-

Row Labels	Count of genres
Comedy Drama Romance	147
Drama	140
Comedy Drama	137
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 Mystery Thriller	41
Crime Drama	40
Action Adventure Sci-Fi Thriller	34
Horror Thriller	33
Horror Mystery Thriller	31
Drama Mystery Thriller	30



## 6. Charts: Used charts and graphs to find different-different insights.

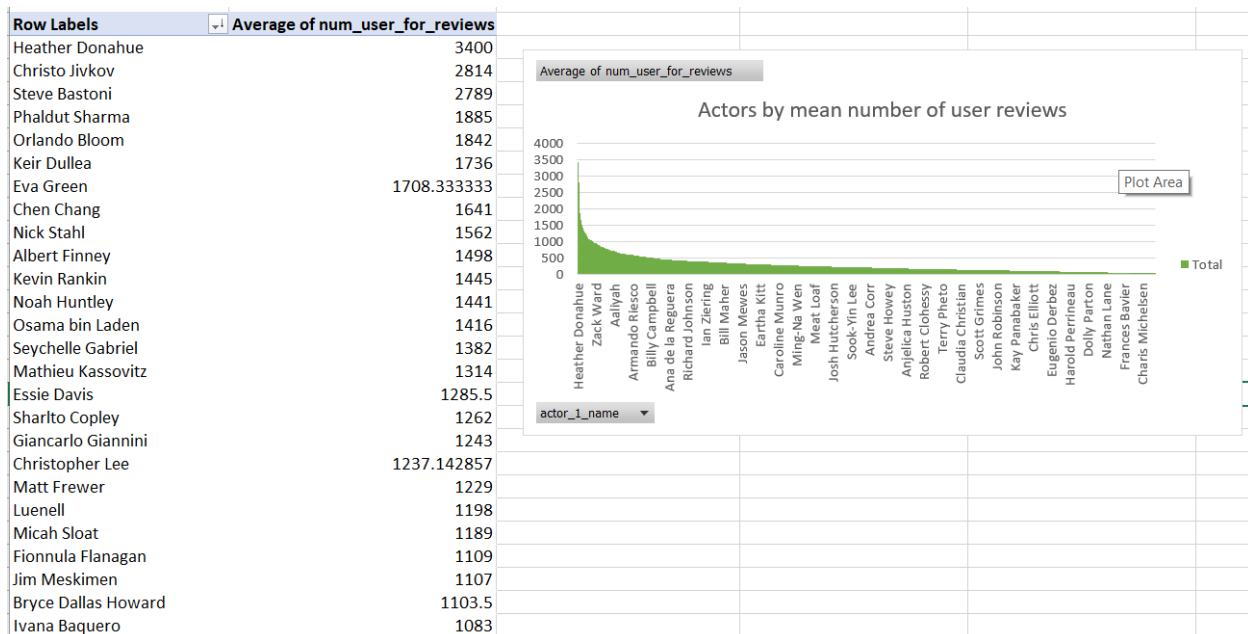
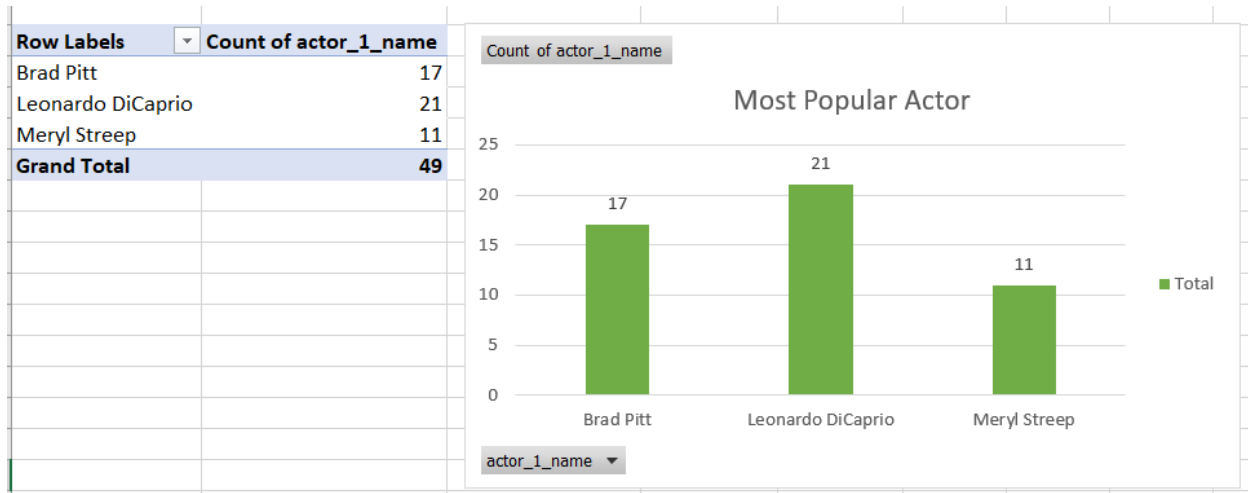
### Excel Formula-

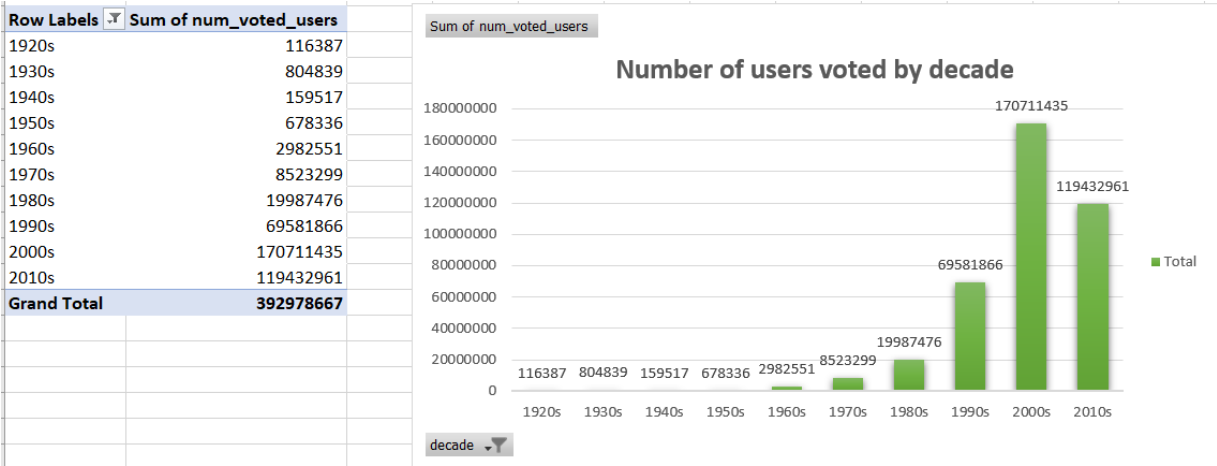
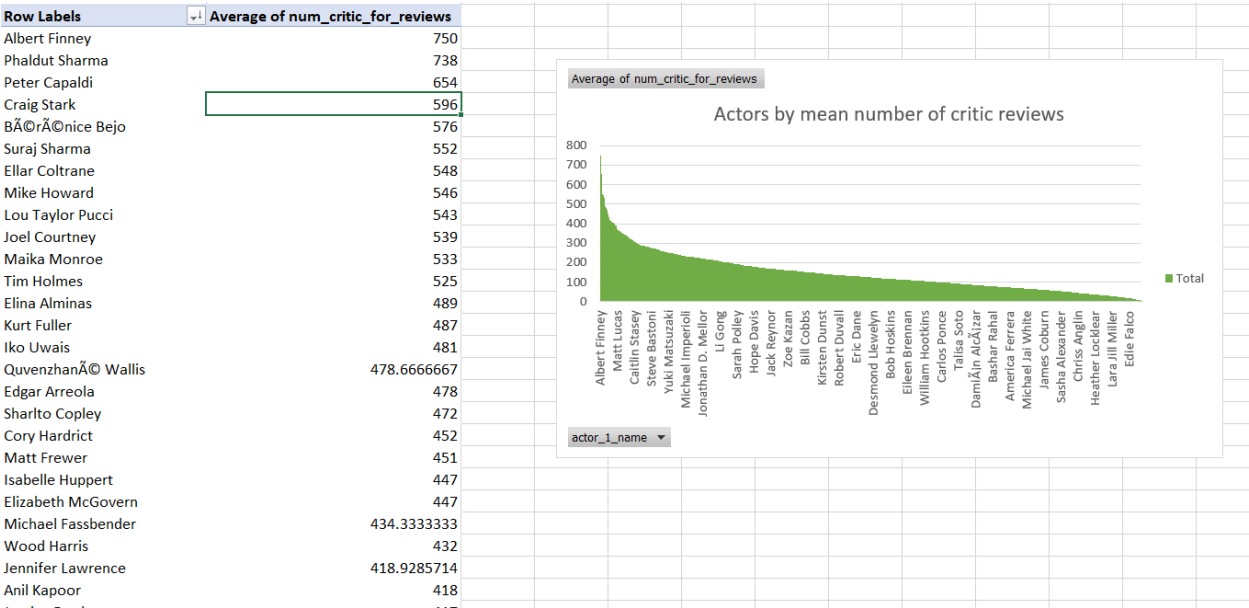
Pivot Tables and Graphs were used to find these insights.

For finding the decade, I used the formula:

`=LEFT(Y2,3)&"0s"`

### Result-







## Project Conclusion

While analyzing the data set provided, several meaningful insights were discovered that could not have been discovered by manually searching the dataset for insights.

We could also leverage the Excel-2021 tool and got a little more experienced in using the tool and also injecting different formulas and pivot tables and graphs to look for insights.