IMDB Movie Analysis

Project Description:

For your Final Project, we had provided with dataset having various columns of different IMDB Movies. We had to Frame the problem. For this task, we need to define a problem we want to shed some light on.

Approach:

The data set and the setup instructions were given using that, working on EXCEL and first and foremost thing was to extract the databases and cleaning process need to be done. Using the advance concepts of the excel made me understand the tool even deeper.

- Dropping the columns which was not needed and removing the null points was our priority. Movies with highest profit was one of the questions answered. With the help of the chart outlier points gave me the idea that the movies given has a greater number of movies which made less profit.
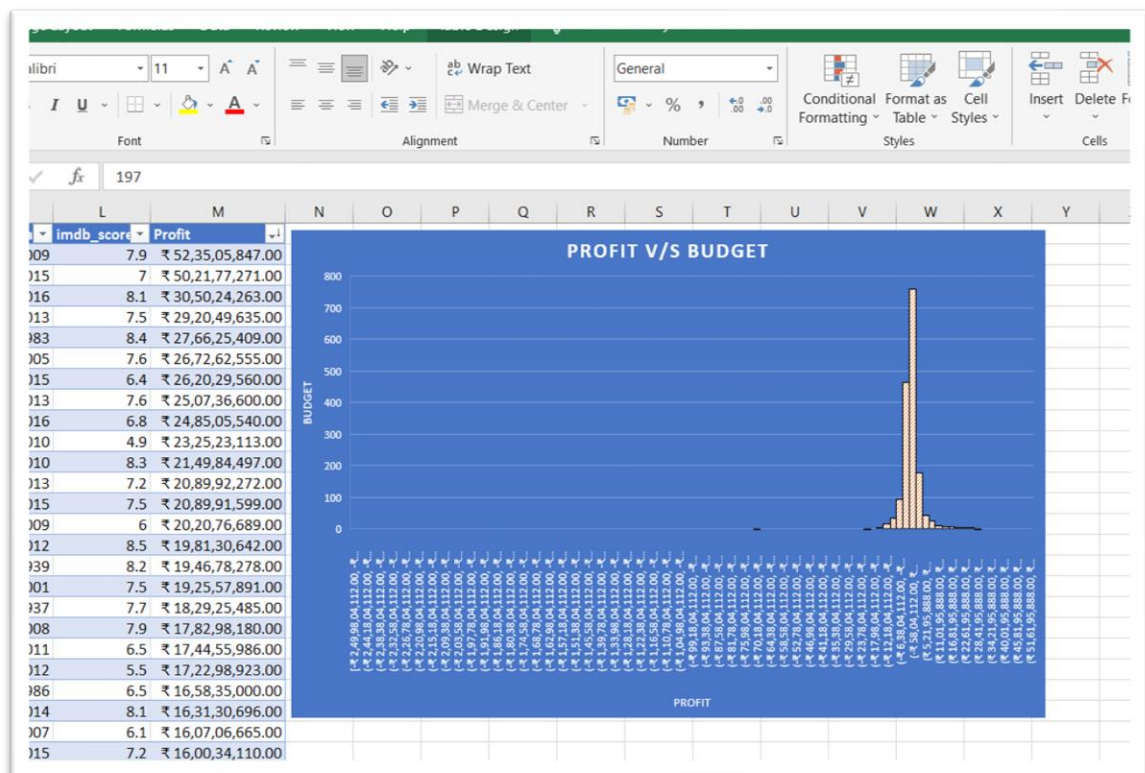

Fig1. The profit chart.

- Finding out the top 250 movies on basis of imdb_scores. The left side table is the list of top 250 movies based on imdb_scores. The right-side table is the list of top 250 movies based on imdb_scores and language (except English).

mdb_score	num_voted_users	Top_Foreign_Lang_Film		language	imdb_score
8.1	491077	Nine QueensÂ		Spanish	7.9
8.1	152089	The ChorusÂ		French	7.9
8	351274	The Red ViolinÂ		French	7.7
8	307639	Good Bye Lenin!Â		German	7.7
7.9	279093	March of the PenguinsÂ		French	7.6
7.8	249688	The Legend of Drunken MasterÂ		Cantonese	7.6
7.6	22105	Instructions Not IncludedÂ		Spanish	7.6
7.3	145422	Jab Tak Hai JaanÂ	TOP 250 ON THE BASIS	Hindi	6.9
7.1	47794	Rumble in the BronxÂ	OF LANGUAGE (EXCEPT	Cantonese	6.7
7	41313	The Secret in Their EyesÂ	ENGLISH)	Spanish	8.2
6.7	173530	The InterpreterÂ		Aboriginal	6.4
6.8	11088	El crimen del padre AmaroÂ		Spanish	6.8
6.6	82331	Maria Full of GraceÂ		Spanish	7.5
6.6	93790	The LunchboxÂ		Hindi	7.8
7.2	12411	Of Gods and MenÂ		French	7.2
5.5	17169	Casa de mi PadreÂ		Spanish	5.5
6.7	2720	The Holy GirlÂ		Spanish	6.7
5.8	1010	One to AnotherÂ		French	5.8
6.4	117096	L'auberge espagnoleÂ		French	7.3
7.2	4391	Mississippi MermaidÂ		French	7.2
7.5	12937	Sarah's KeyÂ		French	7.5
5.9	5699	Sleep DealerÂ		Spanish	5.9
6.3	112167	The Triplets of BellevilleÂ		French	7.8
6.3	40362	TimecrimesÂ		Spanish	7.2
7.7	24921	The Barbarian InvasionsÂ		French	7.7
6.8	1024	La otra conquistaÂ		Spanish	6.8
6.3	56402	PersepolisÂ		French	8
6.3	30325	Kung Fu HustleÂ		Cantonese	7.8
7.2	15267	Le HavreÂ		French	7.2
7.7	9913	Mondays in the SunÂ		Spanish	7.7

| IMDB_Movies | **Top 250 Movies** | Best Directors | Popular generes | Part 1 | Part 2 | Part 3 | ⊕ |

Accessibility: Investigate

Fig2. A snippet of Top 250 Movies.

- Giving the top 10 best directors based on highest mean value of imdb_scores mean.

director_name	imdb_score
Akira Kurosawa	8.7
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Christopher Nolan	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Quentin Tarantino	8.5
Ron Fricke	8.5
Richard Marquand	8.4

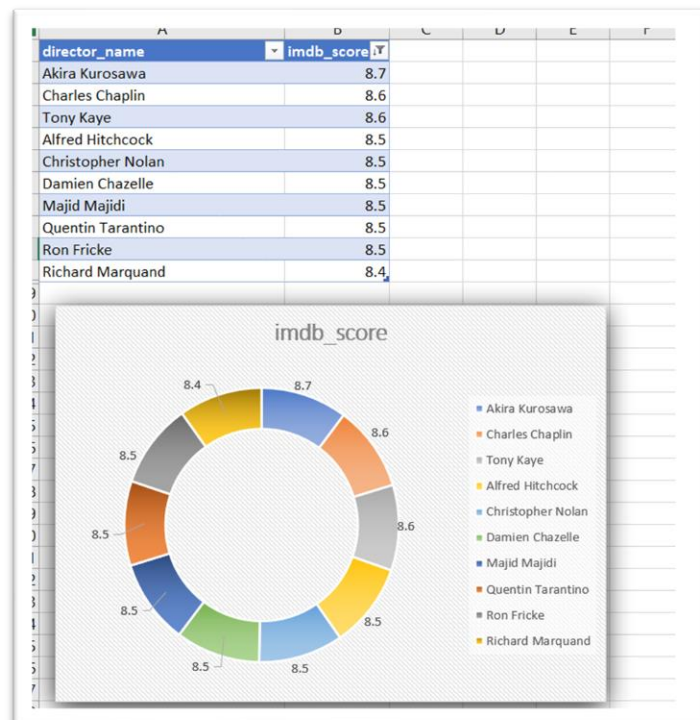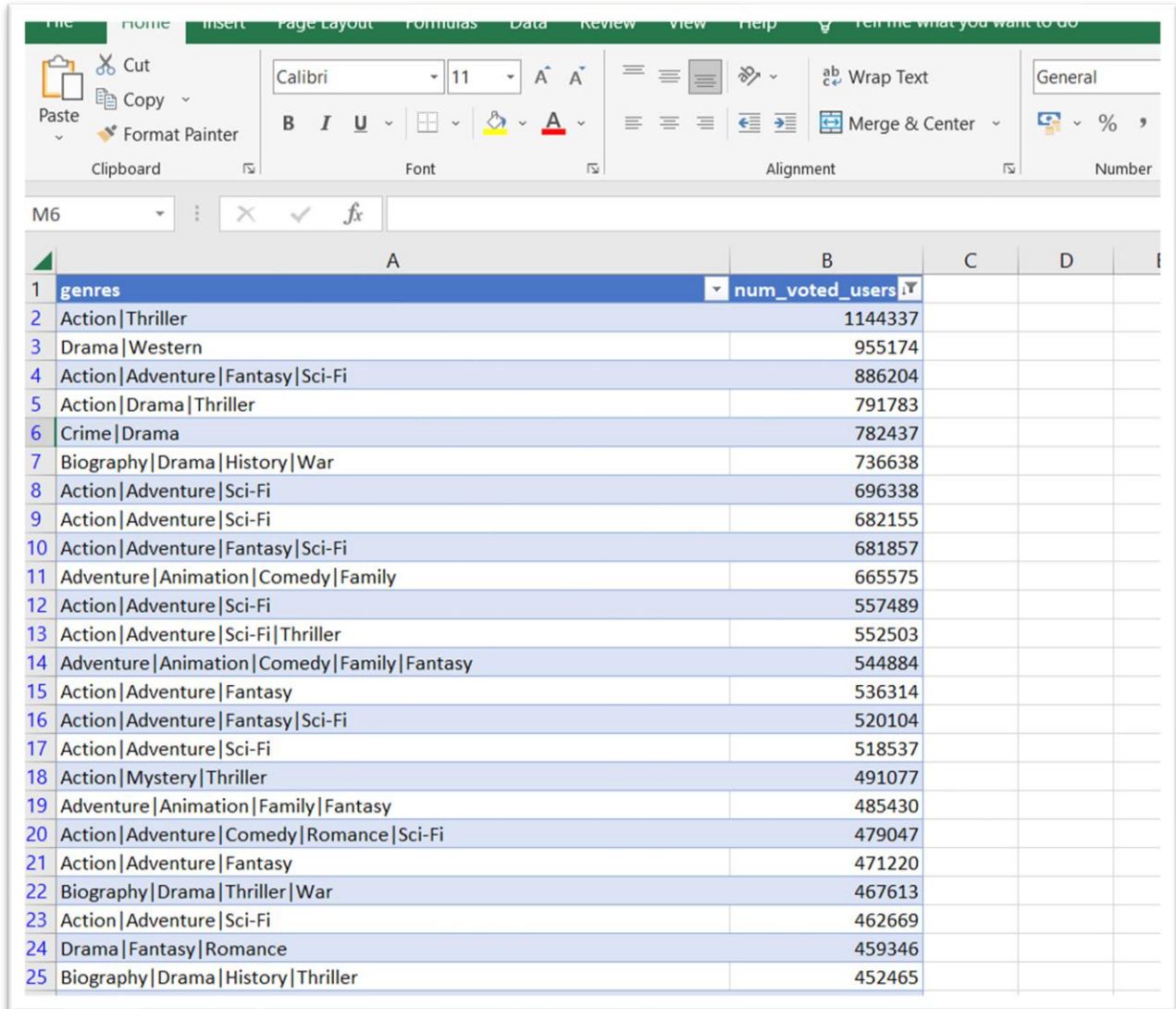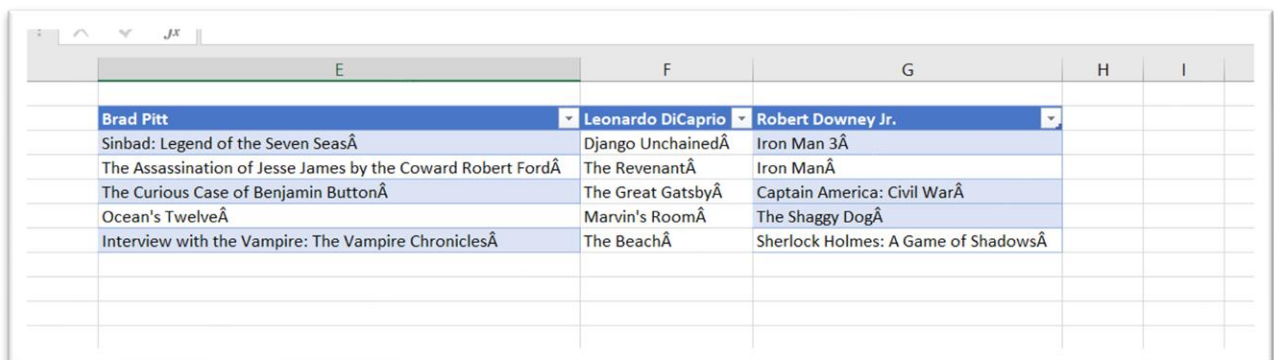*Fig3. A snippet of Top 10 directors.*

- Considering the num_of_users_voted I was able to give the popular genres.

genres	num_voted_users
Action|Thriller	1144337
Drama|Western	955174
Action|Adventure|Fantasy|Sci-Fi	886204
Action|Drama|Thriller	791783
Crime|Drama	782437
Biography|Drama|History|War	736638
Action|Adventure|Sci-Fi	696338
Action|Adventure|Sci-Fi	682155
Action|Adventure|Fantasy|Sci-Fi	681857
Adventure|Animation|Comedy|Family	665575
Action|Adventure|Sci-Fi	557489
Action|Adventure|Sci-Fi|Thriller	552503
Adventure|Animation|Comedy|Family|Fantasy	544884
Action|Adventure|Fantasy	536314
Action|Adventure|Fantasy|Sci-Fi	520104
Action|Adventure|Sci-Fi	518537
Action|Mystery|Thriller	491077
Adventure|Animation|Family|Fantasy	485430
Action|Adventure|Comedy|Romance|Sci-Fi	479047
Action|Adventure|Fantasy	471220
Biography|Drama|Thriller|War	467613
Action|Adventure|Sci-Fi	462669
Drama|Fantasy|Romance	459346
Biography|Drama|History|Thriller	452465

Fig4. A snippet of popular genres.

- The final part was a bit challenging, but the knowledge gained by above process was able give the proper output as expected.

Brad Pitt	Leonardo DiCaprio	Robert Downey Jr.
Sinbad: Legend of the Seven SeasÂ	Django UnchainedÂ	Iron Man 3Â
The Assassination of Jesse James by the Coward Robert FordÂ	The RevenantÂ	Iron ManÂ
The Curious Case of Benjamin ButtonÂ	The Great GatsbyÂ	Captain America: Civil WarÂ
Ocean's TwelveÂ	Marvin's RoomÂ	The Shaggy DogÂ
Interview with the Vampire: The Vampire ChroniclesÂ	The BeachÂ	Sherlock Holmes: A Game of ShadowsÂ

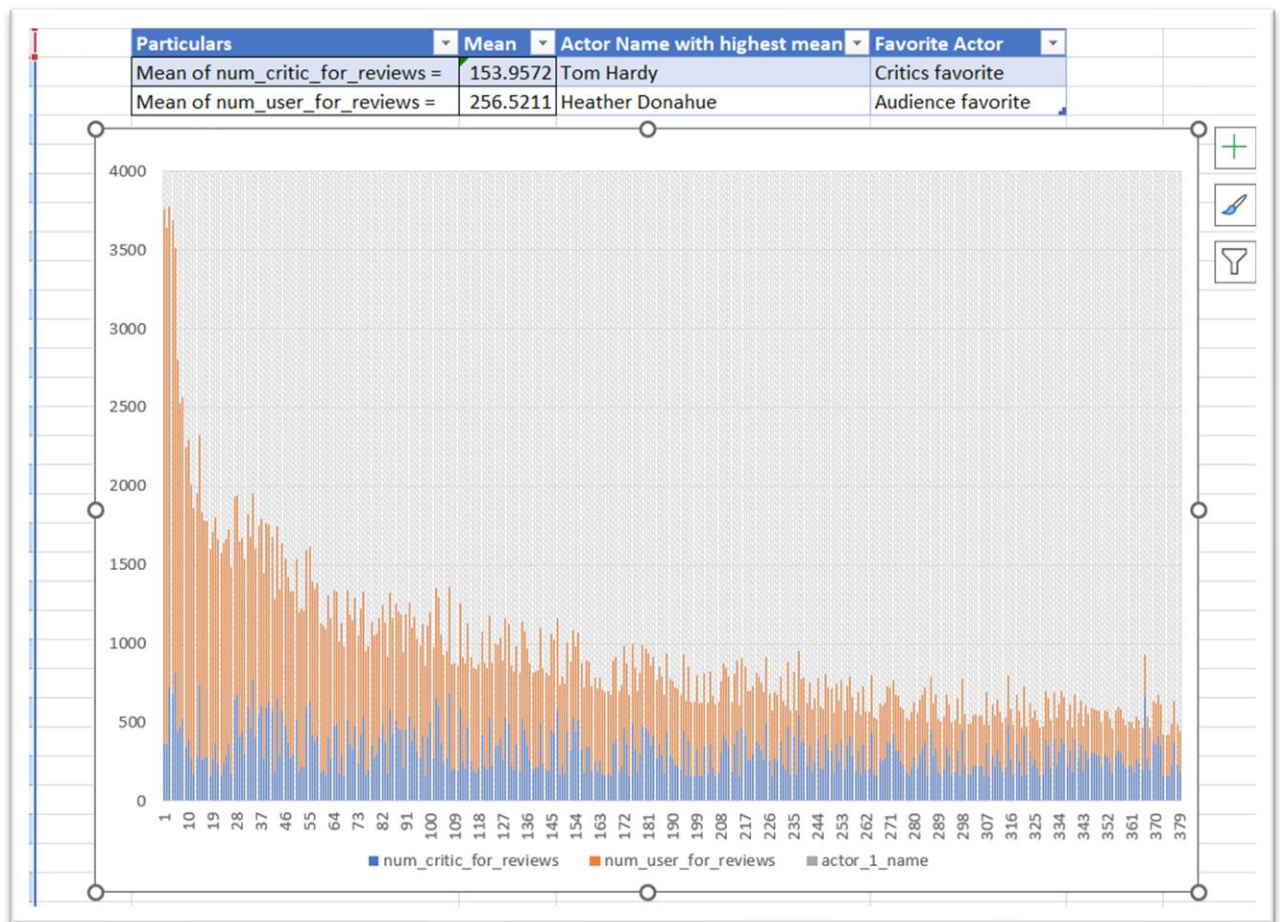*Fig5. A snippet of actors and the movies that they are in the lead.*

Particulars	Mean	Actor Name with highest mean	Favorite Actor
Mean of num_critic_for_reviews =	153.9572	Tom Hardy	Critics favorite
Mean of num_user_for_reviews =	256.5211	Heather Donahue	Audience favorite

.

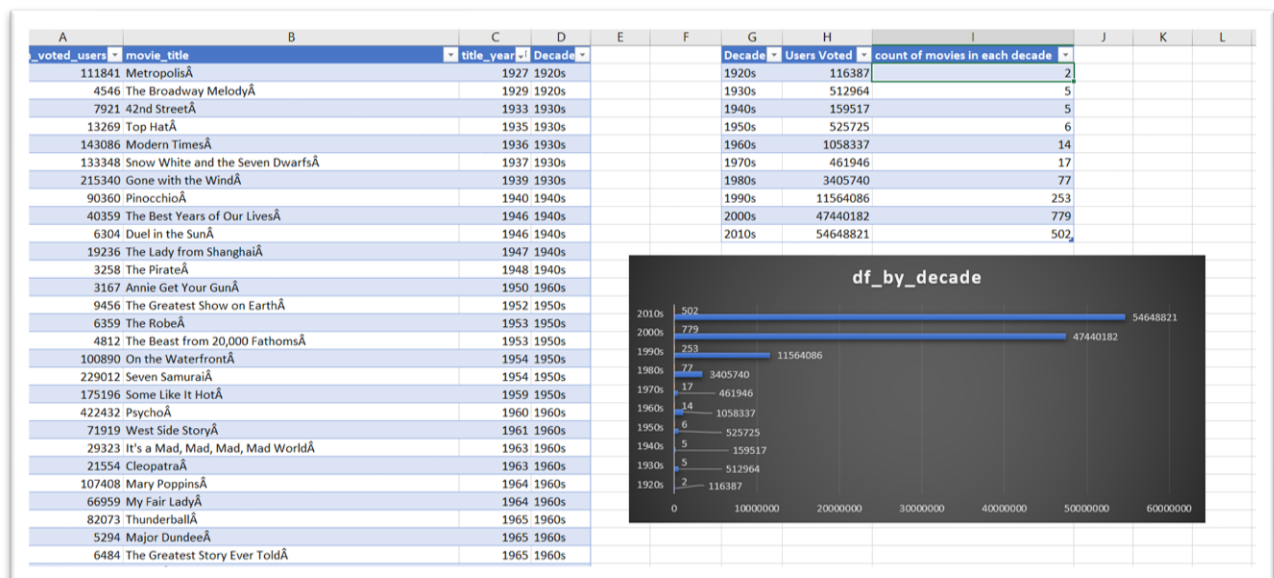**Fig6. A snippet of famous actor according to the critics and audience.**

Fig7. A snippet of users voted in each decade.

Tech-Stack Used:

MICROSOFT EXCEL 360

Insights:

While doing the analysis the main thing is to understand the given database, extracting them accordingly and finding out the output was the agenda, the knowledge on excel by getting trained helped me a lot in clearing all the errors and implementing it helped me to improve my knowledge towards the subject.

IMDB_Movies.xlsx

THE COMPLETE EXCEL FILE (The file is also in the drive link)

From the above file, I was able to find answers for each and questions asked, I was able to get the data from the data set given and exported those output's and I was able to build a visualization for easy understanding.

Result:

Each case study made me understand the concepts even deeper and in the more elaborating way.