IDMB Movies Analysis

Introduction:

From the given data from IDMB_movies we have to see the problems and how to solve that problems and use this information to create graphs, charts, and other visualizations to help understand trends and patterns in the data. For example, we could use IMDb data to create a graph showing the average rating of movies released in a particular year, or to compare the idmb_score, ratings and performance of different movie genres.

Approach:

I prefer using **Pivot table** and **pivot charts** for getting the solution and somewhere I have used formulas like **if, averageifs and INDEX and MATCH .**There were problems that were difficult to solve in excel 2016.

Tech-Stack Used :

I used Excel 2016 to analyze the data .

Insights:

Cleaning the data:: This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)
Your task: Clean the data

I removed some columns and null values that were not required.

color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gro
Color	James Cameron	723	178	0	855	Joel David Moore	1000	760
Color	Gore Verbinski	302	169	563	1000	Orlando Bloom	40000	309
Color	Sam Mendes	602	148	0	161	Rory Kinnear	11000	200
Color	Christopher Nolan	813	164	22000	23000	Christian Bale	27000	448
Color	Andrew Stanton	462	132	475	530	Samantha Morton	640	73
Color	Sam Raimi	392	156	0	4000	James Franco	24000	336
Color	Nathan Greno	324	100	15	284	Donna Murphy	799	200
Color	Joss Whedon	635	141	0	19000	Robert Downey Jr.	26000	458
Color	David Yates	375	153	282	10000	Daniel Radcliffe	25000	301
Color	Zack Snyder	673	183	0	2000	Lauren Cohan	15000	330
Color	Bryan Singer	434	169	0	903	Marlon Brando	18000	200
Color	Marc Forster	403	106	395	393	Mathieu Amalric	451	168
Color	Gore Verbinski	313	151	563	1000	Orlando Bloom	40000	423
Color	Gore Verbinski	450	150	563	1000	Ruth Wilson	40000	89
Color	Zack Snyder	733	143	0	748	Christopher Meloni	15000	291
Color	Andrew Adamson	258	150	80	201	Pierfrancesco Favino	22000	141
Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	26000	623
Color	Rob Marshall	448	136	252	1000	Sam Claflin	40000	241
Color	Barry Sonnenfeld	451	106	188	718	Michael Stuhlbarg	10000	179
Color	Peter Jackson	422	164	0	773	Adam Brown	5000	255

IMDbMovies | Sheet1 | TopForeignLangFilm | CleanedData | HighestProfit | Top 10 Directors ...

Movies with highest profit: Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.
Your task: Find the movies with the highest profit?

MovieName	Profits
Avatar	523505847
E.T. the Extra-Terrestrial	424449459
Jurassic World	502177271
Star Wars: Episode I - The Phantom Menace	359544677
Star Wars: Episode IV - A New Hope	449935665
The Avengers	403279547
The Dark Knight	348316061
The Jungle Book	375290282
The Lion King	377783777
Titanic	458672302

Top 250: Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film. You can use your own imagination also!
Your task: Find IMDB Top 250

Top IMDB movies with English languge

movie_title	num_voted_use	language	imdb_sco	Ranking
The Shawshank Redemption	1689764	English	9.3	1
The Godfather	1155770	English	9.2	2
The Dark Knight	1676169	English	9	3
The Godfather: Part II	790926	English	9	3
The Good, the Bad and the Ugly	503509	Italian	8.9	5
Pulp Fiction	1324680	English	8.9	5
Schindler's List	865020	English	8.9	5
The Lord of the Rings: The Return of the King	1215718	English	8.9	5
The Lord of the Rings: The Fellowship of the Ring	1238746	English	8.8	9
Forrest Gump	1251222	English	8.8	9
Star Wars: Episode V - The Empire Strikes Back	837759	English	8.8	9
Inception	1468200	English	8.8	9
Fight Club	1347461	English	8.8	9
The Matrix	1217752	English	8.7	14
City of God	533200	Portuguese	8.7	14
Goodfellas	728685	English	8.7	14
The Lord of the Rings: The Two Towers	1100446	English	8.7	14
One Flew Over the Cuckoo's Nest	680041	English	8.7	14
Seven Samurai	229012	Japanese	8.7	14
Star Wars: Episode IV - A New Hope	911097	English	8.7	14

Extracted only foreign lang movie from top 250

movie_title	num_v	language	imdb_score	Ranking
The Good, the Bad and	503509	Italian	8.9	5
City of God	533200	Portuguese	8.7	14
Seven Samurai	229012	Japanese	8.7	14
Spirited Away	417971	Japanese	8.6	21
Children of Heaven	27882	Persian	8.5	29
The Lives of Others	259379	German	8.5	29
Amélie	534262	French	8.4	47
Baahubali: The Beginni	62756	Telugu	8.4	47
Oldboy	356181	Korean	8.4	47
Das Boot	168203	German	8.4	47
A Separation	151812	Persian	8.4	47
Princess Mononoke	221552	Japanese	8.4	47
Metropolis	111841	German	8.3	62
Downfall	248354	German	8.3	62
The Hunt	170155	Danish	8.3	62
Pan's Labyrinth	467234	Spanish	8.2	86
The Secret in Their Eye:	131831	Spanish	8.2	86
Incendies	80429	French	8.2	86
Howl's Moving Castle	214091	Japanese	8.2	86
Amores Perros	173551	Spanish	8.1	109

Best Directors: Group the column using the director_name column.
Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.
Your task: Find the best directors

DirectorName	Average of imdb_score
Akira Kurosawa	8.7
Alfred Hitchcock	8.5
Asghar Farhadi	8.4
Charles Chaplin	8.6
Christopher Nolan	8.425
Damien Chazelle	8.5
Majid Majidi	8.5
Richard Marquand	8.4
Ron Fricke	8.5
Sergio Leone	8.433333333
Tony Kaye	8.6

Popular Genres: Perform this step using the knowledge gained while performing previous steps.
Your task: Find popular genres

The popular genres are Action|Drama|Mystery|Thriller

Charts: Create three new columns namely, Meryl_Streep, Leo_Caprio, and Brad_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined.

Group the combined column using the actor_1_name column.

Find the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean.

Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df_by_decade.

Your task: Find the critic-favorite and audience-favorite actors

actor_1_name	(Multiple Items)	
Movies	**Meannum_critic_for_reviews**	**Mean_num_user_for_reviews**
Django Unchained	765	1193
Fight Club	315	2968
Gangs of New York	233	1166
Inception	642	2803
The Departed	352	2054
The Revenant	556	1188
The Tree of Life	584	975
The Wolf of Wall Street	606	1138
Titanic	315	2528
Troy	220	1694