

IMDB Movies

Project Description: The overview of the project is to analyze the given data to know which director has the successful means excellent ratings of IMDB, which genre movies are high and the profit of the movies and so on. Here I am working to get the best results of given data.

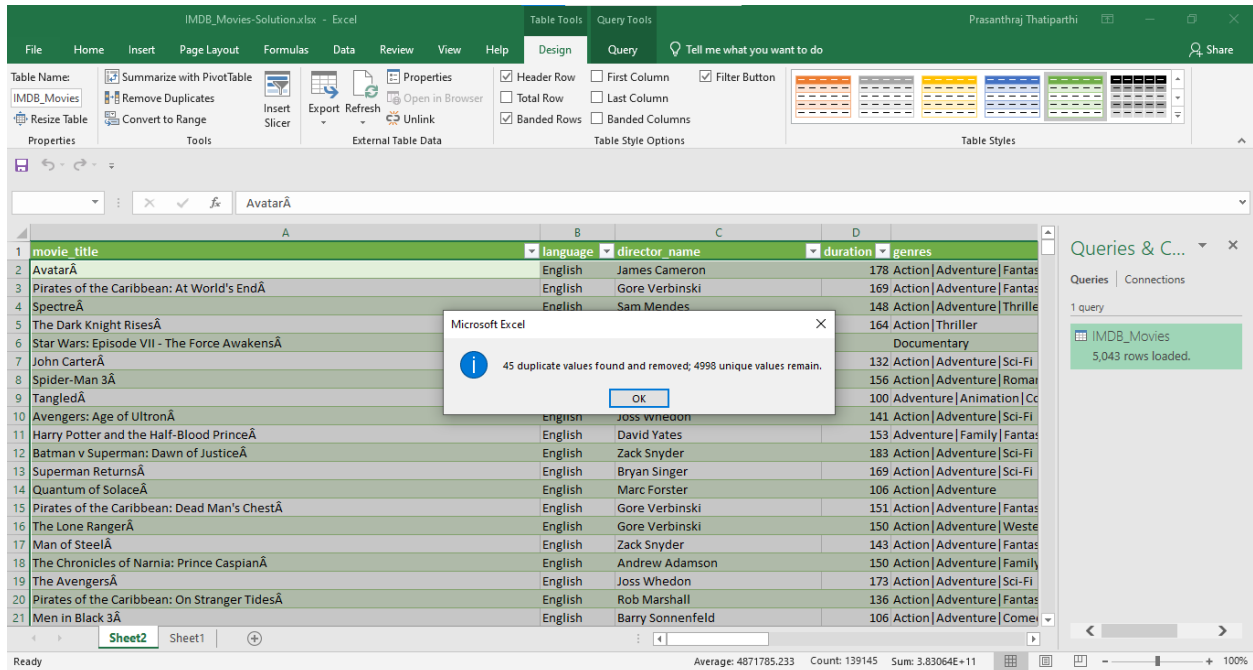
Tech-Stack Used: Here I am using MS Office 365.

Approach:

First of all cleaned the data.

Data Cleaning:

Removing Duplicates:



- Along with this deleted 12 of the null values records with no language.
- Deleted 102 records which did not have the director's name.
- Deleted 12 records which doesnot have the duration.
- Inserted 382 random records for budget using min and maximum budgets.

1	language	director_name	duration	genres	budget	color	num_critc_for_reviews	director_facebook_likes	actor_3_facebook_likes
61	German	Tom Tykwer	119	Comedy Drama Romance	\$697,604,372.00	Color		76	
71	English	Doug Block	90	Documentary	\$11,539,614,627.00	Color		45	
85	English	Mitchell Altieri	87	Comedy Horror Thriller	\$2,463,316,291.00	Color			
120	English	Christopher Guest	91	Comedy Music	\$1,027,507,313.00	Color		97	
137	English	Shane Meadows	90	Comedy Drama	\$5,593,476,445.00	Color		29	
153	English	Ben Stassen	88	Adventure Animation Family	\$12,022,631,787.00	Color		56	
161	Mandarin	Yimou Zhang	95	Comedy Drama	\$11,714,893,402.00	Color		101	
165	English	John Laing	86	Drama	\$1,445,475,165.00	Color			
166	Hindi	Remo	160	Drama Musical	\$11,668,444,882.00	Color		15	
184	English	Max Mayer	99	Drama Romance	\$523,952,989.00	Color		122	
190	English	Greg Mottola	107	Comedy Drama Romance	\$1,920,232,001.00	Color		248	
206	Arabic	Scandar Copti	124	Crime Drama	\$7,450,221,640.00	Color		100	
207	English	Doug Atchison	112	Drama	\$8,908,451,605.00	Color		95	
235	English	Andrew Jarecki	101	Crime Drama Mystery Romance Thriller	\$11,560,799,918.00	Color		140	
237	English	Phil Morrison	107	Comedy Drama	\$2,814,923,879.00	Color		41	
241	English	Bob Fosse	123	Comedy Drama Music Musical	\$2,873,826,019.00	Color		84	
267	English	Patrick Gilles	90	History	\$9,179,003,843.00	Color			
284	English	Shari Springer Berman	101	Biography Comedy Drama	\$4,916,565,308.00	Color		180	
303	English	Davis Guggenheim	96	Documentary	\$2,573,373,074.00	Color		372	
319	French	Jamel Debbouze	101	Adventure Animation Comedy Family	\$11,200,293,223.00	Color		9	

➤ For Gross I have inserted 756 random values based on minimum and maximum values.

1	duration	genres	budget	gross	color	num_critc_for_reviews	director_facebook_likes	actor_3_facebook_likes
2	101	Drama Horror Mystery Thriller	\$1,500,000.00	\$736,545,271.00	Color	35	37	
4	78	Horror	\$1,500,000.00	\$390,601,245.00	Color	252	57	
10	85	Comedy Crime Drama	\$6,000,000.00	\$106,043,621.00	Color	68	9	
11	96	Crime Drama	\$350,000.00	\$142,662,109.00	Black and White	177	0	
20	89	Comedy Romance	\$1,000,000.00	\$534,433,050.00	Color	3	32	
22	168	Drama Family History Musical	\$4,000,000.00	\$319,907,450.00	Color	34	0	
24	142	Action Comedy War	\$35,000,000.00	\$467,989,579.00	Color	56	14000	
25	90	Drama	\$1,000,000.00	\$550,952,348.00	Color	6	66	
28	96	Horror Thriller	\$3,500,000.00	\$721,941,593.00	Color	4	0	
31	127	Adventure Drama Family Fantasy Sci-Fi	\$5,000,000.00	\$291,213,814.00	Color	69	130	
32	127	Adventure Drama Family Fantasy Sci-Fi	\$5,000,000.00	\$717,274,908.00	Color	69	130	
43	96	Comedy Drama Romance Sport	\$290.00	\$730,799,750.00	Black and White	44	222	
49	88	Drama	\$300,000.00	\$577,893,357.00	Color	20	5	
57	80	Comedy	\$3,000,000.00	\$651,209,795.00	Color	29	19	
73	154	Action Adventure Drama History War	\$9,000,000.00	\$182,245,284.00	Color	15	117	
74	90	Drama Thriller	\$2,500,000.00	\$168,535,241.00	Color		37	
85	87	Comedy Horror Thriller	\$8,446,036,823.00	\$463,480,338.00	Color		9	
89	25	Animation Comedy Family	\$150,000.00	\$583,713,891.00	Color	43	36	
91	94	Comedy Family	\$4,000,000.00	\$679,569,311.00	Color	109	84	
96	88	Comedy	\$120,000.00	\$682,252,430.00	Color	8	686	

Here is the Hyperlink file:

[IMDB_Movies-Solution.xlsx](#)

Please use ctrl+click to open the document.

I have 2019 version so to work on this project I have used my official mail. If you face any issues please try to contact me through mail. Prasanthrajthatiparthi123@gmail.com

A.Movie Genre Analysis:

For movie genre to get the particular genre as the given data has the so many concatenations of genre I have used search and isnumber functions and to count them I have used

`=COUNTIFS(K2:K4807, "1")`.

Then I have used sumifs for summation of the data. Then I've calculated descriptive statistics.

And found the most used genre.

B. Movie Duration Analysis:

Here for the movie duration analysis, I have taken the duration of the movie and average idmb score for that movie duration and calculated the descriptive statistics. Along with that I have drawn the scatter plot to know in which criteria or which duration does the idmb score is frequent and high which gives the relation between duration and score.

C. Language Analysis:

Here to get the which language movies are most used

`=COUNTIF('Useful sheet'!B2:B4806, 'C. Language Analysis'!A4)`

Formula. To found the descriptive statistics. Here I have taken a mixed graph of bar and line chart to give the insights of which language has the more movies and which language has the best average of idmb rankings.

D. Director Analysis:

To find the which director is successful. Used the below formulae to find the average score of idmb.

`=ROUND(AVERAGE(FILTER('Useful sheet'!H2:H4806, ('Useful sheet'!C2:C4806='D. Director Analysis'!$A5), "")), 2)`

Then I have found percentile like 25,50,75,90,99 to find the top successful directors. To get the less directors I have used 99 percentile. For that I have used the status column who is given the best movies like good and excellent. And drawn the graph also.

E. Budget Analysis:

Based on the movie , gross and budget taken and found the profit. I have found correlation coefficient and found the movie name which has the more profit.

Please find the attached excel sheet for more information.

Insights: While working this project I have learnt new things and new formulae of excel. Most of all I have learnt how to use excel formula like subqueries in sql. Enjoyed learning and working on it. Here I have found drama genre movies are producing more. Comedy, Crime, Drama, Action, Thriller genre will boom a bit in future because we can say it has more imdb score average. Movie duration is between 90 to 140 minutes is advisable if needed can go up to 180 minutes and can be successful. When it comes to language definitely dubbing or producing in English is must. Telugu language has more imdb score also can produce in telugu also. Christopher Nolan is the best director. Making high budget korean movies will be loss so it is advisable not to invest in Korean movies.

Results: It is very much helpful to me to increase my knowledge in data analytics and I have gained much knowledge through this task. I have felt very difficult to complete this but I have done. Happy to learn this. Using of these charts and graphs with a purpose makes very helpful to be a good data analyst.

Thanks for the knowledge you are provided.