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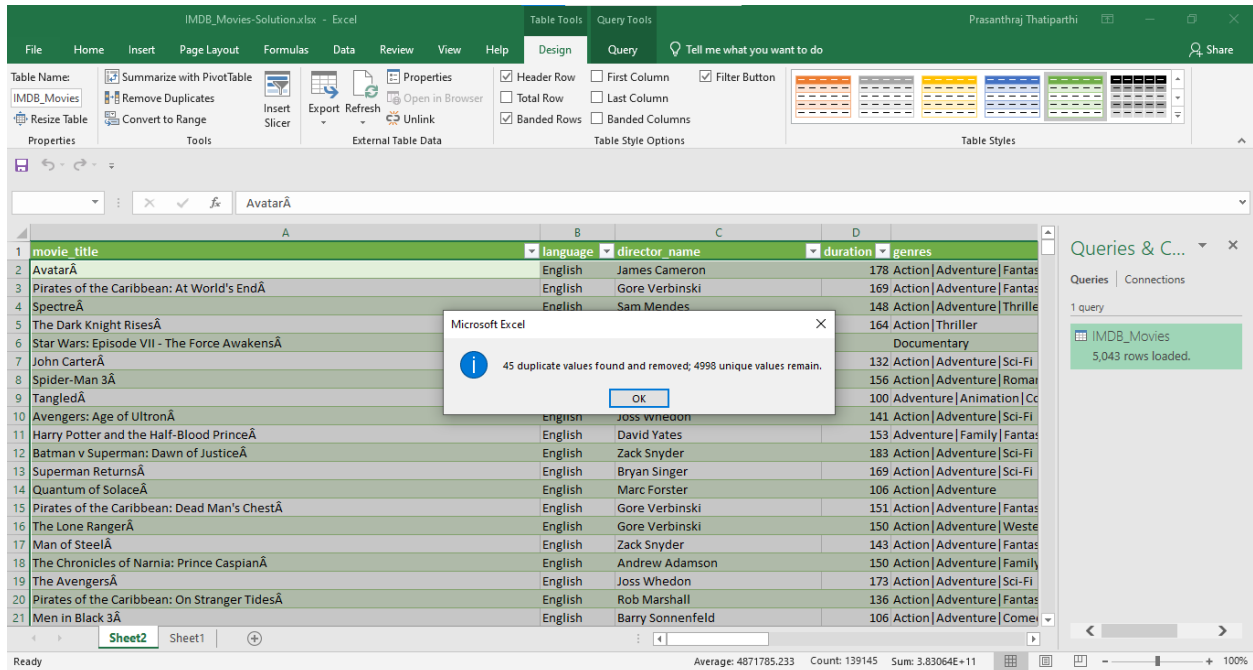
Tech-Stack Used: Here I am using MS Office 365.

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.

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

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 concatinations of genre I have used search and isnumber functions and to count them I have used '=COUNTIFS($K$2:$K$4807,"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 descritptive 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'!$B$2:$B$4806,'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'!$H$2:$H$4806,('Useful sheet'!$C$2:$C$4806='D. Director Analysis'!$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 used 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 upto 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.