**EXPLANATION FOR THE ANALYSIS OF THE MOVIE DATA SET:**

The movie data set used in this analysis is an excel file consisting of 500 rows and 15 columns initially.

The following operations have been performed on the excel data set to arrive at the necessary conclusions and also to perform the necessary analysis:

1)Firstly the gross earnings column had the missing values(Column I) and the missing values of column I have been replaced with the average value(which is calculated in the end of its column)of the gross earnings and all the values have been filled in the column later know as Gross Earnings all values (Column P).

2)Next, the budget is categorized into High, Medium and Low Budget based on the values of the Budget column(Column H) using **ifs**(multiple if conditions statement) like >1000000 its High Budget and >500000 its Medium and <500000 its Low Budget movie.(Column W)

3)The Rating classification has also been included in the data set wherein the Facebook likes, user votes and IMDB score are being considered. If the Facebook likes is >15000, the user votes is >10000 and IMDB score is>6.5 then the rating is Good else it is Average the **if and the and functions are used here**.(Column X)

4)Now the countries with the gross earnings of all the values is being displayed in the excel sheet for this a drop down of the countries(unique values) is considered by first creating a separate column for the unique countries. After that the **index function is used with the aggregate function** to retrieve all the instances of gross earnings all values when a particular country is chosen in the dropdown box. The VLOOKUP function is not used in this case as the VLOOKUP is used to retrieve only single instances of the required result and not all the multiple values.(Columns Q,R,T)

5)Next the charts are used to visualize the data and see the corresponding plots(Charts and Visualization sheet). So here the first **bar chart** is plot between a few German movies along with the user votes and the Facebook likes. And the second chart is a **line chart** that is plot between a few movies and their budget.

6)Finally the third sheet tells about the **pivot tables** through Pivot Table Analysis.

Here basically 5 pivot tables have been plotted and their descriptions are mention besides them.

The pivot table is used instead of the filter as the pivot table itself has the option for filtering the data.

The pivot analysis is done for the following cases:

* Filter by the user votes to retrieve the movie title, the country of origin of the movie, the actor and the director of the movie-So here the details for the highest and the least number of user votes is retrieved and the results are mentioned besides the pivot table analysis.(2 pivot tables totally)
* The next pivot table speaks about the movie title with respect to its duration and IMDB rating like for example in the example the movies starting with letter A have been analyzed as mentioned above.
* The movies with their respective genre and also the count of the genre are also shown in a pivot table.
* Finally a pivot table showing the movie titles with respect to the budget and its count and also the Facebook likes and its count is also been shown (columns K to Y in the top).here with first letter B have been filtered by specifying the label and value condition of less than and greater than for the value and the analysis has been performed.