

Project Title: IMBD Movies Dataset Analysis and Visualization

Acknowledgements

The dataset is downloaded from the [Kaggle](#) website and little changes were made to aid our learning purposes itself. Original dataset can be found [here](#).

Description:

The IMBD Movies Dataset consist of 1000 most popular movies in IMBD from 2006 to 2016. The data points included are: Title, Genre, Description, Director, Actors, Year, Runtime, Rating, Votes, Revenue, and Metascore.

In this project you will be analyzing the IMBD Movies dataset and finding out some possible insight.

Required Steps:

- Use the dataset attach to this file.
- Understand the dataset, types and missing values
- Clean the dataset and handle the missing values, duplicate values, etc.
- Answer the analysis questions
- Perform data visualization on dataset
- Create final summary report with conclusions

Note: This project should be done using Excel and not Power Query

Tasks:

1. Clean each data of any duplicates, inconsistency, outliers and misspelt word.
2. Create new column for "Genre" and "Director" in the **IMDB-Movie-Data** spreadsheet. Look up their corresponding values in the **Genre** and **Director** spreadsheet using VLOOKUP function.
3. Make the table more presentable by converting the dataset info an excel table.
4. Study each column for the right data type and missing value. Choose the right method to fill in missing value.

Analysis and Visualization:

Using pivot table and the appropriate visuals:

1. What is the total number of movies in the dataset?
2. What is the number of movies produced by year?
3. What is the total revenue generated?
4. What is the total revenue generated by Genres?
5. What is the top 5 voted movies?What is the top 5 most rated movies?
6. Find the top 5 profitable movies.
7. Find 5 most rated movies.
8. Show the number of Genres by percentage of total?