

# IMDB 1K Movies dataset from Kaggle.com

## Dataset Description

This dataset is about top 1k movies based on IMDB rating, Meta\_Score, No\_of\_votes, gross, overviews and genre from Kaggle. This will help to find top rated movies for people. There are 1000 rows and 16 columns with heading Poster\_Link, Series\_Title, Released\_Year, Certificate, Runtime, Genre, IMDB\_Rating, Overview, Meta\_Score, Director, Stars, No\_of\_votes, and Gross.

## Problem Statement

Analysis of last 94yrs. (i.e., from 1920 – 2014) movies data IMDB and come up with the success factors of any movie and their correlation.

## Tools & Libraries

• Python • Jupiter Notebook • Pandas • NumPy • Matplotlib

## Data Description

The dataset contains the following Columns:

- **Poster Link:** Link of the poster that IMDB using
- **Series Title:** Name of the movie
- **Released Year:** Year at which that movie released
- **Certificate:** Certificate earned by that movie
- **Runtime:** Total runtime of the movie
- **Genre:** Genre of the movie
- **IMDB Rating:** Rating of the movie at IMDB site
- **Overview:** mini story/ summary
- **Meta score:** Score earned by the movie
- **Director:** Name of the Director Star1, Star2, Star3, Star4: Name of the Stars
- **No of votes:** Total number of votes
- **Gross:** Money earned by that movie

## Data Cleaning

I made the following change in the following variables:

- Removed unnecessary column named Certificate, Stars and renamed Series Title column to Movies Title.

## EDA

By analysing the different trends of the data and below are a few highlights of the analysis;

1. Most of the outliers are at the right side
2. Strong positive correlation is found between IMDB rating & Meta score
3. Somewhat positive correlation is seen between no of votes and the gross
4. There are hardly 53 movies having IMDB rating between 8.5 – 9.3
5. Histogram maximum frequency will get between 70 to 90 meta score
6. Box Graph shows that there are numerous movies rating from 7.75 to 8.25
7. The overall all the IMDb Rating of out of 10 is: 7.9
8. Box plot Interquartile Range (IQR) is 0.37
9. Number of movie releases per year are increasing as we go from 1920 to 2014