

## **TMDb movie data**

**Author: Aminat Owodunni**

### **Overview**

This data set contains information about 10,000 movies collected from The Movie Database (TMDb), including user ratings and revenue.

### **Aims and Objectives**

This project aims to investigate this dataset and determine genres that are most popular from year to year? What kinds of properties are associated with movies that have high revenues? What categories of movies generated high revenues? Complete code can be found on this GitHub [repo](#)

### **Data collection**

The dataset was cleaned from original data on [Kaggle](#).

### **Data importation and pre-processing**

The data after being imported into the jupyter notebook was scrutinized thoroughly. The dataset was seen to contain both categorical and numerical variables. The raw dataset contains about 10865 rows and 21 columns.

### **Columns in the dataset**

Id	- movie id
imdb_id	- movie online database id info.
popularity	- movie rate of demand
budget	- initial amount budgeted for movie
revenue	- initial revenue generated
original_title	- movie title
cast	- featured actors in movie
homepage	- movie homepage
director	- movie director
tagline	- slogan used to advertise movie
keywords	- keyword movie were found for
overview	- movie summary
runtime	- movie duration in minutes
genres	- movie category
production_companies	- company that produced the movie
release_date	- date movie is released to public
vote_count	- compilation of movie vote
vote_average	- movie average rating on a scale 0/10
release_year	- year movie is releases to public
budget_adj	- adjusted budget due to inflation
revenue_adj	- adjusted revenue due to inflation

## Exploratory Data Analysis (EDA)

Exploratory data analysis is performed on the given dataset to gain more insights about the dataset in terms of its summary statistics and visualizations of various variables in the dataset.

The most frequent genres is comedy with a count of **712**.

The most popular genres in the nineties are; Action|Adventure|Science Fiction movies. **Star Wars**, a movie, released in **1977** was the most popular with a score of **12.037933**

Top popular genres/movies in the millennium are: Action|Adventure|Science Fiction|Thriller. **Jurassic world**, released in **2015** was the most popular with a score of **32.985763**

Top rated genres/movies in the nineties are movies of the category: Documentary, Music, Drama, Crime. The two most rated movies are; **Queen \_ Rock Montreal** released in 1981 and **A Personal Journey with Martin Scorsese** released in 1995, both with average rating of **8.5**.

Top rated movies in the millennium are movies of category: Drama, Horror, Mystery, Science Fiction, Thriller, Documentary. The most rated movie of the millennium is **The Story of Film: An Odyssey** released in 2011 with an average rating of **9.2**

The movies that generated high revenues were highly rated and are of genres: **Action, Adventure, Fantasy, Science Fiction**. Most of the movies that generated high income were produced by **Lucasfilm, Twentieth Century Fox Film Corporation, Universal Pictures, Walt Disney, Paramount pictures, Columbia pictures** among others.