**Project 2- Investigate a Dataset- TMDB Movies**

This project was a part of the Udacity Data Analyst Nanodegree program.

**Project Definition:** Analyze a dataset and then communicate your findings about it. Use the Python libraries NumPy, Pandas, and Matplotlib to make your analysis easier. It contains 4 parts:

Step One - Choose Your Data Set

The dataset I have chosen to investigate for the project is Movie Releases. The link to this dataset on Kaggle:  <https://www.kaggle.com/tmdb/tmdb-movie-metadata/data>

Step Two - Get Organized

**Data Wrangling**: In this section of the report, you will load in the data, check for cleanliness, and then trim and clean your dataset for analysis. Make sure that you document your steps carefully and justify your cleaning decisions.

**Data Cleaning:**

* Removing Unnecessary Columns: Columns that are deleted are: imdb\_id, popularity, homepage, keywords, overview, production\_companies, budget\_adj, revenue\_adj
* Removing Duplication: Checking for duplicates and removing it, if found any.
* Checking and Removing rows having 0's or NaN values in the dataset.
* Checking the datatypes and changing the required ones.

Step Three - Analyze Your Data

Decided what questions to ask of the data and carried out explanatory Data Analysis and then drew conclusions.

* **Research Question 1: Top 10 Movies by Rating.**

**Extracting and plotting data from edited dataset and printing Top 10 Movies**

### Research Question 2: Movies with most and least profit.

### Research Question 3: Movies with biggest and lowest budget.

### Research Question 4: Movies with longest, shortest and average runtime.

* **Research Question 5: Total Profits earned by the movies in their respectable Year of release.**

Graphical representation of Total Profits earned Vs Year of the release using Line Chart

* **Research Question 6: The most popular produced genres**

Finding successful movies with respect to Genres and Graphical representation of Movies count by Genres using Bar Chart

* **Research Question 7: The no. of films produced over time.**

Calculating the no. of movies released in that particular year and Graphical representation of Movies count by Release Year

Step Four - Share Your Findings

After performing explanatory Data Analysis, the conclusions of the research questions have been drawn.