**MEMBERS**

* Aabhas Gupta
* Aditi Sharma
* Aishvarya Bandivdekar
* Jagrit Minocha
* Muktha Ramesh

**PROJECT DESCRIPTION:**

The entertainment industry has now found new ways to analyze or predict the statistics of forthcoming movies. Media giants like Netflix are eager to incorporate consumer insights from such analysis to be one step ahead of the others.

Our goal is to study the IMDB dataset and understand the factors that affect movies ratings and viewership. The dataset provides us information about the genres, language, actors, duration, Facebook and/or Twitter likes for each actor or the movie, budget and revenue.

**DATA SOURCE**: Kaggle.com, Secondary data

Through analysis of this dataset we hope to answer the following questions.

Description of dataset

* There are a total of 3085 movies considered for analysis
* Range of data – the movies span from 1916 to 2016
* We have a total of 28 columns describing the movies

Given that thousands of movies are produced each year we have analyzed:

* IMDB ratings Vs important aspect of movies (Actor, Director, Studio)
* Profitability – IMDB score, budget, genre and profitable associations
* Factors determining success – using decision trees
* Trends over time
* Sentiment analysis and text mining

**DATA DICTIONARY**:

| Data Column | Description |
| --- | --- |
| id | ID number given to each movie |
| title | Title of the movie |
| CountryofProduction | Country that the movie was produced in |
| release\_date | Date of release |
| Director | Name of movie director |
| Actor\_1 | Name of lead actor |
| Actor\_2 | Name of second lead actor |
| SocialMedia\_likes | Facebook or Twitter likes of the actor |
| Budget | Budget used for the movie in USD |
| Revenue | Revenue earned by the movie in USD |
| Imdb\_score | Score given to the movie by imdb users |
| Meta\_score | Score given to the movie by movie critics |
| Vote\_count | Count of imdb users that rated the movie |
| Runtime | Duration of the movie in minutes |
| Language ID | Language that the movie was produced in |
| Genre | Three columns that describe the genre of the movie |
| Keyword | Five columns depicting key words used in the movies |
| Studio | Three columns with studio names that produced the movie |
| Status | Status of the movie whether released or in production |
| Tagline | A statement used during promotion of the movie during marketing |