**HYPOTHESIS**

The data set consists of the IMDB rating that consists of movie plots, rating, user rating and reviews. The tables involved in the dataset consists of :-

1. IMDB Table

2. Earning Table

3. Genre Table

The IMDB table consists of the following columns which are used for analysis :-

1. **Movie\_id** - serves as a primary key for IMDB Table and foreign key for Earning and Genre Dataset

2. **Title** - The title of the movies from IMDB Dataset

3. **Rating** - The score of the movie out of 10

4. **TotalVotes** - The total number of votes provided by the user.

5. **MetaCritic** - The weighted average score coming from reputed critics

6. **Budget** - The amount of cost related to development, production and post-production of a movie.

7. **Runtime** - The amount of time that a movie, performance, recording, etc , lasts from beginning to end.

The Earning table consists of the following columns which are used for analysis :-

1.**Movie\_id** - Serves a primary key for the table.

2. **Domestic Earning** - The net revenue generated locally.

3. **WorldWide Earning** - The net revenue generated internationally.

The Genre table consists of the following columns which are used for analysis :-

1. **Movie\_id** - Serves a primary key for the table.

2. **Genre** - The category to which the movie belongs.

The dataset can be used to draw multiple insights, The following information was drawn from the dataset as a part of the project.

1 Finding the highest-grossing movie (i.e domestic earning + worldwide earning) in IMDB database year wise.

2. Finding out the percentage of the budget for each genre in IMDB Movie Dataset.