# **Introduction:**

1. **Dataset chosen for analysis:** TMDb Movies Data. The dataset contains information about 10,000 movies collected from The Movie Database (TMDb), including user ratings and revenue.

# **Data Wrangling:**

1. **Number of rows in the dataset (before cleaning):** 10866
2. **Number of columns in the dataset (before cleaning):** 21
3. **Number of movie enlisted:** 10866
4. **Qualitative variables in the dataset:** Original title, director, homepage, production companies, tagline, overview, genres
5. **Quantitative variables in the dataset:** Budget, popularity, vote\_count, vote\_average
6. **Number of NaN values:** 13434
7. **Columns with missing values:** imdb\_id, cast, homepage, director, tagline, keywords, overview, genres, production companies
8. **Number of duplicate values:** 1

## **Common Problems with the data:**

1. Numerous rows have missing values
2. Some columns of data are not required for investigation and must be dropped
3. Duplicates must be removed

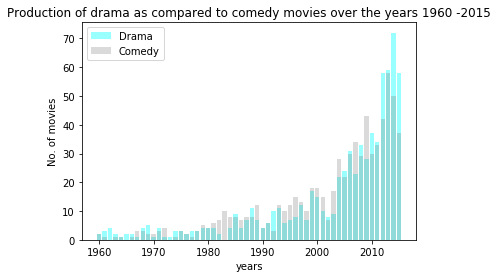
## **General Properties:**

1. **Mean average vote**: 5.975011504832047
2. **Oldest Release year:** 1960
3. **Most recent Release Year:** 2015
4. **Mean Runtime of the movies:** 102.07179015186378
5. **Most Common genre:** Drama (712)
6. **Mean popularity:** 0.6464455549010583
7. **Most common producer:** Paramount Pictures (156)
8. **Most common director:** Woody Allen

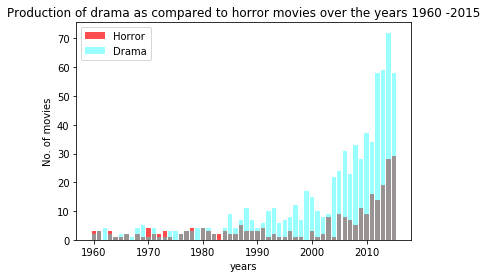
# **Exploratory Data Analysis:**

## Which genres are most popular from year to year?

* Drama has been the most popular genre with a count of 712, followed by Comedy (with or without a supplemental genre) and Horror (with or without a supplemental genre like thriller, sci-fi etc.)
* Horror was the most popular genre in the early 1960s, 1970 and 1978
* In the 1980s and 1990s, comedy was the **most popular genre**

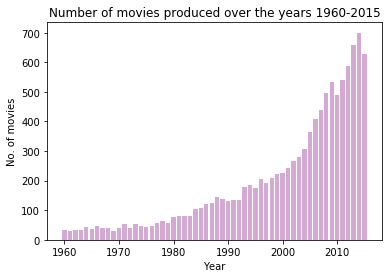


* It can be noted that the number of drama movies slowly increased and surpassed the number of comedy movies. The difference is significant during 2014-2015. Nonetheless, the production of both movies has increased over time



* It can be noted that the number of horror movies produced have increased over the years but they are still significantly lower than the number of drama movies produced over the same time period

1. It has been suspected that globalization has led to increased cross-cultural interactions. This particularly manifests in movies as different culture can be physically rendered via films. Is this true? Has there been an increase in the number of movies over the years?

* It can be clearly seen that the number of movies being produced from 1960-2015 has increased exponentially. Thus, globalization is perhaps one of the causes of this increase.