Group 10 Writeup - “TMDB Movies”

By: Muqsit Momin, Prakhar Jain and Siddha Deshpande

General Focus:
What factors contribute to a movie's success, and can they be predicted before the movie's release?

Main Focus for this Analysis:
Which actor or actress is most likely to bring in the most revenue for a single movie? Using the past history of relevant movies an actor/actress has participated in, we will attempt to understand if there is a correlation between said actor/actress and the revenue a movie earns.

Description of the dataset:
The "TMDB 5000 movies" dataset is a comprehensive collection of data on 5,000 movies, providing information on various aspects such as cast, director, production company, budget, revenue, genre, popularity, runtime, and spoken language. This dataset can be utilized by researchers to explore different variables that contribute to a movie's success and predict a movie's potential popularity and revenue prior to its release. Specifically, we aim to identify the actor or actress who is most likely to generate the highest revenue for a single movie. By examining the dataset, we will compare the various 4,761 actors/actresses with the 5,000 movies in the dataset they participated in, and compare it with the amount of revenue generated by that movie.

Variables:
The "TMDB 5000 movies" dataset contains various variables related to movies, including:

- Title: the name of the movie
- Budget: the budget allocated for making the movie
- Revenue: the revenue earned by the movie
- Cast: the actors and actresses who appeared in the movie
- Director: the director of the movie
- Production Company: the production company responsible for making the movie
- Genre: the category or type of the movie
- Popularity: the popularity of the movie based on ratings
- Runtime: the duration of the movie
- Spoken Language: the primary language spoken in the movie

In conclusion, the main focus is on the relationship between the movie cast variable and revenue earned by a movie. Our main question of focus is: "Which actor or actress is most likely to bring in the most revenue for a single movie?"
To answer this question, we will use cast members as our independent variable (X) and revenue as our dependent variable (Y). By examining the dataset and analyzing the relationship between the two variables, we aim to identify the actor or actress with the highest correlation with movie revenue, thus enabling us to predict which cast members are most likely to contribute to a movie's financial success. We will also consider other variables impacting revenue, such as budget, genre, and popularity to build a more comprehensive understanding of the factors contributing towards a movie’s success.