**Report**

Briefly describing the IMDb dataset, including its source and structure.

Poster\_Link: URL link to the film's poster.

Series\_Title: Title of the film.

Released\_Year: Year of film release.

Certificate: Certification rating of the film (e.g., PG-13, R).

Runtime: Duration of the film in minutes.

Genre: Genre(s) of the film.

IMDB\_Rating: IMDb rating for the film.

Meta\_score: Metascore rating for the film.

Star1: first billed actors/actresses in the film.

No\_of\_Votes: Number of votes/ratings received by the film.

Data Preprocessing

We first read the dataset files from a ZIP file then imported the important libraries then showed the dataset head and tail also the dataset info then we proceeded to change the columns names to make it easier to access by removing spaces and upper case then we checked to see if there is any null values and where then we imported sklearn and labelEncoder then we replaced the nana values with unknow or 0 check the database info again and then made a correlation between the IMDB rating and the no of votes then created a heatmap for them and afterwards we imported other needed libraries and dropped the star2 ,3 and 4 columns then at the metascore column we changed its type to int and removed any null values then done another correlation with IMDB\_reating , metascore and certificate and made sure their were no empty values in any of them then calculated the IQR added more libraries created a table figure then ran the command describe then we created 20 questions and went ahead answering them such as highest rated movie and lowest rated movie and the max and min amount of gross generated by a movie then showed all movies more than 1000000 no of votes then calculated the average rating of movies and average metascore and many other questions made till we reached 20 questions