Integrated M.Sc.
Computer Science
(Artificial Intelligence & Data Science)

Fifth Semester

Project Proposal

21-805-0403: R Programming Lab

Submitted by

NOBLE AUSTINE
(80521014)

DEPARTMENT OF COMPUTER SCIENCE
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY (CUSAT)
KOCHI-682022

August 2023

Dataset : EDA on IMDB Movies Dataset

The dataset contains the details of IMDB top 1000 movies.The datasetcontains the details such as

1. PosterLink: Link of the poster that imdb using

2. SeriesTitle: Name of the movie

3. ReleasedYear: Year at which that movie released

4. Certificate: Certificate earned by that movie

5. Runtime: Total runtime of the movie

6. Genre: Genre of the movie

7. IMDB Rating: Rating of the movie at IMDB site

8. Overview: mini story/ summary

9. Meta Score: Score earned by the movie

10. Director: Name of the Director

11. Star1 ,Star2 ,Star3 ,Star4: Name of the Stars

12. No of votes: Total number of votes

13. Gross: Money earned by that movie

Data Source :

https://www.kaggle.com/code/harshitshankhdhar/eda-on-imdb-movies-dataset

Exploratory Data Analysis Questions :

1. How does the average rating change over different time periods?

2. Is there a correlation between movie budgets and revenue?

3. Are there any specific genres that tend to receive higher ratings or generate more revenue?

4. Are there any actors or actresses who frequently appear in high-grossing movies?

5. Is there a relationship between movie run time and its rating or revenue?

6. How has the number of movies released each year changed over time?

7. How does the language of the movie affect its performance?

8. Are there any interesting correlations between variables like budget, runtime, and revenue?

9. Create word clouds of movie titles, plot keywords, or movie tags to identify common themes and trends.

10. Are there specific years or time periods with a higher number of successful movies?

Inferred Question :

Explore how the popularity of different genres has changed over time. Are certain genres more popular in specific years ?