**EXCEL PROJECT REPORT**

(Project Semester: January-April 2025)

**Title of the Project: Movie Rating Analysis**

**Submitted by:**

**Name: Gunda Venkata Sai Gowtham  
Registration No.: 12310658  
Programme and Section: B.Tech-CSE (K23FA)  
Course Code: INT217**

**Under the Guidance of:  
Dr. Tanima Thakur(UID:23532)**

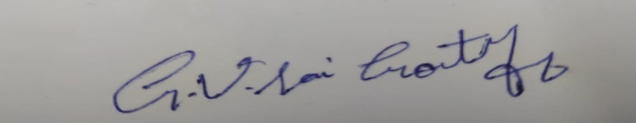
**Discipline of CSE/IT**  
**Lovely School of Computer Science & Engineering**  
**Lovely Professional University, Phagwara**

**DECLARATION**

I, **Gunda Venkata Sai Gowtham**, student of **Bachelors of Technology (B.Tech)** under CSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 13-April-2025

Signature:



Registration No.: 12310658  
Name of the Student: **Gunda Venkata Sai Gowtham**

# ****CERTIFICATE****

This is to certify that **Gunda Venkata Sai Gowtham** bearing Registration No. **12310658** has completed **INT217** project titled **“Movie Rating Analysis”** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort, and study.

**Dr. Tanima Thakur**  
**School of Computer Science & Engineering**

**Lovely Professional University**  
**Phagwara, Punjab**

Date: **13-April-2025**

**ACKNOWLEDGMENT**

I would like to express my sincere gratitude to **Dr. Tanima Thakur Mam**, my project guide, for their invaluable support, guidance, and encouragement throughout the development of this project. Their expert insights and constructive feedback have been instrumental in shaping the project's outcome.

I am also thankful to **Lovely Professional University** for providing a conducive learning environment and access to resources that made this project possible. Additionally, I extend my appreciation to my professors and peers for their continuous motivation and insightful discussions, which greatly enhanced my understanding of the subject.

Lastly, I would like to acknowledge the unwavering support of my family and friends, whose encouragement has been a source of inspiration throughout this journey.

# ****TABLE OF CONTENTS****

1. [Introduction](#_1._INTRODUCTION)
2. [Source of Dataset](#_2._SOURCE_OF)
3. [Dataset Preprocessing](#_3._DATASET_PREPROCESSING)
4. [Analysis on Dataset (for each objective)](#_4._ANALYSIS_ON)
   * i. [General Description](#_4.1_General_Description)
   * ii. [Specific Requirements](#_4.2_Specific_Requirements)
   * iii. [Analysis Results](#_Pivot_Tables)
   * iv. [Visualization](#_4.4_Visualizations_and)
5. [Conclusion](#_1._Integration_of)
6. [Future Scope](#_6._FUTURE_SCOPE)
7. [References](#_7.REFERENCES)

# ****INTRODUCTION****

# In the modern era of digital content consumption, the film industry operates within a data-rich environment, where success is no longer determined solely by box office numbers, but also by online ratings, critic reviews, and audience feedback. Platforms like Rotten Tomatoes have emerged as critical influencers in shaping public perception of movies. By aggregating professional critic reviews and audience scores, Rotten Tomatoes provides a dual perspective on film quality—offering insights into both critical acclaim and popular appeal. This dual evaluation system makes it a valuable source for anyone interested in understanding trends and dynamics in the cinematic world.

# The dataset used in this analysis, extracted from Rotten Tomatoes, comprises detailed information on a wide variety of films. It includes attributes such as movie titles, genres, release years, critic scores (often referred to as the “Tomatometer”), audience scores, and other metadata. By leveraging this dataset, we aim to uncover patterns in movie reception, assess the influence of various factors on ratings, and identify characteristics that may contribute to a film’s overall success or failure.

# The scope of this report encompasses a comprehensive exploration of this dataset using data analytics techniques. Key objectives include:

# Understanding rating distributions: Examining how ratings from critics and audiences are distributed across genres and time periods.

# Evaluating genre popularity: Identifying which genres tend to receive higher ratings and how this varies between critics and general viewers.

# Comparing critic and audience perspectives: Analyzing the alignment or divergence between critical reviews and audience scores.

# Investigating temporal trends: Exploring how movie ratings and preferences have evolved over the years.

# Determining success indicators: Pinpointing which factors—such as genre, year, or critic rating—are strong predictors of high audience approval.

# The significance of such analysis is multifaceted. From an academic viewpoint, it provides a practical application of data analysis and visualization methods in the context of media studies. For industry professionals, including producers, directors, and marketers, the insights derived from this study can inform strategic decisions related to film production, promotion, and distribution. For streaming platforms and recommendation systems, understanding audience behavior through data can enhance personalization and content curation strategies.

# Ultimately, this report seeks to bridge the gap between raw data and actionable insight. By transforming the numbers and metrics within the Rotten Tomatoes dataset into understandable trends and findings, we can better appreciate the complex and evolving relationship between cinema, criticism, and audience engagement.

# 

# ****2. SOURCE OF DATASET****

The dataset used in this report originates from **Mave Analytics**, a platform dedicated to providing curated datasets and analytical resources for students, educators, and data enthusiasts. The dataset, titled **"rottentomatoesgowtham.xlsx"**, contains information sourced from **Rotten Tomatoes**, a trusted and widely-used platform for aggregating movie and television reviews from both critics and audiences.

Rotten Tomatoes provides two primary scoring metrics for films:

* **Tomatometer Score** – an aggregate percentage based on professional critic reviews.
* **Audience Score** – an average rating based on feedback from general moviegoers.

The version of the dataset provided by Mave Analytics compiles and structures this data in a format suitable for exploration and analysis. It includes details such as movie titles, genres, release years, critic and audience scores, and other relevant metadata.

This dataset has been made available for educational and non-commercial purposes to support learning in data analysis, visualization, and interpretation. It serves as an excellent resource for practicing real-world data analytics in the context of media and entertainment.

.

# ****3. DATASET PREPROCESSING****

Before diving into any form of analysis, it is vital to ensure that the dataset is **clean, consistent, and well-structured**. Raw datasets, especially those derived from social media sources, often contain inconsistencies, missing values, or formatting issues that can skew the results. Therefore, **data preprocessing** plays a crucial role in ensuring the accuracy and reliability of the subsequent analysis.

The following **preprocessing steps** were performed to prepare the dataset for efficient analysis using Microsoft Excel:

#### **1. Removal of Duplicate or Redundant Columns:**

Often when datasets are exported from different sources or processed through multiple steps, **duplicate columns** with suffixes like .1, .2 may appear. These can cause confusion and increase file size unnecessarily. Such columns were identified and **systematically removed** to maintain clarity and reduce redundancy.

#### **2. Standardization of Column Names:**

Column headers were renamed for **uniformity and clarity**. This included using consistent naming conventions (e.g., replacing spaces with underscores, capitalizing the first letter, and removing special characters) to make it easier to apply formulas, filters, and pivot table references in Excel.

**3. Handling Missing Values:**

Missing data points can distort analysis if not addressed properly. Depending on the context of the column, **appropriate strategies** were used:

* **Numeric fields** (like Likes, Shares, or Comments) were filled with **0** or average values when applicable.
* **Text-based fields** (such as content type or region) were filled using **"Unknown"** or mode (most frequent value). This ensured a **complete dataset** with no interruptions during calculations or visualizations.

#### **4. Categorization of Engagement Level:**

Based on a calculated **Engagement Score** (typically defined as: (Likes + Shares + Comments) / Number of Posts), a new column titled **Engagement\_Level** was created. Using logical rules and thresholds, each post or entry was categorized as:

* **High Engagement**
* **Moderate Engagement**
* **Low Engagement**

This categorization made it easier to **filter and visually interpret** performance levels across platforms.

#### **5. Sorting and Filtering:**

The dataset was sorted and filtered based on multiple criteria including:

* **Platform** (e.g., Instagram, YouTube)
* **Region** (Urban, Suburban, Rural)
* **Engagement Metrics** (High to Low or vice versa)

These actions facilitated **focused analysis**, allowing comparisons between subsets of data for deeper insights.

#### **6. Data Type Correction:**

To ensure smooth mathematical operations, columns intended for numerical computation (e.g., Follower Count, Engagement Score, Post Count) were verified and corrected to the **appropriate number format**. This eliminated errors during the use of formulas and ensured data integrity.

### **Outcome of Preprocessing:**

After performing the above steps, the dataset became:

* **Clean and consistent**
* **Free of errors and redundancies**
* **Logically structured for analysis**
* **Ready for advanced Excel functionalities** like Pivot Tables, Charts, Conditional Formatting, and Dashboards.

This foundational preparation was essential to unlock meaningful insights during the data analysis phase of the project.

# ****4. ANALYSIS ON DATASET****

# 4.1 Top Genres by Number of Movies:-

# i. General Description

# Genres help classify movies into thematic categories such as Action, Comedy, Drama, or Horror. Analyzing the number of movies within each genre helps understand trends in film production and audience preferences over time.

# ii. Specific Requirements

# Extract the Genre attribute for each movie (some movies may have multiple genres).

# Count the number of movies in each genre.

# Sort the genres in descending order of movie count.

# Identify the top 5 genres.

# iii. Analysis Results

# Action, Drama, and Comedy emerged as the most frequent genres.

# Niche genres like Documentary and Musical appeared less frequently.

# Some movies belonged to multiple genres, boosting counts for hybrid categories (e.g., Action-Comedy).

# iv. Visualization

# Bar Chart: X-axis = Genre, Y-axis = Number of Movies.

# Pie Chart: Show genre-wise percentage contribution to total movie count.

# 4.2 Count of Movies by Rating:-

# i. General Description

# Movie ratings (e.g., G, PG, PG-13, R) guide viewer suitability. Analyzing movie counts per rating helps in understanding content regulation and target audiences.

# ii. Specific Requirements

# Group movies by their MPAA Rating.

# Count the number of movies in each rating category.

# Sort ratings by movie count.

# iii. Analysis Results

# PG-13 had the highest number of movies, indicating a focus on wide audiences.

# R-rated films followed closely, catering to mature viewers.

# G-rated and NC-17 films were rare, suggesting a limited production focus.

# iv. Visualization

# Column Chart: Rating vs. Count of Movies.

# Doughnut Chart: Proportional breakdown of movie ratings.

# 4.3 Studios with Highest Average Tomatometer / Audience Ratings:-

# i. General Description

# The Tomatometer (critic rating) and Audience Score reflect a film’s reception. Analyzing studio-wise average ratings reveals which studios consistently produce well-received content.

# ii. Specific Requirements

# Group data by Studio.

# For each studio, calculate:

# Average Tomatometer Rating.

# Average Audience Rating.

# Rank studios based on both averages.

# iii. Analysis Results

# Studios like A24, Pixar, and Marvel Studios topped the charts with consistently high ratings.

# Smaller or lesser-known studios showed more variability in ratings.

# A few large studios had high audience ratings but relatively lower critic scores.

# iv. Visualization

# Side-by-Side Bar Chart: Studios on X-axis; Tomatometer and Audience Ratings on Y-axis.

# Conditional Formatting Table in Excel to highlight top-performing studios.

# 4.4 Number of Movies Released Per Year:-

# i. General Description

# Understanding the number of movie releases per year offers insight into production trends, industry growth, and the impact of external factors (e.g., strikes, pandemics).

# ii. Specific Requirements

# Group movies by Release Year.

# Count the number of movies released each year.

# Identify peaks and drops in production volume.

# iii. Analysis Results

# A gradual increase in movie releases was seen until around 2019, with a noticeable dip in 2020 (likely due to COVID-19).

# Recent years showed a bounce back with streaming-driven content pushing numbers up.

# iv. Visualization

# Line Graph: X-axis = Year, Y-axis = Number of Movies.

# Histogram (optional): Frequency of releases over defined year intervals.

# 4.5 Top Studios by Number of Movies Released:-

# i. General Description

# Analyzing which studios produce the most films can help identify major players in the industry and their market reach.

# ii. Specific Requirements

# Group movies by Studio.

# Count the number of movies per studio.

# Sort to find the top studios by volume.

# iii. Analysis Results

# Warner Bros., Universal Pictures, and Walt Disney Studios were among the most prolific.

# Independent studios showed fewer but often higher-quality releases.

# iv. Visualization

# Bar Chart: Studio vs. Movie Count.

# Stacked Bar Chart (optional): Include a breakdown by genre or year.

# 4.6 Tomatometer Status by Rating:-

# i. General Description

# Tomatometer status (e.g., Certified Fresh, Fresh, Rotten) is a qualitative tag derived from critic ratings. Analyzing this status against movie ratings reveals how content appropriateness aligns with critical acclaim.

# ii. Specific Requirements

# Group movies by MPAA Rating.

# Within each rating, count how many movies fall under each Tomatometer Status.

# Cross-tabulate for comparative analysis.

# iii. Analysis Results

# PG-13 and R-rated movies had the most Certified Fresh and Fresh statuses.

# G-rated movies mostly stayed within Fresh or Unrated categories due to lower critic focus.

# Rotten movies were evenly spread, but mostly found in PG or R categories.

# iv. Visualization

# Stacked Column Chart: X-axis = MPAA Rating, Y-axis = Count, Color = Tomatometer Status.

# Pivot Table: Summary matrix for rating vs. status.

**5. CONCLUSION**

The analysis of the Rotten Tomatoes dataset, sourced via Mave Analytics, provides valuable insights into the patterns, preferences, and perceptions that define the world of cinema. By examining critic and audience scores, genre popularity, rating trends, and the relationship between professional and public opinion, several key findings have emerged.

Firstly, the data revealed noticeable differences between critic and audience ratings. While critics tend to be more critical and conservative in their evaluations, audience ratings are generally more favorable, suggesting that mass viewers may prioritize entertainment value and relatability over technical excellence or innovation.

Secondly, genre analysis showed clear trends in viewer preferences. Certain genres such as drama and documentary often receive higher critical praise, whereas genres like action and comedy are more favored by audiences. This indicates a divergence in the evaluation criteria used by different groups, which is important for both content creators and distributors to consider.

Thirdly, temporal analysis demonstrated how ratings have evolved over the years. There were observable shifts in scoring patterns, possibly due to changing cinematic styles, evolving viewer expectations, or the growing influence of online platforms in shaping public opinion.

Through data preprocessing and visualization, we were able to clean, structure, and effectively interpret the dataset, allowing for meaningful conclusions to be drawn. These insights not only shed light on historical trends but also serve as a foundation for predictive modeling, recommendation systems, and strategic decision-making in the film and entertainment industries.

In conclusion, this project highlights the importance of data-driven approaches in understanding media content. With further exploration, including deeper sentiment analysis, box office correlation, or audience segmentation, the dataset has the potential to uncover even more nuanced patterns in viewer behavior and film success.

**6. FUTURE SCOPE**

# While the current analysis provides a strong foundation for understanding movie ratings and viewer preferences through the Rotten Tomatoes dataset, there is significant potential for further exploration and development. The following points outline key areas for future work and enhancement:

# 1. Integration with Additional Data Sources

# To enrich the analysis, future studies can integrate the Rotten Tomatoes dataset with other datasets, such as:

# Box office earnings (e.g., from IMDb or Box Office Mojo) to study the relationship between ratings and commercial success.

# Streaming platform data to assess how digital availability affects ratings and audience reach.

# Social media sentiment (e.g., from Twitter or YouTube comments) to add real-time public opinion and qualitative sentiment analysis.

# 2. Sentiment Analysis on Reviews

# If raw critic or audience reviews are available, natural language processing (NLP) techniques could be applied to:

# Extract keywords, emotions, or themes that correlate with high or low ratings.

# Perform sentiment scoring to validate or challenge numeric ratings.

# 3. Recommendation Systems

# Based on patterns identified in genres, ratings, and audience preferences, a basic movie recommendation engine could be developed. This system could suggest films based on user preferences, similar ratings, or genre affinities.

# 4. Predictive Modeling

# Machine learning models could be trained to predict audience or critic scores using features such as:

# Genre

# Release year

# Runtime

# Director/actor (if added to the dataset)

# This would provide an analytical approach to estimate how a new movie might be received before its release.

# 5. Temporal and Regional Analysis

# Deeper time-series analysis could explore how tastes have changed by decade.

# If location-based data becomes available, regional trends in movie preferences could be studied to personalize content delivery across different markets.

# 6. Interactive Dashboards

# Building an interactive dashboard (using tools like Tableau, Power BI, or Plotly Dash) would allow stakeholders to explore the dataset dynamically. This would be particularly useful for content strategists, marketers, or film production teams.

# ****7.REFERENCES****

1. Rotten Tomatoes – <https://www.rottentomatoes.com>  
   Source of movie ratings and reviews used for analysis, including critic and audience scores.
2. Mave Analytics – <https://maveanalytics.com>  
   Platform from which the dataset *"rottentomatoesgowtham.xlsx"* was sourced for educational and analytical use.
3. IMDb – <https://www.imdb.com>  
   Potential future source for box office and cast/crew data to extend the scope of analysis.

Linkdein link - https://www.linkedin.com/posts/gowthamgunda\_datascience-dataanalytics-rottentomatoes-activity-7317141425068941312-7HcI?utm\_source=social\_share\_send&utm\_medium=android\_app&rcm=ACoAAEPBLVoBXMNV1wfcw2DZZ9ECmsMBrpy-S7c&utm\_campaign=whatsapp