Analysis of RottenTomatoes Movie Ratings

Problems and background

Problem:

To perform a detailed data analysis on the Rotten Tomatoes movies dataset, encompassing over 15,000 films, in order to uncover insights regarding film ratings, genre distribution, studio performance, critic-audience discrepancies, and temporal trends in movie releases.

Background:

Rotten Tomatoes is a well-known platform that aggregates movie reviews from critics and audiences, providing ratings that influence both consumer choices and industry perspectives. The dataset includes critical information such as film titles, descriptions, ratings (Tomatometer and Audience), genres, directors, cast, release dates, studios, runtimes, and critic reviews. This project aims to leverage Excel's analytical capabilities to derive meaningful conclusions from the dataset, aiding stakeholders in understanding movie performance and trends.

Solution

Data Analysis:

The dataset comprises over 15,000 movies reviewed by Rotten Tomatoes. Each record includes film details such as title, description, rating, genre, directors, cast, release date, studio, runtime, featured reviews, Tomatometer status, and audience ratings.

Data-Driven Insights:

The comprehensive analysis of the Rotten Tomatoes dataset has yielded several data-driven insights that illuminate the intricacies of the film industry. By examining the distribution of film ratings, it was revealed that while a substantial number of films achieve a Fresh rating, a significant portion still falls under the Rotten category, highlighting potential areas for improvement in movie quality. Genre analysis showed that dramas tend to receive higher critical acclaim, whereas action and adventure films are more prone to negative reviews. Temporal trends indicated a surge in film releases during the mid-2000s, with summer and holiday seasons being peak periods for new releases. Studio performance analysis identified Paramount Pictures and Warner Bros. as leaders in producing highly-rated films, contrasting with lower average ratings from Lionsgate and Columbia Pictures. Furthermore, the exploration of critic-audience discrepancies uncovered films with the largest differences in perception, suggesting areas where studios could better align their productions with audience expectations. Lastly, the frequent use of positive terms such as "engaging" and "captivating" in critics' consensus underscores the language that resonates in well-received reviews. These insights provide a robust foundation for stakeholders to make informed decisions, enhance film quality, and better meet audience demands.

Stakeholder Engagement:

The data analysis of the Rotten Tomatoes dataset offers crucial insights for enhancing stakeholder engagement within the film industry. By understanding the distribution of film ratings, stakeholders can identify areas where improvements are needed to reduce the prevalence of Rotten-rated films and increase the number of critically acclaimed releases. The genre-specific findings reveal that dramas are more likely to receive positive reviews, while action and adventure films often face more criticism, providing guidance for genre-specific strategies. Temporal trends analysis highlights the optimal times for film releases, suggesting that stakeholders should focus their marketing and release efforts during the summer and holiday seasons to capitalize on peak audience interest. The performance analysis of studios reveals that Paramount Pictures and Warner Bros. lead in producing high-rated films, offering a benchmark for other studios to aspire to, while highlighting areas where Lionsgate and Columbia Pictures might improve. Additionally, the discrepancies between critic and audience ratings pinpoint specific films that fail to meet audience expectations, guiding stakeholders in aligning production and marketing efforts more closely with audience preferences. Finally, the common language used in critics' consensus, such as "engaging" and "captivating," suggests that adopting such terminology in marketing could positively influence audience perception. These insights equip stakeholders with the knowledge to make data-driven decisions, enhance film quality, and better engage with their audience, ultimately driving the industry's success.

Project scope

Scope:

The project aims to conduct a thorough analysis of the Rotten Tomatoes movie dataset to uncover insights regarding film ratings, genre trends, studio performance, critic-audience discrepancies, and temporal release patterns. This analysis will provide stakeholders with data-driven recommendations to enhance decision-making in the film industry.

Step-by-Step Process:

(Week 1): Project Kickoff and Data Collection

- Day 1-2: Project kickoff meeting to outline objectives and deliverables.
- Day 3-5: Collect the Rotten Tomatoes dataset and relevant movie details.
- Day 6-7: Review and understand the dataset structure and contents.

Deliverables:

- Project plan and timeline.
- Collected dataset ready for analysis.

(Week 2) : Data Cleaning and Preparation

- Day 8-10: Clean the dataset to handle missing values and inconsistencies.
- Day 11-12: Extract primary genre, year, and month of release from the data.
- Day 13-14: Organize the data into a structured format suitable for analysis.

Deliverables:

- Cleaned and organized dataset.

(Week 3): Descriptive Statistics and Visualization

- Day 15-17: Perform descriptive statistics on film ratings.
- Day 18-20: Create visualizations for rating distributions by overall and primary genre.
- Day 21: Review findings and prepare a summary.

Deliverables:

- Descriptive statistics report.
- Visualizations of rating distributions.

(Week 4): Genre and Studio Analysis

- Day 22-24: Analyze rating distributions by genre.
- Day 25-27: Compare average Tomatometer ratings by studio.
- Day 28: Summarize genre and studio performance insights.

Deliverables:

- Genre analysis report.
- Studio performance comparison.

(Week 5): Critic-Audience Discrepancy Analysis

- Day 29-31: Calculate discrepancies between Tomatometer and Audience ratings.
- Day 32-34: Identify films with the largest rating discrepancies.
- Day 35: Prepare a summary of critic-audience discrepancies.

Deliverables:

- Discrepancy analysis report.

(Week 6): Temporal Trends and Text Analysis

- Day 36-38: Analyze temporal trends in film releases by month and year.
- Day 39-41: Perform text analysis on critic's consensus ratings to identify common language.
- Day 42: Summarize findings from temporal and text analysis.

Deliverables:

- Temporal trends report.
- Text analysis findings.

(Week 7): Compilation and Review of Findings

- Day 43-45: Compile all analysis findings into a comprehensive report.
- Day 46-47: Create visualizations and summaries for the final report.
- Day 48-49: Review the report and prepare for stakeholder presentation.

- Day 50: Conduct a review meeting with the team to finalize the report.

Deliverables:

- Final comprehensive report.
- Presentation slides.

(Week 8): Presentation and Recommendations

- Day 51-52: Present findings to stakeholders.
- Day 53-54: Discuss insights and actionable recommendations.
- Day 55-56: Incorporate feedback and finalize recommendations.
- Day 57-58: Prepare a detailed implementation plan for recommendations.
- Day 59-60: Project closure and documentation.

Deliverables:

- Stakeholder presentation.
- Detailed implementation plan.
- Final project documentation.

If the problem statement is directly pointing to the business:

In the highly competitive film industry, understanding audience preferences and critical reception is crucial for making informed business decisions. This project aims to analyse the Rotten Tomatoes movie dataset to uncover valuable insights into film ratings, genre popularity, studio performance, and critic-audience discrepancies. By leveraging this data, film studios, distributors, and marketers can refine their production and marketing strategies to enhance movie quality, optimize release schedules, and better align with audience expectations, ultimately driving greater commercial success and viewer satisfaction.

For Others:

Project Insights and Recommendations:

The insights and recommendations from the Rotten Tomatoes movie data analysis project can be leveraged across various departments and functions within the organization to enhance processes, improve decision-making, and ultimately address the identified challenges. Here's how each department can benefit:

Marketing Department: The comprehensive analysis of movie ratings, genre trends, and audience preferences provides valuable insights for refining marketing strategies and campaigns. By understanding which genres resonate with audiences and which films have the highest ratings, the marketing team can:

- Target Specific Audiences: Tailor marketing campaigns to appeal to the most engaged and profitable audience segments.
- Optimize Promotional Efforts: Focus promotional activities on films and genres with higher audience and critic ratings to maximize engagement and ROI.
- Enhance Pricing Strategies: Use data on audience demand and critical reception to inform dynamic pricing strategies for movie releases and related products.

Operations and Supply Chain:

Insights into temporal trends and the timing of movie releases can inform supply chain and operations management. By understanding peak release periods and audience demand:

- Optimize Distribution Networks: Ensure efficient distribution and availability of films during peak seasons, reducing stockouts and overstock situations.
- Improve Inventory Management: Align inventory levels with anticipated demand based on seasonal trends, ensuring timely availability of movie merchandise and promotional materials.

Sales and Customer Service:

Evaluating critic-audience discrepancies and understanding audience preferences can help enhance the effectiveness of sales and customer service teams:

- Targeted Training Initiatives: Equip sales and customer service teams with insights into audience preferences and common discrepancies to improve customer interactions and satisfaction.
- Improve Customer Retention: Utilize audience feedback and ratings to address customer concerns and enhance service quality, fostering loyalty and repeat business.

Product Development: Analysis of critic reviews and audience ratings offers valuable insights into product offerings, quality, and innovation opportunities:

- Tailor New Offerings: Develop new film projects and related products that align with audience preferences and critical acclaim.
- Differentiate from Competitors: Use insights to innovate and differentiate product offerings, enhancing market penetration and driving revenue growth.

Stakeholders Summary:

- Management: Provides strategic direction and support, ensuring alignment with organizational goals and objectives.

- Marketing Department: Drives marketing strategies and initiatives based on data-driven insights to enhance customer engagement and campaign effectiveness.

- Sales and Customer Service Teams: Offer frontline insights and implement targeted training and development initiatives to improve customer interactions and satisfaction.

- Operations and Supply Chain Staff: Ensure smooth product flow and optimize distribution and inventory management processes to meet customer demand.

- Customers: Their needs and preferences guide decision-making and product development efforts, ensuring alignment with market demand.

- External Partners: Contribute to project execution and support various aspects of the analysis and implementation process.

Methodology

Data Sources:

SQL: SQL databases serve as a primary data source, containing structured data such as film details, ratings, and reviews. Through SQL queries, we extract relevant datasets for analysis, including film metadata, audience ratings, and critic reviews.

AWS Amazon Web Services (AWS) provides additional data sources, such as cloud-based databases or data lakes. These sources may include real-time data streams, customer interactions from web applications, or logs from operational systems. Utilizing AWS services allows for scalability, flexibility, and accessibility of data for analysis.

Data Scraping: Data scraping techniques may be employed to gather data from external sources, such as competitor websites, social media platforms, or online reviews. Scraped data can provide insights into market trends, competitor strategies, and customer sentiment, complementing internal datasets.

Local Data Sources Local data sources encompass any internal systems or files that store relevant data, such as Excel spreadsheets, CSV files, or proprietary databases. These sources may contain supplementary information, such as promotional campaign data, customer feedback surveys, or product inventory records.

Data Wrangling:

Data Understanding: In this phase, we explore and assess the quality, structure, and relevance of the data from various sources. We identify key variables, understand data distributions, and document any anomalies or inconsistencies

Data Cleaning: Data cleaning involves pre-processing raw data to address issues such as missing values, outliers, or formatting errors. Techniques such as imputation, outlier detection, and standardization are applied to ensure data integrity and consistency.

Data Merging and Joining: Data from different sources are integrated through merging and joining operations. Common identifiers, such as movie IDs or studio names, are used to combine datasets and create a unified view of the data for analysis.

Data Manipulation: Data manipulation techniques are applied to transform and reshape the data as needed for analysis. This may involve aggregating data, creating new variables, or filtering observations based on specific criteria. The goal is to prepare the data for analysis and extract meaningful insights.

Data Analysis:

Finding Trends and Patterns: Using statistical analysis and machine learning algorithms, we uncover trends, patterns, and relationships within the data. This includes identifying rating trends over time, genre-specific trends, studio performance, and correlations between critic and audience ratings.

Data Visualization:

Data visualization techniques are employed to communicate insights and findings effectively. Charts, graphs, and dashboards are created to visually represent trends, patterns, and relationships discovered during the analysis phase. Visualizations enhance understanding, facilitate decision-making, and enable stakeholders to derive actionable insights from the data.

Goals and KPIs (3 - 4 Max)

To measure the success of our project, we will establish key performance indicators (KPIs) aligned with our project goals. Here are the goals and corresponding KPIs:

- **Goal 1:** KPI -----> Achieve a target rating score of 85% or higher on the Tomatometer

- **Goal 2:** KPI Increase the number of Certified Fresh ratings by 15% compared to the previous year

- **Goal 3:** KPI -----> Improve audience engagement and satisfaction, aiming for an audience rating score of 4.5 out of 5 or higher

Technical Processes

VLOOKUP: Relevance to Project: VLOOKUP is used in Excel to search for a value in the first column of a range and return a corresponding value from a different column. In our project, we might use VLOOKUP to match movie IDs from the "Ratings" dataset with movie titles and genres from the "Movies" dataset. This allows us to aggregate and analyse rating data by genre and title.

Pandas Data Cleaning: Relevance to Project: Pandas is a Python library commonly used for data manipulation and analysis. Data cleaning techniques taught in the module, such as handling missing values, removing duplicates, and standardizing data formats, are essential for preparing datasets for analysis. In our project, we would use Pandas to clean the raw data extracted from SQL databases, AWS, or local sources before performing further analysis.

Schema Creation for Database: Relevance to Project: Schema creation involves defining the structure of a database, including tables, columns, data types, and relationships between tables. In our project, we might design a schema for a SQL database to store information such as movie details, ratings, and reviews. The schema ensures data consistency and integrity, facilitating efficient data storage and retrieval for analysis.

Business Concepts Used

Market Understanding: Use: We have conducted market analysis to gain insights into the film industry, including market trends, competitor strategies, and consumer preferences. By understanding the market dynamics, we can tailor our marketing and distribution strategies to better meet audience needs and stay ahead of competitors.

Customer Demographic: Use: We have segmented our customer base based on demographic factors such as age, gender, income level, and location. This segmentation allows us to target specific customer groups with personalized marketing messages and offerings. For example, we may offer family meal deals to households with children or premium pizza options to higher-income customers.

Customer Behaviour:

Use: We have segmented our audience based on demographic factors such as age, gender, and location. This segmentation allows us to target specific audience groups with personalized marketing messages and promotional offers. For example, we may promote family-friendly films to households with children or niche genres to specific demographic segments.

Customer Retention: Use: We have analysed customer behaviour, including viewing patterns, genre preferences, and review ratings. This analysis helps us understand audience preferences and identify opportunities to enhance viewer engagement. For instance, if we notice that certain genres are consistently rated highly, we may focus on producing and promoting more films in those genres.
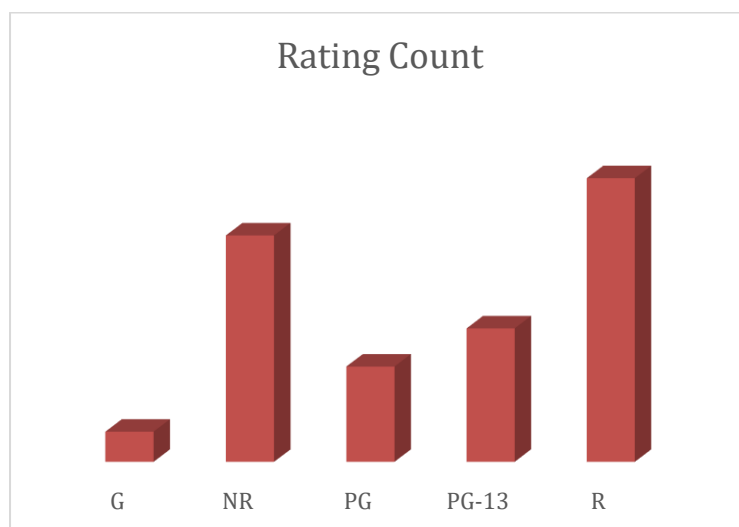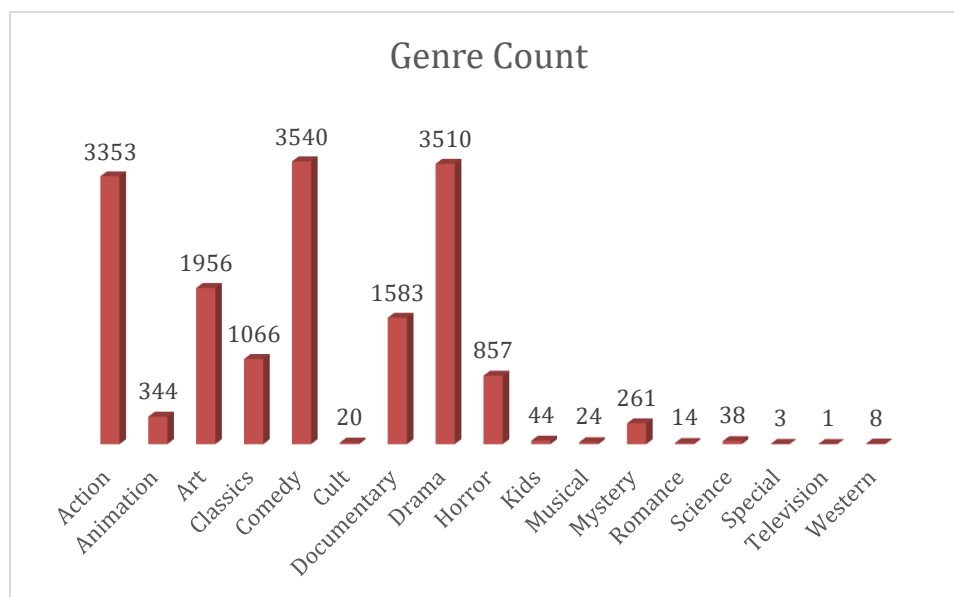
New Customer Acquisition:

Use: We have implemented tactics to attract new viewers, such as targeted advertising campaigns, referral programs, and special promotions for new users. By expanding our audience base, we can increase viewership and grow our platform. For instance, we may offer a "refer a friend" program where existing viewers receive a discount on their next subscription for referring a new viewer to our platform.

Recommended Analysis:

Question 1: What does the distribution of films look like by rating? My primary genre? (hint: use the first genre listed)

This analysis aims to distribution of films that look like b rating. So after analysing it is seen that movies with action genre are having highest rating followed by comedy and drama genre. Also most movies are R rated.

Genre Count

Action 3353, Animation 344, Art 1956, Classics 1066, Comedy 3540, Cult 20, Documentary 1583, Drama 3510, Horror 857, Kids 44, Musical 24, Mystery 261, Romance 14, Science 38, Special 3, Television 1, Western 8

Rating Count

G, NR, PG, PG-13, R

Question 2: What % of films received a Certified Fresh Tomatometer rating? What about Rotten?

This analysis focuses on determining the percentage of films received a certified fresh tamatometre rating. So it is seen that 18% of movies are certified fresh, 43% is rotten and remaining 39% are fresh.
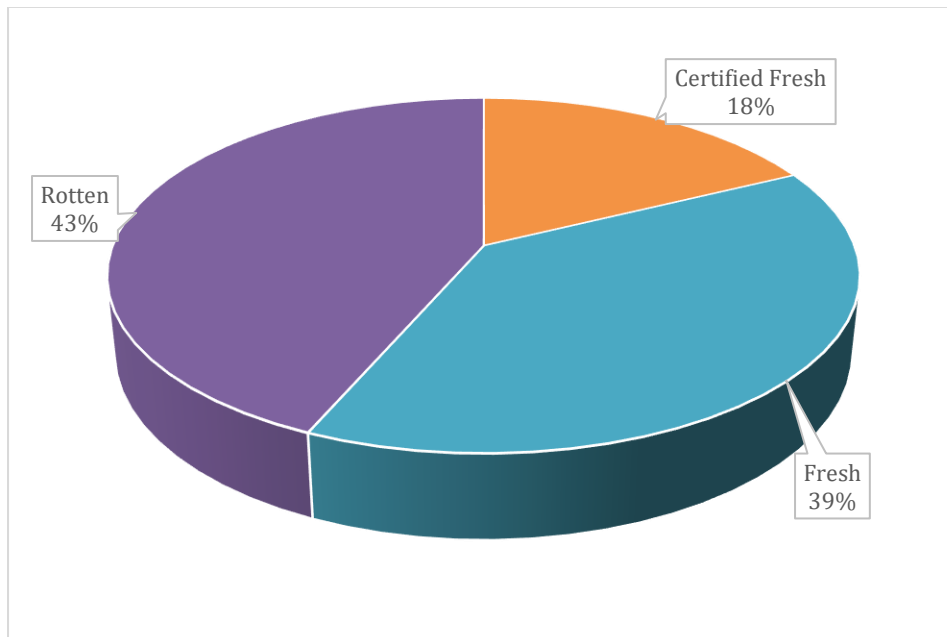
Question 3:. Explore new film releases over time. How has the volume of releases by month trended over time? What year/month were the newest films released?

This analysis involves new film releases overtime. Analysis shows that film releases have increased overtime.
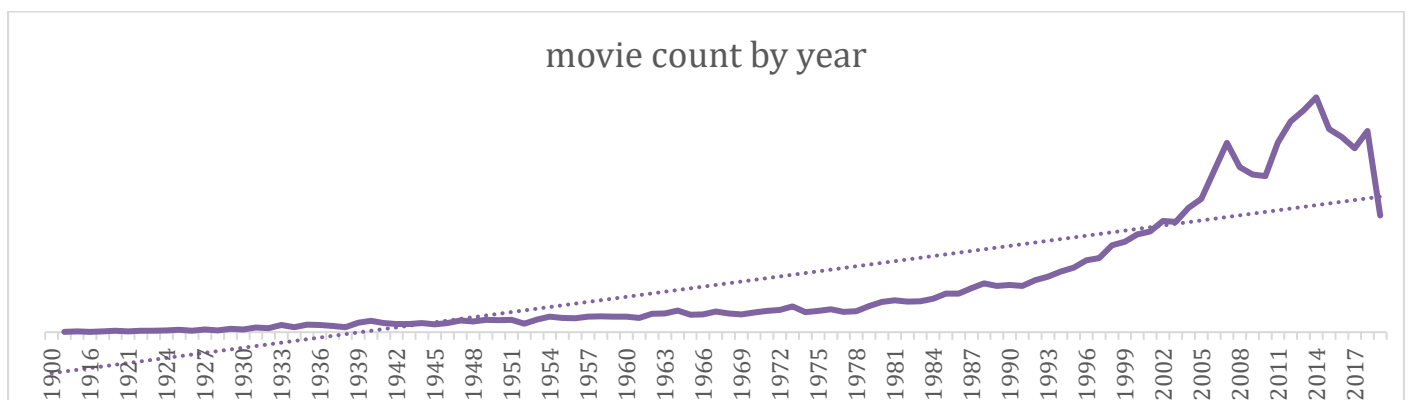
Question 4: Compare average Tomatometer ratings by Studio. Which studios produce the highest-rated films, on average? The lowest?

Lowest rated:

studio	Average of tomatometer_rating	Count of movie_title
Warner Bros. Pictures International	77	1
Warner Bros. Home Entertainment	44	1

Highest Rated:

Row Labels	Average of tomatometer_rating	Count of movie_title
Paramount Pictures	60.69416499	497
Warner Bros. Pictures	56.93902439	492

Question 5:. Compare the Tomatometer ratings against audience ratings. Which films showed the largest discrepancies between audiences and critics?

96 souls has the largest discrepancy between audience and critics.

Process of Working on the Recommended Analysis (Data to Output):

Data Collection: Gather relevant data sources, including sales records, customer transactions, and promotional data.

Data Cleaning and Preparation: Cleanse the data to remove duplicates, handle missing values, and standardize formats. Prepare the data for analysis by structuring it into a format suitable for processing.

Data Analysis: Apply analytical techniques to answer each question, such as calculating daily customer counts, average pizzas per order, total revenue, and identifying seasonal sales patterns. Utilize Excel formulas, PivotTables, and statistical analysis to derive insights from the data.

Data Visualization: Present the analysis results using visualizations such as charts, graphs, and dashboards to communicate key findings effectively.

Insights and Recommendations: Interpret the analysis findings to draw actionable insights and recommendations. Identify areas for improvement, such as menu optimization, pricing adjustments, or promotional strategies, to address any identified issues or capitalize on opportunities.

Impact of Solution:

The implementation of the proposed solution based on the insights gained from the Rotten Tomatoes movie data analysis project is expected to have significant positive impacts on various aspects of the organization's operations, strategy, and performance. In summary, the implementation of the proposed solution based on the insights gained from the Rotten Tomatoes movie data analysis project is expected to drive improvements across various aspects of the organization, leading to increased efficiency, effectiveness, and competitiveness in the marketplace.

Project owner

Name: Samiksha Umesh Jagne

Date: June 10, 2024

LinkedIn: https://www.linkedin.com/in/samiksha-jagne/

Github: https://github.com/samikshajagne

Email: samikshajagne2@gmail.com

Peerlist: https://peerlist.io/samiksha_jagne