

Table of Contents

[Problems and background 2](#_heading=h.gjdgxs)

[Solution 2](#_heading=h.30j0zll)

[Data Analysis 2](#_heading=h.1fob9te)

[Data-Driven Insights 2](#_heading=h.3znysh7)

[Stakeholder Engagement 2](#_heading=h.2et92p0)

[Project scope 3](#_heading=h.tyjcwt)

[Data Collection and Cleaning 3](#_heading=h.3dy6vkm)

[Recommended Data Analysis 3](#_heading=h.1t3h5sf)

[Presentation and Discussion 4](#_heading=h.4d34og8)

[Methodology 4](#_heading=h.2s8eyo1)

[Data Source 4](#_heading=h.17dp8vu)

[Data Wrangling 5](#_heading=h.3rdcrjn)

[Business Concepts Used 5](#_heading=h.26in1rg)

[Consumer Behaviour Analysis 5](#_heading=h.lnxbz9)

[Pricing Strategy 5](#_heading=h.35nkun2)

[Customer Satisfaction and Feedback Analysis 5](#_heading=h.1ksv4uv)

[Risk Analysis 5](#_heading=h.44sinio)

[Data-Driven Decision Making 6](#_heading=h.2jxsxqh)

[Conclusion 6](#_heading=h.z337ya)

# Problems and background

The Rotten Tomatoes Analysis Report is a comprehensive project centered around a rich dataset encompassing over 15,000+ films reviewed by Rotten Tomatoes spanning from 1914 to 2019.

Rotten Tomatoes, the Tomatometer, and Audience scores are the world’s most trusted recommendation resources for quality entertainment. As the leading online aggregator of movie and TV show reviews, provide fans with a comprehensive guide to what’s Fresh – and what’s Rotten – in theaters and at home.  
  
The data includes, title, description, rating, genre, directors, cast, release date, studio duration, tomatomater status, audience ratings and much more.

# Solution

The Rotten Tomatoes Analysis Report aims to provide an in-depth examination of a substantial dataset covering over 15,000 films reviewed on Rotten Tomatoes between 1914 and 2019. By leveraging the Tomatometer and Audience scores—widely recognized as the most trusted metrics for entertainment quality—this project seeks to uncover patterns, trends, and insights into the film industry over more than a century.

The analysis will explore various dimensions such as movie genres, ratings, critical reception, audience opinions, and key contributors (directors, cast, studios) to identify factors that influence a film's success. This report will serve as a valuable resource for industry professionals, researchers, and enthusiasts by providing a detailed understanding of what makes a movie Fresh or Rotten, both in theaters and at home.

## Data Analysis

The analysis will involve a thorough exploration of the dataset, including variables such as movie titles, genres, ratings, release dates, directors, cast, studios, and audience feedback. Advanced statistical methods and data visualization techniques will be employed to identify trends, correlations, and patterns in critical and audience responses. This analysis will help uncover key factors that contribute to a film's success, be it critically acclaimed or popular among audiences.

## Data-Driven Insights

The project will generate actionable insights based on data, revealing the characteristics that make a movie Fresh or Rotten. These insights will be instrumental for filmmakers, studios, and marketers in understanding the elements that resonate with critics and audiences alike. The findings will also provide a historical perspective on how preferences and trends have shifted over time in the film industry.

## Stakeholder Engagement

The report will engage various stakeholders, including film industry professionals, researchers, and entertainment enthusiasts. By presenting data-driven insights in a clear and accessible manner, the project aims to inform decision-making processes and strategic planning in film production, marketing, and distribution. Stakeholders will be equipped with a deeper understanding of the dynamics that influence movie success, enabling them to make informed choices that align with audience expectations and critical standards.

# Project scope

The scope of this project includes analyzing a broad range of data points such as movie titles, descriptions, ratings, genres, directors, cast members, release dates, studios, and both Tomatometer and Audience scores. The analysis will cover over a century of cinematic history, focusing on trends, patterns, and correlations that have emerged over time. The project will not only examine critical and audience reception but also explore how factors like genre, cast, and release timing contribute to a film’s performance. The goal is to produce a comprehensive report that serves as a resource for filmmakers, studios, marketers, and entertainment enthusiasts.

## Data Collection and Cleaning

The dataset for this project has been compiled from Rotten Tomatoes, encompassing detailed information on over 15,000+ films. The data includes variables such as title, genre, director, cast, release date, studio, duration, Tomatometer status, and audience ratings, among others. Before analysis, the data will undergo a rigorous cleaning process to ensure accuracy and consistency. This will involve handling missing values, correcting any inconsistencies, and standardizing data formats. Special attention will be given to ensuring that the dataset is comprehensive and free from errors, providing a solid foundation for reliable analysis.

## Recommended Data Analysis

Analyze the sales data to address the management's questions and objectives:

1. What does the distribution of films look like by rating? My primary genre? (Hint: use the first genre listed) Identify the best-selling pizzas based on sales volume.
   * R-rated films are the most prevalent, making up 37.89% of the total films., NR (Not Rated) films account for 27.25%, PG-13 films represent 17.95%., PG films make up 12.68%, G films constitute 4.01%, NC-17 films are the least common, at 0.22%.
   * This distribution suggests that the majority of films are rated R or NR, with Comedy being the predominant genre.
2. What % of films received a Certified Fresh Tomatometer rating? What about Rotten? Insights and Recommendations
   * Certified Fresh Tomatometer rating was received by 18.51% of the films.
   * Rotten Tomatometer rating was received by 43.84% of the films.
   * This suggests that nearly half of the films were rated as Rotten, while a smaller percentage achieved the Certified Fresh status.
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?
   * Trend of Film Releases Over Time:
     + The volume of film releases has significantly increased over the years, with a noticeable rise from the 1980s onward.
     + The peak in the number of film releases occurred in 2014 with 743 films.
   * Monthly Distribution of Film Releases:
     + January had the highest number of film releases with 1,759 titles.
     + October and June also saw a high volume of releases, with 1,559 and 1,428 films respectively.
     + February had the fewest releases with 1,009 films.
   * Recent Years:
     + The most recent years with notable numbers of new film releases are 2017 to 2019.
     + In 2019, there were 369 new film releases.
   * This data shows a clear trend towards an increasing volume of film releases over time, particularly peaking in the 2010s, with January being a prominent month for new film releases.
4. Compare average Tomatometer ratings by Studio. Which studios produce the highest-rated films, on average? The lowest?
   * Paramount Pictures studio produces highest rated films that has an average tomatometer ratings of 60.70.
5. Compare the Tomatometer ratings against audience ratings. Which films showed the largest discrepancies between audiences and critics?
   * The movie “96 Souls” has been found with the highest discrepancies between audience and tomatometer ratings.
6. Explore the critic's consensus rating: what language is used most often?
   * Positive language is used most often, accounting for 76.09% of the ratings.
   * Neutral language follows at 14.27%.
   * Negative language is the least common, making up 9.64%.
   * This indicates that the majority of the critic's consensus ratings are favourable, with positive terms being predominant in the language used.

## Presentation and Discussion

Prepare a comprehensive report or presentation summarizing the findings and recommendations.

* Present the data analysis results in a clear and understandable format.
* Engage in discussions with the management to ensure a thorough understanding of the insights and recommendations.
* Address any questions or concerns raised by the management and provide further clarification or analysis as needed.

# Methodology

## Data Source

The raw data of 15000+ movies is available in excel files from the year starting from 1914 to 2019.

## Data Wrangling

1. **Data Understanding:** Familiarize yourself with the structure and content of the data obtained/received from the sources. Understand the meaning of each filed or variable to the analysis goals.
2. **Data Cleaning:** We need to clean the data to ensure the accuracy, consistency and completeness. This involves the tasks like handling the missing values, correction of errors and inconsistency of the data, standardizing the formats, and removing the duplicate values and any outliers if found.
3. **Data Merging and Joining:** Involve in the process of creating a unified dataset for analysis. Merge dataset based on common identifiers.
4. **Data Manipulation:** Perform transformation and calculation on the data to derive new variable or outcomes that are relevant to the analysis objectives.

# Business Concepts Used

Several business concepts can be applied to analyze the Rotten Tomatoes dataset, offering insights into various aspects of the film industry. Here are some key concepts:

Market Segmentation:

* **Concept:** Market segmentation involves dividing a broad target market into smaller, more defined categories.
* **Application:** Analyze audience ratings by demographics, genres, or release periods to identify specific segments of moviegoers and how their preferences differ. This can help studios tailor marketing strategies for different audience segments.

## Consumer Behaviour Analysis

* **Concept:** Understanding how consumer makes decisions and what factors influence their preferences.
* **Application:** Examine patterns in audience score to understand what drives positive or negative reactions. Analyze the impact of elements like star power, genre, or marketing campaign on consumer behaviour.

## Pricing Strategy

* **Concept:** Pricing strategy involves getting the right price for the product based on market demand, cost and competitor pricing.
* **Application:** Analyze how timing of a movie’s release (e.g. during holiday or summer) and its pricing (e.g. ticket price, streaming rental fees) correlate with audience rating and box office success.

## Customer Satisfaction and Feedback Analysis

* **Concept:** This concept focuses on measuring and analyzing customer satisfaction to improve product and services.
* **Application:** Use audience scores as measures of customer satisfaction and analyze feedback to identify areas where films meet or fall shorts of expectations. This can guide future content creation to better align with audience desire.

## Risk Analysis

* **Concept:** Risk analysis involves identifying potential risks in business decisions and determining how to manage them.
* **Application:** Evaluate the risk associated with producing films in certain genre or with particular casts based on historical performance data. This can help studios to make informed decisions about which projects to greenlight.

## Data-Driven Decision Making

* **Concept:** Using data and analytics to guide strategic business decision
* **Application:** Utilize the insight from the dataset to make data-drive decision about future film production, marketing strategies, and distribution methods. This ensure that business choices are informed by historical trends and audience behaviour.

# Conclusion

The Rotten Tomatoes Analysis Report provides a comprehensive examination of over a century of film data, offering valuable insights into the factors that drive a movie's success or failure. By leveraging trusted metrics like the Tomatometer and Audience scores, this analysis uncovers critical trends, patterns, and relationships within the film industry. The application of business concepts such as market segmentation, consumer behaviour analysis, and risk analysis further enhance our understanding of the industry’s dynamics.

Through rigorous data analysis and cleaning, the report ensures that stakeholders—including filmmakers, studios, and marketers—receive accurate, actionable insights that can inform strategic decisions. Whether it’s identifying key success factors, analyzing historical trends, or understanding audience and critic discrepancies, the findings from this report are poised to help industry professionals navigate an ever-evolving entertainment landscape.

Ultimately, this project underscores the importance of data-driven decision-making in the film industry, providing a robust foundation for creating, marketing, and distributing films that resonate with both critics and audiences alike.