

# **Exploratory Data Analysis for Microsoft's Movie Studio Initiative**

# Contents

1. Introduction to the Project
2. Data Collection and Cleaning
3. Exploratory Data Analysis
4. Visualizations and Findings
5. Additional Observations
6. Data Analysis Results
7. Implications and Looking Ahead
8. Strategic Recommendations
9. Final Recommendations and Summary
10. Conclusion and Future Scope

Cool



# Introduction to the Project

Microsoft keen on creating a new movie studio requires an in-depth study about which film types are currently performing well at the box office. This project will explore and translate those findings into actionable insights to help head of Microsoft's new movie studio decide on the kind of films to create. It will investigate performance of various films using relevant databases to gather sufficient data ensuring a comprehensive analysis. The ultimate goal is determining a successful formula for creating films resonating with audience and assuring financial success.



# Data Collection and Cleaning

Various Databases and Python Libraries used for project

## Data Sources

Data collected from several renowned movie databases like Box Office Mojo, IMDb, Rotten Tomatoes, TheMovieDB, and The Numbers. Each providing crucial information for comprehensive analysis including box office revenue, movie ratings, reviews, and production budgets.

## Python Libraries

Data manipulation and visualization were achieved via Python libraries like Pandas, NumPy, Matplotlib, and Seaborn providing essentials for data structures and operations, supporting large multi-dimensional arrays, and creating interactive graphical visualizations.

## Cleaning Steps

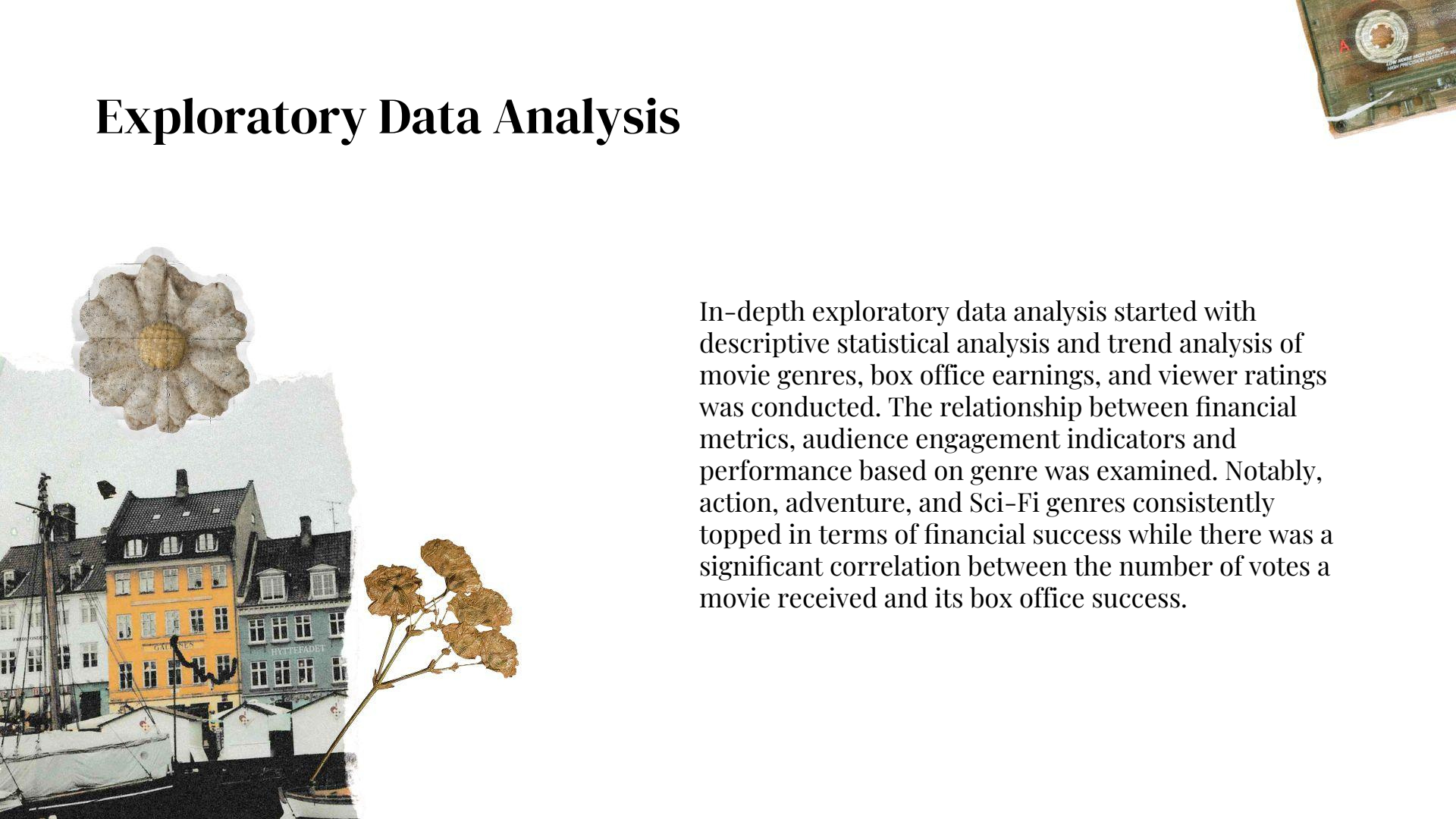
Accuracy and reliability of data ensured by removing duplicates, handling missing values with appropriate data, converting all data to correct format, merging necessary datasets based on common identifiers, and trimming outliers to avoid skewed analysis.





# Exploratory Data Analysis

In-depth exploratory data analysis started with descriptive statistical analysis and trend analysis of movie genres, box office earnings, and viewer ratings was conducted. The relationship between financial metrics, audience engagement indicators and performance based on genre was examined. Notably, action, adventure, and Sci-Fi genres consistently topped in terms of financial success while there was a significant correlation between the number of votes a movie received and its box office success.



# Visualizations and Findings

## Observations from Correlations and Genre Assessment



### 1 Correlation Observations

Movies with ratings between 6 to 8 with relatively lower domestic gross showed that above-average ratings don't necessarily correlate with high box office returns. High domestic gross spread across different average ratings suggesting other factors contribute to financial success.

### 2 Top Grossing Genres

Action/Adventure/Sci-Fi genre revealed the highest total gross revenues (\$2.5 billion), Adventure/Animation/Comedy was the second (\$2.4 billion), while Action/Adventure/Fantasy genre stood third (\$1.5 billion). This shows concentration of earnings at the top.

### 3 Yearly Trend for Genres

Significant year-over-year fluctuations in average total gross for each genre observed. Sci-Fi and Adventure genres showed higher average grosses denoting their strong box office appeal.

# Additional Observations

## Genre Preference, Gross Earnings and Ratings

### 1 Market Share by Genre

Adventure and Action genres clearly dominate market with shares of 23.2% and 20.2%, respectively. Similarly, significant market shares held by Comedy, Drama and Animation genres.

### 2 Comedy Movies Distribution

Significant number of comedy movies seen to earn less in domestic gross revenue indicating that only few achieve high levels of domestic box office success.

### 3 Average Ratings distribution

The market shares of average ratings among top 10 movie genres is quite close, indicating diverse interest among audiences and importance of high-quality content independent of the genre.



# Data Analysis Results

## Correlation Matrix and Genre Ratings

### Correlation Matrix

Films with higher ratings tend to receive more votes indicating that quality drives engagement. Films that perform well domestically also tend to perform well in foreign markets showing global appeal of movies.

### Average Ratings Over Years

The market share of average ratings by top 10 genres is relatively uniform. There is a stable interest and quality in the genres represented, year over year. No single genre or year overwhelmingly dominates the chart.

### Recommendations

Action, Adventure, and Sci-Fi genres consistently show financial success. Understanding trends in years with higher engagement and ratings, delivering high-quality content irrespective of genre, and content that resonates universally can lead to success of the new movie studio.





# Implications and Looking Ahead

The study provides crucial insights into multifaceted industry where genre, quality, and audience engagement intersect to dictate financial success. The genres of Action, Adventure, and Sci-Fi emerge as leading choices at the box office. Consistent high-quality content seems to be of paramount importance for the success of Microsoft's new movie studio. In addition, films that perform well domestically also tend to have strong performance abroad, thus creating content with universal themes and appeal could be key to content development strategy.



# Strategic Recommendations

## Key Points for Microsoft's New Movie Studio

### Genre Focus

Investing in genres with consistent financial success like Action, Adventure, and Sci-Fi. Exploring other genres demonstrating potential through strong engagement metrics could also be beneficial.

### Yearly Trends and Engagement

Understanding why certain years had higher engagement and ratings could uncover opportunities for content that resonates with audiences. It is also crucial to incentivize audience interaction.

### Quality and Global Market Appeal

Emphasis on producing high quality content across all genres is necessary as they tend to get more votes and lead to better financial outcomes. Content should be designed with a universal appeal to capitalize on correlation between domestic and international box office success.



# Final Recommendations and Summary

## Data Visualization Improvements and Long-Term Strategy



### 1 Data Visualization Recommendations

Enhancing clarity of data visualizations can be achieved by clearly associating colours with specific genres and making sure that the visual data connects directly to the data points it represents.

### 2 Long-Term Strategy

A long-term content strategy should be built on the consistent interest in top genres. Optimal length for movies that balances audience engagement with runtime without necessarily aiming for the longest duration is worth considering.

### 3 Data driven Decisions

Cross-analysis of genres, ratings, and engagement data can inform the complex formula for a movie's success. Data-driven decisions can aid in crafting a roadmap for content that is both critically acclaimed and financially successful.

# Conclusion and Future Scope

Microsoft's new movie studio can leverage these insights to navigate the complexities of the film industry. By focusing on audience engagement, investing in high-grossing genres, delivering high-quality irrespective of genre, and harnessing the data at our disposal, the studio is well-positioned to carve out its niche and achieve lasting success in the film industry. We look forward to expanding our research methodology to include other factors impacting movie success and extending our analysis to newer databases for richer insights.





**Thank You**

Your Attention and Time is Greatly Appreciated

MOHAMED  
ALI

