**IMDb Movies Analytics Project Report**

**Cover Page**

**Project Title:** Exploring Movie Trends and Oscar-Winning Patterns (2010–2022)

**Course:** MSIS670 Business Intelligence – Spring 2025

**Instructor:** ProfessorKui Du

**Name:** Jyostna Sree Somisetty

**Student ID:** 02143296

**Date:** May 20, 2025

**Tableau link:** <https://prod-useast-b.online.tableau.com/#/site/msis670b/workbooks/2766826>

**Table of Contents**

1. Executive Summary / Abstract
2. Introduction
3. Data Preparation and Cleaning
4. Methodology
5. Results & Visualizations
   * 5.1 Research Question 1
   * 5.2 Research Question 2
   * 5.3 Research Question 3
6. Discussion
7. User Guide to Tableau Dashboard
8. Conclusion & Recommendations
9. References
10. Appendices

**1. Executive Summary / Abstract**

This project examines significant trends in the global film industry by utilizing the IMDb dataset and concentrating on English-language films released from 2010 to 2022. Through the use of Tableau for interactive visualizations, the project reveals insights into patterns of Oscar recognition, audience engagement (indicated by IMDb votes), and production trends across various continents.

The findings indicate that Oscar-nominated and winning films are predominantly found in the U.S. and Europe. Genres such as Biography and Drama lead in awards, while popular genres like Action and Animation are less recognized. Additionally, the COVID-19 pandemic significantly impacted global film production. An extensive Tableau dashboard enables users to explore these trends interactively.

**2. Introduction**

The film industry significantly impacts culture and the world economy. As thousands of films are released each year, gaining insights into the factors that contribute to critical acclaim and audience popularity can help shape industry strategies. This project explores contemporary film trends using business intelligence methods, addressing three key research questions:

* What are the characteristics of Oscar-winning and nominated movies?
* How do audience engagement metrics vary across genres, years, and regions?
* What are the temporal and regional patterns in movie production?

Target audiences include industry professionals, film researchers, and data enthusiasts.

**3. Data Preparation and Cleaning**

The dataset used contains 33,264 English-language films (2010–2022), sourced from a simplified version of IMDb. Fields include title, year, rating, votes, genre, country, continent, and Oscar award status.

Data Cleaning Steps:

* Filled missing values in Genre, Runtime, Continent, and Award fields.
* Created Award Category field: “Winner”, “Nominated”, or “None”.
* Standardized numeric types for Runtime, Ratings, and Votes.
* Filtered to include only years 2010–2022.
* Removed "Unknown" continents for clarity in trend analysis.

**4. Methodology**

Visual analytics were conducted using Tableau Online. Three main visualizations were developed:

1. A geographic map of Oscar recognition by country
2. A bar chart of average IMDb votes by genre
3. A line chart of film production volume over time by continent

Calculated fields, filters, tooltips, and color legends were used to enhance interactivity. Tableau was chosen for its ease of use and support for exploratory data analysis.

**5. Results & Visualizations**

**5.1. Research Question 1:** Characteristics of Oscar-nominated and winning films

**Visuals Used:** Map of Award by Country, Bar Chart by Genre, Box Plot of Runtime

**Findings:**

* Oscar wins are heavily concentrated in the U.S., UK, and Canada.
* Biography, Drama, and History genres are most recognized.
* Runtime for Oscar-recognized films typically falls between 100–140 minutes.
* Animation and Action genres, though popular, rarely win Oscars.

A map of the world

AI-generated content may be incorrect.

**5.2. Research Question 2:** Audience Engagement Patterns

**Visual Used:** Bar Chart of Average Votes by Genre

**A screenshot of a computer

AI-generated content may be incorrect.**

**Findings:**

* Highest average IMDb votes were for Biography, Action, and Animation.
* Genres like Reality-TV, Musical, and Western saw minimal engagement.
* High vote count doesn’t always translate to critical acclaim (Oscars).

**5.3. Research Question 3:** Production Trends by Continent

**Visual Used:** Line Chart of Production Volume by Continent

**Findings:**

* The Americas (mainly U.S.) produce the majority of English-language films.
* Europe has consistent yearly output (~400–500 films).
* COVID-19 caused visible dips in 2020–2021; 2022 showed recovery.
* Asia, Africa, and Oceania contribute marginally to English film production.

**A graph of a graph

AI-generated content may be incorrect.**

**6. Discussion**

The analysis highlights regional imbalances in award recognition and production. There's a notable misalignment between genres preferred by audiences (Action, Animation) and those rewarded by the Academy (Drama, Biography).

The data also points to the pandemic's measurable impact on global production and reinforces the dominance of Western countries in award circuits. More comprehensive data (e.g., box office, reviews, cast impact) could further enrich these findings.

**7. User Guide to Tableau Dashboard**

* Interactivity: Click on countries, genres, or year markers to filter all charts.
* Legends: Toggle Award Category to isolate nominees, winners, or others.
* Tooltips: Hover over bars, lines, or map regions to view titles, vote counts, and award status.
* Reset View: Refresh or click “Clear All Filters” to return to the full dataset.

**8. Conclusion & Recommendations**

**Summary of Findings:**

* Oscar recognition remains regionally and generically concentrated.
* Action and Animation genres are popular but under-recognized.
* Audience votes do not align with critical acclaim.
* Global film production is centralized and was disrupted by external events like COVID-19.

**Recommendations:**

* Producers: Invest in high-engagement genres but balance creative and award ambitions.
* Analysts: Dive deeper into international and non-English content.
* Viewers: Explore underappreciated films in popular genres or lesser-known countries.

**9. References**

* IMDb Simplified Dataset (2024 Edition)
* Tableau Online & Tableau Public
* MSIS670 Course Materials, UMass Boston
* OpenAI ChatGPT (used for language structuring, not for analysis)

**10. Appendices**

* **Appendix A: Screenshot of Final Dashboard**

**A screenshot of a graph

AI-generated content may be incorrect.**

* **Appendix B: Data Cleaning Summary**

The original IMDb dataset included over 33,000 English-language films released between 2010 and 2022. To ensure analytical accuracy and visualization clarity, the following data cleaning and preparation steps were performed:

**1. Year Filtering**

* **Action:** Filtered the dataset to only include films released from 2010 to 2022.
* **Rationale:** Ensured consistency with project scope and removed outdated or future-dated records.

**2. Best Picture Field Cleanup**

* **Action:** Filled blank values in the Best Picture field with "None".
* **Rationale:** Standardized this field to create a new calculated field (Award Category) and to accurately classify all movies for comparison.

**3. Calculated Field – Award Category**

* **Action:** Created a custom field:

IF [Best Picture] = "Winner" THEN "Winner"

ELSEIF [Best Picture] = "Nominated" THEN "Nominated"

ELSE "None"

END

* **Rationale:** Grouped all movies into three mutually exclusive award statuses for meaningful analysis.

**4. Genre Standardization**

* **Action:** Replaced missing or invalid entries in the Genre field with "Unknown" or "Other".
* **Rationale:** Ensured genre-based analysis could include all entries without error or omission.

**5. Runtime Validation**

* **Action:** Verified and cleaned entries in the Runtime (Minutes) field.
* **Rationale:** Ensured all runtime values were numeric and within a reasonable range for box plots.

**6. Continent & Country Cleaning**

* **Action:** Replaced missing continent values with "Unknown" and removed them from continent trend analysis.
* **Rationale:** Maintained regional clarity in the production trend chart while minimizing data distortion.

**7. Type Conversion**

* **Action:** Converted Votes, Runtime, and Rating columns from text to numeric types in Tableau.
* **Rationale:** Enabled proper aggregation, sorting, and chart creation without type mismatch errors.

**8. Handling Outliers**

* **Action:** Reviewed vote counts, runtime values, and rating distributions for extreme values.
* **Rationale:** Verified that outliers did not skew visualizations significantly; no removals were deemed necessary.

**9. Final Filtering for Visual Quality**

* **Action:** Filtered out entries with "Unknown" continent in the line chart to ensure geographic integrity.
* **Rationale:** Focused trend analysis only on entries with valid continent assignments**.**
* **Appendix C: Tableau Calculated Field Syntax**

IF [Best Picture] = "Winner" THEN "Winner"

ELSEIF [Best Picture] = "Nominated" THEN "Nominated" ELSE "None" END