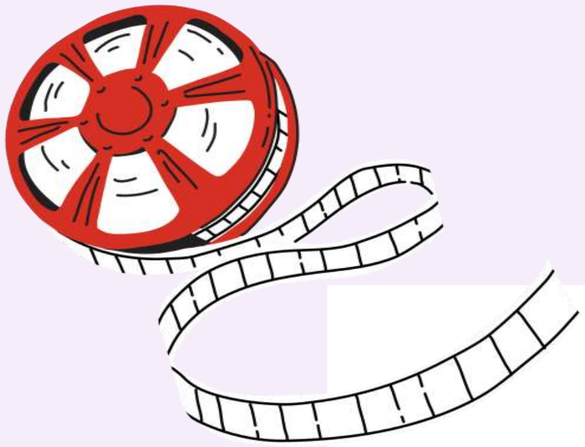


The Movie Success Formula

Data Visualization and Dashboard Development Project

Harshita Agrawal

Executive Summary

The film industry is one of the most vibrant and influential sectors of the global entertainment industry. As new digital platforms arise and audience preferences change, identifying what makes a film a success has become more complex and more critical. In this project, titled "Movie Success Formula", we venture into the field of cinema analytics by utilizing the IMDB and TMDB datasets to uncover patterns, trends, and insights that contribute to the success of a film.

Our examination utilizes Tableau data visualization, and by so doing we can explore key variables such as movie rating, length, genre, budget, revenue, cast, and crew traits. By the creation of interactive dashboards, our aim is to provide an exhaustive and convincing graphical representation of factors that most strongly correlate with the audience rating and the box office.".

This project not only illustrates what elements are consistently present in successful films but also provides strategic suggestions to filmmakers, producers, and marketers who want to maximize content for success in a new market.

Business Context

The global film industry is a billion-dollar world where every movie is an investment of a high risk. From development to production to distribution and release, studios spend heavy amounts of capital on the promise of grabbing viewers' attention and reaping healthy returns. And yet the fate of a film is decided by a myriad of factors—some measurable, others not.

Our project, Movie Success Formula, is on this business requirement. From an analysis of IMDB and TMDB data, we will search for persistent movie success predictors like optimal length, best-performing genres, relationship between budget and revenue, cast or director contribution, and role of genre conventions. These revelations can contribute to strategic decision-making in script development, casting, release timing, and marketing spend.

User Needs

Our primary users are film producers, studio executives, marketing professionals, and content strategists who require data-driven insights to inform decisions across the film lifecycle.

- Forecasting audience reception before production.
- Selecting genres statistically proven to work.
- Deciding on optimal movie lengths and release schedules.
- Comparing past performances to forecast new project outcomes.

Through our Tableau dashboard, we aim to meet these needs by offering:

- Interactive visualizations that reveal patterns in movie performance.
- Comparative analysis of high- vs. low-rated films.
- KPI summaries such as average ratings, budget-to-revenue ratios, and genre-based success metrics.
- User-friendly exploration tools for filtering by genre, year, and other attributes.

Grand Goal

To provide a thorough visual narrative of the economic impact of global cinema, audience preferences, and changing creative trends to enable stakeholders to understand what makes a blockbuster, what fans genuinely appreciate, and how the film business is changing across cultures and time.

Requirements

1. Data Requirements : IMDB and TMDB Dataset
2. Technical Requirements: Tableau for data visualization and transformation, Microsoft Excel, Figma for developing mockups
3. Python Library requirements: Numpy and Pandas.

Processes

Development of the "Movie Success Formula" dashboard was a sequential, multi-step activity that maximized accuracy of the data, applicability of insights, and user-friendly visualizations. The key steps are presented as below:

- Data collection
- Data cleaning and preparation
- Data import into Tableau
- Dashboard design and development
- Analysis and insight generation

Objective 1

To compare the international contributions and analyze the financial performance of the film industry.

Metrics Used
of countries producing films
Box office revenue per country
Budget vs revenue per genre
Revenue of high grossing movie
Country with highest revenue

Objective 2

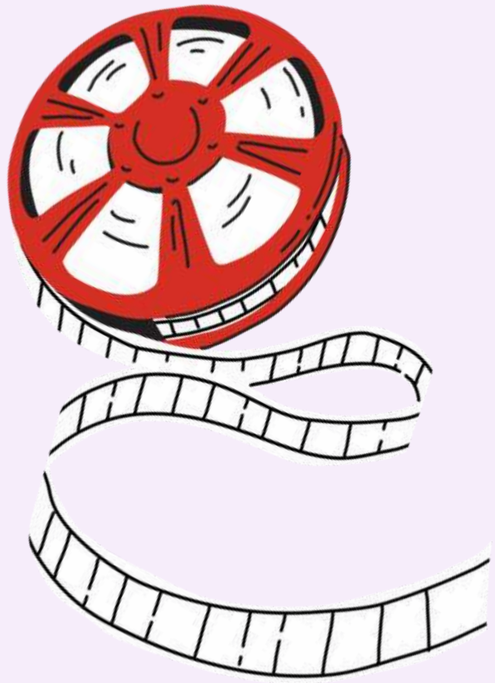
To determine the genres, directors, and themes that resonates most with global audience and to track their influence on success of the film.

Metrics Used
Trending genres and their viewer rating
Popularity of theme through word frequency
Top directors with high performing movies
Vote count and rating of top films
Highest rated movie

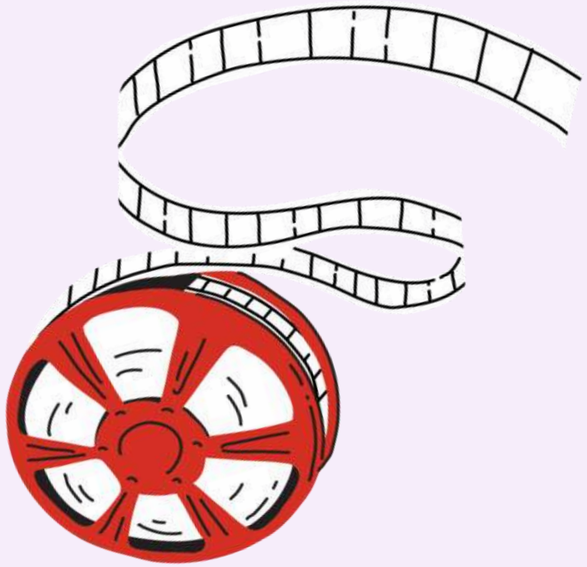
Objective 3

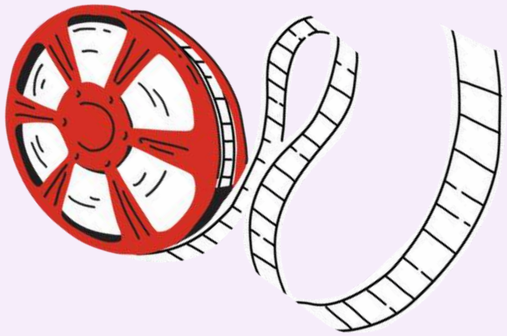
To visualize temporal trend and evolution of cinema with changing audience preference.

Metrics Used
of movies release per year
Average runtime trend with rating
Language specific film count
Genre evolution over time



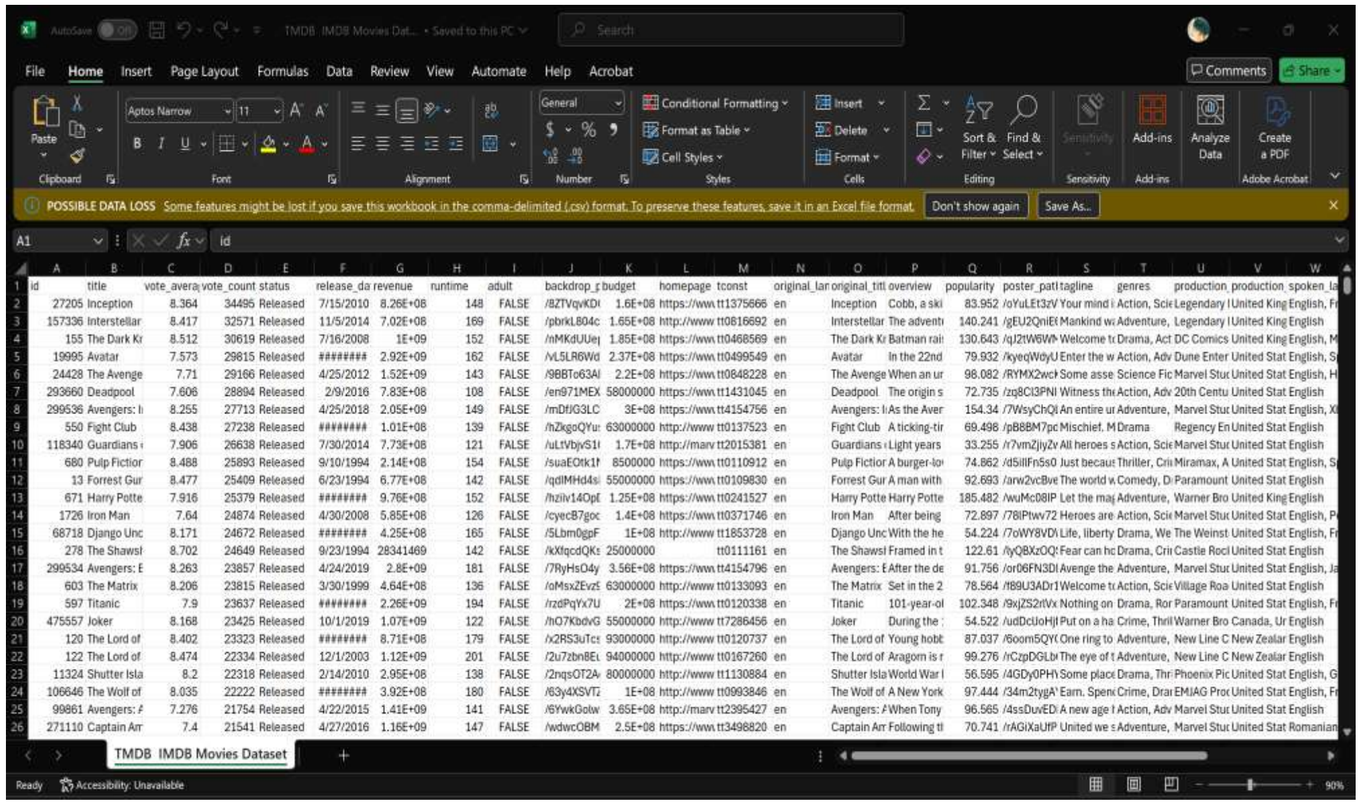
Data Description





Data source background

For this project, we used dataset combining information from IMDb (Internet Movie Database) and TMDb (The Movie Database)—two of the most widely recognized and reliable online film databases. IMDb provides critical data such as user ratings, vote counts, and cast/crew details, while TMDb offers extensive metadata like genres, popularity scores, and revenue figures.



- Link to the dataset : https://www.kaggle.com/datasets/ggtejas/tmdb-imdb-merged-movies-dataset
- Data Format: Dataset is stored in CSV file. Format is tabular, with rows representing individual movies and columns as attributes.
- Number of records: 433900+ Movies
- Number of features: 29

Data set features

We will focus on 19 features in this project, here are their description:

Column Name	Description
Id	Unique identifier for each movie
Title	Movie name
vote_average	Average score from TMDB users
vote_count	Total number of votes cast for the movie on TMDb
release_date	Official release date of the movie
revenue	Total earnings generated by the movie
runtime	Duration of the movie in minutes
budget	Total cost incurred by the movie
popularity	Popularity score based on TMDb analytics
genres	List of genres the movie falls under
production_companies	Name of production company of the movie
production_country	Name of country where movie was produced
spoken_languages	Language of the movie
keywords	Sentiment of the movie
directors	Name of the director of movie
writers	Name of the writer of movie
averageRating	Average rating of the movie on IMDb
numVotes	Number of votes the movie received on IMDb
cast	Names of actors in the movie

Data collection

Two well-liked sources were combined to create the dataset:

- IMDb: Supplied cast and crew details, movie titles, ratings, and vote counts.
- TMDb: Supplied metadata such as budgets, revenues, popularity scores, and genres.

Pre-integrated data was used for consistency and obtained in CSV format from publicly available datasets.

Data Cleaning (1/3)

Step 1: We have dropped the columns that are not required. The columns are:
status, backdrop_path, homepage, tconst, original_title, overview, poster_path, tagline.
This process is done by using python library pandas.

```
data1= data.drop(['status', 'backdrop_path', 'homepage', 'tconst', 'original_title', 'overview', 'poster_path', 'tagline'], axis=1).
```

As the dataset was large so it was very hard for us to work on this. Hence Pandas (Library) helped us to finish this transformation steps easily.

Step 2: We transformed the release_data data type from object to datetime as it was needed for the data visualization process.

```
data1['release_date']=pd.to_datetime(data1['release_date'], errors='coerce')
```

Data Cleaning (2/3)

Step 3: We filtered the dataset to keep movies having release year from 2000 and having orignal language english.

```
data1= data1[data1['release_date'].dt.year>=2000]
data2= data1[data1['original_language']=='en’]
```

Step 4: We are considering the movies which are available for all age groups.

```
data3= data2[data2['adult']==False]
```

We observed that the genres column had inconsistent formatting, with same genre names appearing in baried forms leading to duplication in visualization.

Data Cleaning (3/3)

Step 5: To standardize the entries by sorting them uniformly and removing extra spacing to ensure consistency.

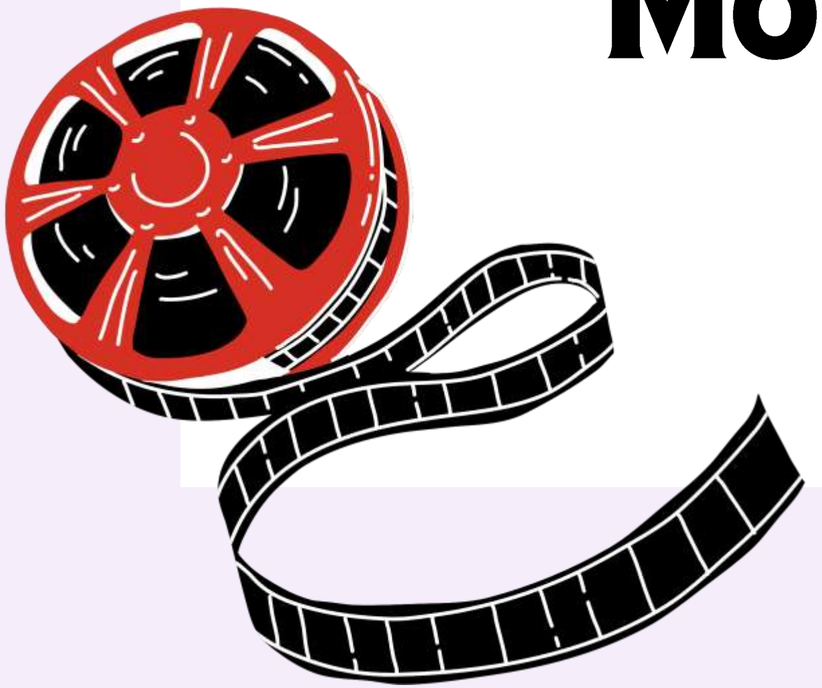
```
def clean_genre(genres):
  if isinstance(genres, str):
   genre_list= genres.split(',')
   clean_list=[]
   for g in genre_list:
     clean_list.append(g.strip())
   clean_list.sort()
    return ' ,'.join(clean_list)
    else: return “ “
```

```
new_column=[]
for value in data3['genres’]:
      new_column.append(clean_genre(value))
data3['genres']=new_column
```

Step 6: To save this data we have used the following command.

```
data3.to_csv('/content/drive/MyDrive/Tableau_IMDB_Project/TMDB+IMDB_New_Data.csv')
```

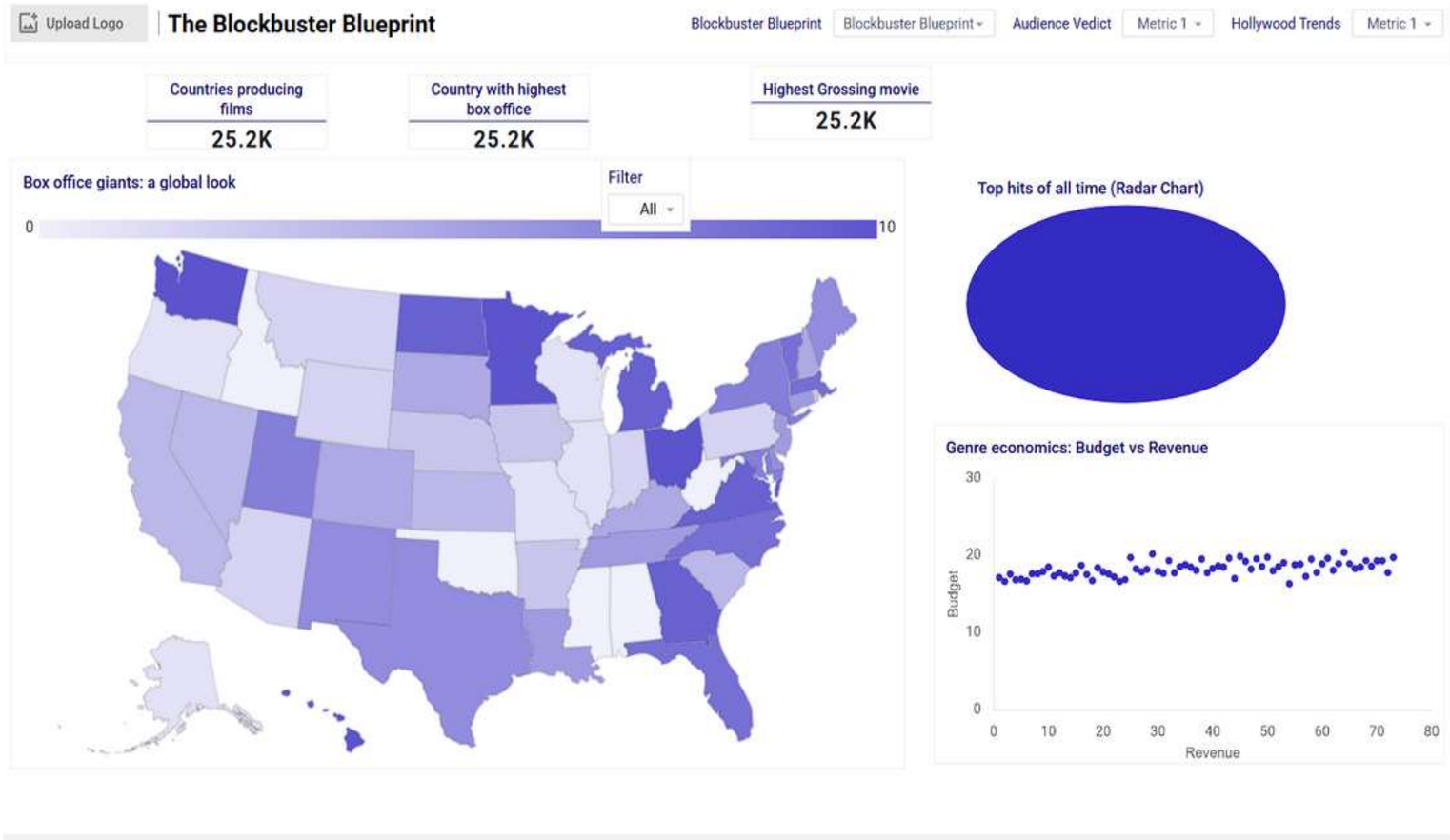
Mockup Design



Mockup for Dashboard 1

Tool: To create this mockup, we have used Figma.

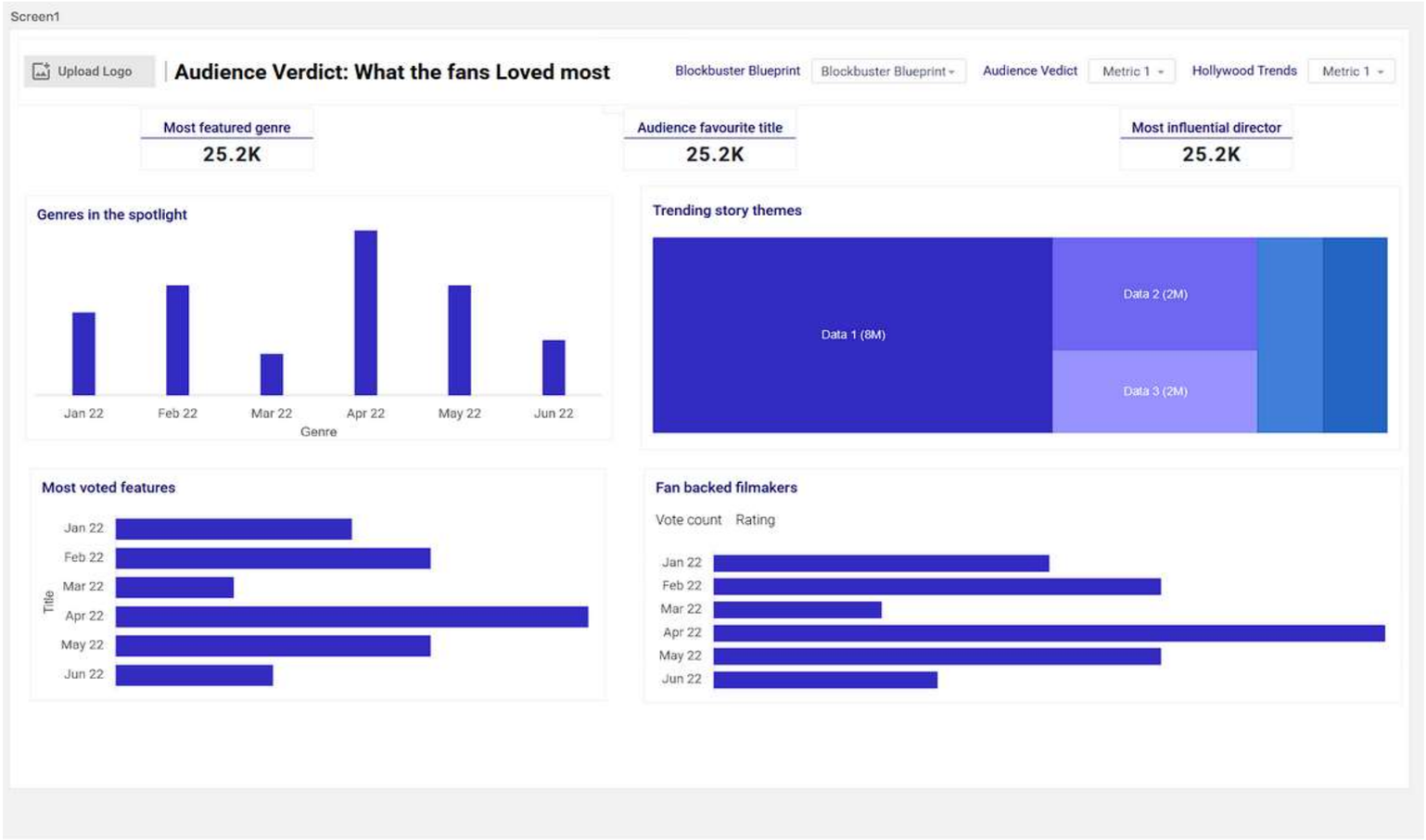
In this dashboard we have used 3 KPI, a global map, a radar chart and a scatter chart. We have created a common filter
with 2 options that can be used to see the trends. The dynamic dimensions have been deployed to surf through the whole tableau dashboards.



Mockup for Dashboard 2

Tool: To create this mockup, we have used Figma.

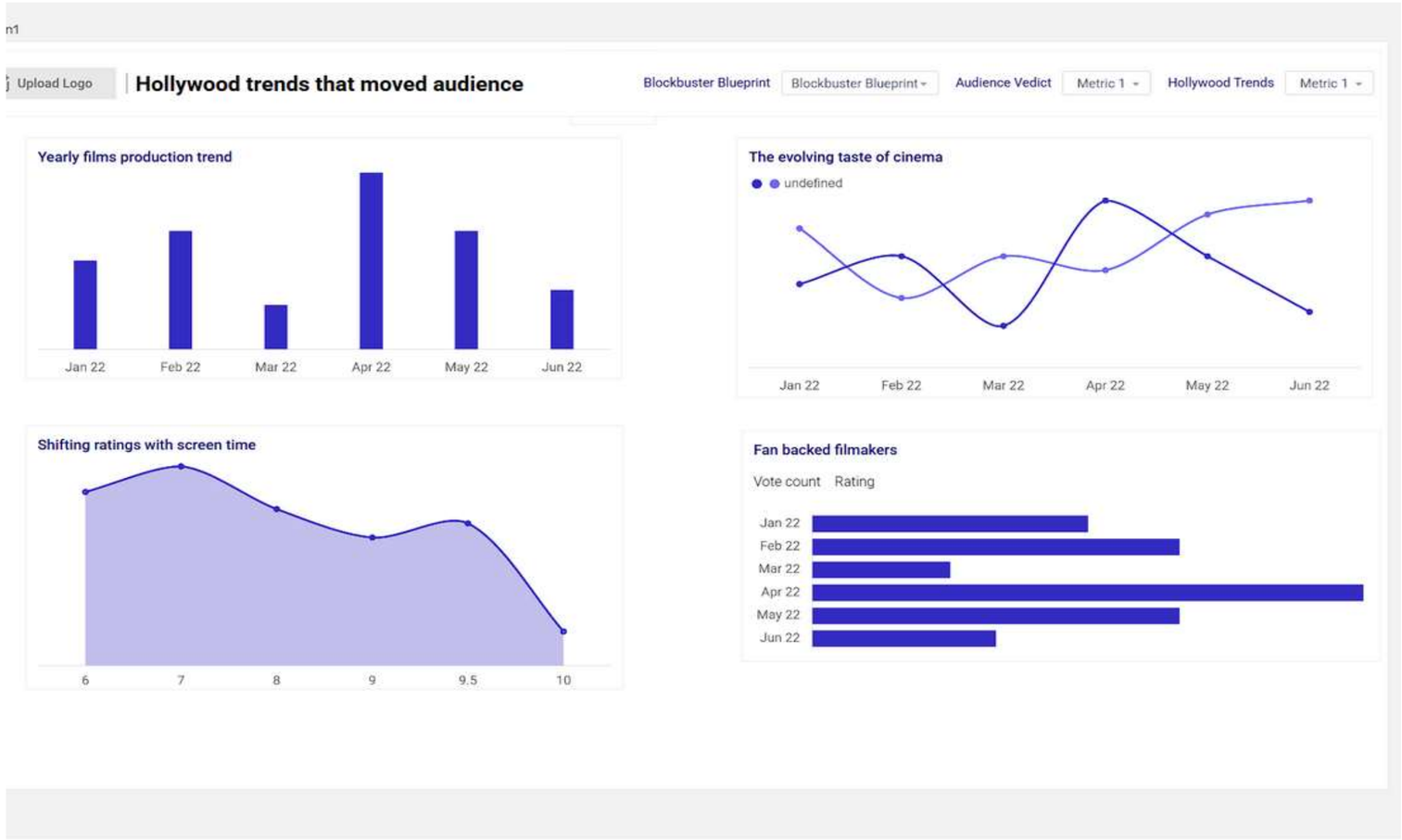
In this dashboard we have used 3 KPI, Vertical bar chart, a tree graph and 2 bar charts. Similar to the first dashboard, we have used 2 filters.



Mockup for Dashboard 3

Tool: To create this mockup, we have used Figma.

In this dashboard we have used a vertical bar chart, an area graph, a line graph and a horizontal bar chart.



Dashboard Design

Dashboard 1

Goal: To compare the international contributions and analyze the financial performance of the film industry.

Layout: The dashboard uses a dark theme with contrasting yellow and white text, aligning with IMDb branding.
The overall layout follows a top section with the logo of IMDB, the title ‘The Economics of Entertainment’, a filter button, three navigation control buttons to switch between three linked dashboards. and three-panel horizontal structure. The design is consistent with standard data storytelling practices—starting global (map), narrowing down to top movies (bar chart), and finishing with genre-specific analysis (scatter plot).

Dashboard 1

Contents:

- **Map- Box Office Giants: A Global Look (Top Left):**
  - Choropleth map with tooltips showing bar graph with revenue and profit per country.
  - Color-coded by revenue metric. Darker color represents higher revenue.

- **Radial Bar Chart- Top Hits of All Time (Top Right):**
  - Displays highest-grossing movies like Joker, Batman vs Superman, etc.
  - Spiral layout showing box office revenue (in billions).
  - Larger the size of spiral higher the revenue.

.

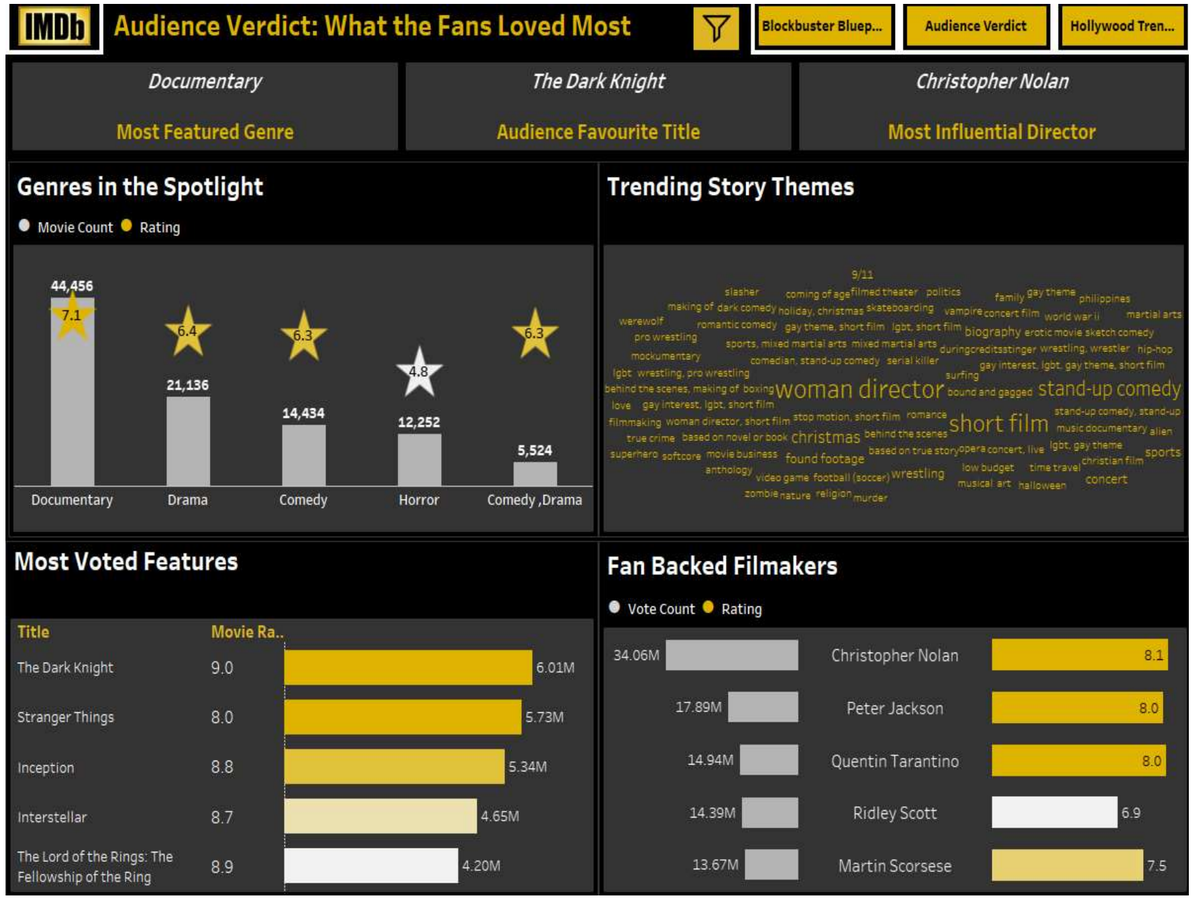
- **Scatter Plot- Genre Economics: Budget vs Revenue (Bottom Right):**
  - X-axis: Revenue
  - Y-axis: Budget
  - Circles represent genres like Drama, Comedy, Horror, etc.
  - Color-coded by profitability. Darker color represents higher profitability.

- **3 Big Number KPIs:**
  - 211 countries producing films
  - Country with highest box office: United States
  - Highest grossing movie: Avatar

Dashboard 2

Goal: To determine the genres, directors, and themes that resonates most with global audience and to track their influence on success of the film.

Layout: The dashboard uses a dark theme with contrasting yellow and white text, aligning with IMDb branding.
The overall layout follows a top section with the logo of IMDB, the title ‘Audience Verdict: What the Fans Loved Most’, a filter button, three navigation control buttons to switch between three linked dashboards and four panel horizontal structure. The design is consistent with standard data storytelling practices- starting with genre in spotlight (bar chart), trending keywords (word cloud), narrowing down to top movies (bar chart), and influence of directors in popularity of film (butterfly chart).



Dashboard 2

Contents:

- **Bar Chart – Genres in the Spotlight (Top Left):**
  - Visualizes Movie Count (bar height) and Average Rating (star overlay) for genre.
  - Rating is represented by shape star and color-coded with yellow. Darker color represents higher rating.

- **Word Cloud – Trending Story Themes (Top Right)**
  - Displays frequent themes based on fan interest and keyword metadata.
  - Font size indicate frequency/prominence.

- **Bar Chart – Most Voted Features (Bottom Left)**
  - Highlights top voted movie names.
  - Two metrics are used: IMDb Rating (shown in yellow bars) and Vote Count (shown as text beside bars).

- **Butterfly Chart – Fan Backed Filmmakers (Bottom Right)**
  - Chart represents top directors by audience support.
  - Left bar indicates vote count and right indicates rating. Right bar is color-coded based on rating. Darker color represents higher revenue.

- **3 Big Number KPIs:**
  - Most Featured Genre: Documentary
  - Audience Favorite Title: The Dark Knight
  - Most Influential Director: Christopher Nolan

Dashboard 3

Goal: To visualize temporal trend and evolution of cinema with changing audience preference.

Layout: The dashboard uses a dark theme with contrasting yellow and white text, aligning with IMDb branding.
The overall layout follows a top section with the logo of IMDB, the title ‘Hollywood Trends that moved Audiences’, a filter button, three navigation control buttons to switch between three linked dashboards and four panel horizontal structure. The design is consistent with standard data storytelling practices—starting with genre in spotlight (bar chart), trending keywords (word cloud), narrowing down to top movies (bar chart), and influence of directors in popularity of film (butterfly chart).

Dashboard 3

Contents:

- **Lollypop Chart- Yearly Film Production Trend (Top left)**
  - Displays number of films produced each year (2000–2023).
  - Values are labeled directly on the bars.

- **Line Chart- The Evolving Taste of Cinema (Top right):**
  - Shows popularity of different genres over the years (2000–2023).
  - Tracks shifts in genre prevalence over time.

- **Area Chart- Shifting Ratings with Screen Time (Bottom left)**
  - Displays relationship between average IMDb rating and screen time ranges.
  - X-axis: Rating scores (6.0 to 9.8)
  - Y-axis: Average screen time

- **Bubble Chart- The Rise of Multilingual Films (Bottom right)**
  - Shows trend towards linguistic diversity in cinema.
  - Helps identify languages that are prominent other than English based on volume and rating.
  - Size of bubble represents volume of movie and yellow color-code represents rating.

Guiding Principles & Best Practices

1. Interactivity: To enable users to interact with charts and delve deeper into insights, filters and highlight actions have been provided.

2. Consistent Color Coding: To enable intuitive interpretation, the same color coding was applied to each genre across all charts.

3. Responsive design: charts can be scaled down without loss of legibility and thus can be utilized across a range of screen sizes.

4. KPI Highlighting: To give instant context, summary cards were positioned at the top.

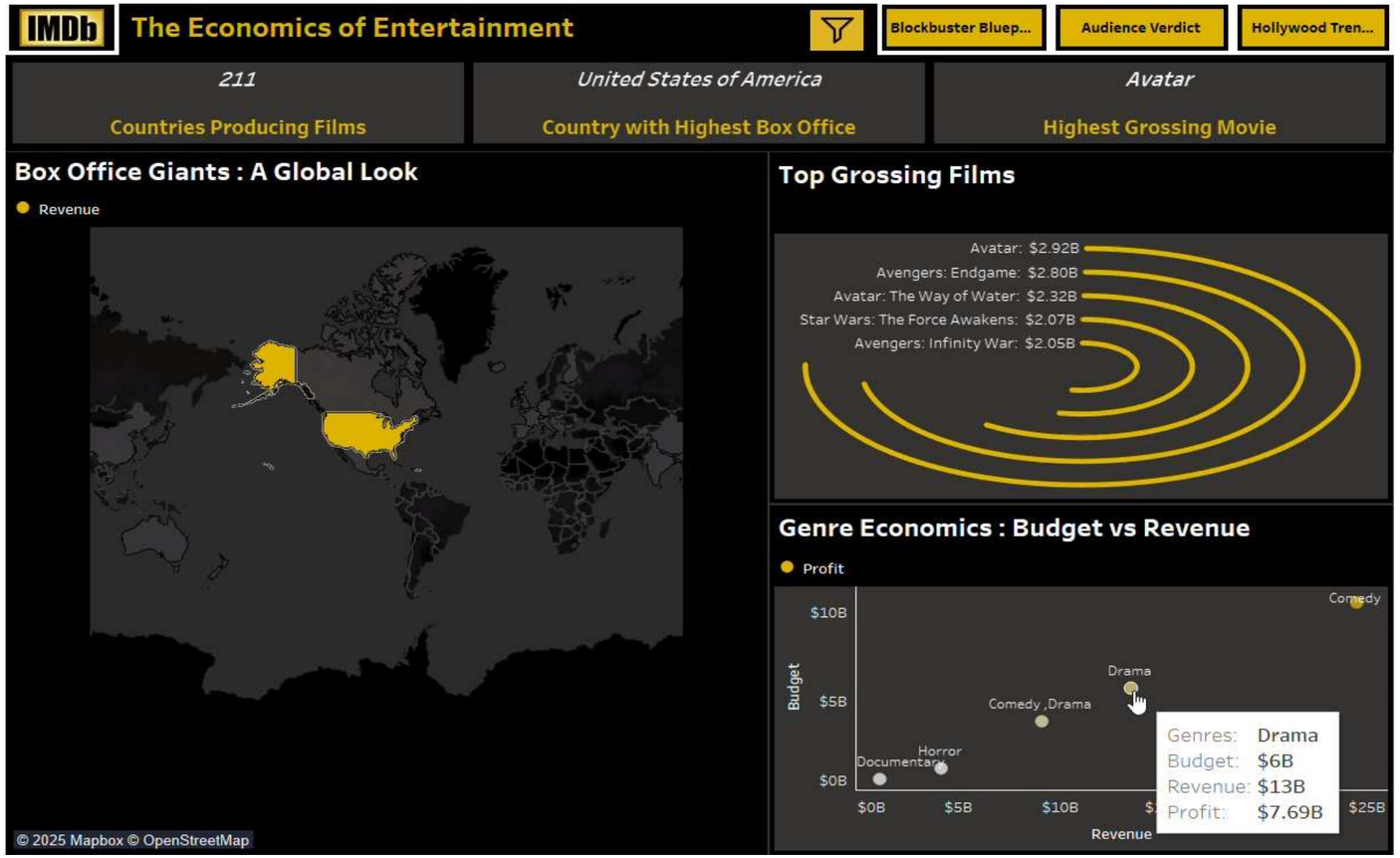
5. Storytelling Flow: Readers can drill into visual analytics after beginning with high-level summaries because of the top-down layout.

Dashboard Application

Video Clip of Dashboard 1 Operation

- **Link to Tableau Public:** https://public.tableau.com/app/profile/harshita.agarwal6987/viz/IMDBMovieRating_17462711124400/Dashboard1?publish=yes

- **Design Implementation:** Tableau has implemented the design exactly similar to what was designed in mockup.



Dashboard 1 Interactions

- **Filter**
  - Filter is at the top represented by yellow funnel button.
  - It opens on click and can be closed using ‘X’ icon.
  - Charts can be filtered using 2 options- ‘Select Top N’ which allows users to view top 5, 10 or 15 data. ‘Production Country’ is the second filter option.

- **Tooltip- For extra-on-demand:**
  - Bar chart is inserted in the tooltip of ‘The Box Office Giants: A Global Look’.
  - When a user hovers over a country, the bar chart displays revenue and profit comparison specific to that country.

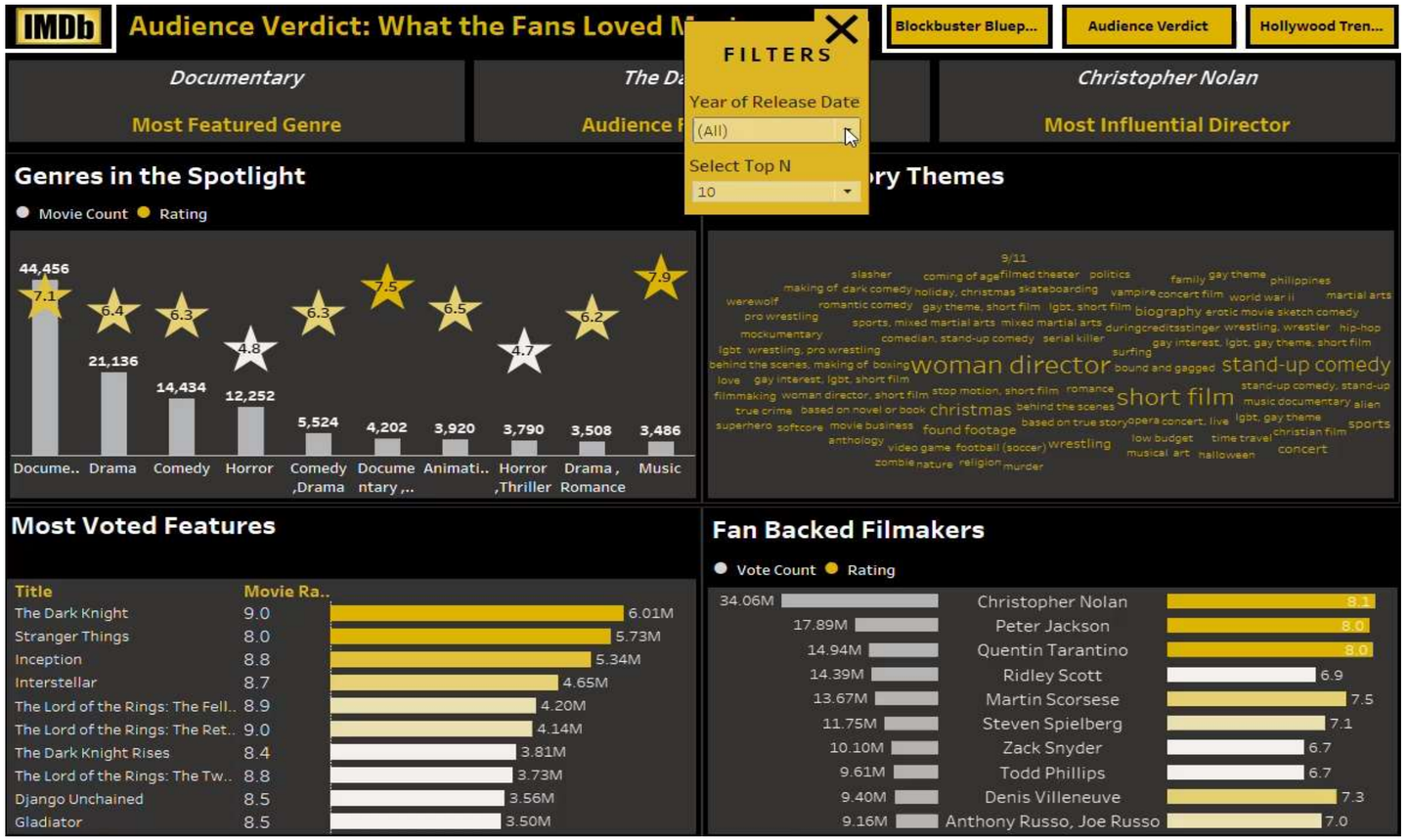
- **Dynamic Interaction- Map driven filtering:**
  - ‘The Box Office Giants: A Global Look’ enables interactive filtering.
  - When a user clicks on a country rest of the dashboard automatically reflects data specific to that country.
  - This enables users to perform targeted analysis without manually applying filters.

- **Hover Tooltip:**
  - Tooltip of map shows name of production country, # of movies and bar chart.
  - Tooltip of scatter plot shows genre name, budget, revenue and profit.

Video Clip of Dashboard 2 Operation

- **Link to Tableau Public:** https://public.tableau.com/app/profile/harshita.agarwal6987/viz/IMDBMovieRating_17462711124400/Dashboard1?publish=yes

- **Design Implementation:** Tableau has implemented the design exactly similar to what was designed in mockup.



Dashboard 2 Interactions

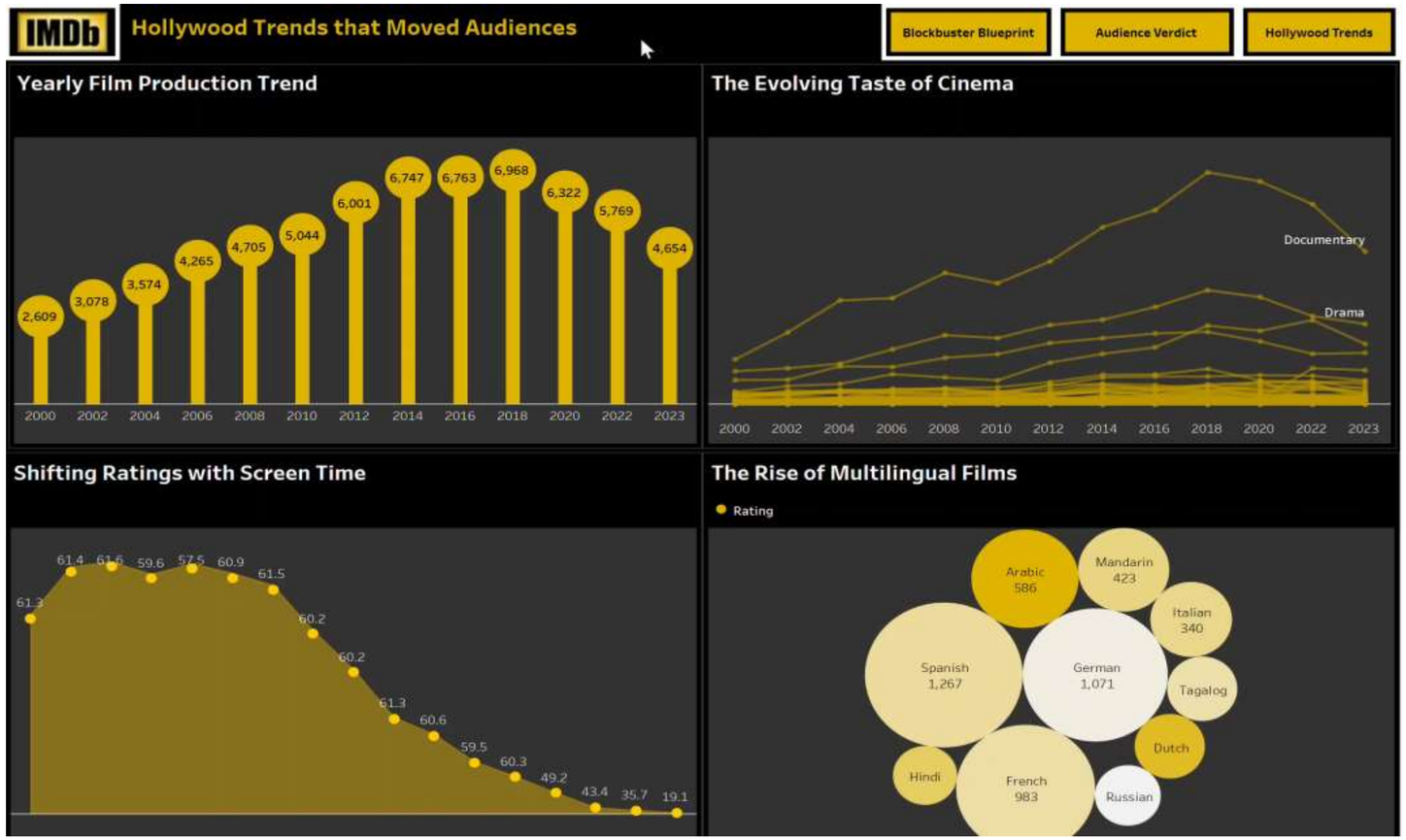
- **Filter**
  - Filter is at the top represented by yellow funnel button.
  - It opens on click and can be closed using ‘X’ icon.
  - Charts can be filtered using 2 options- ‘Select Top N’ which allows users to view top 5, 10, 15 or 20 data. ‘Year of Release Date’ is the second filter option allowing users to filter data based on year.

- **Hover Tooltip:**
  - Tooltip of ‘Genre in spotlight’ shows genre name, # of movies and average rating.
  - Tooltip of word cloud shows theme name, # of movies and average rating.
  - Tooltip of ‘Most voted features’ shows movie name and votes received.
  - Tooltip of ‘Fan backed filmmakers’ shows director name, average rating and votes received.

Video Clip of Dashboard 3 Operation

- **Link to Tableau Public:** https://public.tableau.com/app/profile/harshita.agarwal6987/viz/IMDBMovieRating_17462711124400/Dashboard1?publish=yes

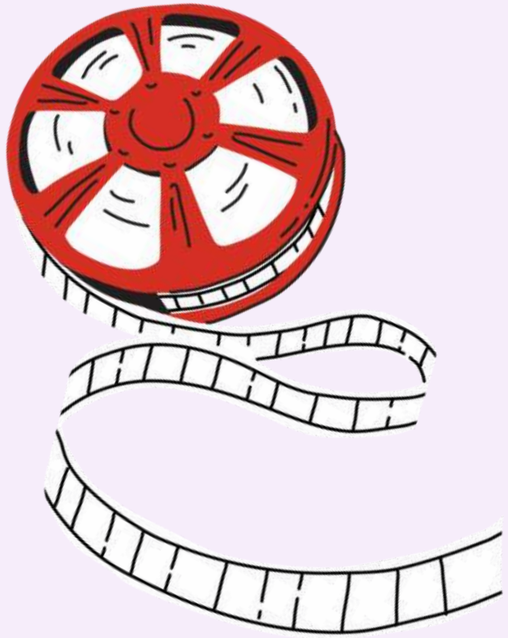
- **Design Implementation:** Tableau has implemented the design exactly similar to what was designed in mockup. However, for 'The rise of multilingual films' we decided to go for bubble chart instead of horizontal bar chart to make the dashboard more appealing.



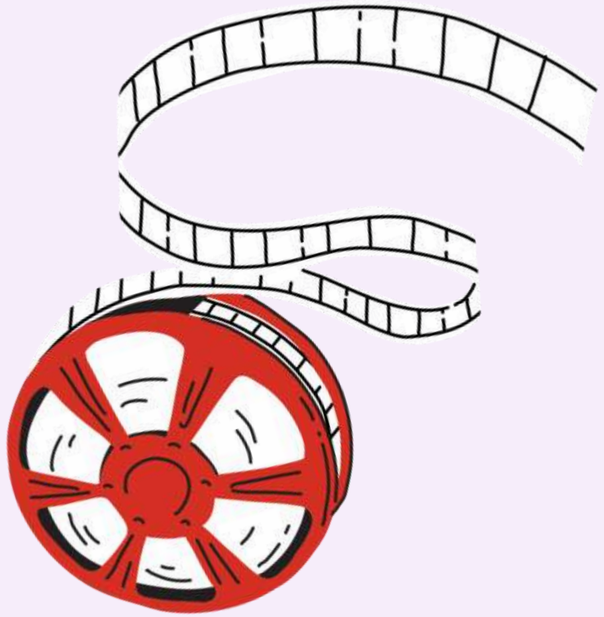
Dashboard 3 Interactions

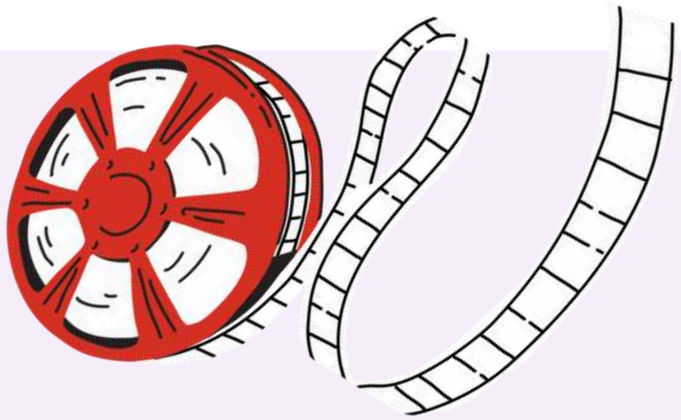
- **Hover Tooltip:**
  - Tooltip of lollipop chart shows year of release date and # of movies.
  - Tooltip of line chart shows genre name and # of movies.
  - Tooltip of area chart shows rating and average runtime.
  - Tooltip of bubble chart shows language name, average rating and # of movies.

- **Navigation:**
  - Top right buttons allow users to switch between linked dashboards to get deeper insights into the film industry.



Conclusion





Conclusion

This project gave us the perfect chance to leverage Tableau to analyze and visualize rich insights from the IMDb movie dataset. We could analyze film production worldwide, genre dynamics, box office performance, and audience preferences in a meaningful visual manner by creating three interactive dashboards: Hollywood Trends, Audience Verdict, and The Blockbuster Blueprint.

Key Learnings

- We discovered how to use design principles and interactive elements to transform unstructured data into meaningful stories.
- To make insights come to life, Tableau's visualization tools—such as word clouds, bar charts, scatter plots, and maps—were used.
- It was enlightening and helped the power of thinking critically and analytically to recognize patterns between such wide-ranging variables as revenue to budget, language changes over the decades, and ratings affected by screen time.

Challenges Faced

- Dealing with data format inconsistencies was a big challenge, particularly while fusing numeric and categorical data for multi-dimensional analysis.
- There was experimentation involved in setting up dual-axis plots and synchronized filters to make sure that dashboard legibility was not compromised through interactivity.
- One of the challenges in design was to display multi-decade trends in a summarized, readable manner without overwhelming the user.
- Separating languages, as movies have been released in multiple languages all over the world was a task.

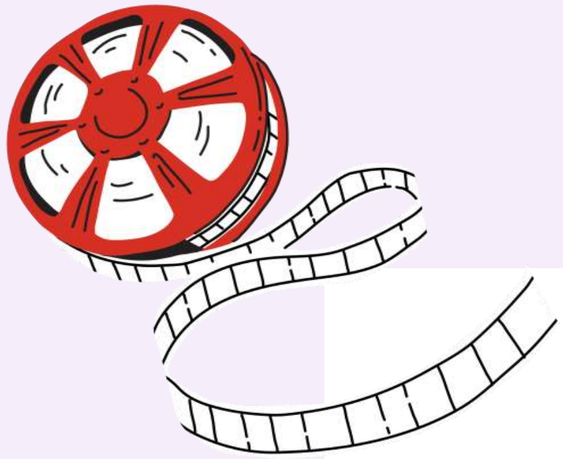
Reflection

- This project highlighted the significance of data storytelling, with design clarity being as crucial as analytical precision.
- We learned hands-on about best practice in dashboard design, including appropriate chart selection, interactivity through tooltips, color coding, and logical layout.
- This project helped in developing skills that will be beneficial for both academic and professional business and analytics roles, particularly the ability to design scalable intuitive dashboards.

Appendix

Learning resources used:

- https://help.tableau.com/current/pro/desktop/en-us/viz_in_tooltip.htm
- https://help.tableau.com/current/pro/desktop/en-us/maps_build2.htm
- https://www.youtube.com/watch?v=OYbPOhK0wPo&t=3s&ab_channel=MavenAnalytics
- https://www.youtube.com/watch?v=0BJnlfzb8Mo



Thank You!