**Project Phase II: Decision Making**

Pavan Kalyan Imadabathini

Sundeep Yalamanchili

Sri Swathi Nimmagadda

Sai Manikanta Maram

IFT 533: Data Visualization & Reporting for IT

Prof: Asmaa Elbadrawy

Nov 19th, 2023

Text

Description automatically generated with medium confidence

**Table of Contents**

[**Section 1** 3](#_Toc151235639)

[**Data Visualization Tools** 3](#_Toc151235640)

[**Section 2** 3](#_Toc151235641)

[**Tableau** 3](#_Toc151235642)

[**Data Preparation & Preprocessing** 4](#_Toc151235643)

[**Section 3** 5](#_Toc151235644)

[**List of the Final Set of Questions** 5](#_Toc151235645)

[**Section 4** 5](#_Toc151235646)

[**List of Plots** 5](#_Toc151235647)

[**Section 5** 11](#_Toc151235648)

[**Dashboard Interactivity** 11](#_Toc151235649)

[**References** 11](#_Toc151235650)

[**Mural Board** 11](#_Toc151235651)

# **Section 1**

## **Data Visualization Tools**

Data visualization tools have undergone a remarkable evolution throughout history, responding to the ever-growing need for effective data analysis and decision-making. Rooted in ancient practices such as maps and charts, the modern era has seen the integration of computers and technology, leading to the development of sophisticated and accessible data visualization techniques. This evolution has been guided by the fundamental role of data visualization in empowering individuals and organizations to glean meaningful insights from complex datasets. By leveraging principles of visual perception, these tools have transformed raw data into comprehensible visual representations, enabling users to identify patterns, trends, and outliers with greater ease than through raw numerical or textual information.

The applications of data visualization span diverse fields, showcasing its versatility and impact. From business intelligence and financial analysis to healthcare analytics and social media insights, organizations leverage these tools to drive informed decision-making. Interactive dashboards, infographics, charts, and graphs serve as powerful instruments in presenting complex data in a user-friendly manner, enhancing understanding, and facilitating more intuitive analysis. Moreover, data visualization finds applications in education, research, and supply chain management, illustrating its widespread influence in enhancing comprehension, communication, and strategic planning. In essence, the history, use cases, and applications of data visualization underscore its pivotal role in transforming raw data into actionable insights across a multitude of domains.

# **Section 2**

## **Tableau**

## In opting for Tableau as the preferred visualization tool, the decision is grounded in its user-friendly interface, expansive data connectivity, and robust visualization capabilities. The intuitive drag-and-drop functionality simplifies the visualization process, accommodating users with diverse technical proficiencies. With compatibility across a wide spectrum of data sources, Tableau ensures flexibility, enabling seamless integration for diverse datasets. The platform's extensive visualization options, coupled with interactive features and drill-down capabilities, empower the creation of dynamic dashboards offering comprehensive insights. Leveraging a supportive user community and rich online resources, Tableau fosters an environment conducive to efficient learning. Its scalability, integration flexibility, and commitment to continuous improvement position Tableau as a reliable choice—from small-scale projects to enterprise solutions—providing a sturdy foundation for diverse data visualization needs.

## **Data Preparation & Preprocessing**

The dataset comprises two main files: main\_data.csv: movie\_metadata.csv:

In the realm of our movie dataset encapsulating details of 45,000 films globally, a pivotal preprocessing phase was meticulously executed to enhance the dataset's suitability for creating interactive dashboards and visualizations. Commencing with an initial dataset of 5044 rows and 28 columns, a discerning examination revealed certain columns to be redundant or extraneous for our specific visualization objectives. Opting for SQL as a proficient preprocessing tool, we judiciously employed it to selectively excise columns that did not contribute substantive insights to our visual representation goals. The curation process involved the targeted removal of specific properties and records, focusing on refining the dataset. Noteworthy exclusions encompassed columns such as actor\_facebook\_likes, director\_facebook\_likes, movie\_facebook\_likes, movie\_imdb\_link, facenumber\_in\_poster, and Num Critic for Reviews. This intricate preprocessing ensures that the dataset now centers on pivotal attributes, paving the way for a more streamlined and purposeful representation in Tableau, aligning seamlessly with our envisioned visualizations and analytical pursuits.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Attribute** | **Description** | **Data Type** |
| 1 | actr\_name\_1 | Actor Names | Nominal |
| 2 | actr\_name\_2 | Actor Names | Nominal |
| 3 | actr\_name\_3 | Actor Names | Nominal |
| 4 | gnres | Genre(s) associated with the movie | Nominal |
| 5 | title\_of\_the\_movie | Title of the movie. | Nominal |
| 6 | dirter\_name | Name of the director of the movie. | Nominal |
| 7 | clr | Indicates whether the movie is in color. | Nominal |
| 8  9 | Dration  gss | Duration (length) of the movie.  Genre code or identifier. | Ratio  Ratio |
| 10 | lang | Language in which the movie is produced. | Nominal |
| 11 | cntry | Country where the movie was produced. | Nominal |
| 12 | Bdgt | Budget allocated for the production of the movie. | Ratio |
| 13 | ttle\_year | Year when the movie was titled or released. | Interval |
| 14 | imdb\_scre | IMDb rating of the movie. | Ratio |
| 15 | apec\_ratio | Aspect ratio of the movie. | Ratio |
| 16 | fb\_likes | Total Facebook likes associated with the movie. | Ratio |

The following will be the final dataset attributes following preprocessing:

# **Section 3**

## **List of the Final Set of Questions**

1. Which film genre is most favored by audiences?
2. What is the highest-rated film in each genre?
3. Which aspect ratio is most appealing to viewers in movies?
4. Is there a relationship between a film's budget and its commercial success?
5. What role does social media play in influencing movie ratings?
6. Which year experienced the highest volume of movie releases?
7. How much has each country invested in movie production over various periods?
8. Who has been recognized as the most acclaimed actor each year?
9. Is there a correlation between a movie's Facebook likes and its IMDb ratings?
10. How does the cast size relate to the budget required for a movie's production?
11. How have the content ratings for movies changed throughout the year
12. Which directors' films have the highest ratings among my preferred genres?

# **Section 4**

## **List of Plots**

1. The most popular movie in the genre is depicted in the image below. The data is displayed using a pie chart. Size and color are the Pre-attentive attributes that are employed in this.

A screenshot of a computer

Description automatically generated

1. Displayed below is a visual representation showcasing the top-rated movie in various genres. In this illustration, distinct colors are assigned to different genres, while the size of each circle corresponds to the number of movies released in that genre. This information is presented through a packed bubble chart, utilizing pre-attentive attributes such as color and size to convey the data effectively. Larger circles indicate a higher movie count within their respective genres.

A screenshot of a computer

Description automatically generated

1. The visual below illustrates the production of movies by various countries. It employs a pie chart to convey this data, where each unique color corresponds to a different nation. The pre-attentive attribute used in this representation is color. By examining the chart, it becomes evident that the United States surpasses all other countries in the number of movie releases

A screenshot of a computer

Description automatically generated

1. The image below illustrates the earnings generated by movies categorized by various content ratings. It displays gross profits on the Y-axis and groups movies by their content rating on the X-axis. Bar charts have been employed to present this data effectively, with pre-attentive attributes of height and position playing a key role in conveying the information

A screenshot of a computer

Description automatically generated

1. The image presented below portrays the IMDb scores attained by movies categorized by various genres. It showcases IMDb scores on the Y-axis and groups movies by their respective genres on the X-axis. This data is effectively represented using bar charts, with the pre-attentive attributes of height and position playing a pivotal role in conveying the information

A screenshot of a computer

Description automatically generated

1. The visual below depicts the distribution of movies categorized by various content ratings. It employs a tree map chart to present this data, utilizing pre-attentive attributes such as color and size. In this representation, distinct colors are assigned to different content ratings, while the size of each rectangle corresponds to the number of movies falling into that category.

A screenshot of a computer

Description automatically generated

1. In the image below, color is used to depict the total budget allocated by each country for their respective movies. The intensity of the color indicates the magnitude of the budget, with darker shades representing higher expenditures. Key pre-attentive attributes employed in this graph include color and position

A screenshot of a computer

Description automatically generated

1. This scatter plot visualization below demonstrates the correlation between IMDb scores and the budgets of individual movies. Pre-attentive attributes used in this plot include position, shape, and height.

A screenshot of a computer

Description automatically generated

1. This chart illustrates the relationship between movie genres and IMDb scores using a bar chart. Pre-attentive attributes utilized in this representation include color intensity and height

A screenshot of a computer

Description automatically generated

1. In the visual, you can observe the quantity of user reviews received by each movie. The Y-axis represents the number of user reviews, while the movie titles are displayed on the X-axis, filtered by the selected genre parameter. This graph utilizes pre-attentive attributes including position, color, and height

A screenshot of a computer

Description automatically generated

1. The image displays a bar chart from a Tableau dashboard, illustrating a comparison between different movie titles and their respective production budgets. The length of the bars represents the budget size, utilizing pre-attentive attributes like length for quick and intuitive data analysis. The chart's intent is to provide a straightforward visual correlation between the financial investment and the individual movies, allowing for an immediate grasp of budgetary disparities within the film industry.

A screenshot of a computer

Description automatically generated

1. This line graph from a Tableau dashboard visualizes the number of movies released each year. The x-axis represents the timeline from 1910 through 2020, while the y-axis quantifies the count of movies produced. The graph uses position along the vertical axis as a pre-attentive attribute, allowing viewers to instantly compare the annual movie outputs. Data points are prominently marked, enabling easy tracking of trends and patterns in the film industry's productivity over the years.

A screen shot of a computer

Description automatically generated

# **Section 5**

## **Dashboard Interactivity**

We've employed various interactive features such as calculated parameters, reference lines, filters, and calculated fields to cater to different scenarios. One of the key features is the calculated parameter, which allows users to select specific genre attributes for data filtering, resulting in clearer and more refined data visualizations.

We've included an 'all' option within the genre parameter to enable the exploration of various genres and access movie-related information. To make this feature function effectively, we've created a calculated field that filters data exclusively for the selected genre. This parameter is also utilized in visualizations tied to genre selection (specifically, Questions 8 and 10).

In addition, we've incorporated internal filters for both the 'Genre' and 'Budget' parameters in our visualizations to display the relationships between IMDb scores and budget, with the budget being represented as a range of values.

The genre parameter's value range is dynamically loaded from the 'Genres' attribute within the dataset, ensuring that all available genres are accessible for selection.

# **References**

## **Mural Board**

* <https://app.mural.co/t/dvproject4830/m/dvproject4830/1699563634432/41164a00ee4d15bf041988db73e8d938b1dbbfb0?sender=u532456dadd8ca56710e31855>
* <https://www.kaggle.com/datasets/bobirino/movie-metadata/data>