



**College of Professional Studies
Northeastern University San Jose**

MPS Analytics

**Course: ALY6070: Communication and Visualization for Data
Analytics**

Assignment:

Assignment 4 — Signature Individual Essay

Submitted on:

March 29, 2023

Submitted to:

Prof: VENKATA DUVVURI

Submitted by:

NIKSHITA RANGANATHAN

ABSTRACT

Tableau and R Shiny are both popular data visualization tools used by data scientists and analysts to create interactive dashboards and visualizations. While they share some similarities, they also have some key differences in terms of functionality, user experience, and programming languages used.

Tableau is a business intelligence tool that helps users to create interactive dashboards and visualizations. It has a drag-and-drop interface, which makes it easy to create different types of charts, graphs, maps, and tables without needing to write any code. Tableau has a wide range of data connectors that enable users to connect to various data sources, including spreadsheets, databases, and cloud services. Tableau has powerful data analysis and exploration features, making it a popular choice for data analysts who want to quickly explore and analyze data without needing to write any code. Tableau has a large and active user community, providing access to numerous resources and tutorials to help users get the most out of the tool.

R Shiny, on the other hand, is a web application framework for the R programming language. It allows users to create interactive web applications and dashboards that can be used for data analysis, modeling, and visualization. R Shiny's flexibility and programmability make it a popular choice for data scientists and analysts who want to customize their dashboards and incorporate more complex visualizations and statistical models. R Shiny also allows users to run R code within the application, providing access to a wide range of statistical and data analysis libraries. R Shiny has a strong user community, with numerous packages and libraries available for use.

One of the key differences between Tableau and R Shiny is the programming language used. Tableau has its own proprietary language for creating visualizations, while R Shiny uses the R programming language. This means that R Shiny users have access to a wider range of statistical and data analysis libraries and can run more complex analyses within the application. Another difference is the user experience. Tableau is designed to be user-friendly, with a drag-and-drop interface that makes it easy to create visualizations without needing to write any code. R Shiny, on the other hand, requires some programming knowledge and expertise to use effectively. This means that R Shiny can be more challenging to use but offers more flexibility and customization options.

In terms of cost, Tableau is a commercial tool that requires a license to use, while R Shiny is an open-source tool that is free to use. This means that Tableau can be more expensive for organizations, while R Shiny can be a more cost-effective option.

INTRODUCTION

Disney is a globally recognized media and entertainment company that has been creating magic through its movies, TV shows, theme parks, and merchandise for over nine decades. Disney has produced some of the most iconic and beloved films and TV shows of all time, including Snow White and the Seven Dwarfs, Mary Poppins, The Little Mermaid, The Lion King, and many more.

In recent years, Disney has expanded its reach by launching its own streaming platform, Disney+. This platform provides subscribers with access to a vast library of Disney movies and TV shows, along with exclusive content produced specifically for the platform. With its expansion into the streaming space, Disney has once again proven its ability to adapt to changing consumer preferences and maintain its position as a leader in the entertainment industry.

About this Dataset: This dataset contains information on 579 movies released between 1937 and 2016, including their titles, release dates, box office revenue, production budgets, genres, ratings, and more. The purpose of the Disney movies dataset is to provide an exploratory analysis of the movies produced by Walt Disney Studios over the years. The dataset aims to answer several questions, including the overall trend of Disney movie production over the years, the most popular genres etc.

In addition to these broad questions, the dataset also allows for more granular analyses, such as examining the performance of Disney movies by decade, the most successful movies in terms of box office revenue, identifying trends in genre popularity over time, etc. The dataset is useful for researchers, analysts, and enthusiasts who are interested in the history and performance of Disney movies and how they have evolved over time. By analyzing the data, we can also gain insights into factors that contribute to the success of Disney movies, such as release dates, and marketing strategies.

Research Questions :

1. What is the distribution of Disney movie releases by season (winter, spring, summer, fall) over the years?
2. How has the total gross revenue of Disney movies changed over the years?
3. How does the total gross of Disney movies vary by the movie rating?
4. Is there any relationship between inflation-adjusted gross and total gross for Disney movies?
5. Which are the top 15 highest-grossing Disney movies of all time?
6. How many Disney movies were released on each day of the week, and which day had the highest number of releases?
7. What are the most popular keywords used in Disney movie titles?
8. What are the top 5 highest-grossing Disney movies for each decade?

ANALYSIS

- **Discuss what you did towards the final group project?**

Team Members : Heejah Roe, Archit Barua, Shyamala Venkatakrisnan and Nikshita Ranganathan

For the final project, we used R shiny as well as Tableau to analyze the Disney movies dataset. While Heejah and Archit focused on creating Tableau dashboards, Shyamala and I were responsible for developing the R Shiny dashboard.

I was responsible for creating an interactive dashboard using the R Shiny framework to analyze the Disney movies dataset. My role involved various tasks such as data cleaning, pre-processing, and understanding the structure of the dataset. I also worked on designing and creating interactive charts, graphs, and other visualizations using ggplot2 and other R libraries.

To start with the project, I first imported the dataset into R Studio and cleaned the data by removing the missing values. I also created new columns for calculating additional metrics that were needed for the dashboard. Once the data was cleaned, I started working on the dashboard design, which involved understanding the Shiny framework and how to use it to create interactive elements such as buttons, sliders, and dropdown menus.

To create the charts and graphs, I used the ggplot2 library, which allowed me to create a variety of visualizations such as scatterplots, bar graphs, and line charts. I also incorporated other R libraries such as dplyr and tidyr to manipulate the data and create dynamic charts and graphs that could be updated in real-time based on user inputs.

During the project, I faced several challenges, such as dealing with data inconsistencies and Shiny code errors. To overcome these challenges, I utilized my problem-solving skills to identify and resolve the issues. I also collaborated with my team members to ensure that we were on the same page and that the project was progressing as expected.

Overall, this project was a great learning experience for me, as it allowed me to further enhance my skills in R programming and data visualization. I learned how to use the Shiny framework to create interactive dashboards, as well as various optimization techniques for improving dashboard performance. Additionally, working in a group environment allowed me to collaborate with others and gain valuable insights into their thought processes and approaches to problem-solving. Throughout the project, our group worked collaboratively and discussed our work together, constantly sharing ideas and giving each other feedback.

- Add screenshots of the graphs you created

Tableau Dashboard:

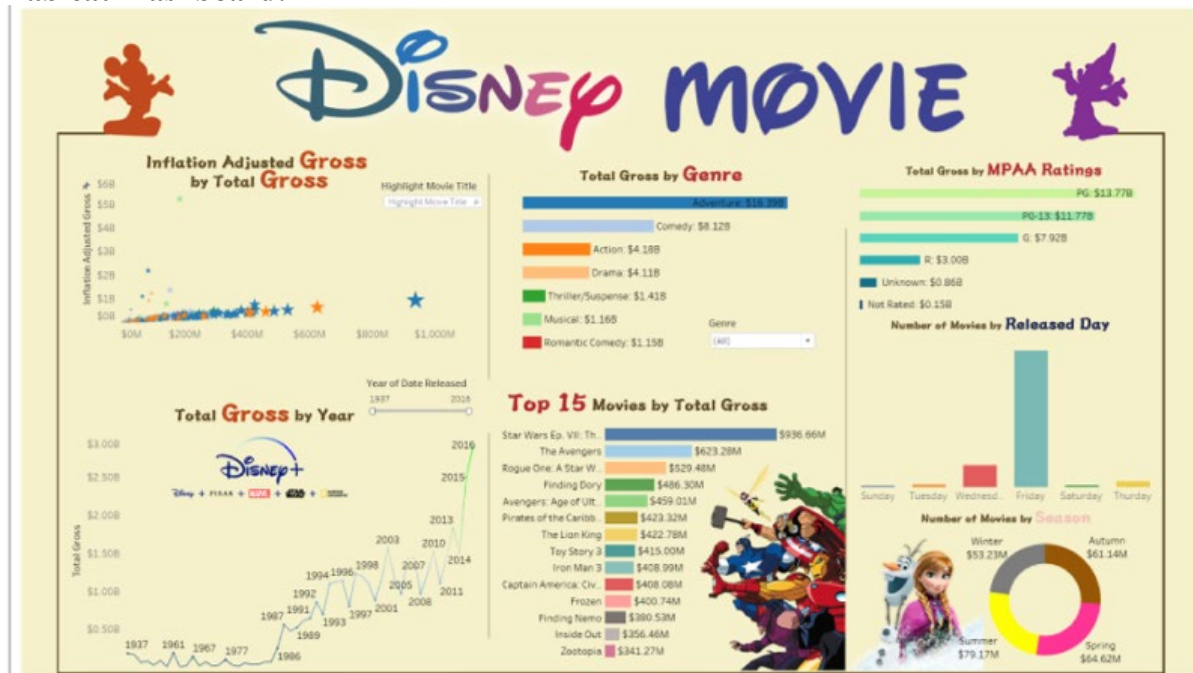


Figure 1 – Tableau dashboard

Some Graphs from R shiny dashboard:

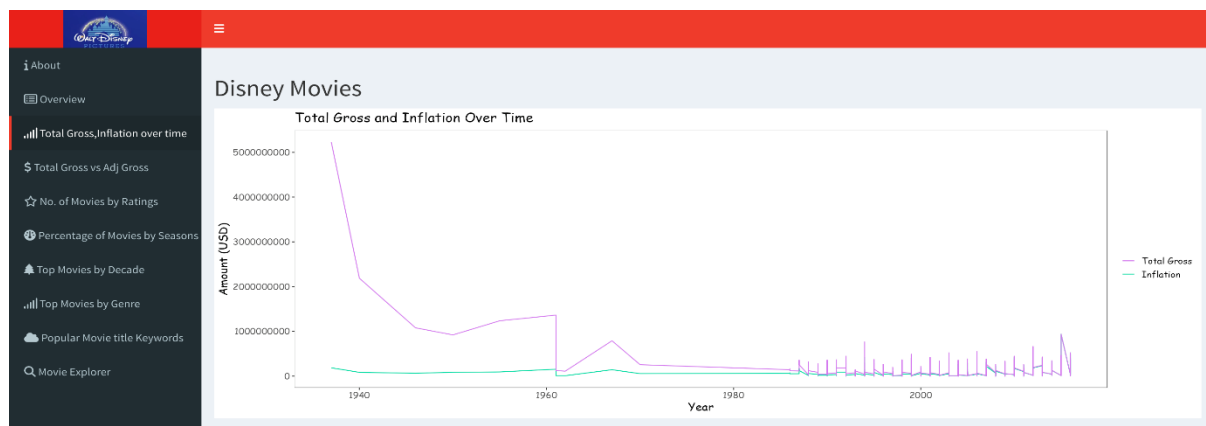
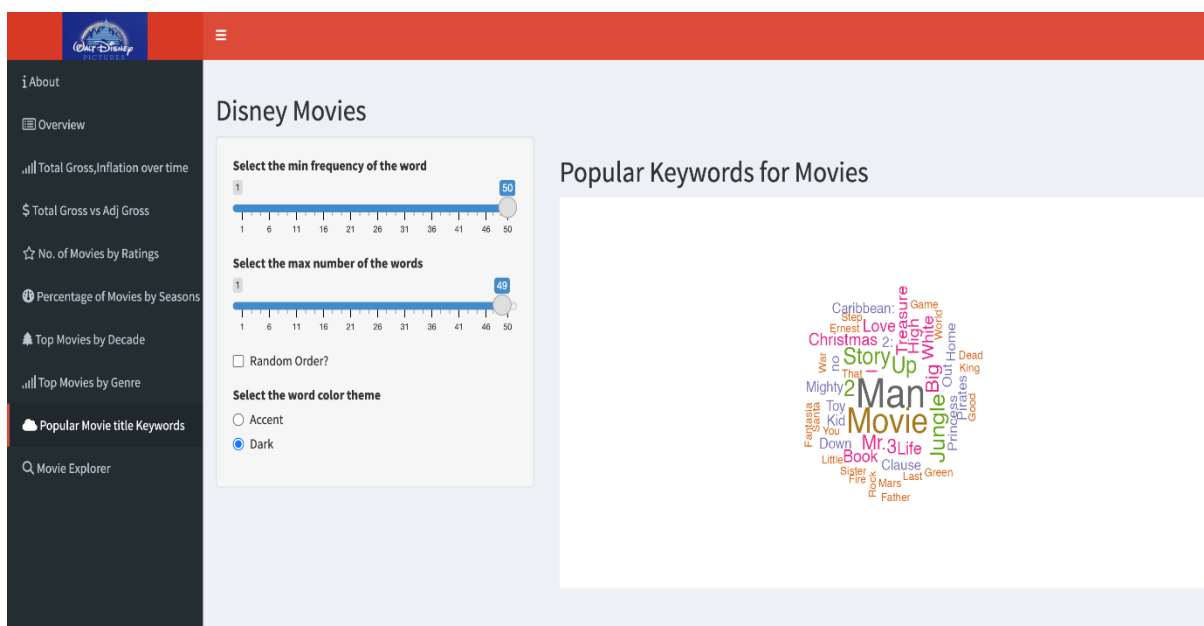
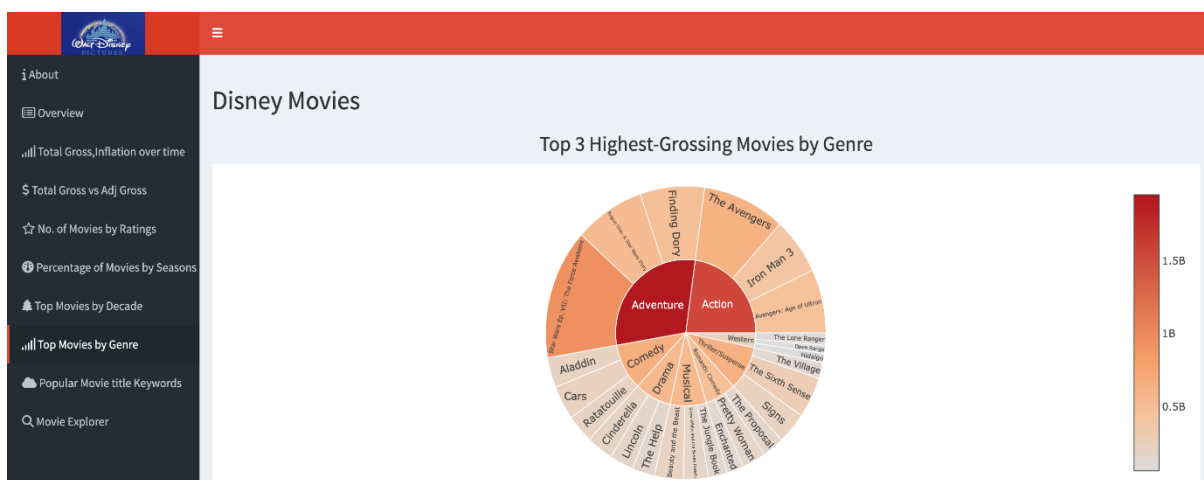
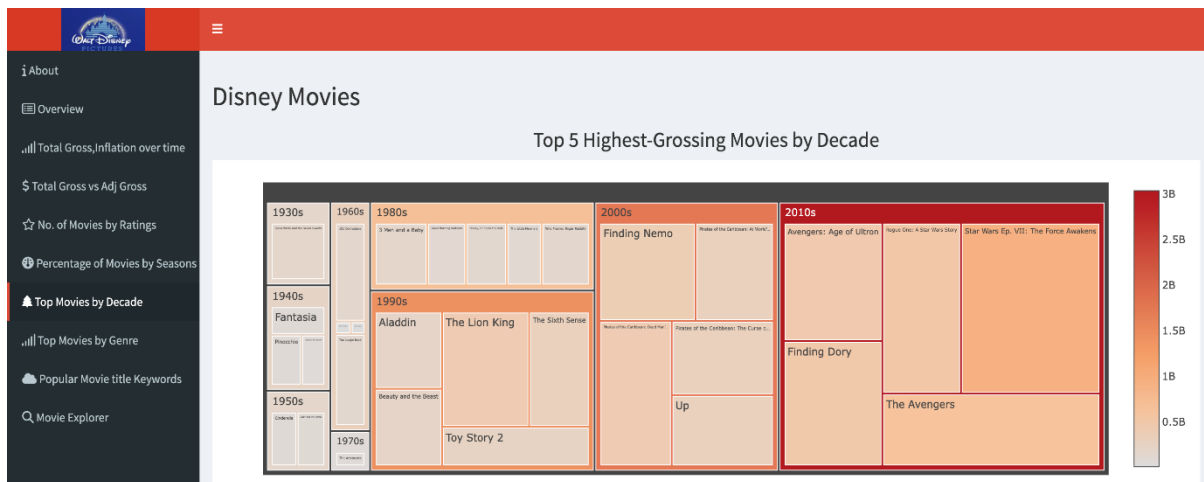


Figure 2 – Total gross and Inflation adjusted gross over the years



- **Explain the analysis you did of the data set as well as visual choices (e.g., design and visual choice)**

For the analysis of the Disney dataset, the dashboards were designed to help users explore and understand trends and patterns in the data. According to me, this can be achieved by properly selecting the right visualization. Choosing the right visualization involves considering factors such as the data type, message, audience, and context.

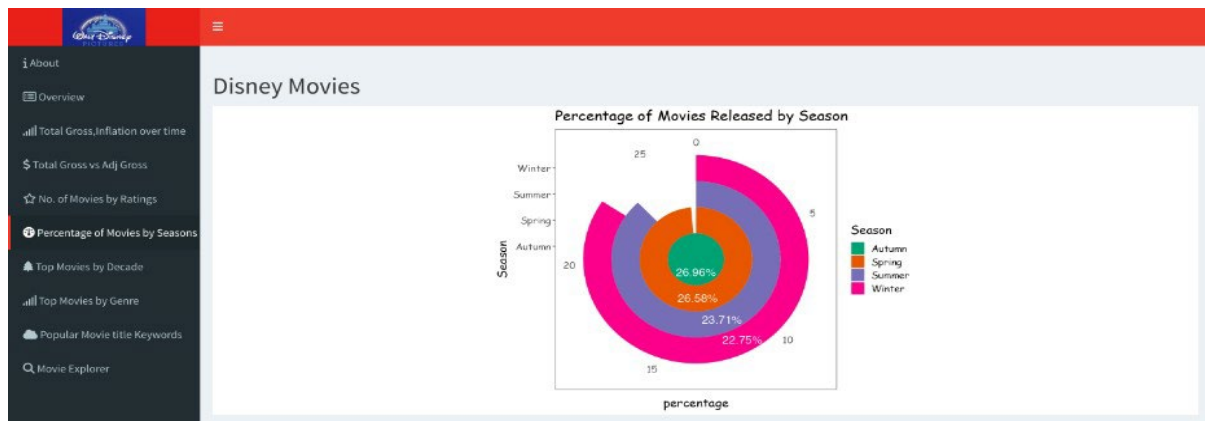


Figure 6 – Percentage of movies in each season

For example - A pie chart was used to understand the distribution of movies in each season. This type of visualization is useful for showing proportions or percentages of a whole, and in this case, it allowed for a clear understanding of the relative frequency of movies released in each season.

In terms of visual choices, we focused on creating a clean and visually appealing dashboard that would be easy to navigate and understand.

One of the key design choices we made was to use a consistent color scheme throughout the dashboard to help users quickly identify different elements. We also used labels and tooltips to provide additional context and information about the data being displayed. Additionally, we used interactive elements such as sliders and dropdown menus to allow users to filter and explore the data in different ways. This also made our dashboards more interactive and user-friendly.

- Present the main findings, patterns and/or statistical results in sentences and graphically

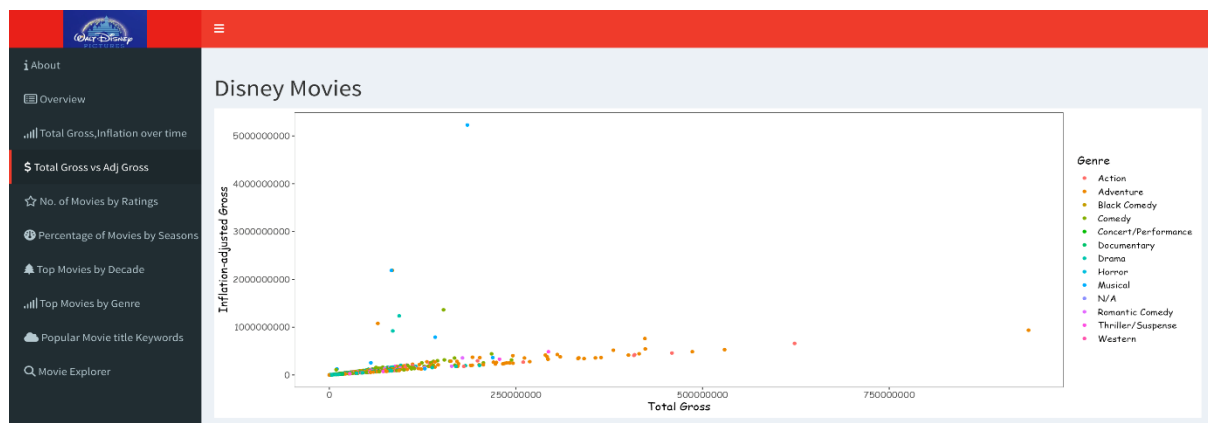


Figure 7 – Total Gross vs Inflation adjusted gross

- From the scatterplot, there appears to be a positive linear relationship between the total gross and inflation-adjusted gross of Disney movies. This indicates that the movies have performed well both in terms of actual earnings and when adjusted for inflation. This also suggests that the success of Disney movies has been consistent over time and continue to resonate with audiences.

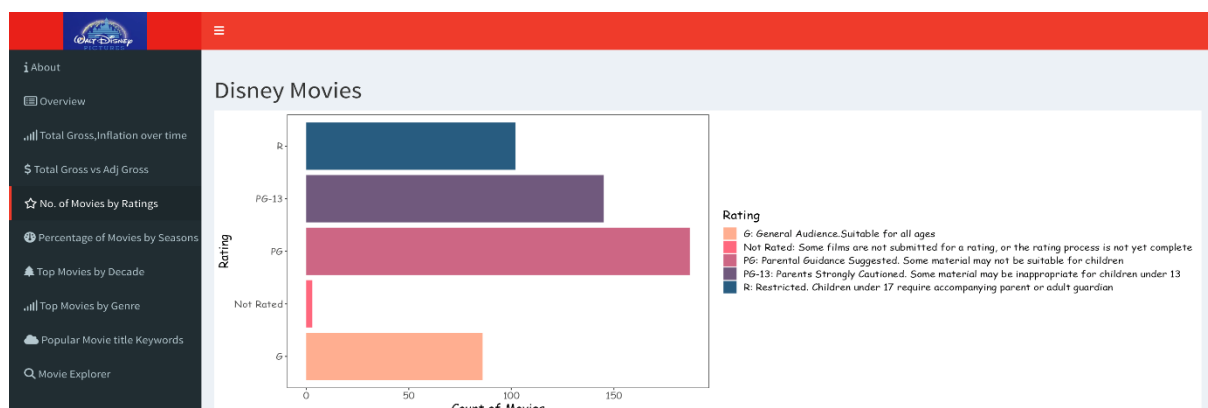


Figure 8 – Distribution of movie ratings

- According to the above bargraph, the majority of Disney movies are rated PG-13 or PG which make up a significant portion of the Disney movie rating distribution. Disney aims to create family-friendly content that can be enjoyed by viewers of all ages, and thus chooses to avoid more mature themes and content that would lead to a higher rating. Additionally, a PG-13 or PG rating may make it easier for Disney to market and distribute their movies, as they are more likely to be shown in theaters and on television.
- From the pie chart in Figure 6, In terms of seasons, we can observe that most Disney movies are released in the autumn and spring, with a comparatively smaller number of releases occurring in the summer and winter.

- The most profitable decade for Disney movies was the 2010s. This can be because of the rise of streaming services like Netflix and Disney+, which have provided new distribution channels for Disney's content. The popularity of 3D and IMAX screenings, as well as the growth of the global box office, may also have contributed to the profitability.
- Adventure and comedy are also popular genres for family-friendly movies, which is Disney's target audience.
- Friday is a popular day for Disney movie releases.

REFERENCES

Resources. (n.d.). Tableau. <https://www.tableau.com/resources>

Shiny. (n.d.). <https://shiny.rstudio.com/>

Disney Character Success. (2022, October 8). Kaggle.

https://www.kaggle.com/datasets/thedevastator/disney-character-success-a-comprehensive-analysis?select=disney_movies_total_gross.csv