COMP7507
Visualization & Visual Analytics
Individual Report
LI Yinghua (3036411281)

Group 14 - Unveiling Movie Magic: Insights into Industry Trends, Audience Preferences, and Box Office Success

1. Overview
In this group project, I was mainly responsible for data collection, visualization analysis, and document writing. Through in-depth exploration of data trends in the film industry, I provided the team with key analytical perspectives and visualization results, and effectively integrated the visualization results of team members to produce an analysis report, which effectively supported the overall completion of the project.

2. Main Tasks
2.1. Data Collection
At the early stage of the project, in order to determine the research direction of the group's topic and the specific content of visualization and analysis, I collected several relevant datasets about movies from open-source data platforms such as Kaggle, which cover a wide range of dimensions from movie budgets and ratings to box-office revenues, and provide a solid data foundation for the subsequent analysis.
2.2. Data Visualization and Analysis
Regarding the specific content of data visualization, I was independently responsible for the part of Trends of the Movie Industry, mainly focusing on the trends of the movie industry's investment and box office returns, referring to the relationship between its budgets, revenues, and returns on investment, etc., combining with various types of visual analysis methods, to provide an overview and connection for the specific analysis of box office by the other members of the subsequent team. The role of this section is to summarize and connect the specific analysis of box office by other members of the group.
At the same time, based on the tableau tool, a variety of interactive and dynamic linking features are used so that users can select aspects of interest on the dashboard in a clickable manner to learn further details and explore the movie industry in more depth.
2.3. Report Writing
2.3.1. Proposal Writing
During the initial project proposal establishment and feasibility analysis phase, proposal writing was completed to envision and write the research objectives, data sources, and expected outcomes. The specific content of which mainly includes:
❖ Research Objective Setting: the project research theme is clearly defined as a market-oriented analysis about the development trend of the film industry, focusing mainly on the correlation between inputs and box office outputs and

audience preferences, and the evaluation indicators are selected to be the industry trend, box office influencing factors, IMDb scores and other key information.

❖ Data Source: After researching the data platforms, Kaggle was chosen as the main data source platform, and 4 main data sets were acquired and screened to summarize their main information.

❖ Expected results: The overall research direction of the project is envisioned based on the research objectives and data sources. Through data cleaning and integration, visualization and analysis, and interpretation of the results, the information on industry trends, rating distribution and box office revenue is visualized and analyzed in the form of charts and text, and interactive analysis is added to improve the readability of users.

2.3.2. Final Report Writing

Based on the visual analysis results completed by the team members, the textual description of the results is transformed and the core conclusions of the project are summarized and outlined.

❖ Textual transformation of visualization results: Based on the visualization forms and interactive operations completed by the team members, the analysis results are further explored and textual descriptions are completed, so that the readers can have a clearer and more concise feeling of the content shown in the visualization charts.

❖ Summary of core conclusions: synthesize the analysis results of each part, and discuss topics such as the overall trend of the movie industry and the determinants of box office.

3. Work Results Display

3.1. Data Visualization Results

3.1.1. Movie Industry Development

The line graph shows the trend of budget and revenue of the movie industry from 1925 to 2020. In the early stages, both budgets and revenues remained low and stable, reflecting the limited investment and returns of the film industry in the mid-20th century. By the 1980s and 1990s, budgets and revenues had gradually increased, reflecting the expansion of film production and the growth of the market. Into the 2000s and 2010s, budgets and revenues show significant growth, reflecting the rise of blockbuster films and technological advances that have resulted in large financial investments and higher box office returns. However, from the 2010s to 2020, despite overall higher budgets and revenues, they show some volatility, which may be related to changes in audience preferences and the rise of streaming platforms.
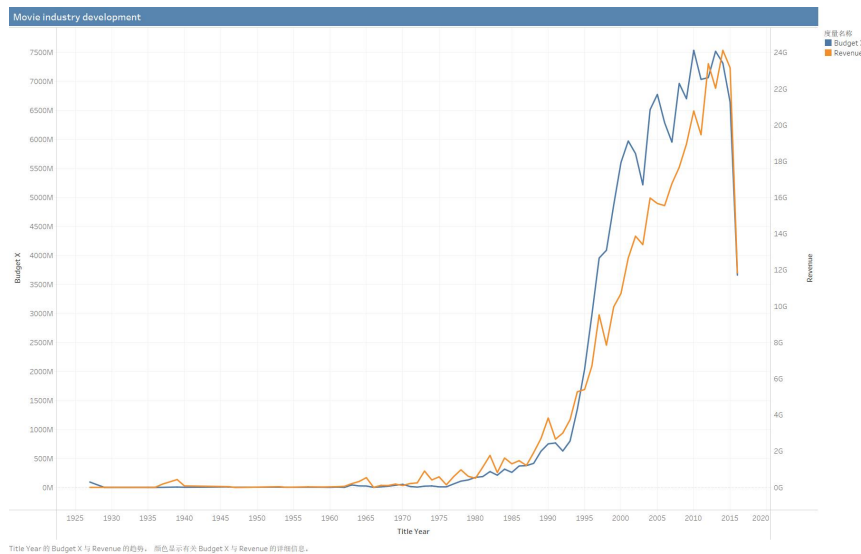
Figure 1: Movie Industry Development

3.1.2. Budget and Revenue

This dendrogram shows the distribution of movie budgets and revenues over time. Each rectangle represents a year, and the size of the rectangle indicates the level of budgets and revenues in that year, with darker colors indicating higher financial inputs and revenues. Larger and darker rectangles are concentrated in the 2000s and 2010s, reflecting the increased investment in the film industry and the significant growth in box office revenues during this period, while smaller and lighter rectangles appear in earlier years, indicating that there was less financial activity in the film industry at that time. This visualization effectively highlights the trend in the financial size of the film industry over time.

The dendrogram is also designed to be interactive, allowing users to click on a specific year to view three detailed charts related to that year. These charts provide an in-depth analysis of the geographic distribution of film budgets and revenues for that year, helping users to understand the distribution of funds and revenue generation patterns in different regions, thus providing important insights into the global dynamics of the film industry.
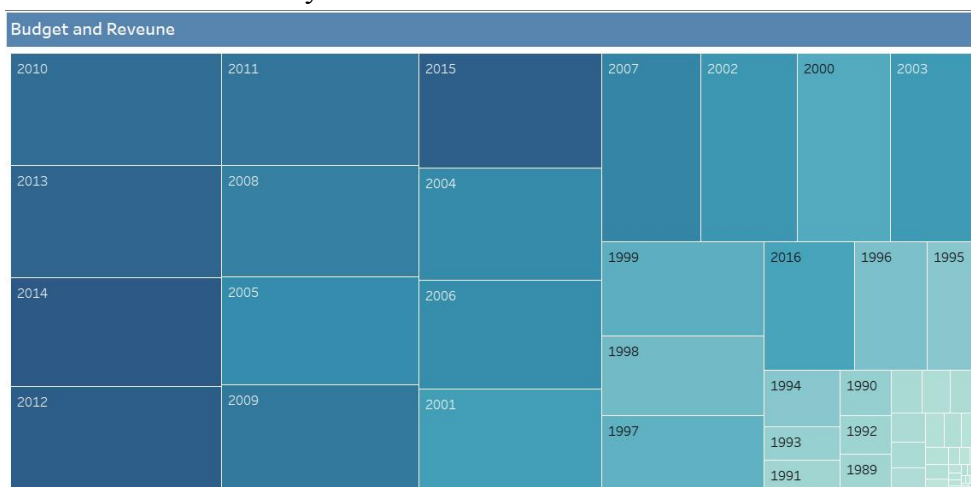

Figure 2: Budget and Revenue

3.1.3. Companies and Revenue

The pie chart shows the distribution of different movie production companies in terms of total revenue. Each sector represents a company's share of revenue, and different colors differentiate the contribution of each company, facilitating a clear comparison of their market influence. As can be seen in the graph, major companies like Walt Disney Pictures and Universal Pictures hold a significant share, highlighting their important role in the industry's revenue generation. This visualization effectively shows the competitive landscape of the film market and the dominance of certain large production companies in the industry.
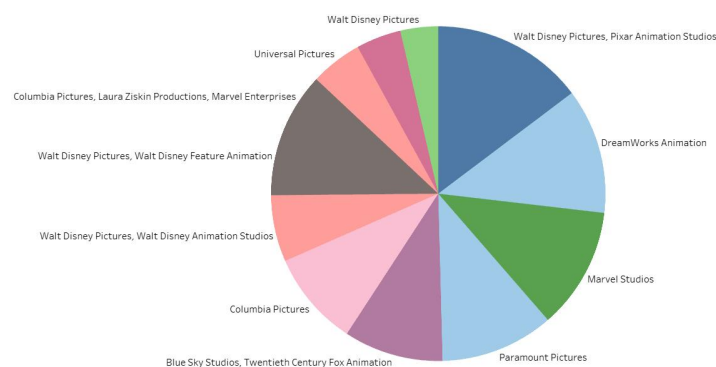

Figure 3: Companies and Revenue

3.1.4. Budget vs. Revenue Scatter Plot

This scatterplot shows the relationship between movie budgets and box office receipts. Each point in the graph represents a movie, the horizontal axis represents budget, and the vertical axis represents box office receipts. The overall trend shows that higher budgets usually result in higher box office receipts, but there is significant variability. Some movies earn significant revenues even with medium budgets, while some high-budget movies do not perform as well. This suggests that while budget is a factor in box office success, it is not the only factor that is decisive.

Figure 4: Budget vs. Revenue Scatter Plot

3.1.5. Geographical Distribution of Budget and Revenue

These two maps show the distribution of movie budgets and box office receipts by country and region, respectively. Darker colors in the maps indicate higher values. The United States is highlighted in the darkest color, reflecting its significant contribution to film budget investment and box office revenues. Other regions vary in their levels of film financial activity, highlighting the globalized nature of film production and revenue generation. These visualizations effectively reveal the geographic differences between regions in the film industry's financial landscape.
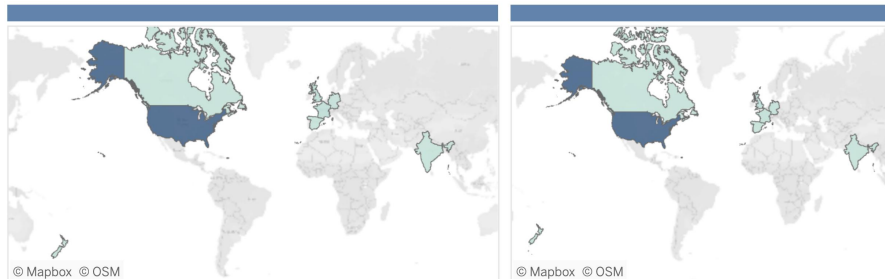

Figure 5: Geographical Distribution of Budget and Revenue

4. Summary

4.1. Project Experience

❖ Project management: by participating in the complete process regarding data collection, cleaning to analysis, visualization and report writing, I enhanced my overall project management and task allocation coordination ability.

❖ Multitasking: Handling data visualization and analysis and text writing at the same time in the project, which enhances the ability to disassemble and plan complex tasks.

4.2. Tool Applications

❖ Skill Enhancement: enhanced my data visualization skills by using tableau tools to implement a variety of complex visualization charts and adding interactive operations to them.

4.3. Data Analysis Skills

❖ Data combing: preliminary acquisition and combing of data, refining key information and selecting key evaluation indicators.

❖ Logical analysis: in the process of data analysis, the correlation between data can construct clear logical relationships and put forward hypotheses on the reasons behind them.

4.4. Industry Understanding

❖ In-depth understanding: a more in-depth understanding of the overall development of the movie industry, especially the advantages and disadvantages of the development trend in terms of finance and market.

4.5. Teamwork

❖ Communication with others: in the whole project process, especially in the report writing process, close cooperation with team members in order to better understand the content expressed in their images and output it accurately in the form of text.