DOUGLASCOLLEGE

Computing Studies and Information Systems

CSIS 4495 002 WINTER 2025

Applied Research Project

Developing a Python-Based Interactive Dashboard for Analyzing Movie Trends and Insights

Submitted to: Prof. Padmapriya Arasanipalai Kandhadai

Submitted by: Lan Dinh

Student number: 300383107

Video Link: video

February 24, 2025

A. INTRODUCTION

Domain Overview and Background

The film industry is a significant cultural and economic force, generating vast amounts of data related to box office performance, audience ratings, critical reviews, and genre trends. With the advent of digital technologies, the volume of movie-related data has grown exponentially, creating opportunities for data-driven decision-making. However, much of this data is unstructured and difficult to interpret without technical expertise. This research aims to bridge this gap by developing a Python-based interactive dashboard that allows users to explore movie trends and insights through dynamic visualizations.

Problem Framing

While vast datasets on movies are readily available, many are presented in raw, unstructured formats, making it challenging to derive meaningful insights without technical expertise. To better understand user needs, interviews were conducted with two distinct types of users: a casual moviegoer (Participant A) and a data analyst (Participant B). Participant A expressed a desire for a dashboard that displays top-rated movies, trending films, and personalized recommendations, catering to their interest in discovering new content easily. On the other hand, Participant B, a data analyst, found existing tools like Power BI and Tableau useful but occasionally overwhelming, particularly when navigating advanced features. These insights highlight the need for a more intuitive and accessible solution for users with varying levels of technical expertise.

Literature Review

Existing research highlights the importance of data visualization in making complex datasets understandable. Stephen Few (2022) emphasizes that data visualization tools, such as dashboards, enable users to derive actionable insights from raw data. Similarly, Genevieve Hayes (2019) identifies Python as a leading programming language for data science, making it an ideal choice for building interactive dashboards. Despite these advancements, there is a lack of user-friendly platforms that integrate multiple movie datasets and provide dynamic visualizations. This research aims to fill this gap by developing a comprehensive, Python-based dashboard.

Initial Hypotheses and Assumptions

The research is grounded in several key assumptions that guide its approach and objectives. Firstly, it assumes that users prefer interactive and intuitive dashboards for exploring movie data, as such interfaces can simplify complex datasets and make them more accessible. Secondly, it posits that Python-based frameworks like Streamlit and Dash are well-suited for building scalable and user-friendly dashboards, given their flexibility and robust ecosystem. Additionally, the research assumes that integrating multiple datasets, such as those from IMDB, Netflix, and Kaggle, will provide a more comprehensive and nuanced view of movie trends, enabling richer insights.

The potential benefits of this research are significant: it aims to enhance accessibility to movie data for non-technical users, empowering them to explore and analyze information without requiring advanced technical skills. For industry professionals, the research could lead to improved decision-making by offering data-driven insights into trends and patterns. Furthermore, the project has the potential to serve as a template for future data-driven applications in other domains, demonstrating how interactive dashboards can be leveraged to make complex data more actionable and insightful across various fields.

B. SUMMARY OF THE INITIALLY PROPOSED RESEARCH PROJECT

The initial research project aims to create an accessible tool that simplifies movie data analysis for a broad audience, including industry professionals, researchers, and movie enthusiasts. Recognizing the complexity of raw, unstructured datasets from sources like Kaggle, the project seeks to transform this data into actionable insights through an intuitive dashboard. Leveraging Python’s data science capabilities—particularly libraries like Pandas, NumPy, and Plotly—the dashboard will feature dynamic visualizations, including bar charts, scatter plots, and heatmaps, allowing users to explore trends, audience preferences, and genre shifts over time.

The research adopts a structured methodology, beginning with data cleaning and storage, followed by exploratory data analysis (EDA) to identify meaningful patterns. Data is sourced from reputable datasets, such as *The Movies Dataset*, *Netflix Movies and TV Shows Dataset*, and *IMDB Movies Dataset*. User needs were also considered through interviews, ensuring the dashboard's design aligns with user expectations for functionality and usability. The project employs Streamlit or Dash for the front end, paired with Flask for back-end operations, to create an interactive, user-friendly interface.

The final product is expected to offer users a streamlined way to analyze movie trends without requiring technical expertise. It will support features like genre filters, year-based sorting, and customizable visualizations. The project also aims to strengthen the developer's technical skills in Python, data visualization, and dashboard development. Ultimately, this research contributes to data-driven decision-making in the film industry and offers a scalable template for similar applications across various sectors.

C. CHANGES TO THE PROPOSAL

Changes in Technology Stack

1. **Integration of Jupyter Notebook**

Justification: Jupyter Notebook was integrated for data cleaning and preparation, as well as data manipulation. Its interactive environment allows for immediate feedback and visualization of results, making it an ideal tool for handling and refining datasets before they are loaded into the dashboard.

2. **Switch from Dash to Streamlit:**

Change: The front-end framework was changed from Dash to Streamlit.

Justification: Streamlit was chosen for its simplicity and ease of use, which accelerated development and allowed for faster prototyping. Its lightweight nature and seamless integration with Python libraries made it a better fit for this project, especially given the need for rapid iteration based on user feedback.

3. **Integration of Plotly for Visualizations:**

Change: Plotly was integrated for creating interactive and dynamic visualizations.

Justification: Plotly offers more advanced visualization capabilities compared to Matplotlib, enhancing the user experience. Justification: Plotly offers more advanced visualization capabilities compared to Matplotlib, such as interactive charts, hover effects, and dynamic updates. This enhances the user experience by making the dashboard more engaging and easier to explore.

Changes in Features

1. **Addition of a Recommendation System:**

Change: A recommendation feature was added to suggest movies based on user preferences and popular genres in selected countries.

Justification: User feedback indicated a strong interest in personalized recommendations. This feature enhances user engagement and provides additional value by helping users discover new content.

2. **Addition of a Welcome page**

Change: A welcome page was added to the dashboard to provide users with an introduction to the tool and its key features.

Justification: The welcome page serves as a user-friendly entry point, offering an overview of the dashboard's capabilities and guiding users on how to navigate the interface.

D. PROJECT PLANNING AD TIMELINE

1. Project Planning

Phase 1: Research Planning and Proposal Development (Week 1 and 2)

Milestones:

- Define the research objectives for the Python-based interactive dashboard (e.g., purpose, target users, and features) (Week 1).
- Identify the tools and frameworks to be used.
- Submit the research proposal and secure supervisor approval (End of January 26, 2025).

Deliverables:

- Approved research proposal.
- Initial requirements document detailing features and functionality.

Phase 2: Design and Prototype Development (Week 3 to 6)

Milestones:

- Create wireframes or mockups for the dashboard layout and features.
- Develop a basic prototype using Python and key libraries.

Deliverables:

- Dashboard wireframes or mockups.
- Prototype with basic functionality (e.g., loading data, basic visualizations).

Phase 3: Implementation and Feature Development (Week 7 and 8)

Milestones:

- Implement core dashboard features, such as data visualization, filtering, and interactivity.
- Add advanced features.

Deliverables:

- Fully functional dashboard with implemented features.
- Progress report highlighting the development process and milestones achieved. (For the midterm progress)

Phase 4: Testing and Optimization (Week 9 and 10)

Milestones:

- Conduct usability testing with client to identify issues and gather feedback.
- Optimize the dashboard for performance, usability, and responsiveness.

Deliverables:

- Usability testing report.

- Optimized and bug-free dashboard ready for deployment.

Phase 5: Documentation and Final Report Writing (Week 12)

Milestones:

- Complete all documentation for the project.
- Finalize the project report and submit it for review.

Deliverables:

- Technical documentation.
- User manual.
- Final project report.

Phase 6: Presentation and Submission (Week 13)

Milestones:

- Deliver a professional presentation of the project.
- Submit all final deliverables.

Deliverables:

- Presentation slides.
- Final deliverables submitted.

The Figure 1 below shows the project management of the research project in Gannt Chart form

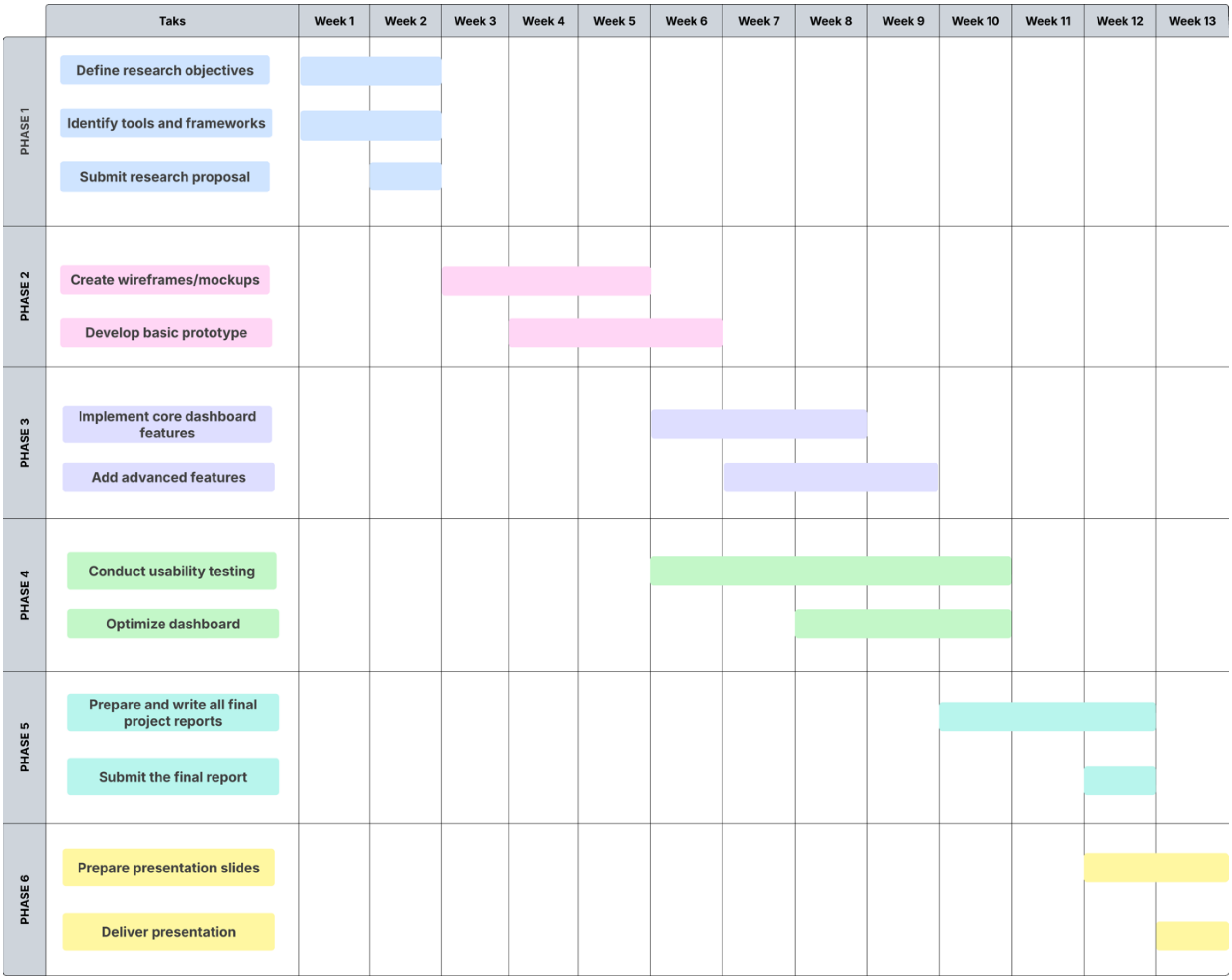


Figure 1- Gannt chart for the research project

E. IMPLEMENTED FEATURES

1. Technical Implementation

Streamlit Framework: The dashboard was built using Streamlit, a Python library designed for creating interactive web applications. Streamlit's simplicity and ease of use allowed for rapid development and prototyping.

Plotly for Visualizations: Plotly was used to create interactive and dynamic visualizations, including bar charts, scatter plots, and choropleth maps. Plotly's advanced features, such as hover effects and color scaling, enhanced the user experience.

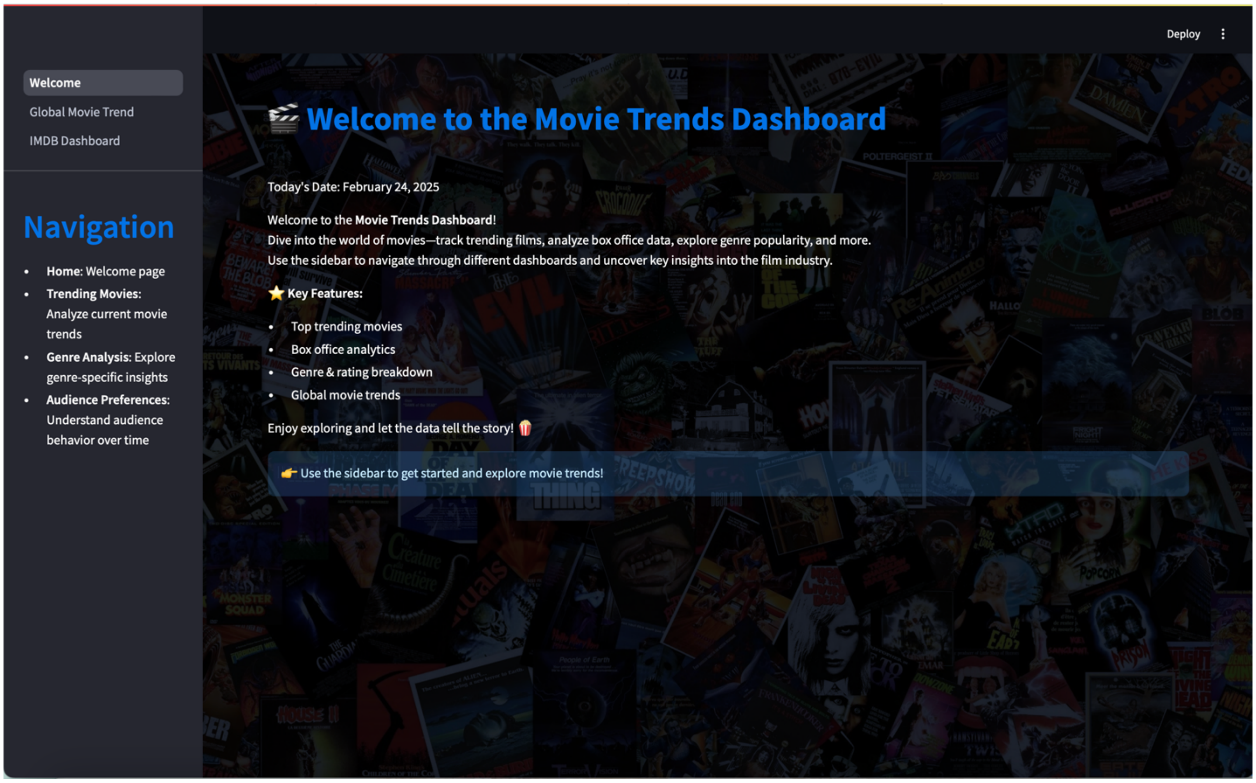
Data Handling: The datasets were cleaned, preprocessed, and prepared for analysis using Jupyter Notebook, an interactive development environment that allows for efficient data manipulation and exploration. With the help of Pandas, a powerful Python library for data manipulation, missing values were handled, and columns were converted to appropriate formats for analysis.

Interactive Filters: Streamlit's built-in widgets, such as sliders, dropdowns, and date pickers, were used to create interactive filters that dynamically update the visualizations based on user input.

2. Welcome Page

Introduction: The dashboard includes a welcome page that provides an overview of the tool and its key features.

Call to Action: Users are encouraged to explore the dashboard using the sidebar navigation.



Navigation

- **Home:** Welcome page
- **Trending Movies:** Analyze current movie trends
- **Genre Analysis:** Explore genre-specific insights
- **Audience Preferences:** Understand audience behavior over time

🎬 Welcome to the Movie Trends Dashboard

Today's Date: February 24, 2025

Welcome to the **Movie Trends Dashboard**!
Dive into the world of movies—track trending films, analyze box office data, explore genre popularity, and more.
Use the sidebar to navigate through different dashboards and uncover key insights into the film industry.

⭐ **Key Features:**

- Top trending movies
- Box office analytics
- Genre & rating breakdown
- Global movie trends

Enjoy exploring and let the data tell the story! 🍿

👉 Use the sidebar to get started and explore movie trends!

Figure 2 – Welcome Page

3. Global Movies Trend Dashboard

Data Loading and Preprocessing

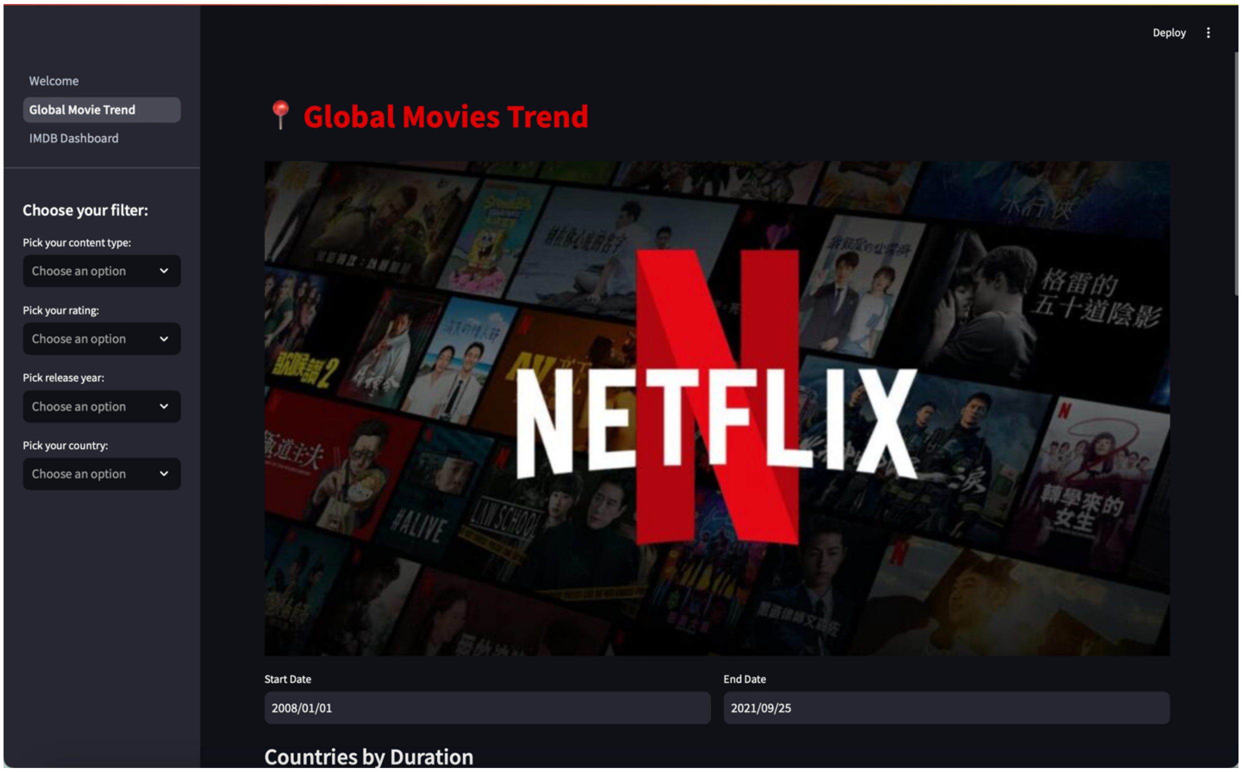
Data Source: The dashboard uses the **Netflix Movies and TV Shows Dataset**, which includes information about titles, release years, ratings, durations, and countries of origin.

Data Cleaning:

- Missing values in the date_added and duration columns were dropped.
- Missing values in the rating, duration, country, cast, and director columns were filled with default values (e.g., "Unknown").
- The date_added column was converted to a datetime format for easier filtering.
- The duration column was extracted and converted to a numeric format.

Interactive Filters

- Date Range Picker: Users can select a start and end date to filter movies and TV shows added to Netflix within a specific time range.
- Content Type Filter: Users can filter content by type (e.g., movies or TV shows).
- Rating Filter: Users can filter content by rating (e.g., PG, R, etc.).
- Release Year Filter: Users can filter content by release year.
- Country Filter: Users can filter content by country of origin. The country column was split into individual countries for multi-country productions.



📍 Global Movies Trend

Choose your filter:

Pick your content type: Choose an option

Pick your rating: Choose an option

Pick release year: Choose an option

Pick your country: Choose an option

Start Date 2008/01/01

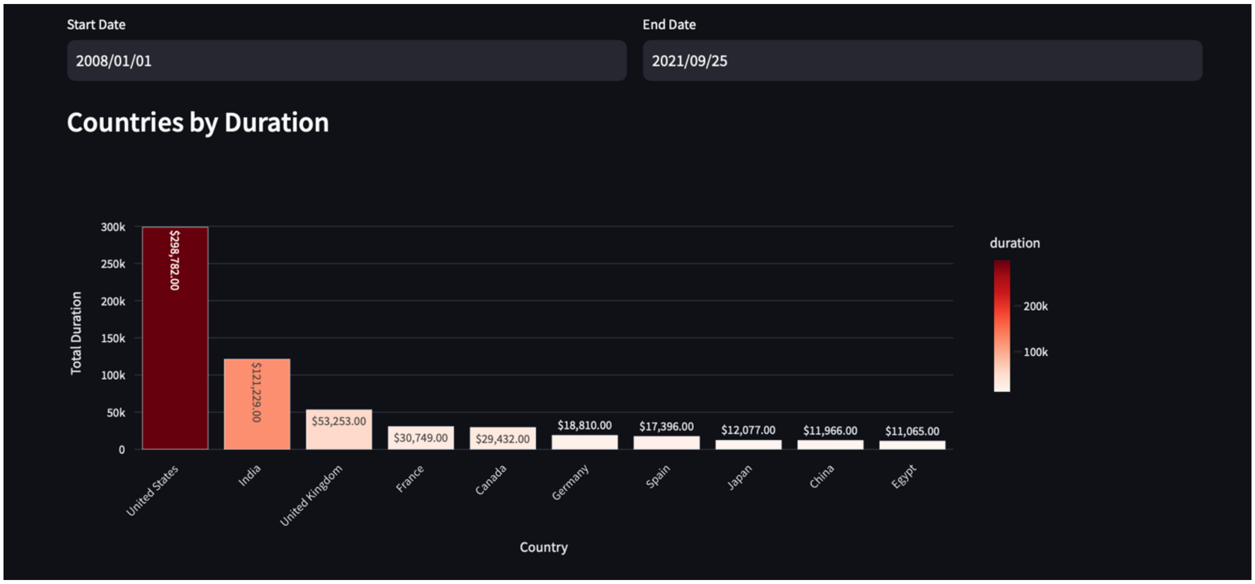
End Date 2021/09/25

Countries by Duration

Figure 3 – Interactive Filters of Global Movies Trend

Visualizations

- **Top 10 Countries by Total Duration**:
  - A bar chart was created to display the top 10 countries based on the total duration of content available on Netflix.
  - The chart is interactive, with hover effects showing the exact duration for each country.
  - The bars are color-coded based on duration, with a gradient scale for better visualization.



Start Date

2008/01/01

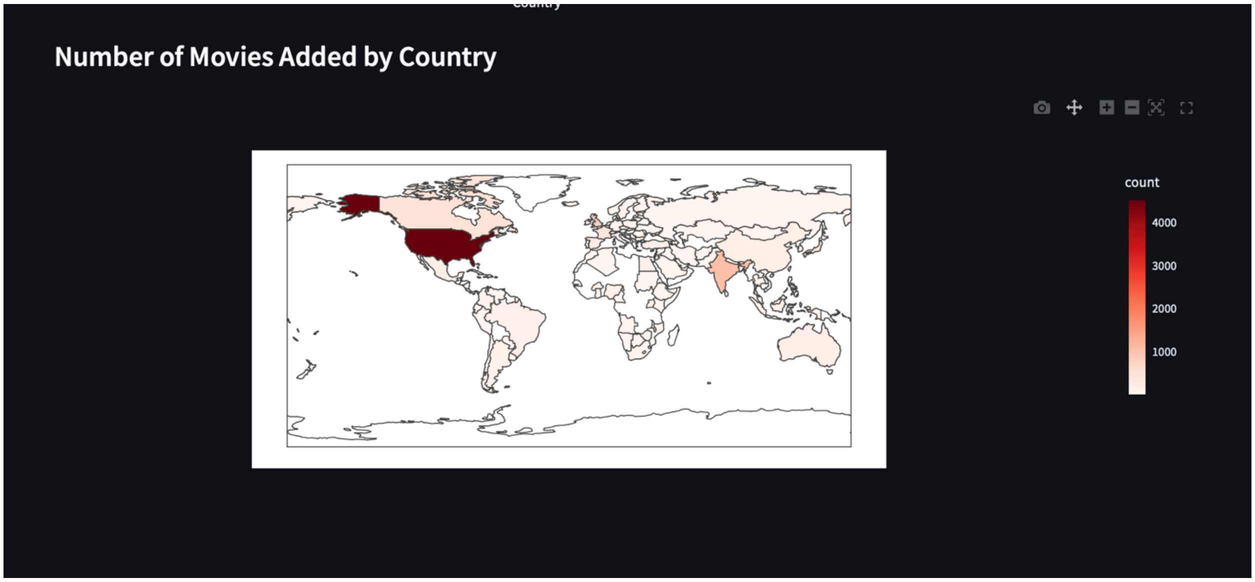
End Date

2021/09/25

Countries by Duration

Figure 4 – Bar chart for top 10 countries by Duration

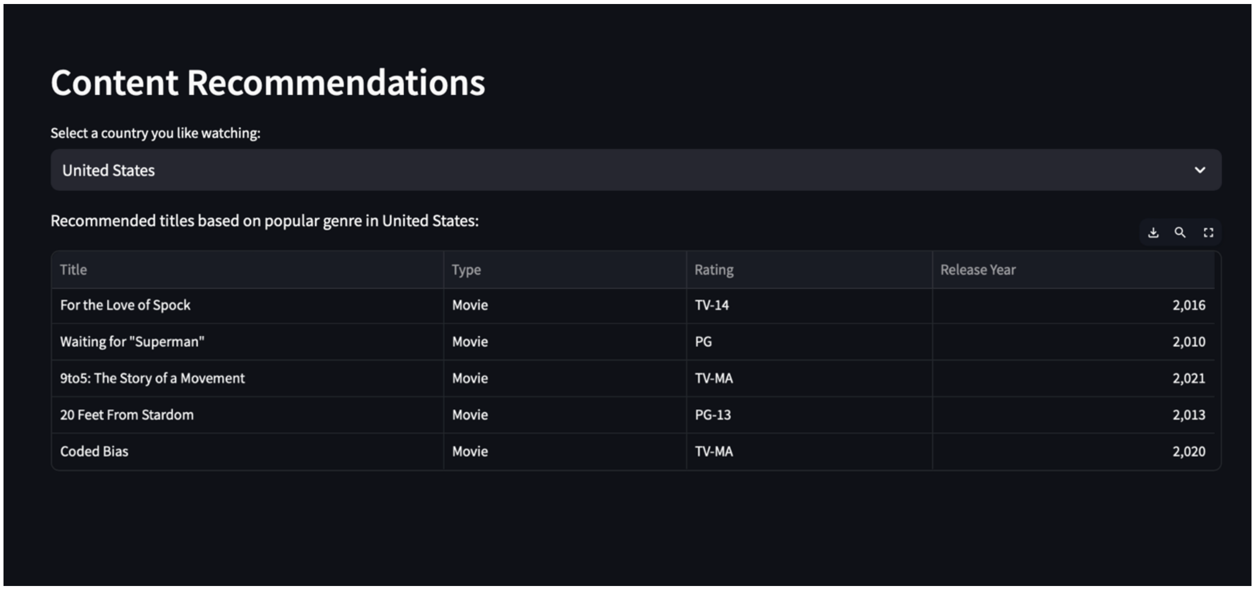
- **Filled Map: Number of Movies Added by Country**:
  - A choropleth map was created using Plotly to display the number of movies and TV shows added to Netflix by country.
  - The map is interactive, with hover effects showing the exact count of titles for each country.
  - The color intensity represents the number of titles, with darker shades indicating higher counts.



Number of Movies Added by Country

Figure 5 – Filled map for Number of Movies Added by Country

- **Content Recommendations**:
  - A recommendation system was implemented to suggest movies and TV shows based on the most popular genre in a selected country.
  - Users can select a country from a dropdown menu, and the system will recommend titles based on the most common genre in that country.
  - The recommendations are displayed in a well-formatted table with columns for title, type, rating, and release year.



Content Recommendations

Select a country you like watching:

United States

Recommended titles based on popular genre in United States:

Title	Type	Rating	Release Year
For the Love of Spock	Movie	TV-14	2,016
Waiting for "Superman"	Movie	PG	2,010
9to5: The Story of a Movement	Movie	TV-MA	2,021
20 Feet From Stardom	Movie	PG-13	2,013
Coded Bias	Movie	TV-MA	2,020

Figure 6 – Content Recommendations for a selected country

4. IMDB Movies Trends Dashboard

Data Loading and Preprocessing

Data Source: The dashboard uses the IMDB Top 1000 Movies Dataset, which contains information about movie titles, directors, actors, ratings, gross earnings, and more.

Data Cleaning:

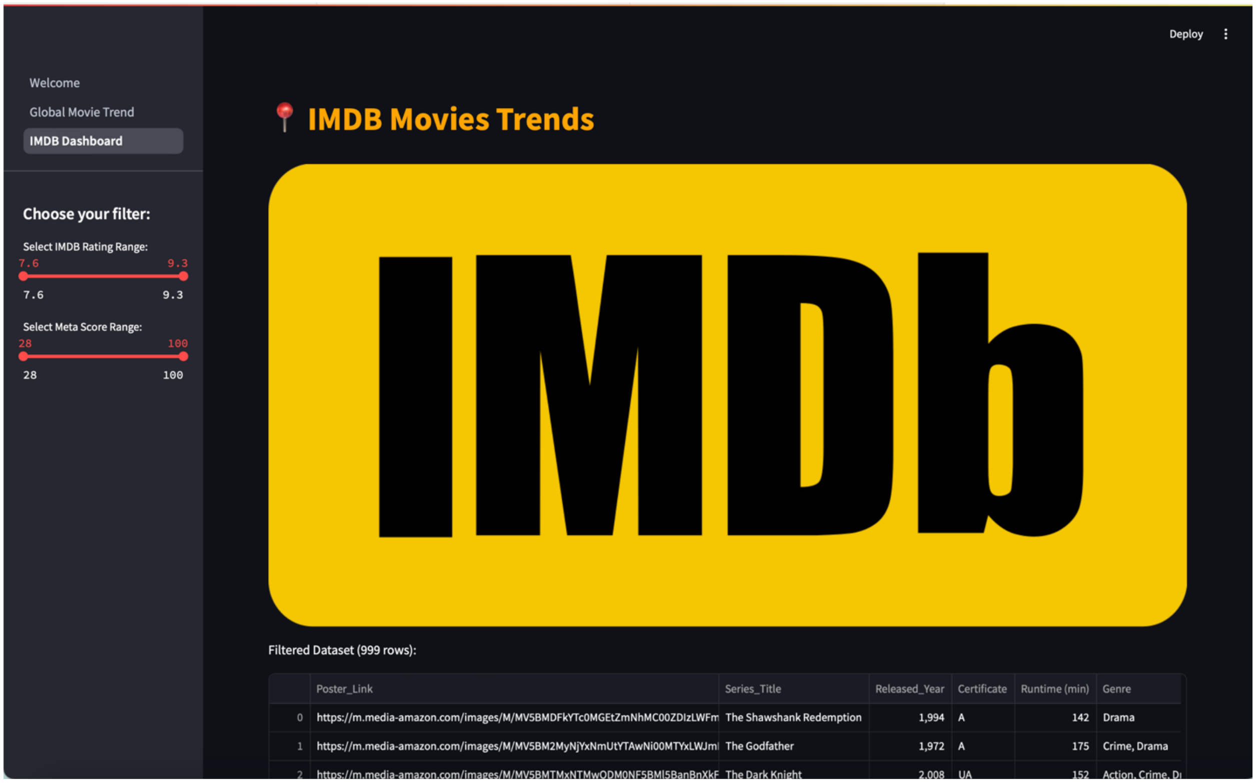
- Missing values in the Certificate column were filled with "Unrated."
- Missing Meta_score values were replaced with the mode of the column.

- The Gross column was cleaned by removing commas and converting it to a numeric format (in millions).
- The Released_Year and Runtime columns were converted to numeric formats for analysis.

Genre Splitting: The Genre column, which contains multiple genres per movie, was split into individual genres for more granular analysis.

Interactive Filters

- IMDB Rating Slider: Users can filter movies based on a range of IMDB ratings using a slider. The slider allows users to select a minimum and maximum rating, dynamically updating the dataset displayed on the dashboard.
- Meta Score Slider: A similar slider was implemented for the Meta Score, allowing users to filter movies based on critic ratings.



Choose your filter:

Select IMDB Rating Range: 7.6 – 9.3

Select Meta Score Range: 28 – 100

📍 IMDB Movies Trends

Filtered Dataset (999 rows):

	Poster_Link	Series_Title	Released_Year	Certificate	Runtime (min)	Genre
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc0MGEtZmNhMC00ZDIzLWFm	The Shawshank Redemption	1,994	A	142	Drama
1	https://m.media-amazon.com/images/M/MV5BM2MyNjYxNmUtYTAwNi00MTYxLWJmI	The Godfather	1,972	A	175	Crime, Drama
2	https://m.media-amazon.com/images/M/MV5BMTMxNTMwODM0NF5BMl5BanBnXkF	The Dark Knight	2,008	UA	152	Action, Crime, Dı

Figure 7 – Interactive Filters of IMDB Movies Trends

Visualizations

- **Top 10 Directors by Gross Earnings**:
  - A bar chart was created using Plotly to display the top 10 directors based on the total gross earnings of their movies.
  - The chart is interactive, with hover effects showing the exact gross earnings for each director.
  - The bars are color-coded based on gross earnings, with a gradient scale for better visualization.

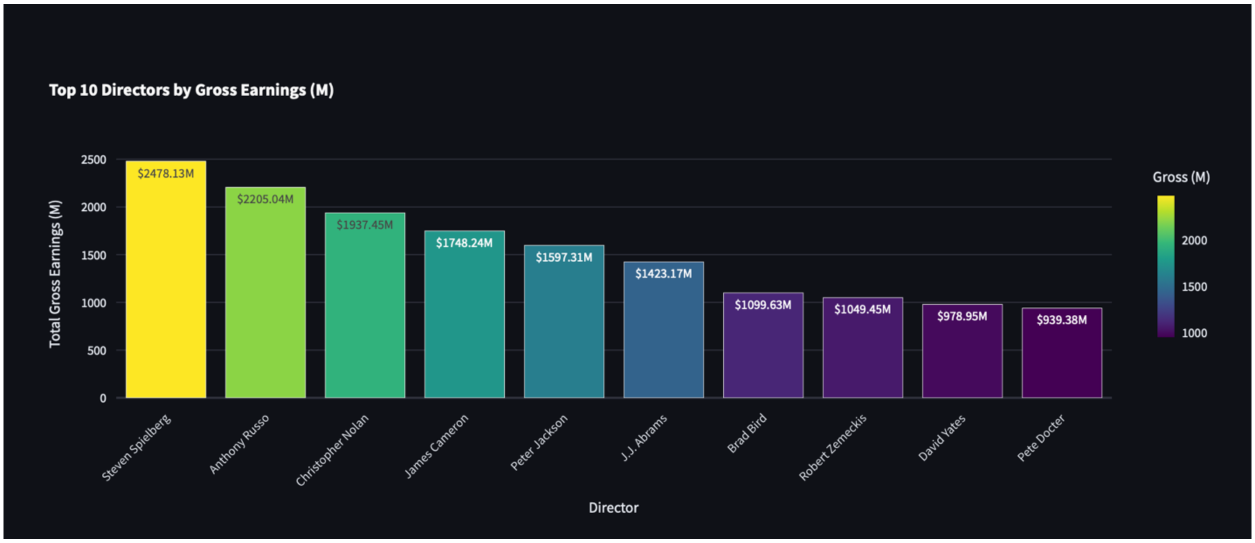


Figure 8 – Bar chart for top 10 Directors by Gross Earnings

- **Top 10 Actors by Gross Earnings**:
  - Another bar chart was created to display the top 10 actors based on the total gross earnings of the movies they starred in.
  - The chart aggregates earnings from all movies where the actor appeared in one of the four starring roles (Star1, Star2, Star3, or Star4).
  - Similar to the directors' chart, this visualization is interactive and color-coded.

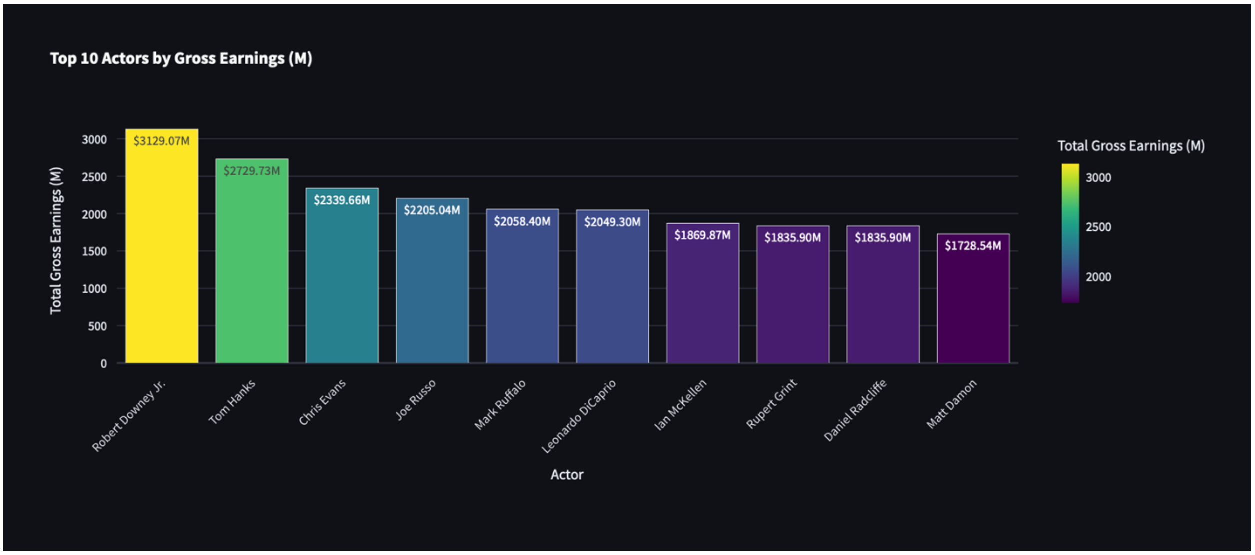


Figure 9 – Bar chart for top 10 Actors by Gross Earnings

F. HOURS LOGS

ACTIVITY	DATE	TOTAL TIME	STATUS	NOTES
Doing Interest and Expertise Survey to find out what skills and technologies want to work on	January 12, 2025	1:00	Complete	Survey completed successfully. Identified interest in Mobile Developments and Building Dashboards by Tableau and PowerBI
Finding the topic for the research project, looking for paper research and previous works	January 14, 2025	2:00	Complete	Explored multiple topics. Narrowed down to Activity Reminder Mobile App.
Creating and adding the Client in the Repo	February 1, 2025	0:00	Complete	Client added to the repository. No time spent as it was instantaneous.
Writing the draft proposal for the Client	January 20, 2025	6:00	Complete	

Discussing to the Client about the choosing topic is Activity Reminder Mobile App for Douglas College students	January 23, 2025	0:30	Complete	Client provided feedback. Topic was approved initially but later changed.
Changing the topic to the Dashboard Development by Python for analyzing movie trends and insights	January 24, 2025	0:30	Approved	Topic changed based on client's suggestion. New focus on Python dashboard developments.
Rewriting the proposal	January 24, 2025	7:00	Complete	Proposal rewritten to reflect the new topic. Submitted for approval. And it was approved
Cleaning database (Netflix database) by using JupyterNotebook	January 26, 2025	2:00	Complete	Data cleaned and prepared for analysis.
Writing the Progress Report	January 29, 2025	5:00	Complete	
Writing questions for an interview	February 1, 2025	1:00	Complete	Questions are being developed for Participant A and B.
Drafting the ideas for the dashboard	February 2, 2025	1:00	Complete	Initial ideas for dashboard layout and functionality drafted.
Coding for the dashboard by Python	February 6, 2025	10:00	In Progress	Coding in progress. Focused on data visualization and user interface.
Doing the interview with Participant A	February 8, 2025	0:30	Complete	Still finding the potential participant and scheduling. The deadline is Feb 10
Doing the interview with Participant B	February 10, 2025	0:30	Complete	Still finding the potential participant and scheduling. The deadline is Feb 12

Analyzing interview results	February 14, 2025	2:00	Complete	Analyze feedback from Participants A and B to refine dashboard features.
Writing the Interview Questions and Reponses	February 14, 2025	2:00	Complete	Interview questions and responses documented and uploaded to GitHub.
Cleaning database (IMDB database) by using JupyterNotebook	February 16, 2025	1:00	Complete	Data cleaned and prepared for analysis.
Developing data visualizations for the dashboard		0:00	In Progress	
Testing the dashboard functionality		0:00	On Hold	

G. REFERENCES

Few, S. (2022). *Stephen Few on Data Visualization: 8 Core Principles*. Tableau Software.

Hayes, G. (2019). *Which Programming Language Should Data Scientists Learn First?*. Towards Data Science.

Amar, N., & Baliga, S. (2022). *Design and Development of Analytical Dashboard*. Journal of Current Research in Engineering and Science.