**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Programme Title:** | HDip Data Analytics for Business |
| **Module Title:** | Data Visualisation Techniques |
| **Assessment Title:** | CA2 |
| **Lecturer Name:** | Mr. David McQuaid |
| **Student Full Name:** | Selvi Ganesh Kumar |
| **Student Number:** | Sba23395 |
| **Assessment Due Date:** | 25th May 2025 |
| **Date of Submission:** | 22nd May 2025 |
| **GitHub Link:** | <https://github.com/CCT-Dublin/machine-learning-for-business-ca2-Analytics-Selvi> |
| **Streamlit Link:** | <https://dashboard5py-gitkvut7ugh2ueftkkettr.streamlit.app/> |

**(Selvi Ganesh Kumar)**

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| 1 | Objective | 3 |
| 2 | Data Preparation and Experimental Methods | 3 |
| 3 | Experiment 1: Feature engineering & Visualisation: | 3 |
| 3.1 | Visualisation 1: Average Ratings and Count of Ratings by Year | 3 |
| 3.2 | Visualisation 2 & 3: Top Movies by Number of Ratings & Release year | 3 |
| 3.3 | Visualisation 4 & 5: Top Genres by Average Rating & Rating count | 3 |
| 3.4 | Visualisation 6: Most Frequent Movie Tags (WordCloud) | 4 |
| 3.5 | Visualisation 7: Time-Based Analysis – Monthly and Weekly Trends | 4 |
| 4 | Experiment 2: Baseline Dashboard Layout Using Genre | 4 |
| 5 | Experiment 3: Outlier Panel Visualization | 4 |
| 6 | Experiment 4: Animated Genre Popularity Chart | 4 |
| 7 | Experiments 5–7: Genre Table to Dual Dropdown Dashboard | 4 |
| 8 | Experiments 8–12: Color Scheme Optimization | 5 |
| 9 | Experiments 13–15: Advanced Filters – Genre Popularity and Tags | 5 |
| 10 | Experiments 16–18: Framework Comparison – Panel and Streamlit | 5 |
| 11 | Conclusion | 5 |
| 12 | Appendix | 6 |
| 13 | References | 26 |

**Objective**The primary goal of this project is to design and implement an interactive dashboard using a movie dataset containing movie titles, user ratings, genres, tags, and timestamps. Targeted at a young adult audience (ages 18–35), the dashboard emphasizes usability, interactivity, and visual appeal. A sequential experimental approach was used for feature engineering, visualisation design, layout structuring, and color scheme testing. Each stage's visual outputs are documented in the Appendix. The final dashboard achieves a balance between intuitive design and insightful data exploration, serving both casual users and machine learning applications.

**Data Preparation and Experimental Methods**To build a meaningful and user-friendly dashboard, series of experiments were conducted across data preparation, visualization techniques, and user interface design. The datasets (ratings, movies, and tags) were merged into a unified DataFrame to provide a holistic view of user behavior and enable cross-feature analysis. Data cleaning followed, removing nulls and duplicates, to ensure reliability. Timestamps were converted to datetime format to extract year and month components, supporting seasonal trend analysis. Genres were split and one-hot encoded to allow filtering and comparison across multiple genre categories. Movie release years were extracted from titles for chronological analysis.

Line and bar charts were selected for their clarity and effectiveness, especially for the target demographic of 18–35-year-old casual users. Rating count was prioritized over average rating as it provides a more intuitive and impactful measure of popularity. Visual aesthetics, including color schemes and layout structure, were refined through iterative testing to enhance readability and engagement.

**Experiment 1: Feature engineering & Visualisation:**

**Visualisation 1: Average Ratings and Count of Ratings by Year**To identify engagement trends over time, the ratings was grouped by release year to compute average ratings and total count of ratings. Line plots were used to visualize these metrics. The count plot effectively captured peaks in user activity, particularly in the 2000s, and was more visually impactful for the 18–35 demographic. Conversely, the average ratings relatively offered limited variation for casual users. The count plot highlights data-rich periods beneficial for training robust models and identifies sparse years that may need special handling. (Fig 1.)

**Visualisation 2 & 3: Top Movies by Number of Ratings & Release year**A bar plot of the top 20 movies by rating count (Fig. 2) highlights the most frequently rated titles, offering immediate insight into viewer favorites and culturally iconic films, particularly engaging for younger users. These movies, due to their popularity and high engagement, serve as strong training data for recommendation systems in machine learning. Complementing this, a line plot of total movies by release year post-1950 (Fig. 3) reveals a steady rise in film production, peaking around the 1990s–2000s. This trend not only illustrates the industry’s growth but also identifies data-rich periods useful for temporal feature engineering, helping models better capture evolving viewer preferences.

**Visualisation 4 & 5: Top Genres by Average Rating & Rating count**To explore user preferences, genres was transformed into a long format and grouped to calculate average ratings and ratings count. This guides both user exploration and high engagement. For average ratings, Animation and Film-Noir consistently receive high ratings while rating counts Drama, comedy and thriller showed high. This will be useful for personalized recommendation for the users and to identify sparser genres. (Fig 4. & 5.)

**Visualisation 6: Most Frequent Movie Tags (WordCloud)**Focusing on films after 2000, a word cloud was generated from user-generated tags. This visualization highlights frequently used descriptors like "romantic," "action," or "drama." For users, it offers an engaging summary of popular themes. For machine learning, frequent tags serve as valuable content-based features, supporting clustering and sentiment-aware recommendations. (Fig 6.)

**Visualisation 7: Time-Based Analysis – Monthly and Weekly Trends**Temporal features were extracted to analyze rating activity across months and weekdays. The analysis revealed user activity peaks in July and November, and high engagement on Sundays. While average ratings remained stable, count plots provided actionable insights for content scheduling. These features also enhance time-aware recommendation systems and inform strategies for cold-start content. (Fig 7.)

**Experiment 2: Baseline Dashboard Layout Using Genre**

The initial dashboard was built using Dash with a singular filter based on the "Genre" column. This allowed for interactive filtering to observe trends in movie ratings over time. The use of genre as a filter aligned well with younger viewers’ casual browsing behavior, encouraging exploration through familiar categories. The simplicity of a genre-based dropdown with a corresponding line plot offered both usability and analytical insight, providing a foundational structure for further experiments. (Fig 8.)

**Experiment 3: Outlier Panel Visualization**

A second layout was created using Panel to visualize outlier movies based on user-defined rating thresholds. By highlighting movies with ratings above a chosen value, users could identify critically acclaimed films across decades. While useful for analytical purposes and valuable for machine learning tasks such as outlier detection and temporal feature engineering, this approach was better suited for niche users. Given the project’s focus on the broader 18–35 demographic, who prefer trend-based and straightforward interfaces, the outlier dashboard was ultimately excluded in favor of more intuitive visualizations. (Fig 9.)

**Experiment 4: Animated Genre Popularity Chart**

This experiment explored an animated bar chart using animation\_frame='year' to visualize genre popularity over time. Although visually dynamic, the rapid transitions and inconsistent data representation led to poor interpretability. Genres appeared and disappeared too quickly, resulting in visual noise. While animation can appeal to younger users, the lack of clarity outweighed the aesthetic benefits. The baseline genre plot proved more effective for analytical comparison and user experience. (Fig 10.)

**Experiments 5–7: Genre Table to Dual Dropdown Dashboard**

A genre-based dashboard was iteratively enhanced by introducing a table for top movies (Experiment 5), replacing it with a bar plot for top movies (Experiment 6), and finally incorporating dual dropdowns for year and genre (Experiment 7) followed by a rating plot. These stages progressively improved the user interface by offering both detailed (movie-level) and aggregated (trend-level) insights. The final layout displayed top-rated movies, trending titles, and genre performance trends using a structured, intuitive format. Dropdowns allowed users to dynamically explore genre and year combinations, enhancing personalization. The color scheme, particularly a black background with white and blue plots, provided high contrast and appeal, though further refinement was needed. (Fig 11, 12, 13.)

**Experiments 8–12: Color Scheme Optimization**

Multiple experiments were conducted to optimize the visual theme. Complementary color schemes were found most effective for clarity and engagement. A white background with blue text offered readability but lacked vibrancy. A black background with gold text (Experiment 10) emerged as the most effective, providing contrast, visual hierarchy, and a sense of cinematic prestige well-suited for the target audience. Red text on black background (Experiment 11) created energy, offering a strong alternative. A gold background with black text (Experiment 12), though legible, lacked depth and immersion. The black background -gold scheme was selected for the final design due to its modern, engaging aesthetic. (Fig 14, 15, 16.)

**Experiments 13–15: Advanced Filters – Genre Popularity and Tags**

Experiment 13 introduced a dynamic genre popularity plot, which adjusted based on year selections. This provided a macro-level overview of engagement trends. In Experiments 14 and 15, a “Tags” dropdown was implemented to allow filtering by specific descriptors like "suspenseful" or "romantic." This allowed users to refine movie lists based on nuanced interests. The dashboard updated all relevant plots based on genre, year, and tag selections, giving users personalized and granular control over their viewing analysis. These enhancements also offered deeper input for feature engineering in machine learning models. (Fig 17, 18, 19.)

**Experiments 16–18: Framework Comparison – Panel and Streamlit**

To test deployment flexibility, the final Dash dashboard was replicated using Panel (Experiment 16) with identical features — genre, year, tags filters, and core plots — organized into intuitive tabs. The design and color scheme remained consistent, confirming that Panel could deliver a comparable user experience. In Experiment 17, the dashboard was also recreated using Streamlit and deployed as a web application. This baseline model was included due to time constraints, this version demonstrated strong potential for future integration of predictive analytics and recommendation engines. (Fig 20, 21, 22.)

**Conclusion**

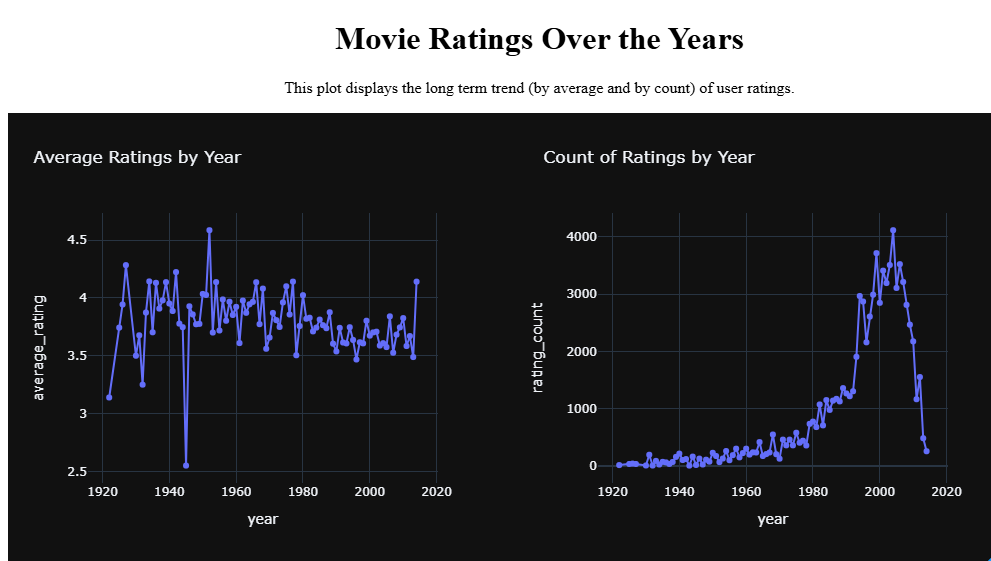
Across 16 experiments, the dashboard was developed through a systematic, user-centered approach focused on the 18–35 age group. The final product features dynamic filters, clear navigation, and visually engaging plots, balancing accessibility with analytical depth.

The dashboard adopts a clean, tabbed layout to reduce cognitive overload, allowing users to digest information in manageable sections. A clear hierarchy guides users from the title to interactive widgets (Year, Genre, and Tags), then through sequential visual tabs: starting with “Top Movies,” followed by trending movies, genre trends, rating analysis and temporal analysis. Captions placed discreetly below plots offer contextual cues without distracting from the visuals.

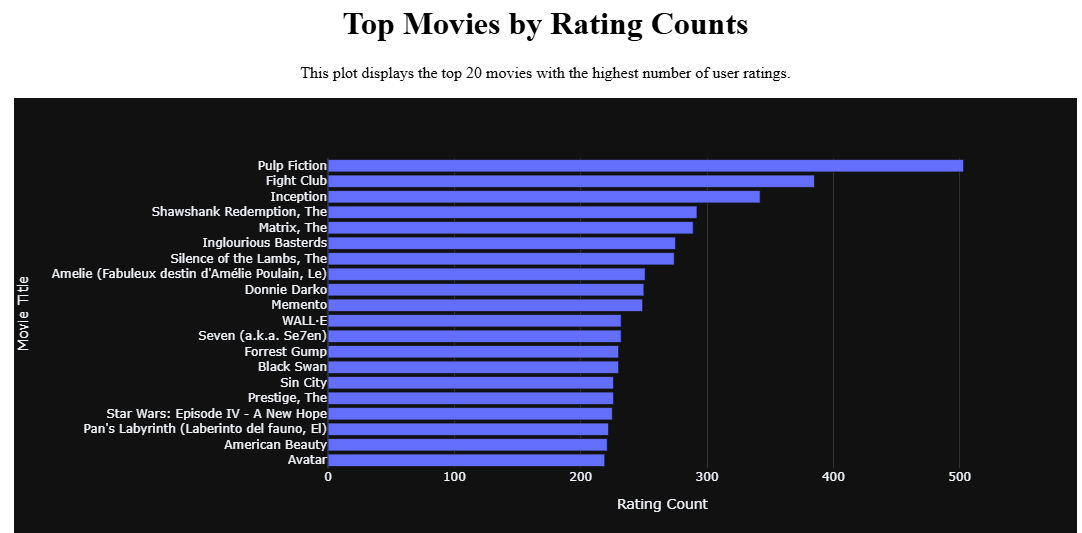
Visual hierarchy is emphasized through strategic design: primary attention is drawn to data representations (bars and lines), followed by plot titles and axis labels. Widgets enhance interactivity but remain optional, enabling both casual exploration and targeted analysis. The high-contrast black background with vibrant yellow text was chosen for readability and to reflect the energetic aesthetic of the entertainment industry. This design encourages efficient exploration and confident decision-making for both casual users and those interested in deeper data insights.

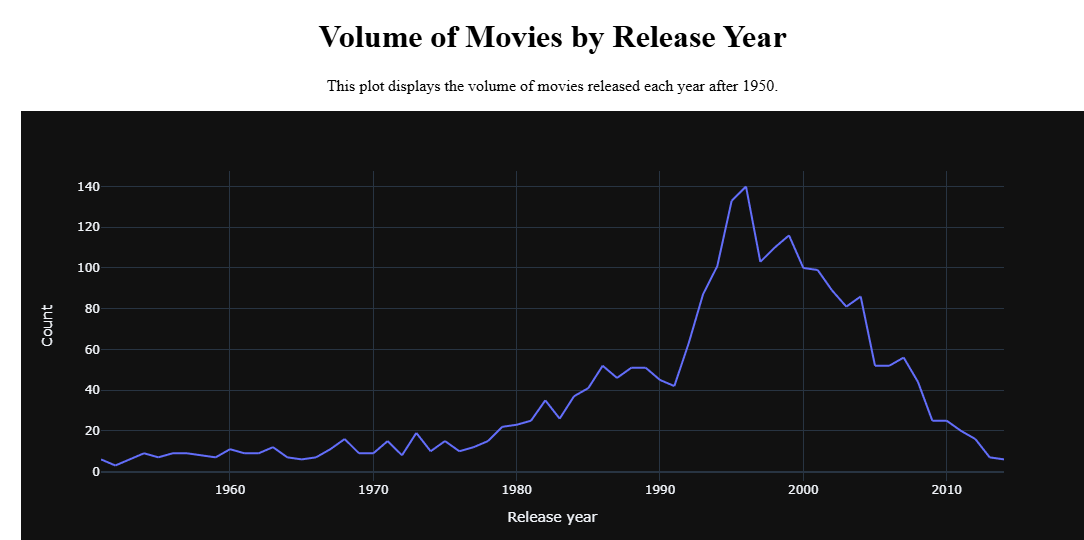
The final dashboard was a step-by-step process followed using an experimental approach, testing various features, designs, and layouts before selecting the final structure. Given the time constraints and the fact that a few aspects were beyond the current project's scope, the dashboard represents a solid foundation rather than a fully mature product. However, it marks an important starting phase of discovery. With additional time, data sources, and resources, this dashboard can be further enhanced, both in terms of functionality and design. It has the potential to evolve into a powerful tool for data analysis and user engagement insights.

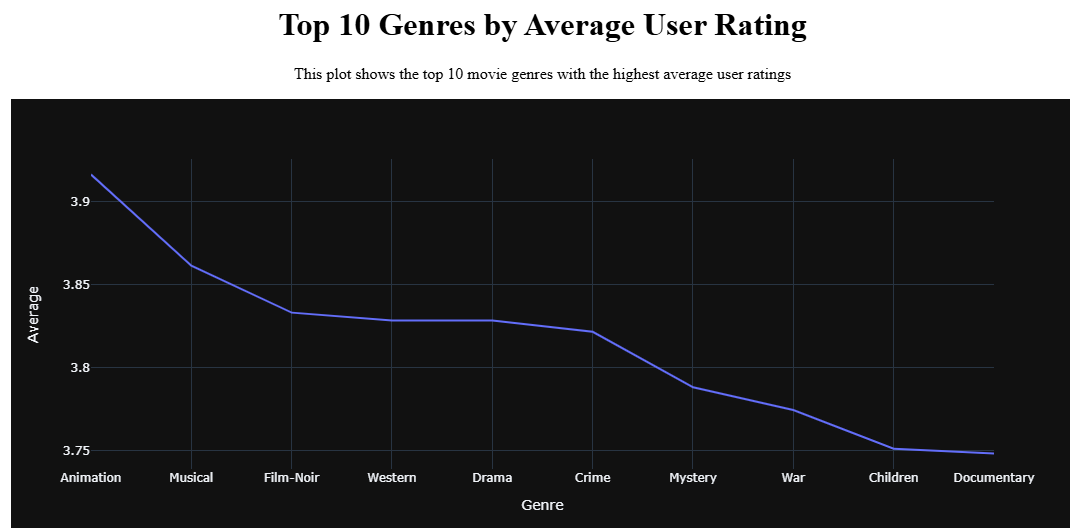
**Appendix**

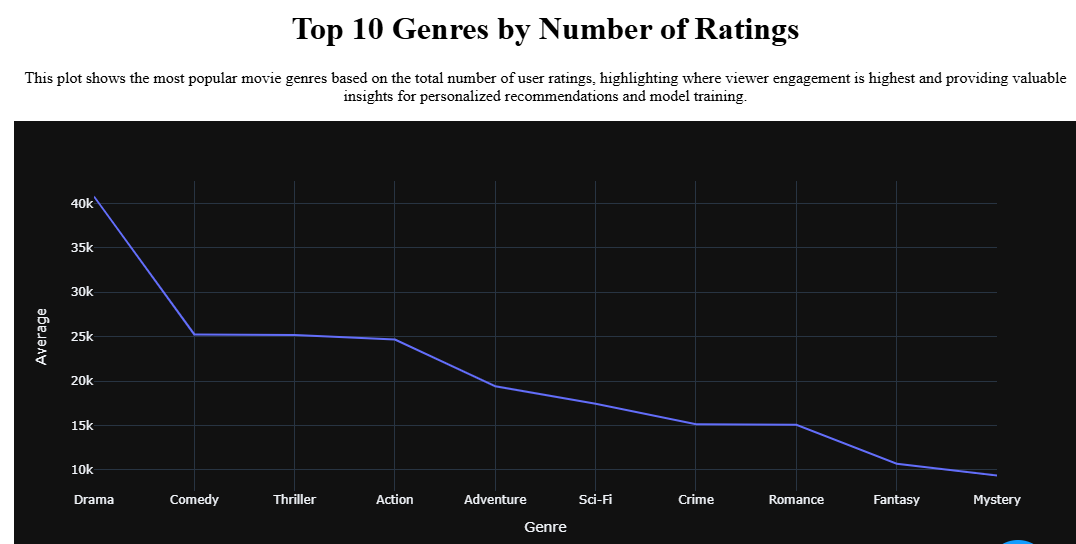
**Fig 1: Visualisation 1: Average Ratings and Count of ratings by Year:**

**Fig 2. Visualisation 2 - Top movies by number of ratings**

****

**Fig 3. Visualisation: 3. Total Movies by release year**

**Fig 4. Visualisation 4 - Top Genres by average rating**

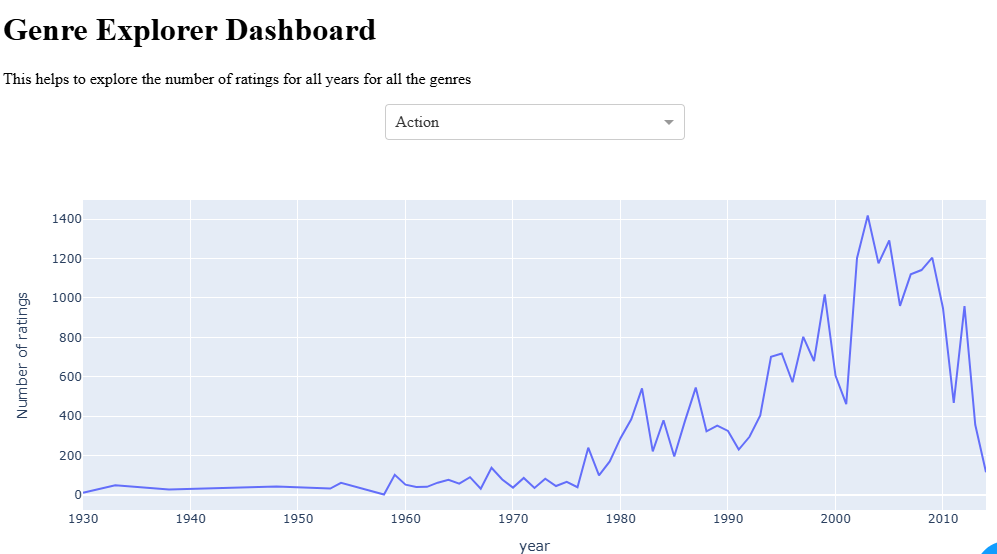
**Fig 5. Visualisation 5 - Top Genres by Number of Ratings**

**Fig 6. Visualisation 6 - Most frequent movie tags : WordCloud**

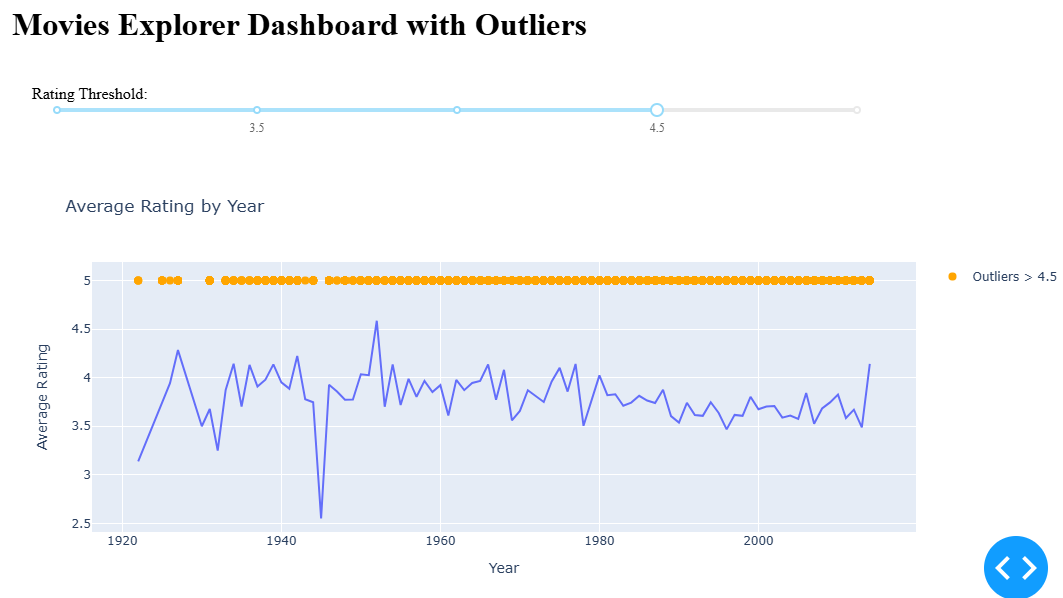


**Fig 7. Visualisation 7 - Time based analysis - Monthly and weekly trends:**

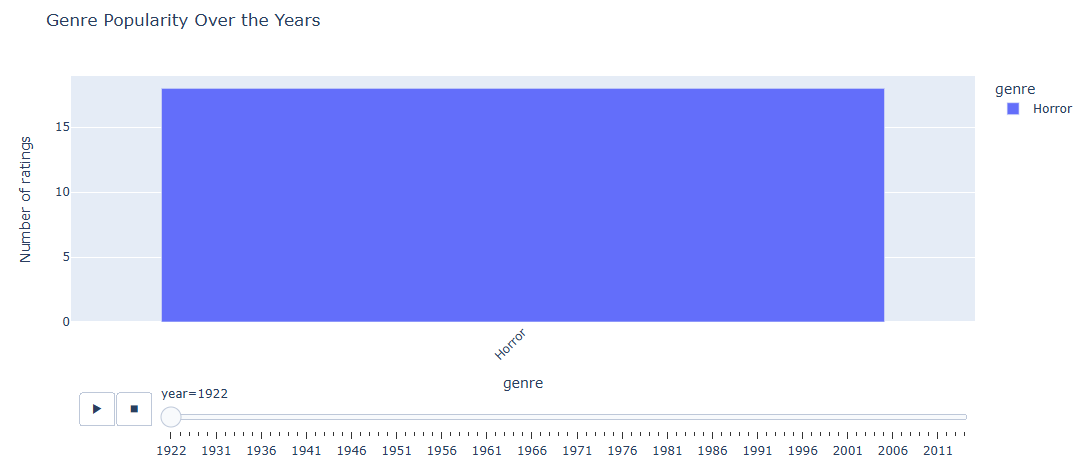
**Fig 8. Experiment 2: First Testing: Baseline Dash Layout using "Genre":**

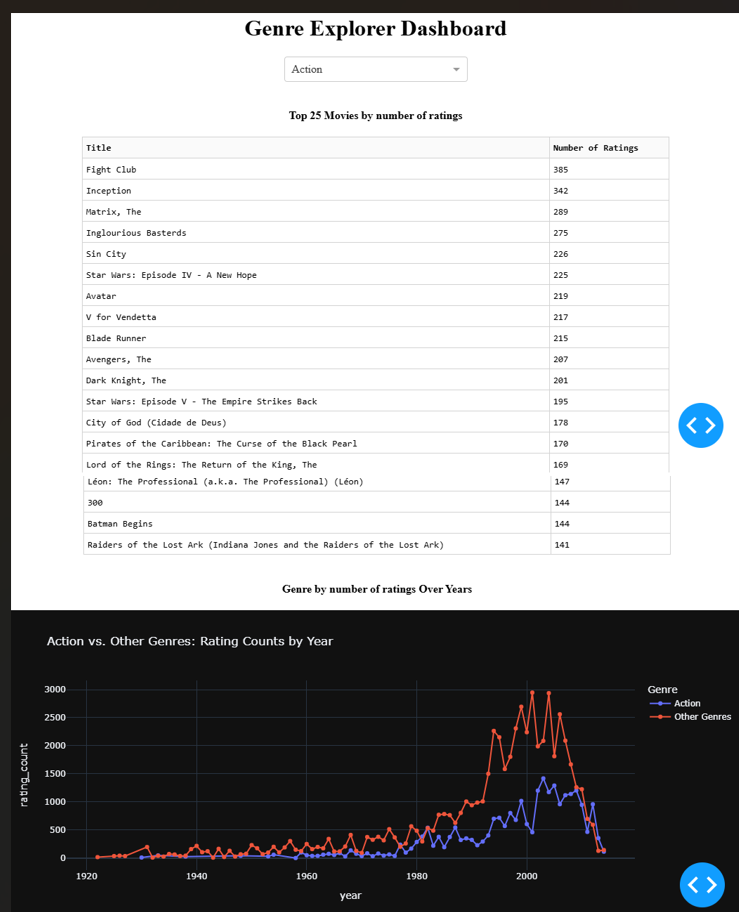
****

**Fig 9. Experiment 3: Baseline Panel (with Outliers) layout testing with Dash:**

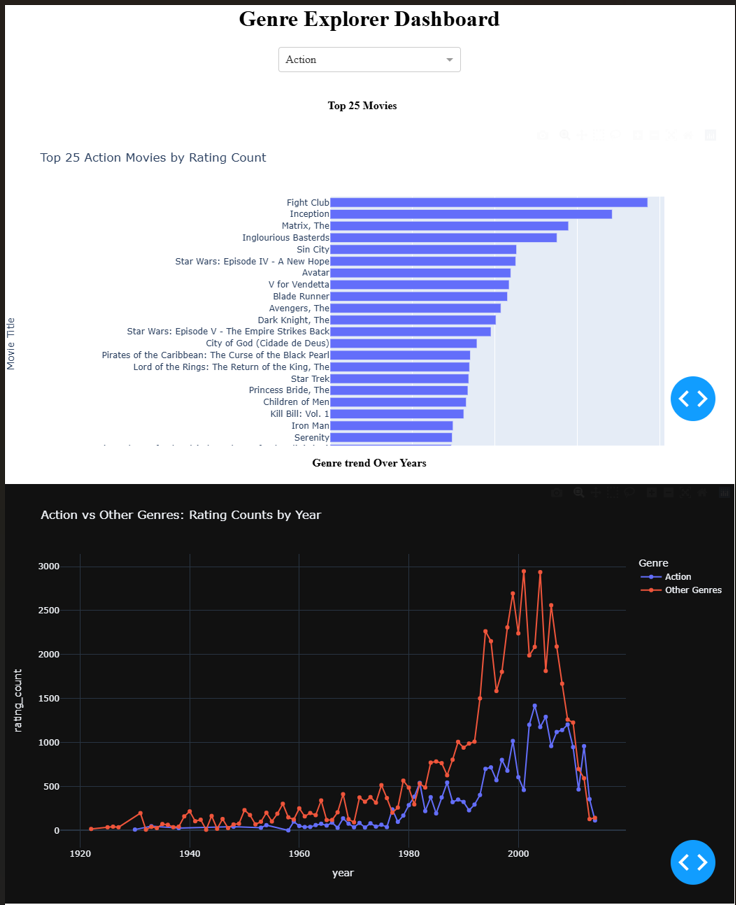
****

**Fig 10. Experiment 4: Popularity of movie genres by Year - Baseline model using Dash and Animation Frame by Year**

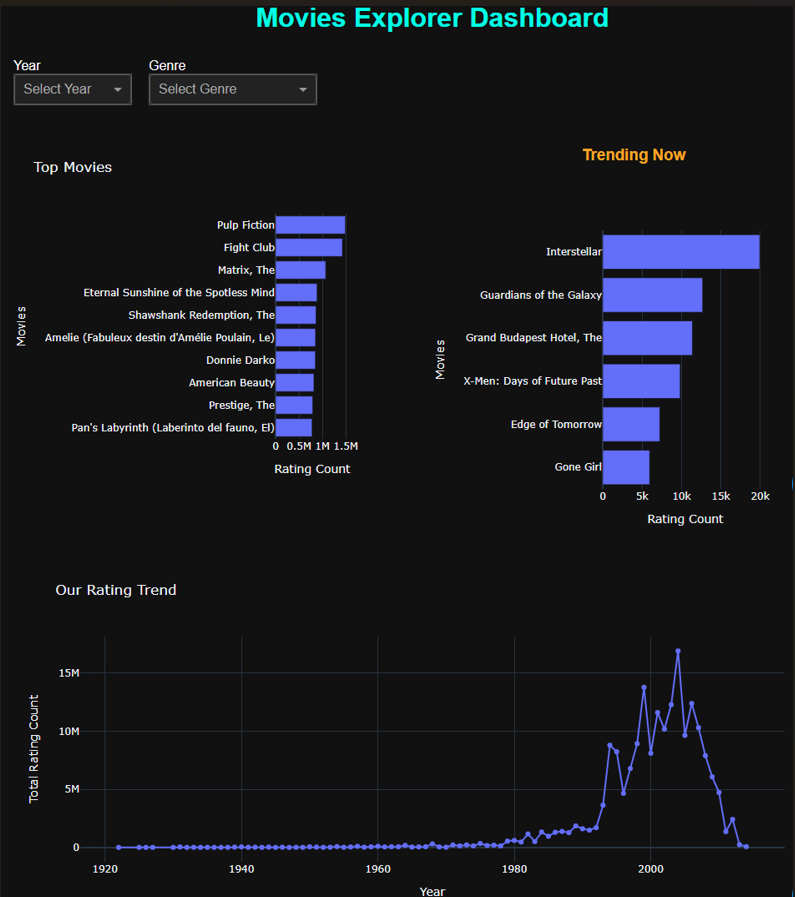
****

**Fig 11. Experiment 5: Dash Dashboard with Table and Lineplot:**

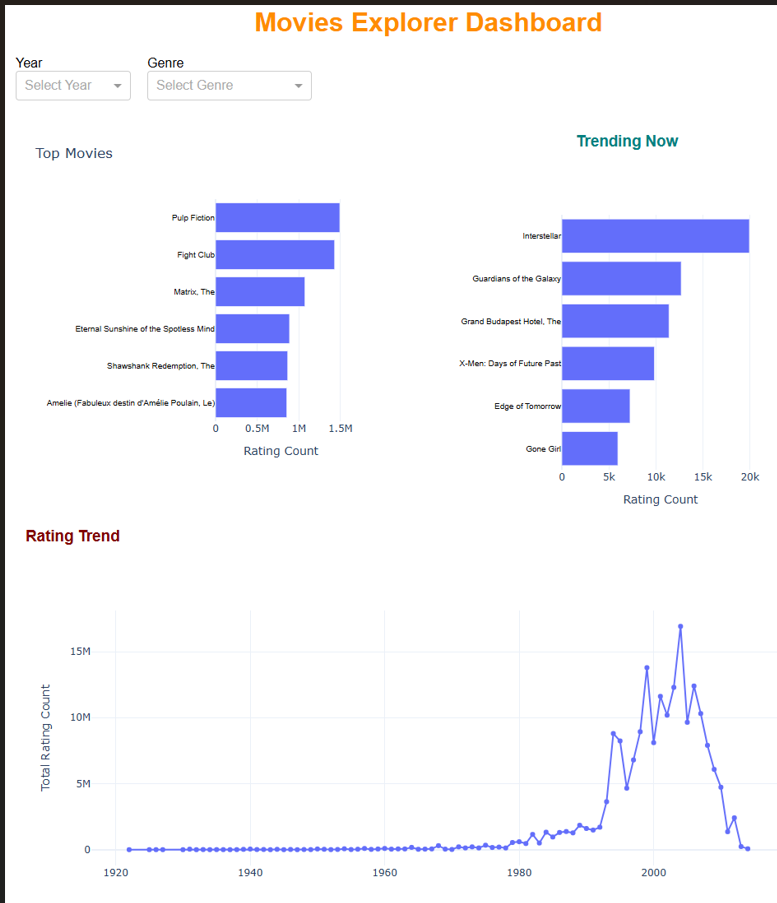
**Fig 12. Experiment 6: Genre Dashboard with Bar Plots and Line Plots using Dash:**



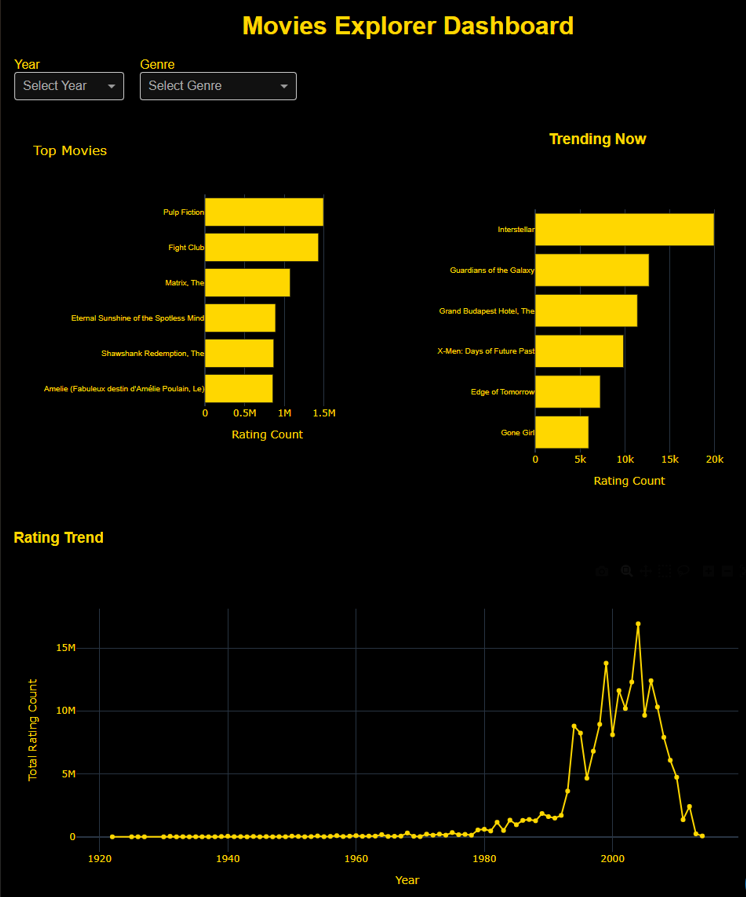
**Fig 13. Experiment 7: Enhancing the Experiment 6 Dashboard with Two Dropdowns (Year and Genre)**



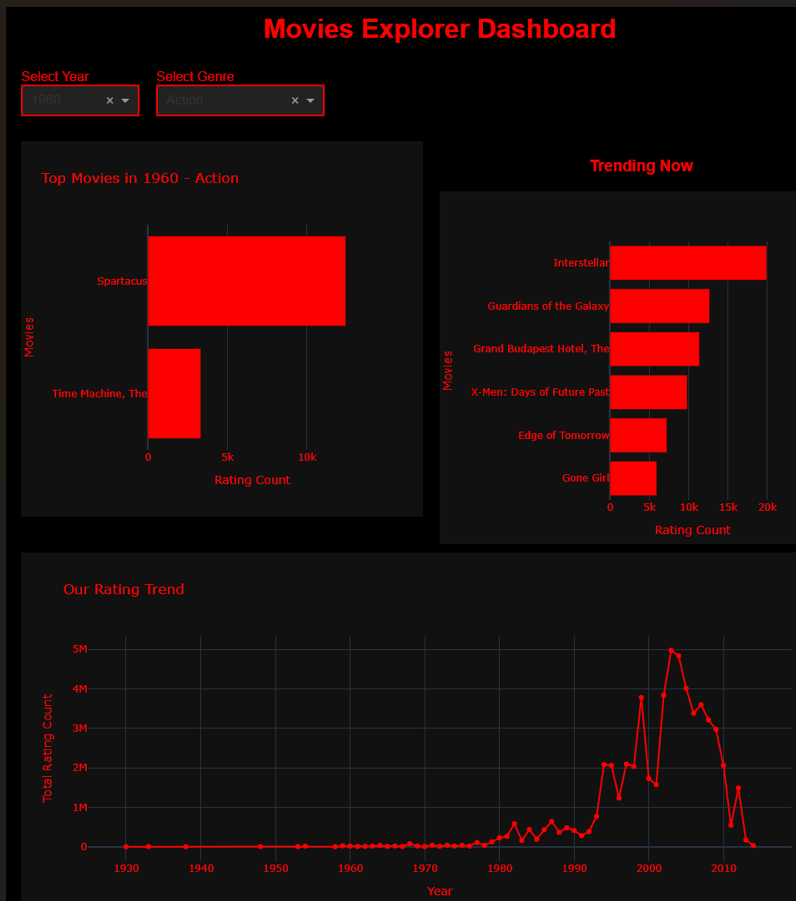
**Fig 14. Experiment 10: Dashboard Color Scheme: White background with blue texts:**

****

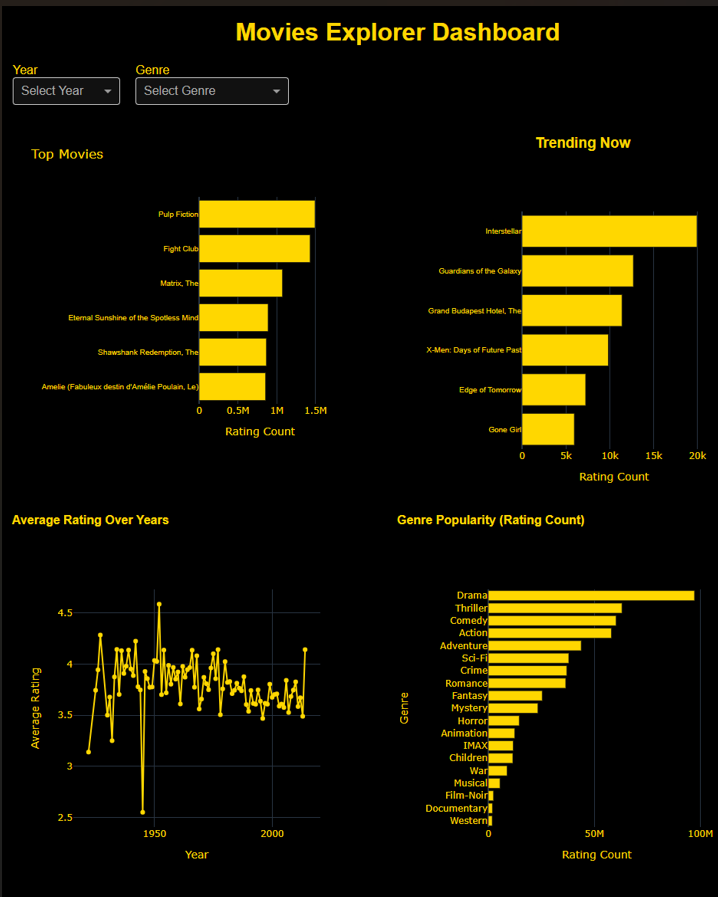
**Fig 15. Experiment 11: Dashboard Color Scheme: Black background with gold texts::**

****

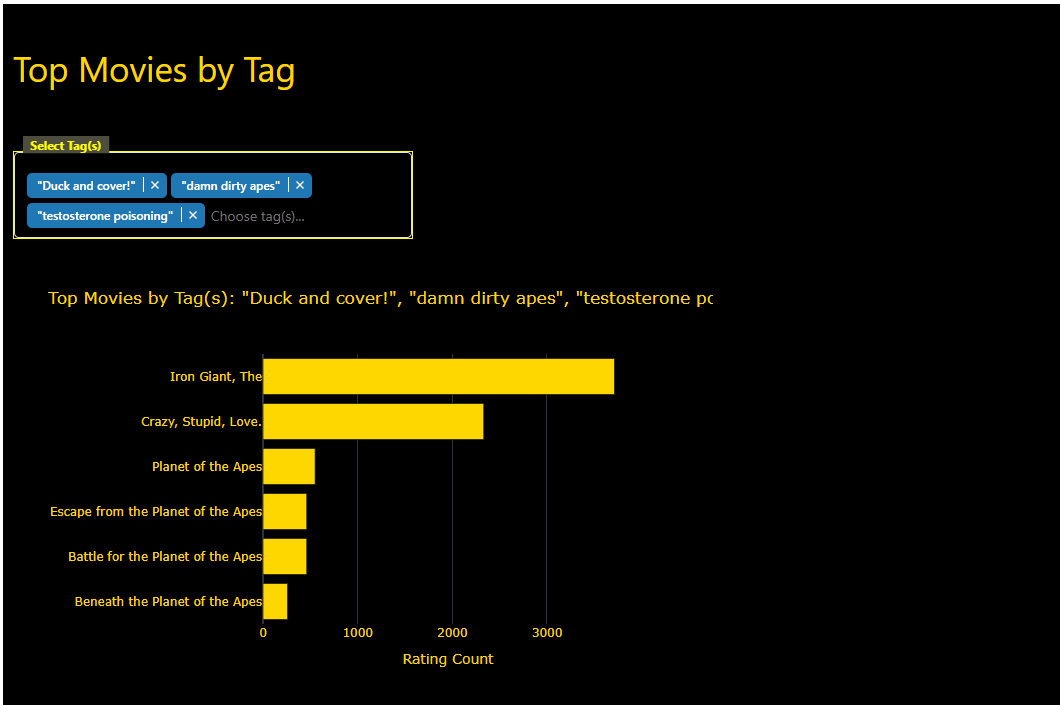
**Fig 16. Experiment 12: Dashboard Color Scheme: Black background with red texts:**

****

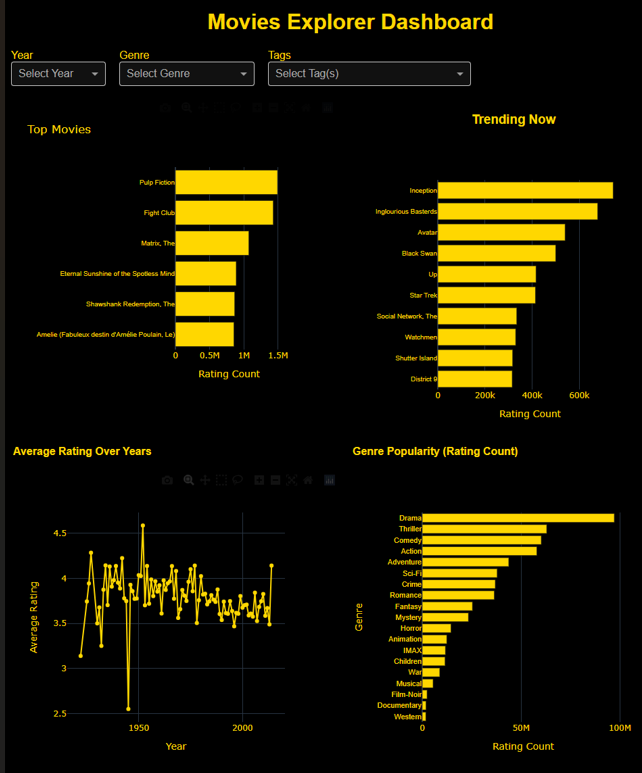
**Fig 17. Experiment 13: Adding plot "Genre popularity" to the dashboard**

****

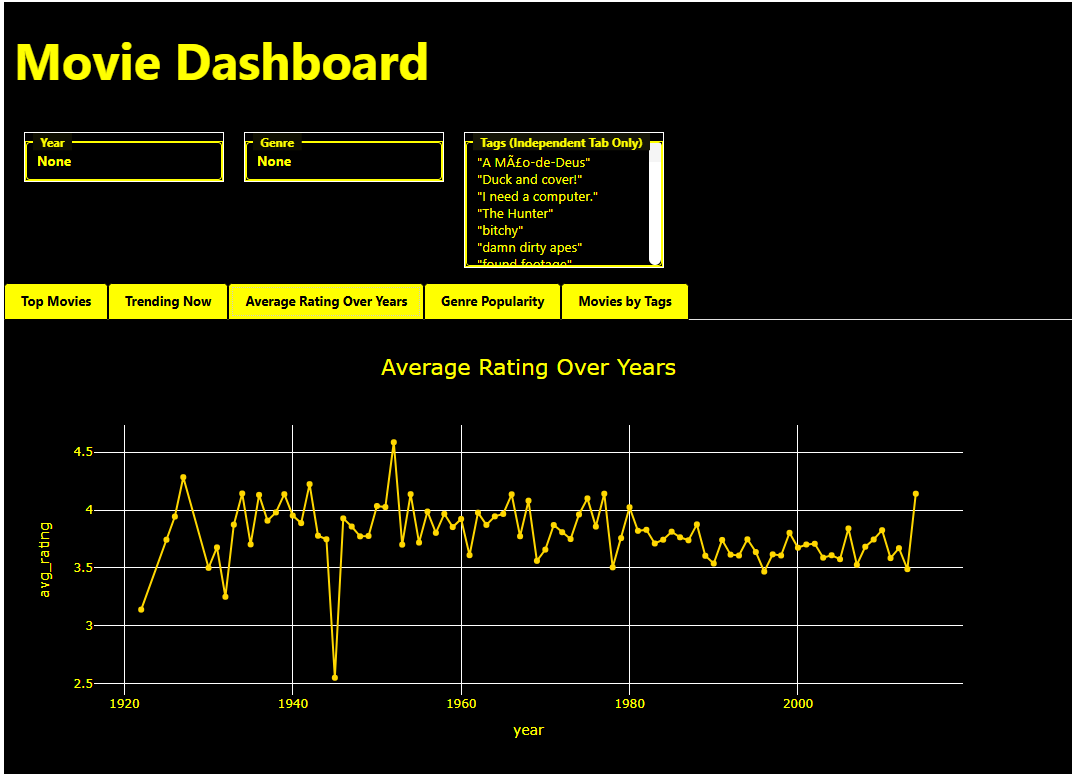
**Fig 18. Experiment 14: Adding Tag to the Dashboard:**

****

**Fig 19. Experiment 15: Additional dropdown "Tags" added to the dashboard**

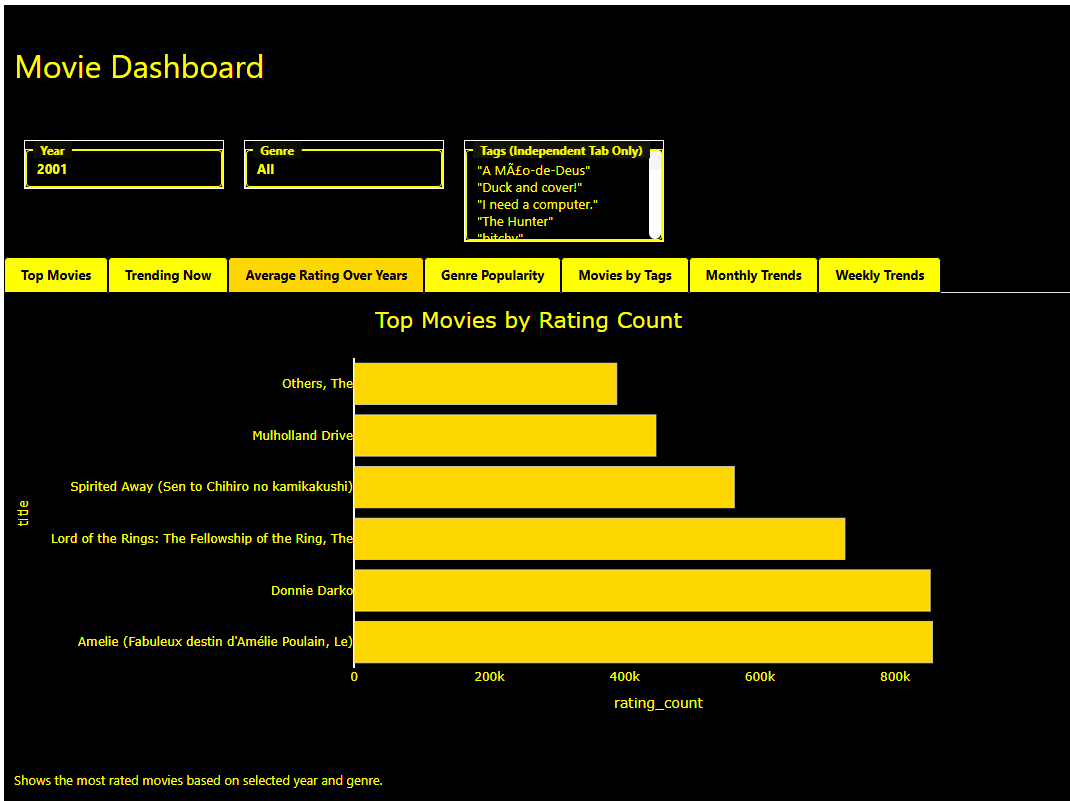
****

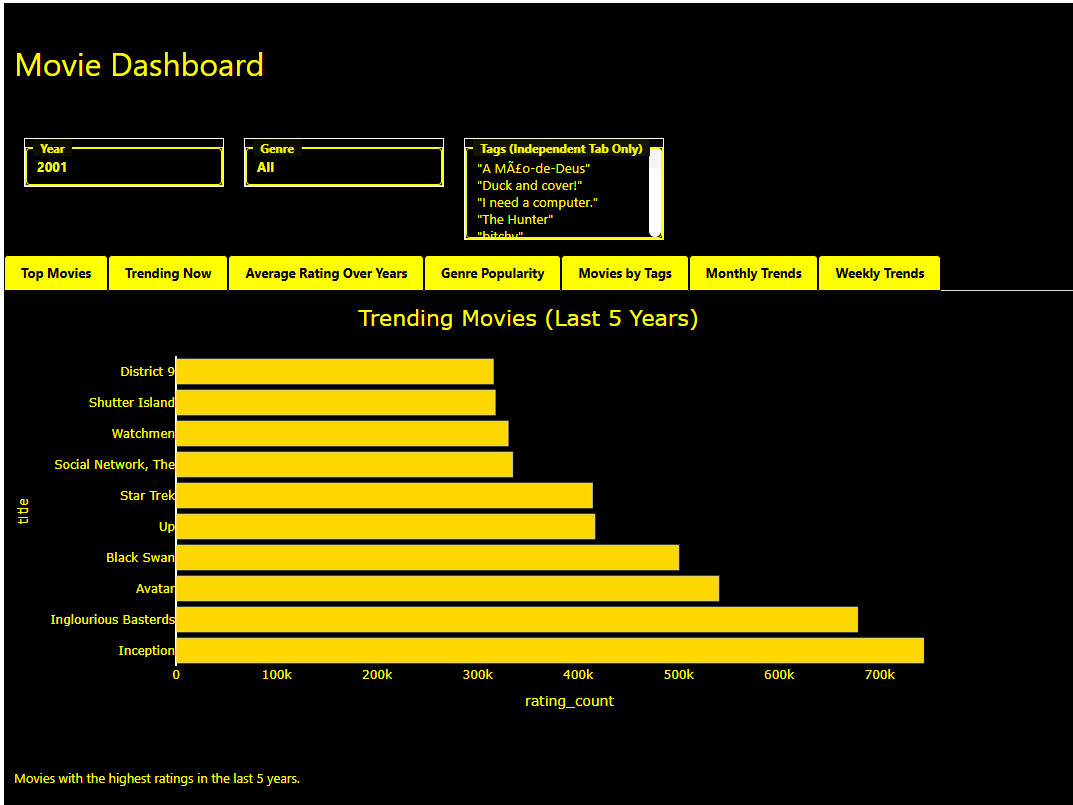
**Fig 20. Experiment 16: Using Panel for Dashboard with the same design**

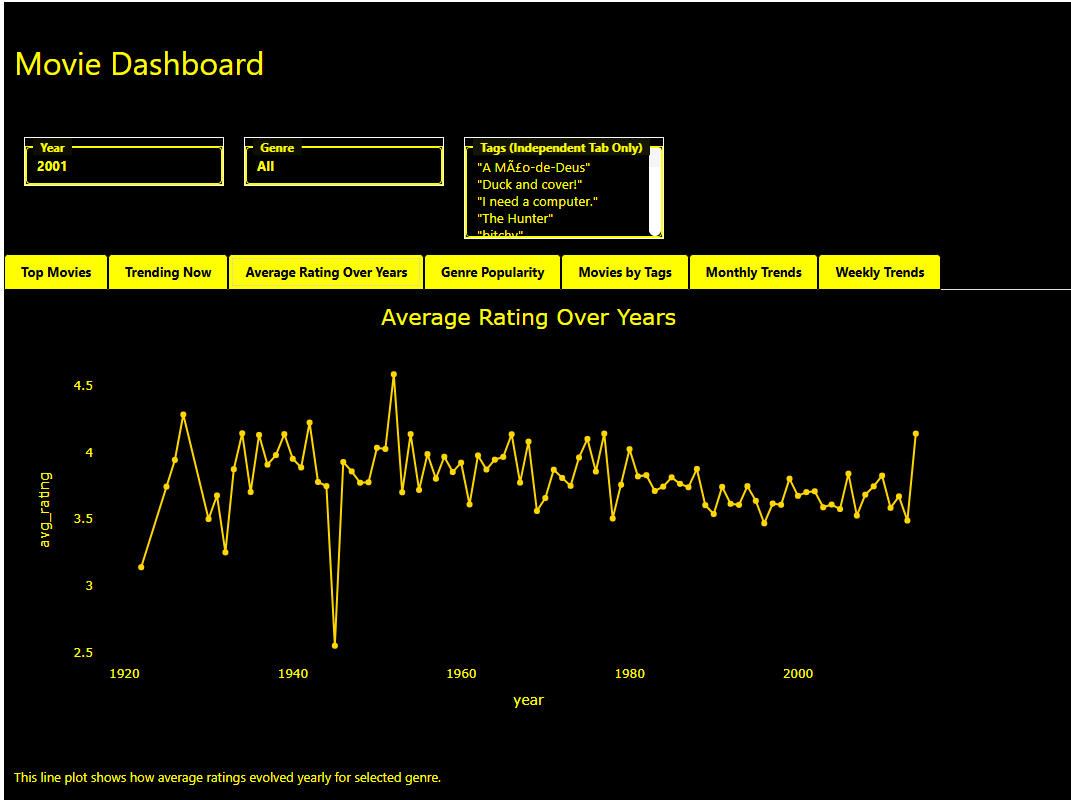
****

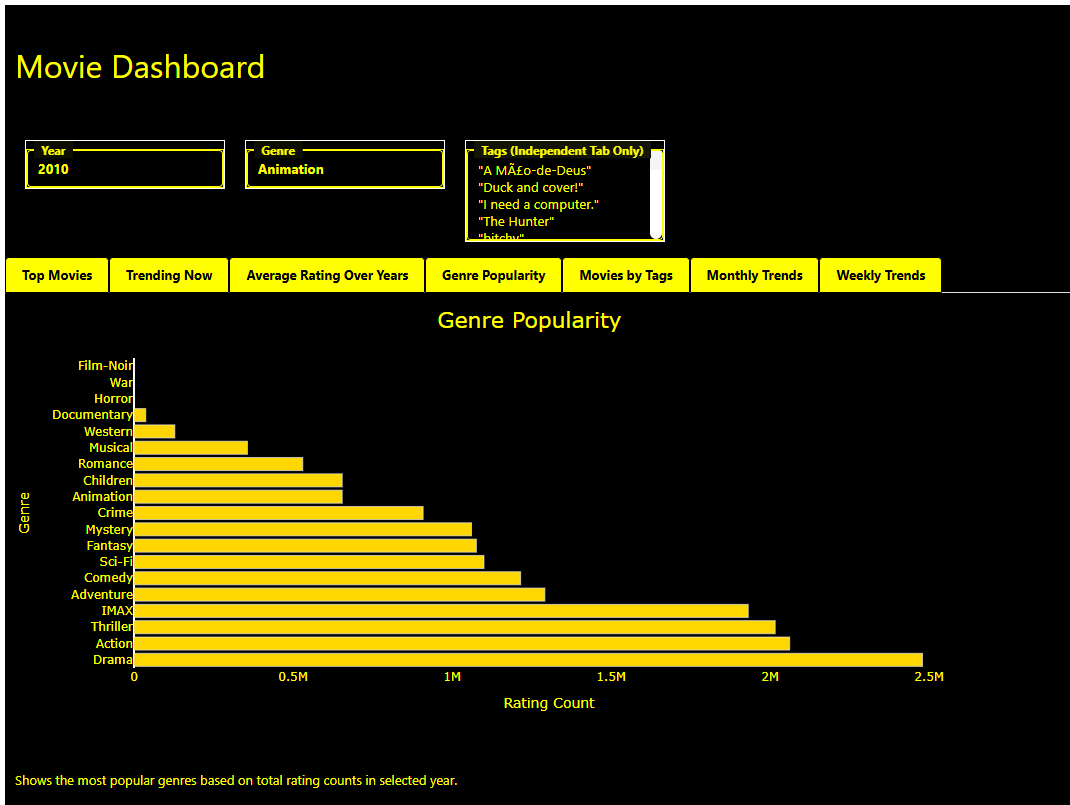
**Fig 21. Experiment 17: Adding Monthly and Weekly Trend Tabs to the Dashboard:**

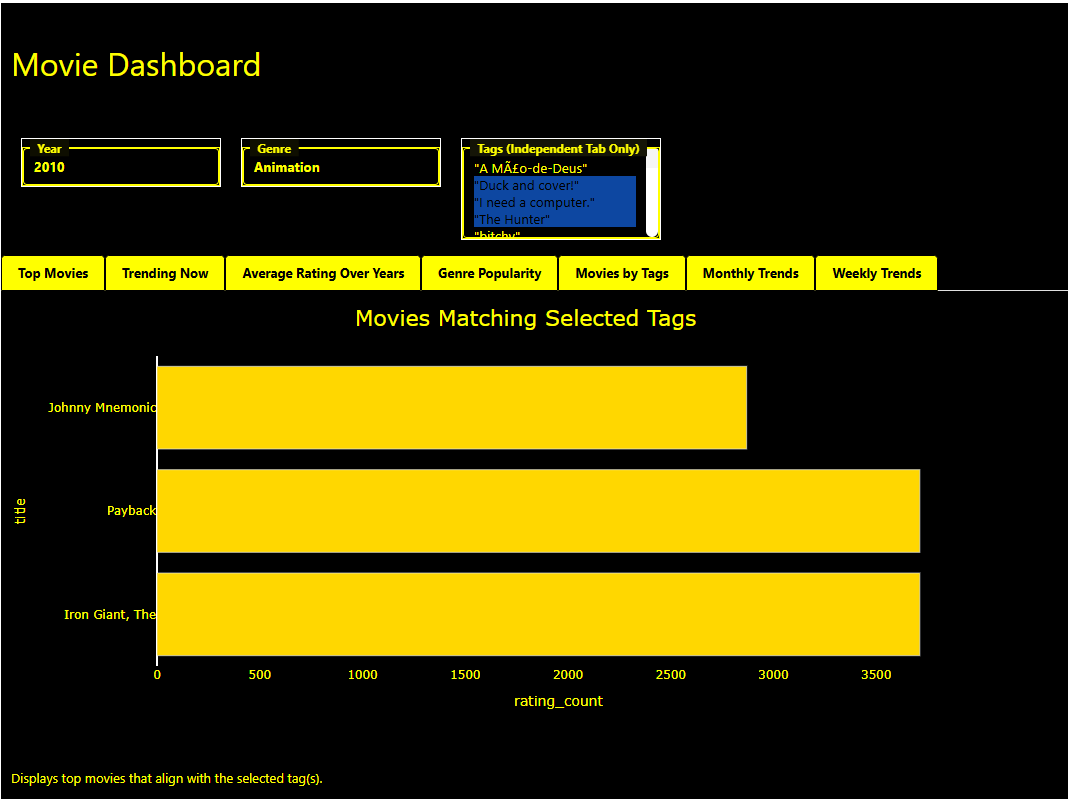
**Final Dashboard: Few screenshots of the functionality of the Dashboard:**

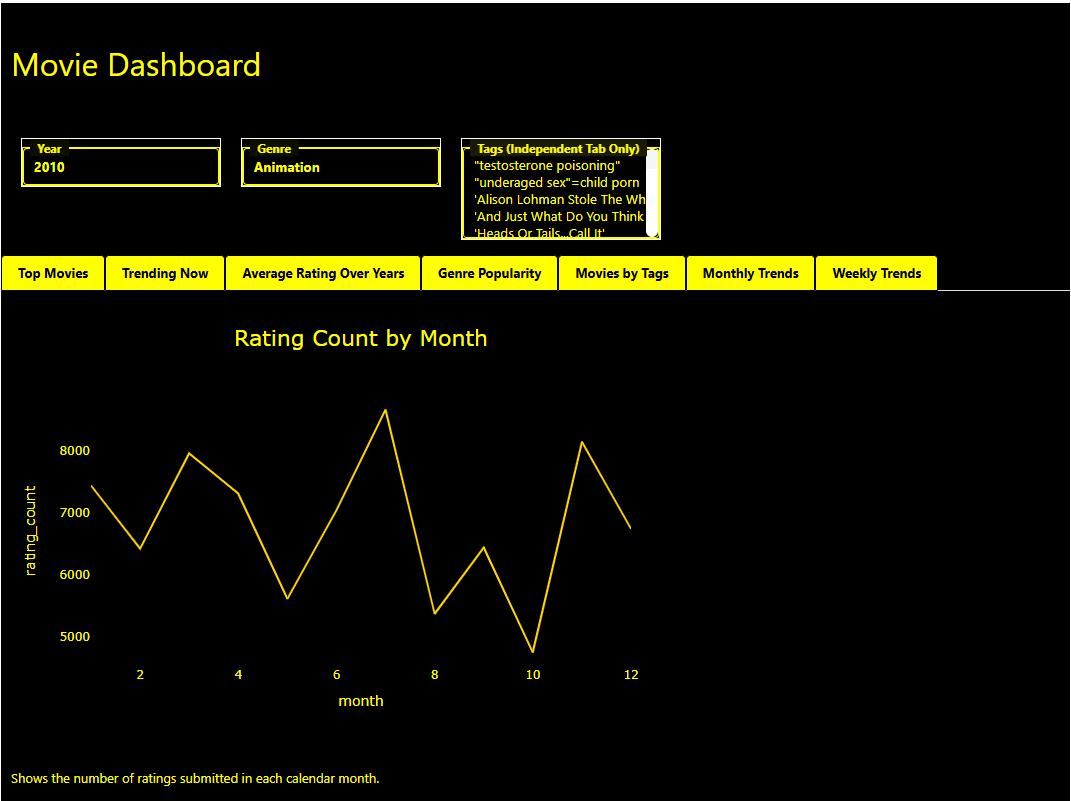
****

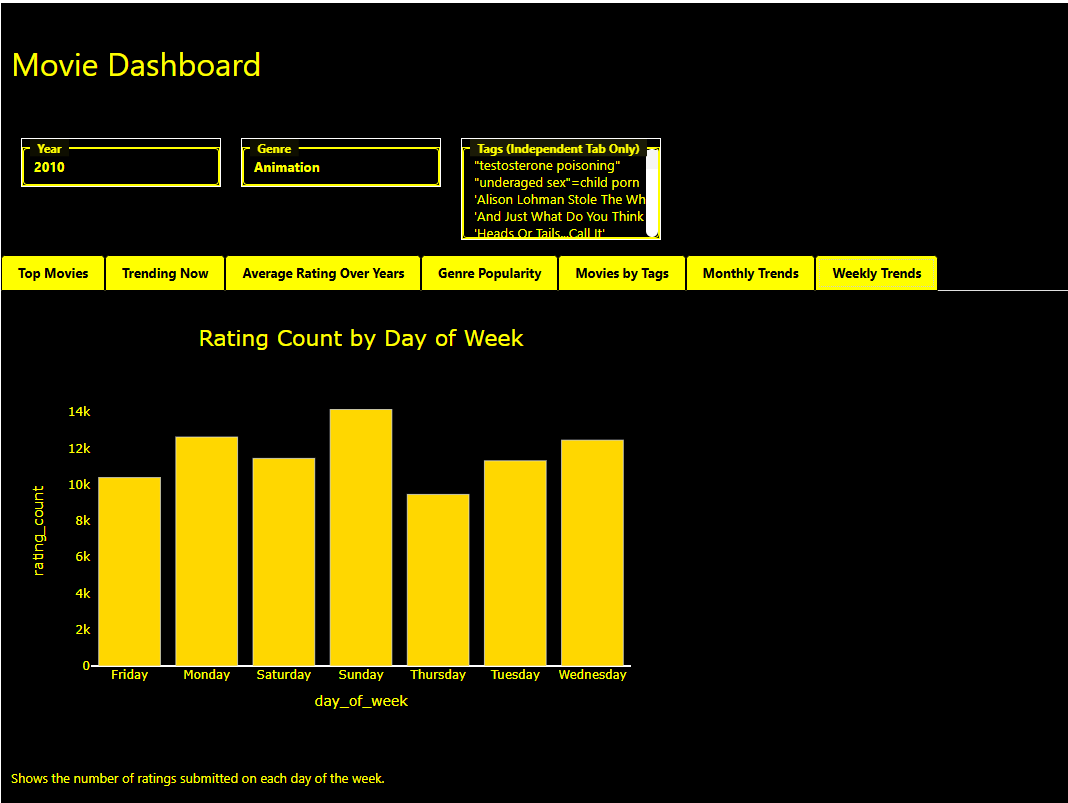
****

****

****

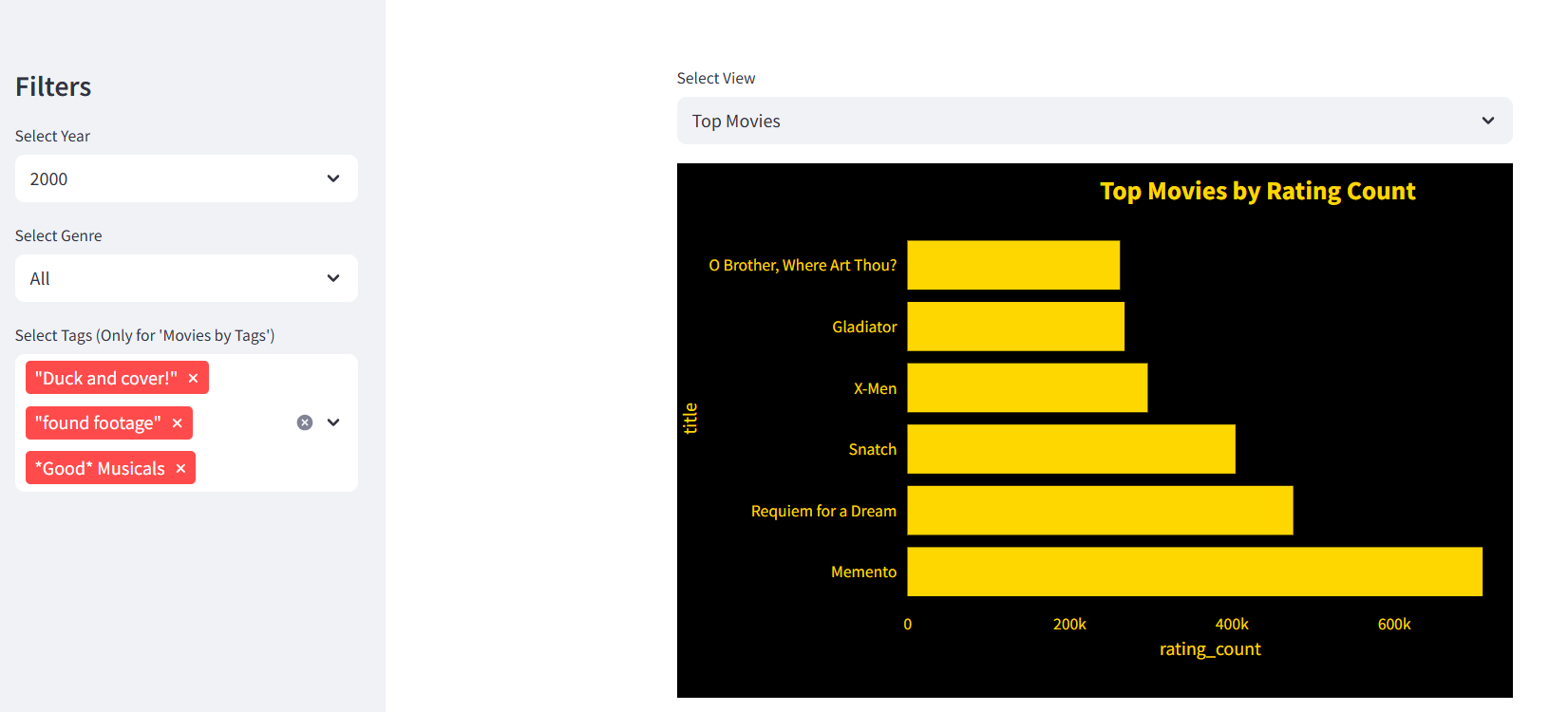
****

****

****

**Fig 22. Experiment 18: Streamlit Dashboard**

[**https://dashboard5py-gitkvut7ugh2ueftkkettr.streamlit.app/**](https://dashboard5py-gitkvut7ugh2ueftkkettr.streamlit.app/)



**References:**

1. (Create your first Streamlit app, Mr. McQuaid)
2. (Avocado\_dasboard, Mr. McQuaid)
3. (Creating an Interactive Dashboard from Jupyter Notebook with Voila, Mr. McQuaid)
4. (Dates and Times in Python.pdf, Mr. McQuaid)
5. (How to Create a GUI in Python using Tkinter, Mr. McQuaid)
6. (Colour Theory, Mr. McQuaid)
7. (The 9 Principles of Design, Mr. McQuaid)
8. (What Visualisation should I use v3, Mr. McQuiad)