BEMM461

Visualisation Project Report

ID: 710079698

Introduction

Choosing what movie to watch is always a conundrum for anyone let alone film-buffs. It can be overwhelming to locate a movie that suits your personal tastes and preferences among the millions of films listed on IMDb. Here is where an interactive dashboard for IMDb genre-based movie ratings is useful. Users of such a dashboard would be able to quickly explore and filter movies depending on their favourite genres and star ratings, giving them a unique and customised movie-watching experience. From the interactive dashboard created, the user can choose the genre and then one could see what the best rated movies are in each genre, the percentages of different certificates given to movies in each genre, the runtime for each movie in minutes for each genre, the ratings of each movie based on each director for each year and a comparison between the gross revenue and Imdb rating for each movie in each genre can also be seen. Ultimately, a genre-based interactive dashboard for IMDb movie ratings would offer movie fans a user-friendly and individualised experience, making it simple for them to select and watch films they'll adore.

Purpose and Target Audience

An IMDb movie ratings dashboard's function is to offer customers a platform that allows them to browse and examine the movie ratings information that is available on the IMDb website. This dashboard allows users to select and display data according to various genres, directors, ratings, release years, and other relevant features. Users of the dashboard can use it to find well-liked films and filmmakers, comprehend how ratings are distributed, and investigate connections between various elements. Everyone interested in examining and evaluating movie ratings data, including movie aficionados, businesspeople, critics, and researchers, might use such a dashboard.

Data

The data for this dashboard has been collected from the Imdb movies dataset from Kaggle.com. This dataset contains information about movies from 1920 to 2020. The key KPIs present in this dataset are Poster Link, Series title, Released Year, Certificate, Run Time, Genre, Imdb Rating, Overview, Meta Score, The director’s name, Star1, Star2, Star3, Star4, Number of Votes and Gross Revenue for each movie.

As a part of data cleaning, the columns Poster Link, Overview, Star1, Star2, Star3 and Star4 were removed as they were not necessary for any visualizations on the dashboard. The column Run Time had data in the form of minutes + ‘min’. This had to be changed to just minutes in integer format. We also had to remove any null values existing in the dataset.

The following code snipped was run in python for the data cleaning and pre-processing:

```
a = pd.read_csv("imdb.csv")
df=pd.DataFrame(a)
# Cleaning the data to remove the null values
df.dropna(inplace=True)
#Data Precprocessing
df['Runtime'] = df['Runtime'].str.extract('(\d+)').astype(int)
df = df.drop(['Poster_Link', 'Overview', 'Star1', 'Star2', 'Star3', 'Star4'], axis=1)
```

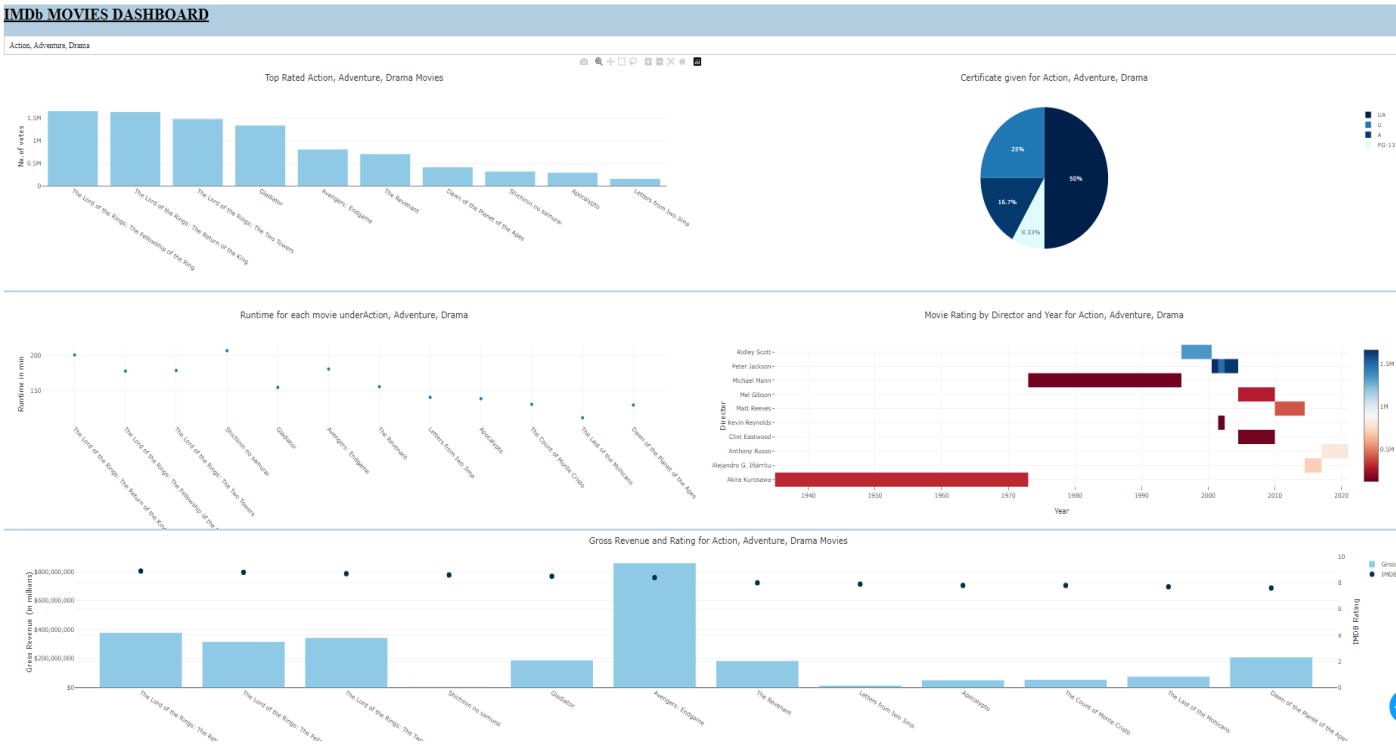
Here is a sample of the dataset:

Series_Title	Released_Year	Certificate	Runtime	Genre	IMDB_Rating	Meta_score	Director	No_of_Votes	Gross
The Shawshank Redemption	1994	A	142	Drama	9.3	80.0	Frank Darabont	2343110	28,341,469
The Godfather	1972	A	175	Crime, Drama	9.2	100.0	Francis Ford Coppola	1620367	134,966,411
The Dark Knight	2008	UA	152	Action, Crime, Drama	9.0	84.0	Christopher Nolan	2303232	534,858,444
The Godfather: Part II	1974	A	202	Crime, Drama	9.0	90.0	Francis Ford Coppola	1129952	57,300,000
12 Angry Men	1957	U	96	Crime, Drama	9.0	96.0	Sidney Lumet	689845	4,360,000

Visualisation Tools used: Python code was run on Visual Studio Code to create the html interactive dashboard. Plotly and dash libraries were used in the code.

The dashboard was evaluated after the original design was finished to find any usability problems or potential improvement areas. The dashboard has an intuitive and user-friendly structure and navigation. This entails dividing the dashboard into distinct areas with uniform colour schemes and typefaces. The design was improved through trial and error to better suit user needs. Also, keeping an eye on the data sources, updating the visualisation software as needed, and taking care of any error or problems as they developed was done.

Here is a snippet of the dashboard under genre Action, Adventure, Drama:



Design Process

Idea Generation:

In the initial stage of my design process, I brainstormed different ideas on how to visually represent the data. I focused on identifying the key insights that users should be able to obtain from the dashboard, selecting the appropriate types of charts and graphs that would effectively communicate the information, and presenting the data in a user-friendly and intuitive way. By considering these factors, I aimed to create a dashboard that would enable users to extract meaningful insights easily and quickly from the data.

Domain Validation:

The dataset provides all the required properties to examine distinct trends and ratings for numerous movies across different genres over the years, and the problem that consumers face at the domain level is well stated. During the idea creation phase, graphs and concepts are produced that are in line with the data domain and user needs. As a result, the visualisations created for the dashboard are suitable for effectively communicating insights to the users, and the data abstraction process may be successfully completed with the chosen charts and graphs.

Abstraction Validation:

I carefully chose and structured the pertinent data during the data abstraction process for my dashboard. This includes choosing how to organise or summarise the key data points from a bigger dataset and determining their importance. For instance, it involved aggregating data into relevant ranges and formatting columns to the necessary format. I used functions and developed code to perform the necessary computations and generate new columns with pertinent data. The columns Poster Link, Overview, Star1, Star2, Star3, and Star4 from the original dataset were removed as part of abstraction. The primary goal was to guarantee that the dashboard only used the most pertinent data and that it was meaningfully organised throughout the data abstraction phase. Instead of having target users complete an abstract job set by the system's creators, the system must be tested by actual users doing their own work. A target user's discovery is a strong justification for the usefulness of this dashboard.

Idiom Validation:

Using well-known perceptual and cognitive concepts, I have carefully justified the design of the idioms. Systematic checks were made to make sure the design does not contravene any accepted standards. I chose the right chart styles, colours, fonts, and other design components to properly display the data in my dashboard. The chart types used were a bar chart for the highest-rated films over time, a pie chart for certificate percentages for each genre, a scatter plot for the length of each film in minutes for each genre, a heat map for movie ratings for each director per year for each genre, and a combination of a bar chart and scatter plot for the comparison of box office receipts and movie ratings for each genre. These chart kinds were chosen based on how well they conveyed information to the viewer and how well they complemented the data being represented. I took care

to choose colours that would improve the data's legibility and clarity. The dashboard's colour scheme is visually appealing and easy on the eyes. I tried to refrain from utilising too many vibrant or striking colours that can be upsetting or draw the viewer's attention away from the data itself that is being visualised.

Algorithm validation:

The algorithm's performance in terms of time or memory may not be at its best. One defence against this vulnerability is to display still images or movies produced by the implemented algorithm. To develop a unified and efficient dashboard, this validation entails figuring out how to connect the data with the selected idioms in the best possible way. The dashboard must be user-friendly, simple to use, and give useful information on Istanbul's shopping trends thanks to the algorithm. To create this dashboard, I utilised Python as my programming language and a variety of data processing modules, including Pandas, Dash and Plotly. I've used these libraries proficiently in previous jobs, and I have some experience programming in Python. Given my proficiency with these tools, I chose to use Python to build this dashboard to take use of its robust data manipulation and visualisation features.

Charts used:

Visualisation	Chart type	Evaluation	User Interaction
Top rated movies of all time for each genre	Bar chart	This visualization allows users to easily see the top-rated movies for a specific genre, and compare their popularity based on the number of votes they have received.	Users can choose the genre by clicking on the dropdown menu on top of the visualisation. By hovering the cursor over the bars corresponding to each movie, they can see the number of votes received.
Percentages of each type certificates for each genre	Pie Chart	This visualization allows users to easily see the distribution of movie certificates for a specific genre in a clear and concise manner.	On choosing what type of genre they want, the users can see what percent of each type of certification is given for movies of that genre
Runtime for the movie for each genre	Scatter Plot	This visualization allows users to easily see the runtime of each movie in a selected genre and	On selecting the type of genre, the user can see what the total runtime of each movie is by hovering the

		compare them based on the length of the movie.	cursor over each dot on the scatter plot
Movie rating per each year for each director for each genre	Heat map	This visualization allows users to easily see how the movie ratings for each director have changed over the years in a selected genre, with a clear indication of which directors are more popular in terms of the number of votes they received.	After selecting the type of genre from the drop-down menu on top, the user can see the comparison of the popularity of each director for each year by hovering the cursor over the heatmap
Comparison of gross revenue and rating for the movies for each genre	Combination of Scatter Plot and Bar graph	the visualization also allows users to easily compare the gross revenue and IMDB ratings of movies in a specific genre.	By selecting a different genre from the dropdown menu, users can update the graph to show the gross revenue and ratings for a different set of movies.

We have made the following accommodations to retain these fundamental design principles:

Interactions that are simple and only require mouse clicks, streamlined headers and structural layout and to prevent the user from losing focus, consider using beautiful design elements that complement the colour of the visualisations.

Recommendations

One potential limitation of this dashboard is that it only shows the top movies in the selected genre by gross revenue, which may not be representative of all movies in that genre. IMDb ratings and reviews can be influenced by various factors, such as popularity, marketing, and the demographic of the users. This may result in biased ratings and reviews. Ratings and reviews on IMDb are subjective and based on the opinions of the users. This means that the ratings and reviews may not accurately reflect the quality of a movie. IMDb provides information on movies but does not provide context or analysis of the broader movie industry or culture, which may be important for some users.

This IMDb dashboard primarily focuses on movies, but it may not provide information on other aspects of the entertainment industry, such as TV shows, theatre, or live performances.

Conclusion

The initiative was designed to help anyone with an interest in movies learn how movies have changed and evolved over time. For both movie buffs and industry insiders, IMDb continues to be a well-liked and reliable source of movie information. To make the dashboard accessible to those without technical expertise, it was designed simply. Effectiveness was ensured through the assessment and evaluation procedure. As a result, we were able to effectively communicate the value and applications of various visualisations through this project.

References:

- Harshit Shankhdhar. (2019). IMDB Dataset of Top 1000 Movies and TV Shows. Kaggle. https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows?resource=download
- Munzner, T. (2014). Visualization Analysis and Design. CRC Press.
- GeeksforGeeks. (2021, July 8). Python Plotly Tutorial. https://www.geeksforgeeks.org/python-plotly-tutorial/