Project Report: Movie Data Analysis and Genre Rating Visualization

Introduction

For this project, movie data will be gathered via an external API, analysed, and the results will be visualised. The objective is to assess the average ratings for various movie genres for a list of films in which Denzel Washington starred.

This project's main goal is to:

1. Get movie data for films, including Title, Genre, Year, IMDb Rating, and Votes.
2. To extract and clean up valuable information, process the data.
3. Analyse the information, focusing on the average ratings of the film genres.
4. Visualise the findings as a bar chart with the average movie ratings.

Data Collection

The data is collected via the **OMDb API** (Open Movie Database API).

API Details:

- **Base URL**: `http://www.omdbapi.com/`
- **API Key**: `66******`
- **Endpoint**: The key API endpoint used in the code is `?t=<movie_title>&apikey=<api_key>`, where the title of the movie is passed as a parameter.

For each movie, details such as **Title**, **Genre**, **Year**, **IMDb Rating**, and **IMDb Votes** are fetched from the API.

Data Processing and Transformation

Fetching Data

The `get_movie_data()` function sends a GET request to the OMDb API for each movie title in the list and returns it as JSON. Things like **Title**, **Genre**, **Year**, **imdbRating**, and **imdbVotes**, are extracted for further processing.

Data Cleaning

- The **imdbRating** is converted from a string to a float, but only if the rating is not `'N/A'` (i.e., if a valid rating is provided).
- The **imdbVotes** is also converted to an integer by removing commas. If the value is `'N/A'`, it is replaced with `None`.

Grouping by Genre

A new column **`Genres`** is created in the DataFrame to store these individual genres.The DataFrame is grouped by the **Genres** column, and the average **Rating** for each genre is computed. The genres are then sorted in descending order based on their average rating.

Data Analysis

Calculation of Average Ratings

After grouping the DataFrame by **Genres**, the **mean rating** for each genre is calculated using the `.groupby()` and `.mean()` functions. This allows us to find the average IMDb rating for all movies within a particular genre.

Data Visualization by **Plotting a Bar Chart**

Using **Matplotlib**, the **average ratings by genre** are plotted as a bar chart.The bar chart will display genres along the x-axis and average ratings along the y-axis, with genres on the left side of the chart having higher average ratings.

Challenges

- Understanding how API works
- Choosing an idea for the project
- Handling Multiple Genres
- Visualization Challenges
- Integration of External Libraries
- Writing a project report

Conclusion

This project gives a way to explore movie ratings and compare the performance of different genres. Future improvements could include adding more movie data, refining the genre categorization, or additional categories like budget or box office earnings.