Python for Data Science – Mini Project Report

Group details:

1. Piyush Sharma 1911104 B3-Batch
2. Rishabh Kothari 1911110 B3-Batch
3. Rushil Shah 1911122 B4-Batch

Problem Statement: Data understanding and analysis using NumPy, Pandas and Matplotlib on chosen dataset.

Dataset chosen: Movies streaming on OTT platforms. (MoviesOnOtt.csv)

The chosen dataset is a csv file containing details of various movies available on OTT platforms. It includes following columns:

1. **ID**
2. **Title** – name of the movie
3. **Year** – releasing year of the movie
4. **IMDb** – IMDb rating of the movie
5. **Rotten Tomatoes** – rating of movie on Rotten Tomatoes
6. **Netflix** – binary value 0 or 1 depending on its availability on Netflix
7. **Hulu** – binary value 0 or 1 depending on its availability on Hulu
8. **Prime video** – binary value 0 or 1 depending on its availability on Prime Video
9. **Disney+** – binary value 0 or 1 depending on its availability on Disney+
10. **Directors** – name of directors of the movie
11. **Genres** – genre of the movie
12. **Country** – country where movie is made
13. **Language** – audio language of the movie
14. **Runtime** – runtime of movie (in minutes)

Why this Dataset?

We have selected this data to analyse the details of various movies that are streaming on various online television platforms like **Netflix, Prime Video, Disney+ and Hulu**. We know in recent years there is a huge shift of audience from normal Televisions to smart TVs/smart Phones providing online streaming platforms where people can watch thousands of movies, series and shows as per their interests and convenience. Also, in the last 6-8 months there is a significant increase in the subscribers of various OTT platforms. That's why we realised the importance of analysing details of various movies that are being currently streaming on these platforms. Analysis will give us the information about these movies in terms of their **ratings, releasing year, age rating, genre, language,**

country of production and runtime. It will provide us the details of **most popular genres** and **availability of movies** on different platforms.

Functionalities and Details of various Analyses performed

First of all, we have cleaned the dataset. **Cleaning dataset** includes

- Replacing nan values of age by '13+'
- Replacing nan values of IMDb by 0.0
- Replacing nan values of Rotten Tomatoes rating by 0%
- Replacing nan values of Directors, Genres, Country and Language by 'unknown'
- Deleting of all the records where runtime was nan
- Deleting entire unnamed first column due to its insignificance and avoiding confusion.
- Deleting column titled 'Type' because of its insignificance as its value is 0 for all records.

Various analyses:

- Finding movies which are suitable for all age groups or family friendly movies.
- Finding movies which are restricted or suitable only for 18+ age group.
- Finding movies with maximum runtime and minimum runtime.
- Finding movies(s) which is/are available Netflix, Prime Video as well as on Disney+.
- Finding movies having highest IMDb ratings.
- Finding all the movies available with Hindi audio.
- Finding all Indian movies on various OTT platforms.
- Finding Indian movie with highest IMDb rating.
- Pie chart to find movie share between various online platforms.
- Finding oldest movie available online.
- Finding all 'Action' movies.
- Finding all 'Animated' movies.
- Finding all Indian comedy movies.
- Finding number of movies for each country.
- Finding number of movies under each genre.

Graphic visualization of Data:

We have visualized the data of movies that are released after 2015 by finding the distribution of movies among various platforms.

Next, we have analysed the movies on the basis of genres on different OTTs.

We have plotted a pie chart for each platform showing the percentage of movies for the top genres and a bar graph for average IMDb of top genres in movies.