Capstone Project-3 Exploratory Data Analysis(EDA) using Python

Netflix movies titles dataset

Final Project Summary: Netflix TV Shows and Movies Dataset Analysis

This project involves a comprehensive analysis of the Netflix TV shows and movies dataset, aiming to extract meaningful insights and visualize various aspects of the data. The dataset is sourced from Kaggle and includes information about Netflix's content offerings.

Datasource link: Netflix Movies and TV Shows (kaggle.com)

Modules used:

1.Numpy-Numerical operations

2.Pandas-Data Manipulation & Cleaning process

3.Matplotlib-Visualization

4.Seaborn-Visualization

Key Steps and Findings:

1.Data Loading and Initial Exploration:

*The dataset was loaded using Pandas, and initial exploration involved printing the first and last few rows to understand its structure.

2.Data Cleaning:

* Missing values were handled by filling them with appropriate defaults (e.g., 'TV-MA' for ratings and 'United States' for country).

* The column 'listed_in' was renamed to 'genre' for clarity.

3. Descriptive Statistics and Dataset Information:

* The dataset's info and descriptive statistics were printed, focusing on the `release_year` column to understand the distribution of release years.

*The dataset contains information from 1966 to 2021.

4.Content Type Analysis:

*The distribution of movies and TV shows was analyzed, revealing the balance between these two types of content on Netflix.

5.Null Values Check:

*A check for null values was performed, and appropriate replacements were made to ensure data integrity.

<u>Visualizations and Insights:</u>

1.Mean Release Year:

*The mean release year was calculated and visualized using a horizontal bar plot, indicating the average age of the content on Netflix.

2.TV Shows vs Movies on Netflix:

*A bar chart showed the distribution of TV shows and movies, highlighting Netflix's content strategy.

3.Type of Shows Percentage:

*A pie chart displayed the percentage distribution of TV shows and movies, providing a clear view of the content mix.

4.Netflix Rating Distribution:

*A count plot visualized the distribution of different ratings, offering insights into the target audience for Netflix content.

5.Top 5 Netflix Country Distribution by Type:

*A bar graph illustrated the distribution of movies and TV shows in the top 5 countries, aiding in understanding regional content preferences.

6.Top 10 Genres of Movies and TV Shows:

*A bar chart highlighted the most popular genres, revealing viewer preferences and trends.

7.Best Time to Release a Movie/TV Show:

*A heatmap visualized the best time to release content based on the month and year added, helping optimize release schedules.

Conclusion

This project provided a detailed analysis of Netflix's content, offering valuable insights into the distribution, characteristics, and trends of TV shows and movies available on the platform. The visualizations helped in understanding various aspects of the data, from content type distribution to regional preferences and optimal release times. These insights can inform strategic decisions for content acquisition, production, and marketing on Netflix.