Capstone Project 3

Netflix Shows Dataset Analysis Report

Dataset Link: https://www.kaggle.com/datasets/shivamb/netflix-shows

Modules needed: Numpy, pandas, matplotlib, seaborn

Introduction

This report provides a comprehensive analysis of the Netflix Shows dataset available on Kaggle. The dataset includes information about movies and TV shows available on Netflix, including their titles, release years, genres, and more. The analysis is performed using Python libraries such as Numpy, Pandas, Matplotlib, and Seaborn.

Data Loading and Initial Exploration

Load the File: The dataset is loaded using Pandas.

First 5 Rows of Data: Display the first 5 rows to get an initial understanding of the dataset.

Last 5 Rows of Data: Display the last 5 rows to check the end of the dataset.

Data Cleaning

Cleaning the Data: Handle missing values and null values to ensure the dataset is clean and ready for analysis.

Dataset Information: Print the dataset information to understand the data types and non-null counts.

Data Description: Provide a statistical summary of the dataset, focusing on the release_year column as it contains numerical values.

Dataset Overview

Columns in the Dataset: List all the columns present in the dataset.

Shape of the DataFrame: Display the shape of the DataFrame to understand the number of rows and columns.

Years of Data: Calculate the range of years (1966-2021) covered by the release_year column.

Movies vs TV Shows: Analyze the distribution of movies and TV shows in the dataset.

Handling Missing Values

Check for Null Values: Identify columns with null values.

Replace Null Values: Replace null values in the rating column with “TV-MA” and in the country column with “United States”.

Rename Column: Rename the column listed_in to genre for better clarity.

Data Visualization

Mean Release Year:

Calculate the mean of release_year and visualize it using a horizontal bar plot titled “Netflix Release Year”.

TV Shows vs Movies: Create a bar chart to show the distribution of TV shows and movies on Netflix.

Type of Shows Percentage: Use a pie chart to visualize the percentage distribution of different types of shows on Netflix.

Netflix Rating Distribution: Visualize the distribution of Netflix ratings using a count plot.

Top 5 Countries with Highest Movies/TV Shows: Display the top 5 countries with the highest number of movies/TV shows using a pie chart.

Country Distribution by Type: Show the distribution of Netflix content by country and type (movies vs TV shows) using a bar graph.

Top 10 Genres: Visualize the top 10 genres of movies and TV shows.

Best Time to Release Content: Use a heatmap to determine the best time to release movies and TV shows.

Conclusion

This analysis provides insights into the Netflix dataset, including the distribution of content types, popular genres, and release patterns. The visualizations help in understanding the trends and patterns in the data, which can be useful for content strategy and decision-making.