ABOUT THE PROJECT:

Performed exploratory-data-analysis (EDA) for netflix using the following tools: MS-Excel, PowerBI and Python Programming.

I have done the following data analysis on different platforms:

1) Microsoft Excel:

a) Frequency of each content-type (represented on a horizontal bar-graph)

b) Yearly contribution of titles to the netflix content-library. (represented on a pie-chart and donut-chart)

c) Top Peforming Genres (represented on a vertical bar-graph)

d) All types of ratings (represented on a vertical bar-graph)

e) Countries having the most contribution to the netflix content-library (represented on a vertical bar-graph).

f) Monthly contribution of titles to the netflix content-library (represented on a vertical bar-graph).

Added slicers based on follwing parameters:

a) Months

b) Genres

c) Release Year

d) Country

e) Rating

f) Content Type

2) Microsoft PowerBI:

(data cleaning done using Python in this notebook: cleaning_data_through_python.ipynb)

Dashbaord Existing:

a) Overview

b) Genre & Rating Analysis

c) Country Insights

d) Trends

Inside Overview Dashbaord:

a) How many movies and TV-Shows exist in our dataset (visualized using donut-chart)

b) Total no. of contents (including both Movies and TV-Shows) added per year (visualized using Stacked column-chart)

Inside Genre & Rating Analysis Dashboard:

c) Genres existing in our database (visualized using Tree-map)

d) All types of Ratings along with the highest frequency (visualized using Clustered column-chart)

e) Total no. of Genres in each content-type (i.e. Movies and TV-Shows) (visualized using Clustered column-chart)

Inside Country Insights Dashboard:

f) Total no. of Contents contributed by each country (Along with top-performing countries). (visualized using Filled-map)

Inside Trends Dashboard:

g) Yearly Trend of contents-being added to netflix content-library. (visualized using both Line-chart and Area chart)

cards for:

i) Total Movies

ii) Total TV-Shows

iii) Total no. of Titles in Netflix (including both Movies and TV-Shows)

iv) No. of Titles added per year

3) Python Programming AND ML EDA (i.e. Machine Learning EDA):

ML EDA PART:

i) Movie (and / or) TV-Show recommendation (using a GUI (Graphical-User-Interface)) [NOTE: but will recommend only upon existing entries in the dataset]

ii) Predicting whether any new entry is a Movie / TV Show.

(along with brief information about Netflix as a company)

a) Frequency of each content-type (i.e No. of movies v/s No. of tv-shows) (represented on a horizontal bar-graph)

b) No. of releases in each year (represeted on a lineplot)

c) Top 5 Genres generating the most traffic (represented on a vertical bar-graph)

d) Which ratings are most oftenly given to the netflix content-library? (represented on a vertical bar-graph)

e) Which countries have contributed to the netflix-library (using plotly)

f) Who are the top-10 actors by unique content-types?

g) Correlation between content's release-year and year (in which it was added to netflix) by unique titles (using python pandas)

h) Try to predict the content type (Movie vs. TV Show) using features like rating, release_year, and country (using Random Forest)

i) Clustering titles by description (using K-Means and Principal Component-Analysis) to identify thematic groups

j) Listing down these on seperate word-clouds:

j.1) directors

j.2) actors

j.3) titles

4) SQL (Structured Query Language):

DATABASE-NAME: netflix_titles

TABLE-NAME: netflix_titles_details2

(used SQLAlchemy)

a) Distinct Content-types in the netflix dataset

b) Frequency of contents produced by each country

c) Frequency of Rating-types

d) Frequency of each unique genre-type

e) Frequency of each unique actor (in the netflix dataset)

f) Top 5 Years in which most content was added to netflix dataset.

 # PROJECT PERFORMED BY:

Name: Om Satyawan Pathak
Email: omsatyawanpathakwebdevelopment@gmail.com (or) omsatyawanpathakgit@gmail.com
Github Repository Link: https://github.com/omsatyawanpathakgit/netflix-data-analysis.git