**Netflix Content Analysis and Visualization (using Python)**

**Project Overview**

* **Title:** Netflix Content Analysis and Visualization using Python
* **Author:** Saloni Pal
* **Date:** 18th July 2024

**Introduction**

**Objective:**

* Analyze the Netflix dataset to gain insights into their content offerings, trends, and focus areas using Python libraries like pandas and numpy.
* Develop a data visualization to showcase findings using matplotlib.

**Scope:** This document covers the data sources, data wrangling, visualization, and key insights derived from the analysis.

**Data Sources**

This data source is taken from Kaggle. [Link](https://www.kaggle.com/datasets/shivamb/netflix-shows/discussion?sort=hotness)

**Data Preparation**

* **Data Wrangling:** Imported the dataset into a Pandas dataframe.
* **Data Cleaning:**
  + Handled missing values and changed data types where necessary.
  + Extracted genres and casts
* **Data Analysis:**
  + Performed EDA
  + Calculated the content-type distribution.
  + Identified the top genres, directors, actors, and companies in the dataset.

**Data Visualization**

* + Showed Content-Type Distribution [Movies vs. TV Shows], Top directors, actors, and companies using Bar Chart
  + Trend in Movies vs. TV Shows over Years using Line Chart
  + Used Pie Chart for Genre and Rating Distribution.

**Key Insights**

* The Audience prefer Movies (6131 count) over TV Shows (2676 count).
* The top Genres are International Movies and Drama.
* The top director is Rajiv Chilaka with 19 movies and the top actor is Anupam Kher with 39 movies.
* Top 5 Actors from US are Jeff Dunham, Kevin Hart, Kraig Sechler Actor, Samuel West, Jim Gaffigan.
* Top 5 Production Companies are country United States, India, United Kingdom, Japan, and South Korea.
* TV-MA , TV-14 , and TV-PG are the top ratings.

**Learnings**

* **Data Wrangling and cleaning:** Gained practical experience in handling real-world data issues such as handling missing values, converting data type and extraction.
* **Python Libraries:** Developed hands-on experience in using python libraries such as Pandas, Numpy, Matplotlib, Seaborn, etc.
* **Data Analysis:** Improved ability to analyze large datasets to extract meaningful insights and trends.
* **Data Visualization:** Learned best practices for designing user-friendly and interactive Visualization using matplotlib and seaborn.

**Recommendations**

* **Content Strategy:**
  + Focus on Movies: Given the audience's preference for movies over TV shows, Netflix should continue to invest in producing and acquiring high-quality movies.
  + Diversify Genres: While International Movies and Drama are the top genres, Netflix should consider expanding its offerings in other genres to cater to diverse audience preferences.
  + Collaborate with Top Creators: Partner with top directors like Rajiv Chilaka and actors like Anupam Kher to produce content that appeals to a broad audience.
* **Targeted Marketing:**
  + Focus on US Audience: Given the top 5 actors and production companies are from the US, Netflix should focus on marketing its content to the US audience.
  + Target Specific Age Groups: With TV-MA, TV-14, and TV-PG being the top ratings, Netflix can target specific age groups with content that caters to their preferences.
* **Content Acquisition:**
  + Acquire Content from Top Production Companies: Netflix should consider acquiring content from top production companies like those from the US, India, UK, Japan, and South Korea.
  + Expand International Content: Given the popularity of International Movies, Netflix should continue to acquire and produce content that appeals to a global audience.