# Netflix Data Analysis and Visualization

## ❖ Project Overview

**Objective**:
To analyze the content available on Netflix using exploratory data analysis and visualizations, uncovering trends in genres, countries, release years, and content types (Movies vs TV Shows).

**Project Type**: Data Analytics / Visualization

**Domain**: Entertainment / Streaming Media

**Tools & Technologies Used**:

- **Languages**: Python
- **Libraries**: Pandas, NumPy, Matplotlib, Seaborn, Plotly
- **Tools**: Jupyter Notebook / VS Code, Tableau (optional)

**Dataset Source**:
Netflix dataset available from Kaggle

**Dataset Description**:

- Rows: 8807
- Columns: 12
- Key Columns:
    - `show_id`, `type`, `title`, `director`, `cast`, `country`, `date_added`, `release_year`, `rating`, `duration`, `listed_in`, `description`

## ❖ Data Understanding & Cleaning

**1. Data Loading**:

```
df = pd.read_csv('netflix_titles.csv')
df.shape  # (8807, 12)
```

**2. Null Value Handling**:

- Columns like `director`, `cast`, and `country` had missing values.
- Replaced nulls with "Unknown" or used forward fill.

**3. Data Type Corrections**:

- `date_added` converted to datetime format.
- Split `duration` into numerical (`duration_int`) and type (`duration_type`).

**4. Sample Data Overview**:

```
| Title               | Type  | Country    | Release Year | Duration |
Genre         |
|--------------------|-------|-----------|-------------|----------|------
---------|
| Stranger Things    | TV    | United States | 2016      | 3 Seasons| Sci-
Fi, Drama |
| Bird Box           | Movie | United States | 2018      | 124 min  |
Thriller       |
```

**5. Feature Engineering**:

- Created `year_added`, `month_added` from `date_added`
- Extracted `main_genre` from the `listed_in` column

## ❖ Exploratory Data Analysis (EDA)

**1. Content Type Distribution**:

- ~70% Movies, ~30% TV Shows
  (Pie chart visualization)

**2. Top Countries Producing Content**:

- US, India, UK, Canada, Japan
  (Bar chart)

**3. Most Common Ratings**:

- TV-MA, TV-14, R, PG-13
  (Histogram)

**4. Content Over Time**:

- Netflix has consistently added more titles each year, peaking around 2019
  (Line chart: Number of titles by year)

**5. Frequent Directors & Actors**:

- **Top Director**: Rajiv Chilaka (Indian Animation)
- **Popular Cast Members**: Anupam Kher, Rami Malek
  (Word Cloud or Bar Chart)

**6. Genres Breakdown**:

- Top Genres: Dramas, Comedies, Documentaries, International Movies
  (Multi-bar chart of genres)

## ❖ Visualization Insights

**1. Yearly Addition of Content**

```
df['year_added'].value_counts().sort_index().plot(kind='bar')
```

- □ Netflix ramped up production significantly from 2016 onwards.

**2. Movie Duration Distribution**

```
df[df['type'] == 'Movie']['duration_int'].plot(kind='hist')
```

- Most movies are between 80 to 120 minutes.

**3. TV Show Seasons Distribution**

- 1–2 season TV shows dominate the catalog.

**4. Country vs Genre Heatmap**

```
sns.heatmap(pd.crosstab(df['country'], df['main_genre']), cmap='Reds')
```

- US dominates in almost every genre; India shows high content in dramas and romance.

**5. Ratings Trend by Year**

```
df.groupby('release_year')['rating'].value_counts().unstack().plot()
```

- TV-MA and TV-14 have become increasingly common in recent years.

## ❖ Conclusion & Recommendations

## Key Insights:

- **Movies dominate** the Netflix platform, but **TV Shows have grown rapidly**.
- **US and India** are the top countries for content production.
- **Drama** is the most prevalent genre across all types.
- Netflix has **increased content additions** each year post-2016.
- Most content is geared toward **mature audiences (TV-MA/TV-14)**.

## Business Recommendations:

1. **Localization Opportunity**:
    - Invest in regional content, especially for countries like India, Brazil, and South Korea.
2. **Diverse Genres**:
    - Increase content in underrepresented genres like horror, animation, or sci-fi.
3. **TV Show Expansion**:
    - Encourage multi-season development for top-performing shows.
4. **User Personalization**:
    - Use genre + duration preferences to tailor UI/UX recommendations.
5. **Strategic Partnership**:
    - Collaborate with trending directors/actors to attract their fan base.

## Future Work:

- Perform sentiment analysis on user reviews or social media data for Netflix originals.
- Use clustering to group similar shows for better recommendations.
- Explore predictive analytics for forecasting viewer trends.