**PROJECT EXPLANATION**

A detailed guide on how to perform EDA on the Netflix dataset, demonstrating the process step-by-step.

#### 1. **Loading the Data**

The journey begins with loading the dataset into a Pandas DataFrame. This step is crucial as it allows for efficient manipulation and analysis of the data.





Once the dataset is loaded, it's important to understand its structure and content.

#### 2. **Understanding the Structure**

Begin by exploring the first few rows of the dataset and checking the data types of each column. This gives a preliminary overview of the information available.



Common columns in the Netflix dataset include:

- **Title**: The name of the movie or TV show.

- **Genre**: The genre(s) of the title (e.g., Action, Comedy).

- **Director**: The director of the title.

- **Cast**: Main actors featured.

- **Country**: The production country.

- **Release Year**: The year the title was released.

- **Rating**: Audience ratings.

- **Duration**: The length of movies or the number of seasons for TV shows.

- **Description**: A brief synopsis of the title.

- **Language**: The primary language(s) of the title.

- **Date Added**: The date the title was added to Netflix.

- **Type**: Specifies whether the content is a movie or a TV show.

#### 3. **Data Cleaning and Preprocessing**

Data cleaning is a critical step in EDA. Before analysis, it’s essential to handle missing values and ensure data types are correct. Start by checking for any missing values across the dataset.



Depending on the amount and significance of missing data, you might choose to fill or drop these entries.Next check for duplicates.



Dropping the duplicates as they are not significant.





Convert necessary columns to appropriate data types. For instance, ensure that duration is numeric by stripping out text and converting it to an integer.



#### 4. **Exploratory Data Analysis**

Conducted an exploratory analysis of Netflix’s content, uncovering insights on key series like *House of Cards*, release trends over the years, regional exclusives (e.g., India), director impact, and genre preferences (e.g., UK comedy). The analysis also examined actor prominence (e.g., Tom Cruise), Netflix’s rating systems, content duration, and identified the country with the highest number of TV shows.

#### 5.**Feature Engineering**

Creating new columns for further analysis.Here creating a date column from Release\_year column.





#### 6. **Univariate Analysis**

Univariate analysis focuses on one variable at a time. Start by exploring the distribution of the release years to identify trends over time.



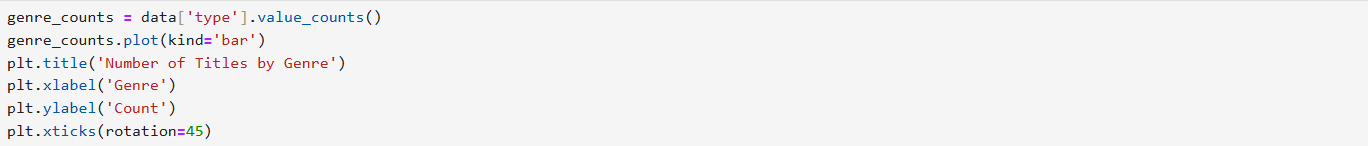




This bar graph provides insight into how many titles have been released in each year, revealing trends in content availability.

#### 7. **Genre Analysis**

Next, analyze genre popularity. A bar chart can effectively show how many titles fall under each genre.



This visualization helps identify which genres dominate Netflix's library, guiding content acquisition strategies.