Netflix Data Analysis Project - Summary & Code Explanation

A. Project Summary

This project analyzes a dataset of Netflix Movies and TV Shows to understand trends in content type, release year

The key goals:
- Clean and understand the dataset
- Identify content trends (e.g., type, year, rating)
- Visualize the top genres and producing countries
- Use simple, effective plots to communicate insights

B. Libraries Used and Their Purpose

pandas: Data loading, cleaning, transformation, and analysis
matplotlib.pyplot: Creating basic visualizations (bar, pie, line charts, etc.)
seaborn: Simplified and beautiful statistical visualizations (countplots, etc.)
collections.Counter: Counting and analyzing genre frequencies

C. Graphs Used

Bar Chart: Show frequency (e.g., missing values, countries, ratings)
Count Plot: Compare counts of Movies vs TV Shows
Histogram: Show distribution of content releases over years
Pie Chart: Show proportion of genres or content types

D. Detailed Code Explanation

1. Importing Libraries:
- pandas, matplotlib.pyplot, seaborn, Counter from collections

2. Read CSV File:
- df = pd.read_csv('netflix_titles.csv')

3. Dataset Overview:
- df.shape, df.info(), df.head()

4. Missing Value Analysis:
- df.isnull().sum() -> bar plot

5. Drop Critical Nulls:
- df.dropna(subset=['title', 'date_added', 'release_year'])

6. Count Plot (Movies vs TV Shows):
- sns.countplot(...)

7. Histogram of Releases:
- sns.histplot(...)

8. Top 5 Genres (Pie Chart):
- Use Counter on listed_in

9. Top 10 Countries:
- df['country'].value_counts().head(10) -> barplot

10. Top 10 Ratings:
- df['rating'].value_counts().head(10) -> bar chart