# ■ Netflix Dataset Mini Project

This is your first project using Pandas & NumPy. It covers dataset cleaning, transformation, and building a mini-dashboard.

## Step 1: Import & Load Dataset

```
import pandas as pd
df = pd.read_csv("netflix_titles.csv")
print(df.shape)
print(df.head())
```

## Step 2: Explore the Data

```
print(df.info())
print(df.describe())
print(df.isnull().sum())
```

## Step 3: Handle Missing Values

```
df = df.dropna(subset=['title'])
df['director'] = df['director'].fillna("Unknown")
df['country'] = df['country'].fillna("Unknown")
df['cast'] = df['cast'].fillna("")
```

## Step 4: Data Type Conversion

```
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
df['year_added'] = df['date_added'].dt.year
df['month_added'] = df['date_added'].dt.month
```

## Step 5: Clean & Transform Columns

```
df.columns = df.columns.str.lower().str.replace(" ", "_")
df['content_type'] = df['type']
df[['duration_value', 'duration_unit']] = df['duration'].str.extract(r'(\d+)\s*(\w+)')
df['duration_value'] = pd.to_numeric(df['duration_value'], errors='coerce')
```

## Step 6: Handle Duplicates

```
print("Duplicates:", df.duplicated().sum())
df = df.drop_duplicates()
```

## Step 7: Analysis - Movies vs TV Shows

```
print(df['content_type'].value_counts())
```

## Step 8: Analysis - Top 10 Countries

```
print(df['country'].value_counts().head(10))
```

## Step 9: Analysis - Content by Year

```
print(df['year_added'].value_counts().sort_index())
```

## Step 10: Analysis - Longest Content

```
print(df[['title','duration_value','duration_unit']].sort_values(by='duration_value', ascending=
```

## Step 11: Analysis - Top 10 Directors

```
print(df['director'].value_counts().head(10))
```

## Step 12: Interactive Mini Dashboard

```
def dashboard_menu():
    print("\n--- NETFLIX MINI DASHBOARD ---")
    print("1. Movies vs TV Shows")
    print("2. Top 10 Countries")
    print("3. Content by Year")
    print("4. Longest Content")
    print("5. Top 10 Directors")
    print("6. Exit")

while True:
    dashboard_menu()
    choice = input("Enter your choice: ")
    if choice == "1":
        print(df['content_type'].value_counts())
    elif choice == "2":
        print(df['country'].value_counts().head(10))
    elif choice == "3":
        print(df['year_added'].value_counts().sort_index())
    elif choice == "4":
        print(df[['title','duration_value','duration_unit']].sort_values(by='duration_value', as
    elif choice == "5":
        print(df['director'].value_counts().head(10))
    elif choice == "6":
        print("Exiting Dashboard...")
        break
    else:
        print("Invalid choice!")
```