+ Code + Text

```
pip install pandas
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

```
print("Initial Shape:", df.shape)
```

```
Initial Shape: (8807, 12)
```

```
print("\nMissing values:\n", df.isnull().sum())
```

```
Missing values:
 show_id           0
type             0
title            0
director       2634
cast            825
country         831
date_added       10
release_year      0
rating            4
duration          3
listed_in         0
description       0
dtype: int64
```

```
df = df.drop_duplicates()
```

```
df['country'].fillna('Unknown', inplace=True)
df['rating'].fillna('Unknown', inplace=True)
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
df = df.dropna(subset=['director', 'cast'])
```

```
<ipython-input-12-9aa4cb5c6665>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignm
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting valu

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].me


  df['country'].fillna('Unknown', inplace=True)
<ipython-input-12-9aa4cb5c6665>:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignm
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting valu

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].me


  df['rating'].fillna('Unknown', inplace=True)
```

```
df.columns = df.columns.str.strip().str.lower().str.replace(' ', '_')
```

```
text_cols = ['type', 'title', 'director', 'cast', 'country', 'rating', 'listed_in']
for col in text_cols:
    df[col] = df[col].astype(str).str.strip()
```

```
print("\nCleaned Shape:", df.shape)
```

```

Cleaned Shape: (5700, 12)
```

```
print("\nData types:\n", df.dtypes)
```

```

Data types:
 show_id                 object
type                   object
```

```
title                   object
director                object
cast                    object
country                 object
date_added      datetime64[ns]
release_year             int64
rating                  object
duration                object
listed_in               object
description             object
dtype: object
```

```
df.to_csv('cleaned_netflix_titles.csv', index=False)
```