1. **Import libraries**

* This line imports the Pandas,Numpy library, which helps us handle and clean tabular data (like CSV files)

1. **Read the csv file using pandas**

 This step is done to load Netflix dataset from the data/raw/ folder into a DataFrame called df.

 df is at the present working table of data.

1. **Finding duplicate values**

 df.duplicated() finds any **exact copies of rows**.

 It shows how many rows are repeated.

 duplicate\_rows.shape[0] gives the number of duplicates.

1. **Removing duplicate rows**

* df = df.drop\_duplicates() / This removes all exact duplicate rows from the dataset.

1. **Finding missing values**

 df.isnull() marks cells that are empty ().

 .sum() counts how many missing values are in each column.

1. **Calculating the percentage of missing values**

 This calculates the **percentage** of missing values in each column.

 (missing / len(df)) gives the fraction, multiplied by 100 for percent.

1. **Summary of missing value**

This creates a new table called missing\_summary that shows:

* Total missing values
* Percentage of missing values (rounded to 2 decimals)

1. **Showing columns that has missing value**

 This only displays columns that **have missing data**.

 It hides the ones that are already complete.