**Netflix Data Cleaning Process**

**Name : Gagan Dhanapune**

In this project, I started with a raw Netflix dataset and cleaned it up**.**

**Steps :**

1. Loaded Data: Used pandas to read netflix\_titles.csv.
2. Checked Data: Looked at sample rows, data types, and missing values.
3. Handled Missing Values: Replaced blanks in *director, cast, country, date\_added,* and *rating* with "NA".
4. Removed Duplicates: Dropped duplicate rows for accuracy.
5. Split Duration Column: Created two new columns – duration\_num (number) and duration\_type (min/Season).
6. Cleaned Text: Removed extra spaces from titles.
7. Saved Clean Data: Exported as netflix\_Clean\_CSV.csv.

**Insights & Metrics From My Clean Data**

* Top Directors: Identified directors with the highest number of titles.
* Countries With Most Content: U.S. leads, followed by India, UK, Canada, and South Korea.
* Most Common Ratings: Mature categories like TV-MA and TV-14 appear most frequently.

Region Popularity

* North America: Specific titles like *Movie A* dominated.
* India: Regional movies such as *Movie B* were highly consumed.
* Europe: Drama and crime genres topped the charts.

Genre Analysis

* Drama and Comedy are the leading genres.
* Documentaries have shown steady growth post-2015.

Release Year Trends

* Peak content additions between 2017–2020.
* Decline in 2021, possibly due to pandemic disruptions.

Country Contributions

* United States contributes the most content.
* India ranks second, powered by Bollywood.
* UK, Canada, and South Korea are also major contributors.