**Date:**20.09.2025 to 4.10.2025  
**Milestone 1: Netflix Dataset**

**1. Data Loading**

* Imported the **pandas** library.
* Loaded the dataset **netflix\_titles.csv** into a DataFrame.
* Displayed the dataset to confirm successful loading.
* Verified structure using:
  + **df.head()** → showed the first 5 rows.
  + **df.info()** → provided column details, data types, and null counts.

**Output Insight:**

* The dataset contains movie and TV show details like title, director, cast, country, release year, rating, and genres.

**2. Data Cleaning**

* **Missing Values**:
  + Replaced null values in **director** with "**Unknown".**
  + Replaced null values in **rating** with "**Unknown".**
* **Duplicates**:
  + Checked for and removed duplicate records.
* **Data Formatting**:
  + Converted **listed\_in** (genres) into lists by splitting on , .
  + Converted list-type entries into tuples to maintain consistency.

**Output Insight:**

* After cleaning, the dataset became consistent with no duplicates.
* All missing values in key columns were handled.

**3. Exploratory Data Analysis (EDA) – Initial Checks**

* Used .**head()** and .**info()** to explore:
  + Distribution of rows and columns.
  + Column data types (strings, integers, lists).
  + Presence of null values.

**Output Insight:**

* Confirmed that dataset was ready for further analysis after basic cleaning.

**Conclusion**

Milestone 1 successfully completed the **data preparation phase**:

* Dataset was loaded correctly.
* Missing values and duplicates were handled.
* Columns were reformatted for consistency.
* Dataset is now ready for detailed Exploratory Data Analysis in the next milestone.