**📌 Report on Data Cleaning & Preprocessing – 1000 Books Dataset**

**📍 Overview**

The **1000 Books Dataset** was cleaned and preprocessed to ensure accuracy, consistency, and usability for further analysis. The steps included **handling duplicates, standardizing column names, and converting data types**.

**✅ Step 1: Loading the Dataset**

* The dataset was successfully loaded into Pandas.
* It contained **1000 rows** and **6 columns**.
* No missing values were detected.

**✅ Step 2: Handling Duplicates**

* Checked for duplicate rows using df.duplicated().sum().
* **No duplicates were found**, so no records were removed.

**✅ Step 3: Standardizing Column Names**

* Column names were **converted to lowercase** and **spaces were replaced with underscores** for consistency.
* Example: "Star Rating" → "star\_rating"

**✅ Step 4: Converting Data Types**

* **Star Rating** was converted from text to numeric using:

python

df["star\_rating"] = pd.to\_numeric(df["star\_rating"], errors="coerce")

* Other columns were checked for incorrect data types, but most were already in the correct format.

**✅ Final Outcome**

* The dataset is now **clean, structured, and ready for analysis**.
* The processed dataset can be used for **visualization, recommendations, and machine learning models**.