Date : 19/02/2026

Global Sales Data – Python Task Report

In this task, Data cleaning of Global Sales dataset was done using Python in a Jupyter Notebook. The main objective of this work was to improve the quality of the data.

The dataset contains **1,000 rows and 6 columns**. It includes columns such as **transaction ID, transaction date, store location, sales amount, currency, and server timestamp**. Python libraries like Pandas and NumPy were used for all data cleaning and processing tasks.

First, the dataset was loaded into a Pandas **DataFrame** to understand its structure. The total number of rows and columns was checked using the **df.shape** function. Column names, data types, and missing values were reviewed using df.info(). Basic summary verified by using **df.describe()**.

After that, the dataset was checked for missing values in all columns. Duplicate records were also checked . There were no missing values as well as duplicate rows.

The **store_location** column also required cleaning because some text values were not displaying correctly due to encoding issues. To fix this, the column was first encoded using **latin1** format and then decoded into **utf-8** format

Next, the **transaction date** column was cleaned and converted into a standard datetime format using Python.

The **server timestamp** column required cleaning as it contained timezones like **PST**, **EST**, and **UTC**, which can cause issues during datetime conversion. To fix this, the column was converted to string format and the timezone names were replaced with their numeric values, such as **PST to -08:00, EST to -05:00,** and **UTC to +00:00**. This made the timestamp values consistent and easy to convert into datetime format.

The **server timestamp** column was first converted into string format to allow text operations. After that, the numeric **time zone** value **+00:00** was replaced with **UTC** to make the **time zone** easier to read

Finally, the dataset was checked again after all cleaning steps were completed. This was done to make sure that no data was lost and that all columns had the correct data types. The final cleaned dataset, with **1,000 rows and 6 columns**, is now well-organized and in correct format.