

Day 6 - Report

Task - Encoding and Timezones

Day - Thursday

Date - 19th Feb 2026

Dataset – Global_Sales_Messy_Data.

As per provided dataset I have completed yesterday's task on that dataset.

Dataset Overview -

This dataset shows **Sales records** from stores of different **Countries**. Each row has different sales records along with Date, Time and Place.

- **Columns** of the dataset are Transaction_ID, Transaction_Date, Server_Timestamp, Store_Location, Sales_Amount, Currency.
- Checked for the column names and basic data by using **MS-Excel** first.
- As per visual observation dataset doesn't look like it has null values, but data is messy and not cleaned, as its name says.
- It contains **1000 rows** and **6 columns**.
- Imported dataset into **Jupyter Lite** for the data analysis and cleaning purpose.
- Imported necessary libraries and checked the structure of the data by using basic commands.
- Checked for the null and duplicated values, also checked the data type of each column.

- Changed the **datatype** and **date format** of the **Transaction_Date** column as it has variable date formats and data type as objects.
- Also formatted the **Server_Timestamp** column into one format that is **UTC**. So it would be easy to understand and analyze data.
- The column **Store_Location** has **encoded values** which represent the different country names. Cleaned it to a human readable format.
- After cleaning the whole dataset saved it with the **.xlsx** format because when I saved it as **.csv** and opened it through **Excel** the **Store_Location** column did not get cleaned and showed the **old values**.