

Day – 5

Task – The Regex Nightmare

Day – Tuesday

Date – 17th Feb 2026

Time – 7:00 AM – 10:00 AM

Dataset – Salary_data

Source – The dataset was downloaded from Kaggle.

- As per the task requirement, searched lot of Datasets but I didn't find the exact dataset which have combined columns.
- So, to perform task using Regex I decided to merge columns in **Excel manually**, and I merged the **4 Columns** in one column. For further task.
- This dataset is about the Job salaries data of US.
- Before merging the original dataset, it has the **6 Columns**, after merging it becomes **3 Columns** and **376 Rows**.
- **Columns** are Age, Gender, Education_Level, Year_of_Experience, Job_title, Salary.
- Uploaded the dataset in **Jupyter Lite** and started performing operations.
- I tried to clean the dataset using Regex operations. The dataset had combined column values, so I attempted to split them into separate columns. But while applying the regex, I encountered some errors. I tried debugging the issue by checking the pattern but due to time bound I was unable to resolve the error.