Project Requirements (Your Tasks)

Hello,

I need you to clean, transform, and analyze the given datasets (employee_data.csv and sales_data.csv).

Your final output should be a well-structured dataset that I can use for reporting and analysis.

Here are the tasks you need to complete:

1. Data Cleaning

- Check for missing values in Salary, Performance_Rating, and Product. Handle them properly.

- Identify and remove duplicate rows if they exist.

- Ensure that Salary and Quantity have only positive values.

2. Data Type Conversion

- Convert Date_of_Birth and Date_of_Joining columns from text to date format.

- Convert Salary column to a numeric format.

3. Add New Columns

- Create a new column Age by calculating employees' age from Date_of_Birth.

- Create a Status column based on Performance_Rating with the following rules:

  - "Excellent" to "Good"

  - "Good" to "Average"

  - "Average" to "Needs Improvement"

4. Data Merging

- Merge the employee_data.csv with sales_data.csv using Employee_ID.

The final dataset should include all employee details along with their sales data.

5. Formatting and Validation

- Check for any inconsistent or incorrect data (e.g., invalid dates, wrong salary values).

- Ensure all column names are properly formatted (no extra spaces, correct spelling).

6. Data Analysis

- Calculate the total sales made by each employee.

- Show the distribution of employee ages (a simple count of employees in different age ranges).

- Create a pivot table to display total sales per region and product.

Final Deliverable:

- A cleaned and transformed CSV file with all required changes.

- A short report explaining how you handled missing values, data types, and transformations.

- Basic visualizations (if possible) for age distribution and sales per region.

Once done, please submit the final CSV file and the report. Let me know if you have any questions.

Thanks!