

HR Report

Puranjoy Patra,

Sending you these tasks following up on our call.

The HR dataset can be found in the file "attendance_data.csv". Your tasks are divided into two parts: data cleaning and data analysis.

Task 1: Data Cleaning

1. Check for duplicates in the dataset and remove them.
2. Standardize the date values to the format YYYY-MM-DD and extract the month name and day type from them.
3. Remove any extra characters, such as special characters, from the employee ID values. Some IDs may contain a '@' character at the end, which can be cleaned and brought to a common format.
4. Standardize the capitalization of names. Convert all names to title case, which means capitalizing the first letter of each word.
5. Map the corresponding values in the status column with the given abbreviations:
 - Work From Office --> WFO
 - Work From Home --> WFH
 - Birthday Leave --> BL
 - Menstrual Leave --> ML
 - Paid Leave --> PL
 - Sick Leave --> SL
 - Weekly Off --> WO

Task 2: Ad Hoc aka Analysis

1. What is the total count of distinct employee names within the dataset?
2. Calculate the work-from-home (WFH %) percentage in the month of May.
3. Determine which day of the week had the highest attendance percentage in the month of June.
4. Find out the number of employees who had a WFH percentage greater than 10% in the month of April.

You are free to use any tool of your preference, such as Pandas, Excel, Power BI, etc., to complete this task.

If you have any follow-up questions on the task, you can ask your seniors here: [\[discord link\]](#)

But they are busy, try to solve it on your own as much as possible.

Good luck with your task!

Best regards,



Hem

Head of Data Analytics | AtliQ Technologies Pvt. Ltd.

Website: www.atliq.com