

◆ TASK 15: Data Analysis Using Pandas

Tools:

- Python
- Pandas
- Jupyter Notebook / VS Code

Hints / Mini Guide:

1. Create data_analysis.py.
2. Load CSV dataset into Pandas DataFrame.
3. Explore dataset using head, info, describe.
4. Handle missing values.
5. Filter and sort data.
6. Group data and calculate aggregates.
7. Add new calculated columns.
8. Export cleaned data to CSV.
9. Interpret insights.

Deliverables:

- Analysis script
- Cleaned dataset

Final Outcome:

- Intern can analyze real datasets.

Interview Questions Related To Above Task:

- What is DataFrame?
- Handling missing data?
- GroupBy use cases?
- Pandas vs Excel?
- Performance limitations?

Task Submission Guidelines

-  **Time Window:**

You can complete the task anytime between 10:00 AM to 10:00 PM on the given day. Submission link closes at 10:00 PM

-  **Self-Research Allowed:**

You are free to explore, Google, or refer to tutorials to understand concepts and complete the task effectively.

-  **Debug Yourself:**

Try to resolve all errors by yourself. This helps you learn problem-solving and ensures you don't face the same issues in future tasks.

-  **No Paid Tools:**

If the task involves any paid software/tools, do not purchase anything. Just learn the process or find free alternatives.

-  **GitHub Submission:**

Create a new GitHub repository for each task.

Add everything you used for the task — code, datasets, screenshots (if any), and a short README.md explaining what you did.

Submit Here:

After completing the task, paste your GitHub repo link and submit it using the link below:

-  [\[Submission Link\]](#)

