Softsinc Internship Program 2025

Week 2 – Data Manipulation & Visualization

◆ Topic 1: Deep Dive into OOP & File Handling

📚 Concepts:

- Advanced inheritance and polymorphism
- Composition vs. inheritance
- Context managers (`with` statement)
- Working with binary/text files

🛠 Tasks:

1. Expand your `User → Intern → Mentor` hierarchy to include:
    - `Admin` and `HR` roles using composition
    - Polymorphic behavior for different roles
2. Implement file logging:
    - Each class action (e.g., login, message sent) logs to a file using context managers
3. Create a reusable file reader/writer module

🎯 Challenge:
Design a role-based logging system:

- Each role logs different types of actions (e.g., Intern logs tasks, HR logs hiring)
- Logs saved to different files (`intern.log`, `hr.log`, etc.) using context managers

---

◆ Topic 2: Data Cleaning with Pandas

📚 Concepts:

- Detecting and handling missing values
- Converting data types (e.g., strings to dates, floats to integers)
- Replacing, filling, and dropping data

🛠 Tasks:

1. Load a messy dataset with missing or inconsistent entries
2. Perform the following:

    - Identify missing data
    - Convert incorrect data types
    - Drop/fill missing rows with explanation
3. Summarize the cleaned dataset (shape, types, basic stats)

🎯 **Challenge:**
Write a reusable `clean_data(df)` function that:

- Detects and fixes missing/invalid entries
- Returns a cleaned DataFrame with a log of what was changed

◆ Topic 3: Data Visualization with Matplotlib & Seaborn

📚 **Concepts:**

- Creating line, bar, scatter, and histogram plots
- Customizing titles, labels, legends
- Seaborn for distribution and categorical plots

 Tasks:

1. Visualize your cleaned dataset using:
    - Bar plot (e.g., category-wise counts)
    - Line chart (e.g., trend over time)
    - Histogram (e.g., age distribution)
2. Use Seaborn to plot:
    - Heatmap for correlation
    - Boxplot for outlier detection

🎯 **Challenge:**
Create a `visualize_data.py` module:

- Takes a DataFrame and generates 3 plots
- Saves plots to a folder `plots/` with timestamped filenames

◆ Topic 4: Reporting & Summarization

📚 **Concepts:**

- Summary statistics (mean, median, std, count)
- Timestamp logging

- Exporting to `.txt` and `.csv`

 Tasks:

1. Write a script that:
    - Generates a summary report of cleaned and visualized data
    - Includes timestamp and file info
2. Save summary to both `.txt` and `.csv`

Challenge:
Automate the entire process:

- Load → Clean → Visualize → Summarize → Export
  Use a main script (`run_pipeline.py`) to manage the flow

◆ Topic 5: Git & GitHub – Continued Workflow

Concepts:

- Pull Requests (PRs)
- Issues and project boards
- Markdown tables, images, and code blocks

 Tasks:

1. Work in `dev` branch for each task
2. Create a PR to `main` after peer review
3. Update README with:
    - Weekly progress table
    - Embedded plot screenshots
    - Links to script files
4. Configure a **GitHub Action** to:
    - Auto-generate a report when `main` is updated
    - Lint and validate your `.py` files using `flake8` or `black`
5. **(New)** Upload your Week 2 progress to **LinkedIn**:
    - Write a short post about what you learned
    - Mention **Softsincs** and use the hashtag **#softsincs**
    - Include screenshots or links to your GitHub project

Challenge:
Make your LinkedIn post engaging and professional—highlight real skills gained, tag Softsincs, and use proper hashtags for visibility.