Internship Program: Data Analysis Task

Task Title: Data Analysis with Python & Dashboard Creation

Introduction:

In this task, interns are required to perform data analysis using a dataset of their choice. They can either use their own dataset or download one from Kaggle or other open data sources. The task involves data manipulation using Python and visualization using Power BI or Tableau. This task will help interns enhance their skills in data processing, analysis, and dashboard creation.

Technologies to Use:

- **Programming Language: Python**
- **Libraries & Tools:**
    - **pandas (for data manipulation)**
    - **numpy (for numerical operations)**
    - **matplotlib / seaborn / plotly (for visualization in Python)**
    - **openpyxl / xlsxwriter (for writing to Excel files)**
    - **Power BI / Tableau (for dashboard creation)**

Task Explanation:

1. **Data Collection:**
    - **Choose a dataset from Kaggle or another open-source platform, or use a self-collected dataset.**
    - **Load the dataset into Python for processing.**
2. **Data Processing & Manipulation:**
    - **Clean the dataset by handling missing values and duplicates.**
    - **Perform data transformations and aggregations using pandas and numpy.**
    - **Save the processed data in CSV or Excel format for further analysis.**
3. **Data Analysis & Visualization:**
    - **Perform exploratory data analysis (EDA) using pandas, matplotlib, seaborn, or plotly.**
    - **Generate meaningful insights from the data.**
4. **Dashboard Creation:**
    - **Use Power BI or Tableau to create an interactive dashboard.**

- **Visualize key metrics and trends from the dataset.**
- **Ensure the dashboard is user-friendly and visually appealing.**

Conclusion:

This task will allow interns to apply their Python programming skills in real-world data analysis. It will help them understand how to clean, manipulate, analyze, and visualize data effectively. Additionally, working with Power BI or Tableau will enhance their skills in creating interactive dashboards for data presentation.

Deliverables:

- **Source code in a GitHub repository (or shared via a zip file).**
- **A short report/documentation explaining the implementation approach.**
- **A Power BI or Tableau dashboard file showcasing the analysis.**
- **A demonstration video (if possible) showcasing the dashboard.**

Good luck! Happy coding! 🚀