Lab 3 | Version Control & Automation

Introduction

GitHub Actions is a powerful feature within GitHub that enables the automation of software workflows. By allowing developers to define custom workflows directly within their repositories, GitHub Actions streamline the integration and deployment process, enhancing productivity, and ensuring consistent quality code.

Practical Exercises

- You're automating ETL (Extract, Transform, Load) pipeline with GitHub Actions to run **daily**, the steps are as follows:
  - Extract data from any free API of your choice
  - Clean and transform the data, and make some insightful aggregations
  - Store the output in a CSV file
- Your tasks:
  - Write some data validation tests (you can add any other tests if needed):
    - Test the API Response (status_code=200)
    - Test the transformations on sample data with same schema of the API data
  - Prepare your dependencies in a **requirements.txt** file
  - Write the extraction, transformation, and loading code in 3 separate files
  - Enhance the debugging with logging functionality in your code
  - Write the YAML file for the workflow

Additional Resources

- [Git Commands Documentation](#)
- [Git Cheat Sheet](#)
- [GitHub Actions - Writing Workflows](#)