Saylani Mass IT Training

Cloud Data Engineering - Hackathon Documentation

Hackathon Project Report

Name: Muneeb Babar

Roll No: CDE:375264

Course: Cloud Data Engineering

This document contains a summary and functional breakdown of the tasks completed as part of the Hackathon project.

Task 1 - TradingView Stock Market Scraper

This task involves building a web scraper using Selenium to extract stock market data from TradingView's 'All Stocks Market Movers' page. The script simulates scrolling and clicking the 'Load More' button until all stock entries are loaded. For each stock, the scraper captures 13 fields: Symbol, Security Name, Price, Change %, Volume, Relative Volume, Market Cap, P/E Ratio, EPS (TTM), EPS Growth, Dividend Yield, Sector, and Analyst Rating.

Functionality:

- Dynamically interacts with the page to reveal all data.
- Extracts and parses financial data from each table row.
- Handles delays using waits and exceptions.
- Exports the final dataset into a structured CSV file.

Task 2 - Sarmaaya Mutual Funds Data Scraper

This task automates data extraction from Sarmaaya.pk's mutual funds listing. It uses Selenium to iterate over paginated tables and collects fund metrics by verifying the visibility of each data cell. Only rows with exactly 11 columns of visible data are recorded.

Functionality:

- Identifies visible table rows and extracts meaningful values.
- Iterates through pagination using the 'Next' button.
- Exports data to a CSV file containing Fund Name, RP, PM, TER, MF, SAM, MTD/YTD Returns, NAV, Date, and AUM.

Task 3 - Documentation and Structuring

This task covers the preparation of clean and modular documentation for the web scraping tasks performed. It outlines each task's purpose, methods used, and functionality. The documentation is formatted into a professional PDF report to be presented for the hackathon submission.

Functionality:

- Generates a styled, structured PDF summarizing the hackathon work.
- Provides clarity and reproducibility of all scraping operations.
- Ensures best practices in script design, output format (CSV), and handling dynamic content.