Project Report: Infinite Analytics - Coca-Cola Product Price Extraction and Analysis

Introduction

In this project, we aimed to extract prices of various Coca-Cola products in Japan from five different websites. The objective was to organize this data into a structured CSV format, implement a logging mechanism to track price changes, store the data in a MySQL database, and expose it through a FastAPI in JSON format for integration with the Sherlock website. The project was developed using VS Code and leveraged various Python libraries such as Selenium, Pandas, MySQL Python Connector, and FastAPI.

Project Scope

The primary focus of this project was to develop a robust data extraction and analysis pipeline for Coca-Cola products. The secondary objective was to make the code flexible and adaptable for extracting prices of other products from the specified websites.

1. Design a web scraping tool to fetch Coca-Cola product prices from five websites.
2. Structure the data into a CSV file with date, product name, and price.
3. Implement a logging mechanism to monitor price changes.
4. Create a MySQL database to store the extracted data.
5. Build a FastAPI to convert the data into JSON format for integration with Sherlock website.
6. Ensure flexibility in the code to accommodate other products and websites.

Implementation Details

1. Web Scraping with Selenium

To extract Coca-Cola product prices from the websites, we utilized Selenium with Chrome driver. The automated browser visited each of the five specified websites and extracted the prices of Coca-Cola products using the provided list of links.

2. Data Reframing with Pandas

The extracted data was organized into a more structured and suitable format using the Pandas library. We arranged the data into a CSV file with columns for date, product name, and price. The CSV format made it easy to handle and analyze the data.

3. Logging Price Changes

To monitor price fluctuations over time, we implemented a logging mechanism. The tool calculated the percentage change in price from the last run and recorded this information in a log file. This log file allowed us to track product price trends efficiently.

4 .MySQL Database Integration

For efficient data management, we transformed the structured data from the CSV file into a MySQL database using the MySQL Python Connector. Storing the data in a database provided a scalable solution for handling large amounts of data and allowed for easy retrieval and analysis.

5. FastAPI for JSON Conversion

To expose the data to the Sherlock website, we built a FastAPI. The API converted the data from the MySQL database into JSON format, making it suitable for integration with the Sherlock website. Companies could access the data through the API endpoint and use it for customer insights and decision-making.

Liquor Websites and Mondelez Products Extraction

In addition to the Amazon web scraping tool, we further expanded the data extraction capabilities to include liquor websites and Mondelez products.

Liquor Websites

We designed a separate web scraping tool using Selenium to extract liquor product prices from multiple websites. This enabled companies to analyze consumer behavior and trends in the alcoholic beverage market.

Mondelez Products

Similarly, we created another tool to fetch prices of Mondelez products. This provided insights into consumer preferences in the confectionery and snack industry.

Conclusion

The successful implementation of the Coca-Cola product price extraction and analysis project showcased the effectiveness of the developed web scraping tool. The tool efficiently fetched Coca-Cola product prices from multiple websites, organized the data into a structured format, and enabled tracking of price changes over time through the logging mechanism.

By leveraging Selenium, Pandas, MySQL Python Connector, and FastAPI, we created a flexible and adaptable solution for extracting and analyzing product prices. The project's success demonstrates its potential applicability beyond Coca-Cola products, as the code can be easily adapted to extract prices of other products from the specified websites.

This project opens up opportunities for companies to gain valuable consumer behavior insights and make data-driven decisions to acquire customers more effectively. The data obtained from the scraping tool can provide significant value in understanding consumer preferences and market trends. Overall, the project contributes to enhancing the data analytics capabilities of businesses, empowering them to stay competitive and make informed decisions for growth and success.