Bovespa Index Composition ETL

By Kevyn A. Marcelino

São Paulo – Brasil

Hello, my name is Kevyn, I’m from São Paulo, Brasil and my cs50p final project is a data fetcher from B3, the main stock exchange in Brazil. It retrieves data from the bovespa index composition available in the website . In short, it’s an index made with the most liquid stocks in Brazil.

The inspiration for this project is that I work at the financial market and like to create my own databases. One challenge I faced is that some crucial data from B3 is hard to find for free, as the exchange only provides the current and next day's data and does not maintain a freely accessible historical archive. To solve this, I decided to create an ETL (Extract, Transform, Load) project. My program scrapes the data, transforms it into a useful format, and saves it into a file. This program is designed to run as a cron job, automatically executing every business day.

This project is pretty straightforward, but the most challenging part was understanding how the website worked **(Change to the b3 site).** Specifically, I had to find a "hidden API" that feeds the site's data in JSON format. Once I located the data source, I needed to figure out how to read it, transform it into a structured format using pandas, and finally save it. For this demonstration, I chose to save the data as an Excel file. However, in a professional setting, I could use this same process to populate a SQL database for more robust and scalable data storage.

**(Change to the code, run and check the file saved)**

Let’s run the program to see how it works and check the saved file.

O melhor seria no final mostrar o arquivo salvo