Aryana Bellamkonda, Mariana Mendez, Caroline Nguyen

DS 2002

24 March 2025

One of the major challenges faced during implementation of this project was finding a reliable API that provided free and clean data. Many APIs either required payment or didn't include historical data that made it difficult to extract meaningful correlations. As a result, we had to change our research topic multiple times before discovering Alpaca, which proved to be the most user-friendly and accessible option. One aspect that was easier than expected was asking the user to pick the format they wanted the merged data set in. We were able to use if and else statements along with the to_fileformat (to_csv) function to convert the data set into the required format.

Aside from finding a suitable API, another challenge that we did not consider was handling date formats across different sources. This process required careful adjustments to ensure proper merging. The stock data used timestamp data while the csv file column was labeled as "Day". We merged the two data sets based on "date" and deleted the redundant timestamp/day columns from the individual data sets.

Given that there seemed to be a negative correlation between the data in the CSV file (about covid cases) and the data fetched from the API (about stock prices for an airline), this is very useful information for people wishing to invest in companies during uncontrollable situations, such as the COVID-19 pandemic. There are many more types of stocks to invest in and also catastrophes that can arise at any given moment, so people seeking to invest in specific companies can conduct their own data projects to determine if they will be experiencing negative or positive returns.