

Project Scope Update:

Originally, my idea was to use VOO as my baseline for the US Economy's health. However, I found that SPY has existed for much longer and can provide more data points. Therefore, I switched from VOO to SPY. They are both popular ETFs that track the S&P 500, so changing them should not impact the project.

I also originally planned to download the unemployment rate dataset from FRED's website. However, I have found a free API that allows me to extract the data.

I also originally planned to measure the correlation between unemployment rate and the raw stock closing prices. After playing with the data, I found that I need a method to standardize the stock prices to measure correlation. I found that switching to percent change standardizes the values to percentages. I am hoping that I can find a more meaningful correlation between my datasets after standardizing them.

Data sources:

I originally downloaded the data off FRED and used the panda function `pd.read_csv()` to parse the dataset, but I have changed it to use the FRED API. I created a free API key and stored it in my `.env` file. I created a python function that has 5 parameters to help me obtain the data. The parameters are the dataset name, observation start date, observation end date, API key, and the frequency. Using these parameters, I made a request to the FRED API and obtained the unemployment rate per month starting from 2000-01-01 to 2025-11-01. Afterwards, I added column headers, and I made the date column into a datetime data type.

For my SPY and NVIDIA stock prices datasets, I used the Yahoo Finance API. The Yahoo Finance API does not require an API key. I created a python function with 5 parameters to help me obtain the data. The parameters are the dataset name, observation start date, observation end date, interval, and auto adjust. Auto Adjust is a boolean parameter that controls whether the price data is adjusted for stock splits and dividends. Using these parameters, I made a request to Yahoo Finance and obtained the SPY & NVIDIA monthly stock prices from 2000-01-01 to 2025-11-01. Afterwards, I reset the index and selected only the data and close columns. I also turned the date column into a datetime type. Afterwards, I used the panda's function `pct_change` to get the percent change by month.

Issues / difficulties:

The only issue I am running into is organizing and managing my GitHub. I suspect that as I continue to add more code, I will need to keep my GitHub even more clean and organized. I am also expecting to run into some issues with the different time periods of AI. I suspect that the majority of the data will be uncorrelated since AI has only become relevant in more recent times. I want to show the difference in correlation between and after the AI Boom, but I may be creating a lot of bias by choosing a specific date.