DSCI 510 Final Project Progress Report

Student Name: Sangwoo Choi
Professor: Dr. Alexey Tregubov
Date: November 11, 2025
Project Title: Building a Set and Forget ETF Strategy for Graduate Students

1. Project scope update
My project keeps the same title and goal. I want to use data to help graduate students understand how simple investing can grow over time. I will study a few well-known exchange-traded funds. I will compare growth funds, dividend funds, and a blended approach. I also added a clear plan to compare dollar cost averaging with a single lump sum. I will calculate basic metrics that we learned in class. These include compound annual growth rate, volatility, Sharpe ratio, and maximum drawdown.

2. Data sources
I use Yahoo Finance to collect historical prices for the funds. I will work with adjusted close prices and then resample to monthly data. I also plan to use the Federal Reserve data for inflation and the ten-year Treasury rate. This lets me show results in today's dollars and gives context for risk. Both sources are public and easy to access in Python.

3. API used
I am using the Yahoo Finance interface in Python through the yfinance library. I call the download function for each ticker and store the result in a pandas series. I then convert daily prices to monthly data for analysis and charts. I will also call the Federal Reserve endpoint with the requests library to pull CPI and the ten-year rate. This part is additional work, but it helps me show real returns and a basic risk-free rate.

4. Issues and difficulties
The first issue was occasional missing data when I requested long time periods. I fixed this by loading each ticker one by one and adding simple error handling. The second issue was turning the formulas from class into code. I wrote small functions for compound annual growth rate, annual volatility from monthly returns, Sharpe ratio, and maximum drawdown. I tested them on one fund and the numbers look reasonable. My next steps are to finish the first set of charts and run a quick test on dollar cost averaging.

Summary
The repository is set up and public. The data pipeline works for one ticker and will be applied to the rest. The analysis plan matches the course goals and uses the tools we learned. I will complete the figures and the short slide deck next. I plan to finalize visualizations and test functions before submission.

GitHub repository: https://github.com/hittoai/dsci510_finalproject_etf