Building a Set-and-Forget ETF Strategy for Graduate Students

Student: Sangwoo Choi

DSCI 510 — Principles of Programming for Data Science

Professor: Dr. Alexey Tregubov

25 November 2025

Motivation: Why do many graduate students avoid investing?

- Many students think investing is too complicated or risky
- Most believe small contributions don't matter
- Waiting until "later in life" feels safer — but costly
- **Goal of this project:** use real ETF data to show how simple automated investing can realistically build long-term wealth

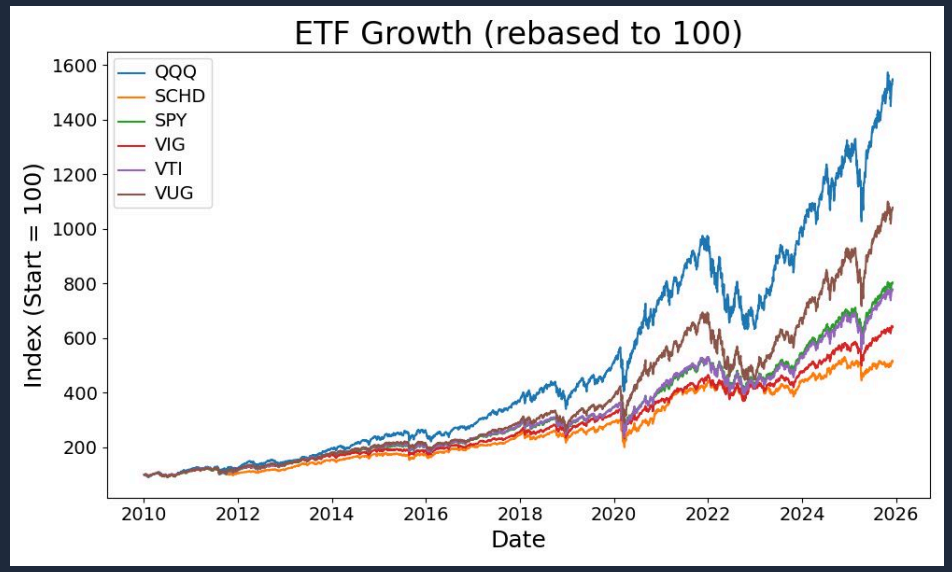
Data Sources and Methodology

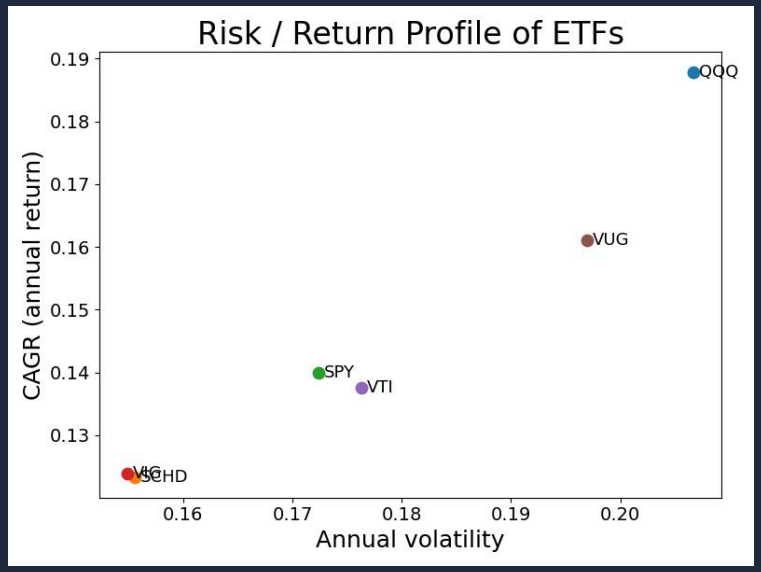
Data Source	Description	How Processed	Purpose
Yahoo Finance (ETF Prices)	Daily price history for QQQ, SPY, VIG, VTI, VUG, SCHD. ~3,800 trading days per ETF × 6 ETFs ≈ 22,800 total price records	Loaded via yfinance → cleaned/matched dates → saved to CSV	Real market data for ETF growth analysis
Derived Monthly Metrics	Monthly returns + CAGR + volatility + Sharpe	Resampled daily → calculated metrics → saved to CSV	Compare performance & risk across ETFs
Simulation Models	DCA, Lump Sum, start-early vs start-late	Python compounding functions (no external data)	Test long-term investing strategies for students

Simulation Model Description:

These simulations use deterministic monthly compounding based on constant average annual returns. No predictions or machine learning models are used. Portfolio values are calculated using fixed contribution formulas from age 22 to 65.

ETF Return Comparison: Growth vs Dividend vs Blended





Growth ETFs (QQQ, VUG)

Highest returns, higher volatility, technology sector focus

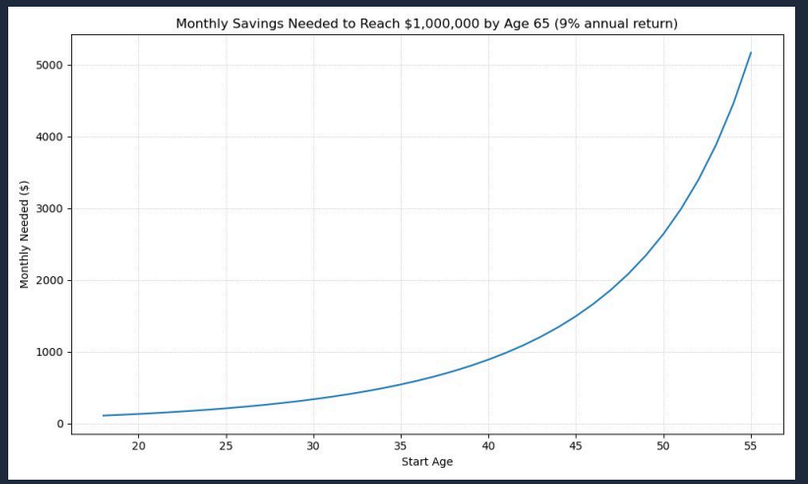
Dividend ETFs (SCHD, VIG)

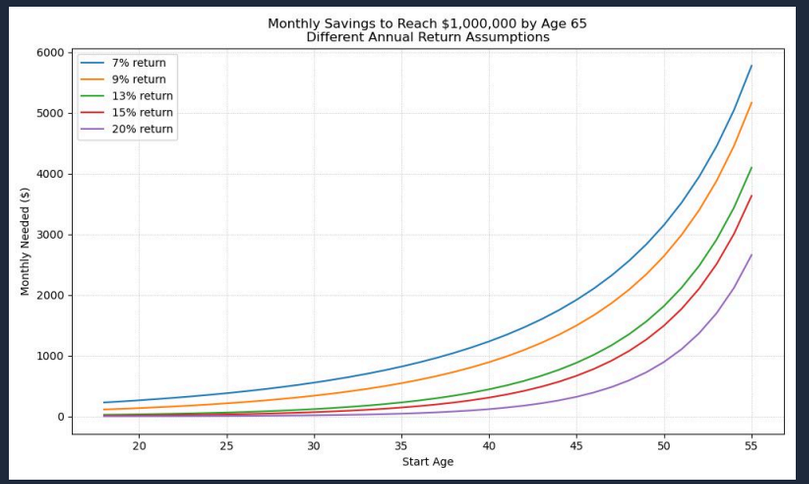
Smoother returns, lower volatility, income-focused

Blended ETFs (SPY, VTI)

Balanced growth, moderate volatility, broad market exposure

Monthly Savings Needed to Reach $1M by Age 65





🔵 The Early Advantage

Starting at age 22 → only ~$120/month; Starting at age 30 → ~$270/month (more than 2× increase)

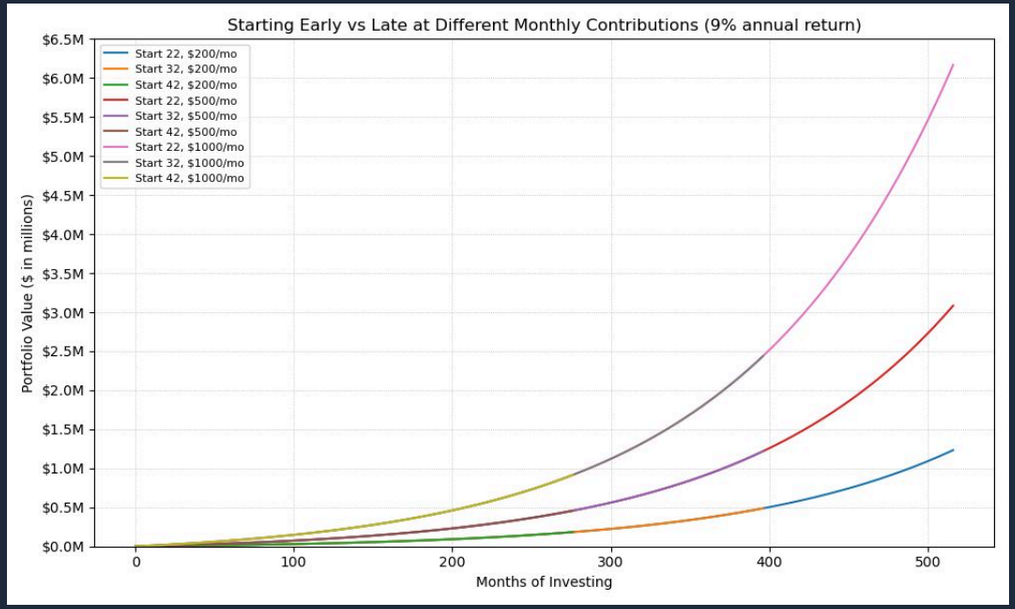
🟢 Exponential Growth With Delay

Starting at age 40 → ~$650/month (about 5× increase vs age 22); Starting at age 50 → ~$1,900/month (about 16× increase vs age 22)

🔴 Why This Matters

The later you start, the steeper the cost; Time in the market matters more than contribution size; Even small early deposits beat large late deposits

Starting Early vs Late at Different Monthly Contributions



Time in Market Beats Timing

Starting at 22 with $200/mo exceeds starting at 32 with $500/mo

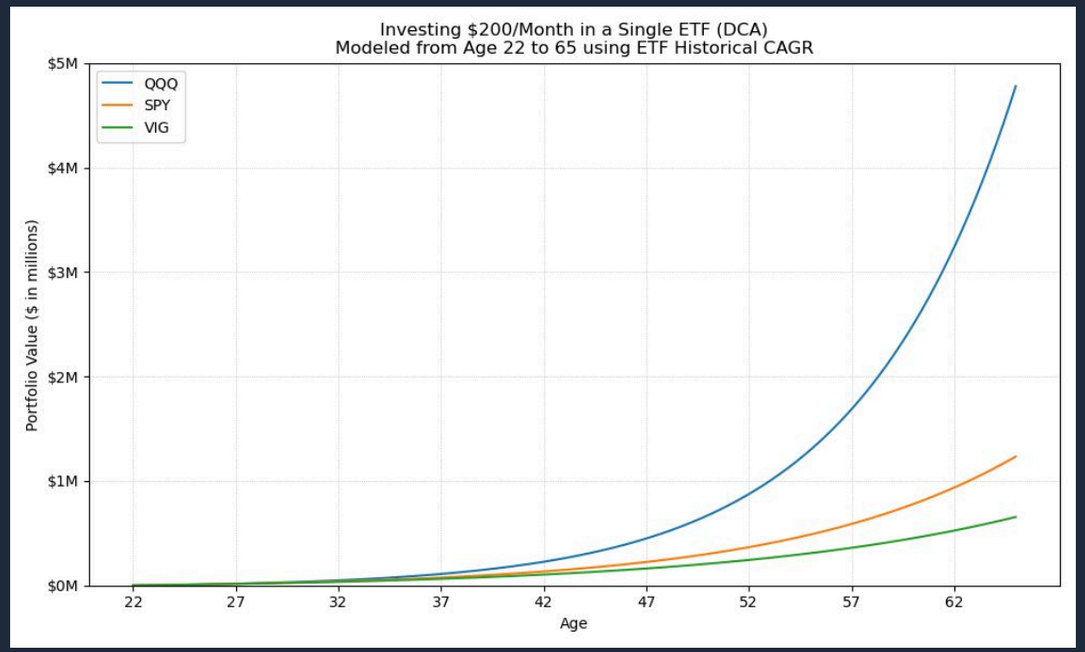
Power of Compounding

Exponential growth accelerates in later years; $1,000/mo at age 22 → over $6M by age 65

Strategic Implication

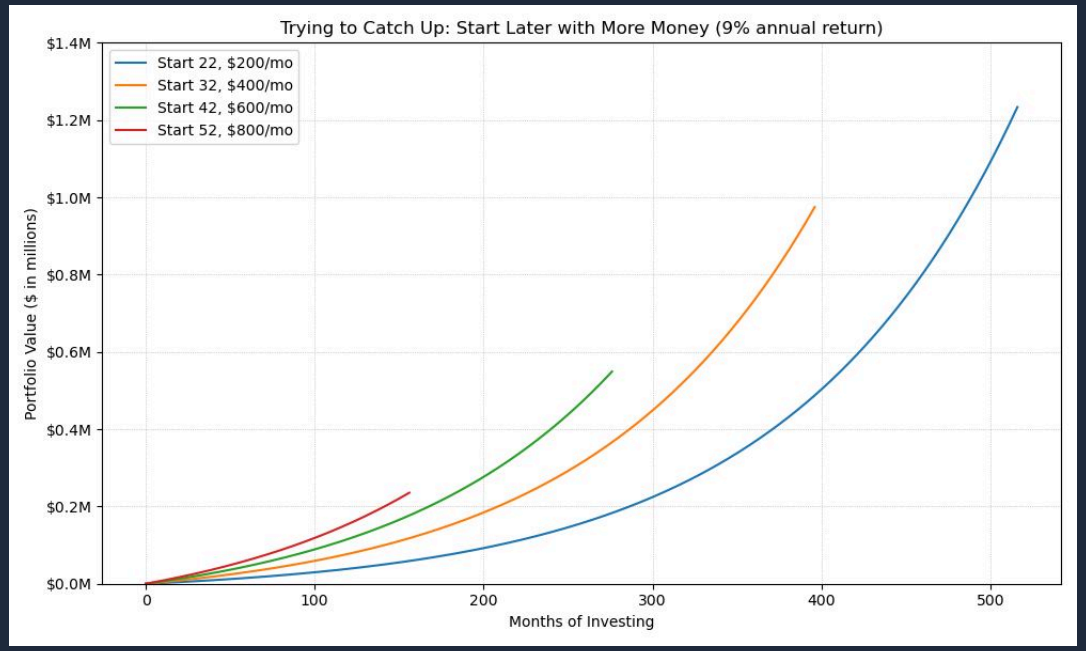
Start now with any amount you can afford; consistency beats waiting for larger contributions

Impact of ETF Choice With the Same Monthly Investment



Investing $200/month from age 22 to 65 shows dramatic differences based on ETF selection. QQQ's higher growth rate produces significantly more wealth than SPY or VIG, demonstrating the importance of choosing ETFs aligned with your time horizon and risk tolerance.

Trying to Catch Up: Start Later with More Money



Even with substantially higher monthly contributions, starting late cannot fully compensate for lost compounding time. Early start advantage is difficult to overcome.

A Practical Set-and-Forget Strategy

Risk Preference	Suggested ETF Mix	Key Characteristics
Conservative	60% SCHD + 40% VOO	Lower volatility with dividend income. Smoother returns during market turbulence. Ideal for shorter time horizons.
Standard	60% VOO + 40% QQQ	Balanced growth and risk. Core S&P 500 with tech tilt. Suitable for most graduate students.
Aggressive	100% QQQ	Highest long-term returns. Technology sector concentration. Best for 25+ year horizons and high risk tolerance.

Dollar-Cost Averaging (DCA) Implementation Principles

1. Set up automatic monthly transfers to your brokerage account
2. Invest on the same day each month regardless of market conditions
3. Rebalance portfolio annually to maintain target allocations
4. Use tax-advantaged accounts when possible (Roth IRA, 401k)
5. Start as early as possible with any amount you can afford
6. Maintain consistency over decades, ignoring short-term fluctuations

Challenges & Future Improvements

Project Challenges

1. Data Cleaning & Processing
2. Return Assumptions
3. API Reliability
4. Visualization Challenges

Future Improvements

1. Inflation Adjustment
2. Account Type Comparison
3. Asset Allocation Strategies

Thank You

Questions?