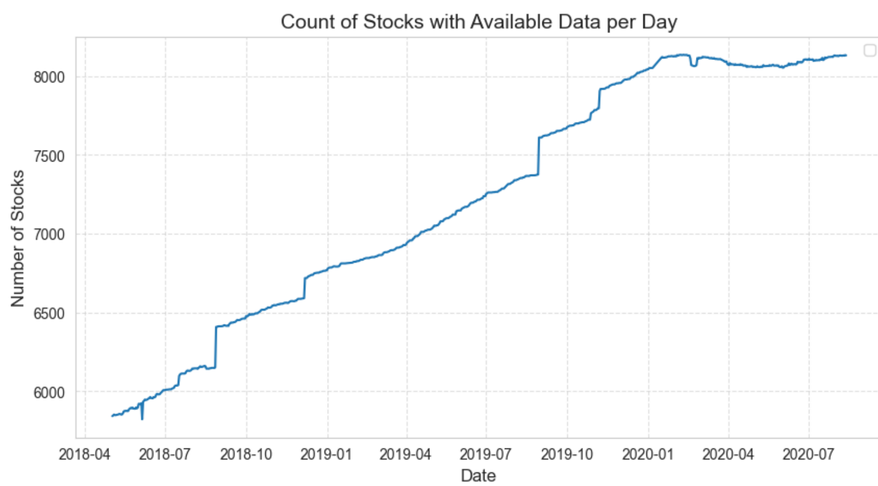
1 Robinhood Dataset

1.1 Description of the Dataset

This data is retrieved from https://robintrack.net/, the creator retrieved data from the official Robinhood API.

The dataset contains the number of Robinhood users holding at least one share of 8,221 securities. The available data spans from February 5, 2018, to August 13, 2020, covering 818 days (data is available also for non-trading days). Although the data was originally recorded hourly, I aggregated it to a daily frequency by computing the average number of holders per day to simplify computations due to the dataset's size. This aggregation can be easily reversed if needed.



Handling NaNs The original dataset contains missing values for 3,331 securities, primarily in the earlier periods. In some cases, assets appear in the dataset only after a certain date, despite being publicly traded before. It is important to distinguish between missing values and zero values, as they represent different concepts. Some securities exhibit a sudden increase from zero to a larger number of holders, but interpreting these as errors would impose an assumption on investor behavior.

Additionally, 1,248 securities have at least one recorded zero in the number of holders. The majority of missing data corresponds to small-cap stocks, which collectively account for at most 3 percent of total market capitalization. Given the limited impact of these securities on overall retail activity, I opted to remove all securities with missing values to ensure consistency in the dataset.

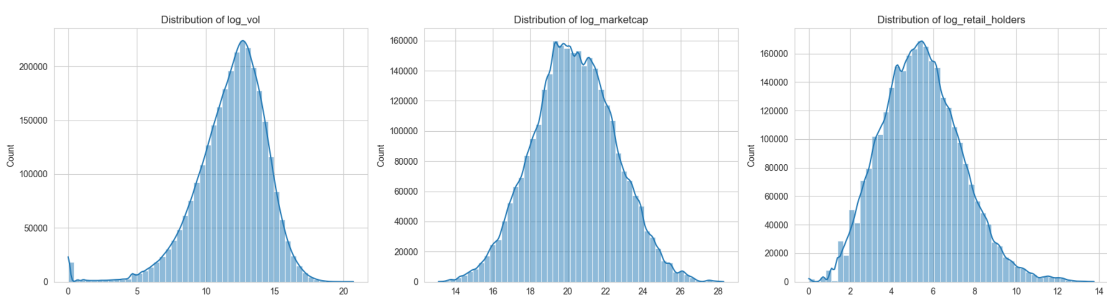
1.1.1 Distribution of Key Features (Log-Transformed)

The distributions of trading volume, market capitalization, and retail holders were initially highly skewed, with a few extreme values dominating the dataset. To address this, I applied a logarithmic transformation: $x' = \log(1 + x)$.

This transformation reduces the impact of outliers, enhances interpretability by making the data more symmetric, and facilitates comparisons between stocks of different sizes.

The key observations after applying the transformation are:

- **Trading Volume:** The distribution appears approximately normal, centered around a peak, with a slight left tail. While most stocks have relatively low trading volume, a few highly traded stocks, such as large-cap or meme stocks, exist but no longer dominate the distribution.
- **Market Capitalization:** The transformed market capitalization data exhibits a bell-shaped curve, suggesting a more balanced spread across small, mid, and large-cap stocks. However, some large-cap stocks remain in the extreme right tail, indicating that a few companies, such as Apple and Microsoft, are significantly larger than the majority.
- **Retail Holders:** The number of retail holders follows a roughly log-normal distribution, confirming that a small number of stocks attract massive retail participation while most remain relatively unpopular. The left tail suggests that many stocks have very few retail holders, reinforcing the notion that retail trading is concentrated in a subset of securities.



1.2 Comparing the Portfolios

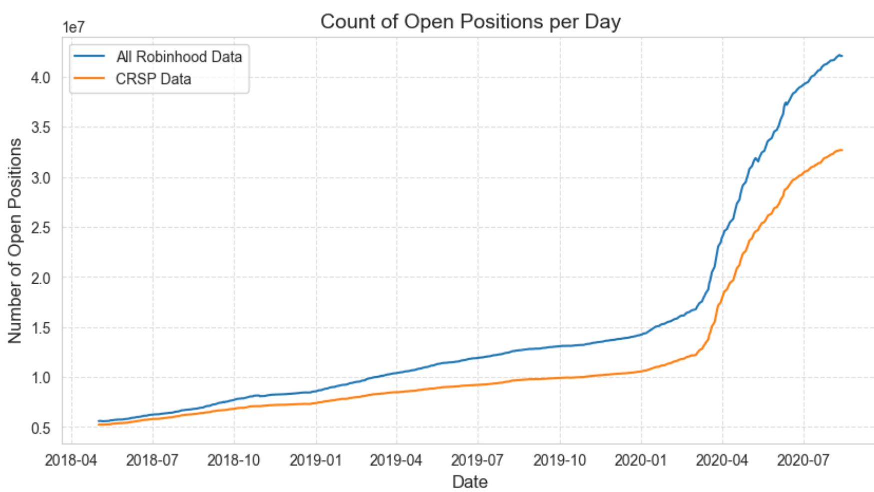
1.2.1 Methodology and Overview

To build a representative portfolio of the average Robinhood investor, it is necessary to retrieve the prices of the securities. A capitalization-weighted approach can be used, multiplying the price of each security by the number of users who hold it. This approach assumes that all Robinhood users hold a similar number of shares for a given ticker, or that the distribution of shares held per user follows a normal distribution.

Over the years covered in the dataset, Robinhood has gained a significant number of users. Data on active users is available on Statista[1], though only on a yearly basis. Comparing the Statista figures with Robinhood's reported numbers for 2023 suggests that the active user count corresponds to December 31 of each year. This data could later be used to normalize the number of users and build a reference portfolio.

The total number of open positions can be computed as the sum of all investors who hold at least one security in each asset, effectively a row-wise sum of the dataset.

Market data for all securities was retrieved from the CRSP[2] database, accessed via WRDS. However, only 8,099 securities are available in CRSP, as it focuses exclusively on American assets. The difference in open positions between the full dataset and the CRSP subset is minimal. If, instead, all securities with missing values are dropped, leaving only 5,221 securities, the gap widens.



The graph illustrates the count of open positions per day on Robinhood from April 2018 to mid-2020, showing a steady increase over time, with a sharp acceleration in early 2020. This surge aligns with the onset of the COVID-19 pandemic, which likely drove a significant influx of new retail investors seeking market opportunities amid economic uncertainty and stimulus checks.

[1] https://www.statista.com/statistics/822176/number-of-users-robinhood/

[2] The Center for Research in Security Prices, based at the University of Chicago, provides high-quality historical market data widely used in finance research and investment analysis.

1.2.2 Retail Investors Prefer "Famous" Stocks

The majority of the securities are common shares, representing about 57.9%. ETFs represent about 23.7% and other funds are the 9.2% of the dataset. Other structured investments, REITs, and ADRs cover the remaining part.

Analysing the securities by market capitalisation about 82.9% is represented by stocks and 9.6% by ETFs. If we look at the "Retail Market Cap" (i.e. number of positions times price), 89.2% of securities are stocks and 5.8% are ETFs.

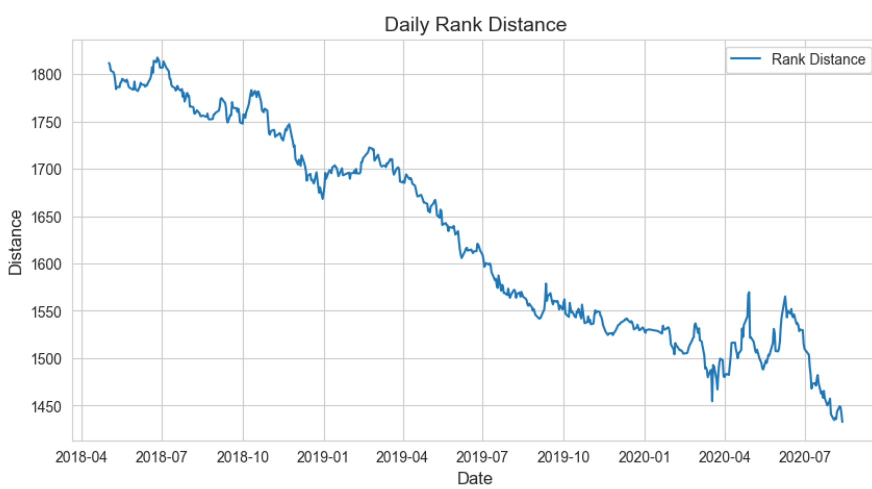
Looking at the securities Robinhood users prefer holding, ranked by "Retail Market Cap", investors prefer holding smaller cap stock. A qualitative analysis shows "famous" stocks, such as Tesla, Starbucks, and Nvidia to name a few, to appear among the most popularly owned.

1.2.3 Possible Measures of Divergence

Rank Distance To describe the preference of retail investors for smaller cap stock I propose the following measure:

$$d_R = \sum_{i=1}^{N} \frac{R_i^{\text{Mkt}} - R_i^{\text{RH}}}{R_i^{\text{RH}}}$$

Where R_i^{Mkt} is the rank of the i^{th} security by market cap, and R_i^{RH} is the rank by retail market cap. The normalization by R_i^{RH} reduces the impact of small-cap stocks with minor ranking differences.



The plotted Daily Rank Distance suggests a clear downward trend from early 2018 to mid-2020, indicating that the ranking of stocks by retail market cap has become increasingly aligned with the ranking by total market cap. Initially, the distance is above 1800, gradually declining towards 1450. This implies that retail investors, who originally exhibited a stronger preference for smaller-cap stocks, have progressively shifted towards stocks that are more representative of the broader market.

Between 2018 and 2019, the decline is relatively steady, reflecting a gradual change in retail investment behavior. However, the trend accelerates in 2019 and 2020, suggesting a more pronounced shift. The beginning of 2020 shows increased volatility, with occasional upward spikes, which could be attributed to market disruptions, possibly linked to the COVID-19 crash and the subsequent retail trading boom. The rapid expansion of retail investing during this period, fueled by stimulus checks and zero-commission trading, may have led to temporary deviations, but the overall trend continues downward.

A sustained decrease in rank distance suggests that retail investors have moved closer to institutional preferences, potentially increasing their exposure to large-cap stocks or index-tracking assets. If this trend persists, it would indicate a continued assimilation of retail behavior into the broader market structure. Conversely, a reversal in this pattern could signal renewed speculative activity or a shift back to small-cap stocks.

2 Analysing Returns

2.1 Building the Robinhood Portfolio

As explained above, the biggest limitation of the Robintrack dataset is that it counts the number of users holding a certain security and doesn't provide any information on the amount invested in a particular security.

A possible solution is building a portfolio in which we assume that all positions hold the same equivalent amount of money in a certain security. This means weighing stocks by their "popularity":

$$Pop_{i,t} = \frac{n_{i,t}}{\sum_i n_{i,t}}$$

where $n_{i,t}$ is the number of users holding a certain stock on a given day.

We can therefore define the value of the Robinhood Portfolio as follows:

$$V_{RH,t} = \sum_{i=1}^{N} Pop_{i,t} \cdot P_{i,t}$$

In comparison, the value of the reference index is simply the sum of the market cap for each security present in the Robinhood dataset.

2.2 Comparing Returns and Risk measures

Having defined the value for the Robinhood Portfolio and reference index we can analyze their performance, also comparing them to the S&P500 as a general market proxy.

Returns for a given time frame are computed as the sum of the daily log returns for the previous days.

Keeping this in mind we can obtain the moving average and cumulative returns for all these indices[3], as well as their distributions.

At short horizons (5 to 15 days), all three portfolios show relatively similar performance, with the Robinhood portfolio slightly more volatile. This suggests that over short periods, retail behavior does not drastically deviate from market movements. However, a notable divergence emerges as we move to medium and long horizons.

At the 30- and 60-day horizons, the Robinhood portfolio begins to exhibit both higher peaks and deeper drawdowns relative to the market and VOO. This pattern reflects increased exposure to volatility and suggests that retail traders may chase momentum or respond to events with delayed timing, amplifying fluctuations. Interestingly, around mid-2020, retail returns spike significantly above the benchmarks, likely reflecting gains

[3]VOO is the ticker of the Vanguard S&P500 ETF

from tech or meme-stock rallies.

In the 120-day panel, the Robinhood portfolio consistently underperforms the benchmark indices for most of the observed period, particularly during downturns such as the COVID-19 crash. This underperformance suggests poor diversification or overexposure to high-risk assets. However after the COVID market crash, the Robinhood Portfolio gained momentum reaching the highest return of all three.

The 564-day horizon highlights a substantial divergence: while VOO steadily increases and outpaces the market benchmark, the Robinhood portfolio lags behind both at first. Despite some recovery post-COVID, retail investors failed to fully capture the long-term gains available in the broader market, possibly due to suboptimal stock selection or timing, although being able to outperform the reference index.

Looking at the distribution of returns for the same periods other insight can be drawn[4].

At shorter horizons (5-30 days), all three distributions are tightly clustered and symmetric, with slight differences in tails. The Robinhood distribution tends to have fatter tails, suggesting more extreme short-term gains and losses.

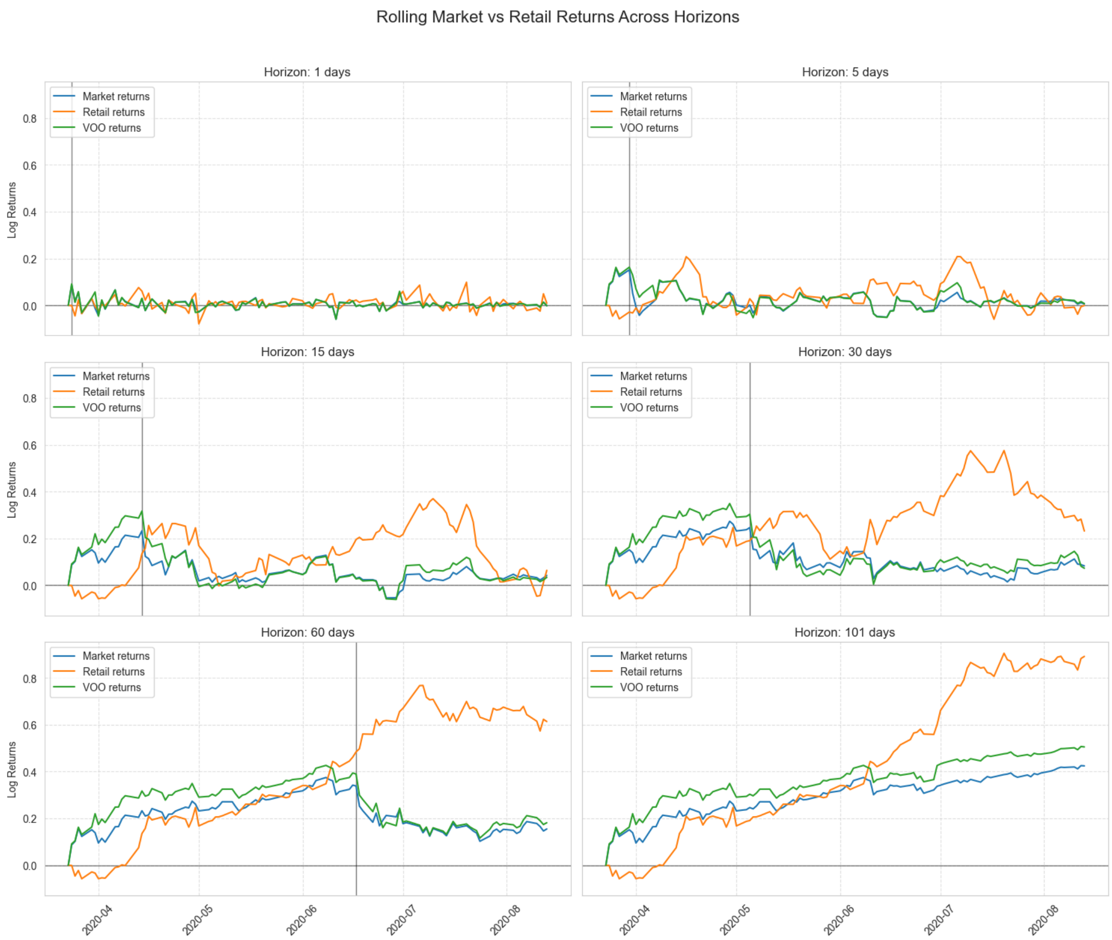
As the horizon increases (60-564 days), the Robinhood return distributions become wider and more dispersed, with heavier tails and lower peaks, indicating higher return volatility and greater exposure to extreme outcomes. In contrast, VOO and the market benchmark maintain tighter, more concentrated distributions, consistent with diversified, lower-risk portfolios.

[4]The details of the distribution are in table 1

At the longest horizon (564 days), the Robinhood distribution is notably left-skewed, consistent with the earlier finding of long-term underperformance. VOO, on the other hand, exhibits a clear rightward shift, reflecting higher and more consistent long-term returns.

These plots reinforce earlier conclusions: the Robinhood portfolio is more volatile, more prone to extreme outcomes, and delivers worse performance over long horizons relative to both the market and VOO.

What Happened during the Pandemic? Analyzing the same indices from February 3rd, 2020 we can find some relevant insights.



At very short horizons (1-5 days), performance across all portfolios is similar and volatile, with no consistent pattern. However, from the 15-day horizon onward, a clear divergence emerges: Robinhood portfolio returns (orange) begin to outpace both the market (blue) and VOO (green).

This trend becomes particularly evident at 30, 60, and 101-day horizons, where the Robinhood portfolio exhibits significantly higher cumulative gains, peaking around July 2020. In contrast, VOO and the market index recover more gradually, with smoother return paths and lower cumulative gains. This reflects broader diversification and reduced exposure to speculative stocks.

3 Literature Review

The rise of commission-free trading platforms such as Robinhood has significantly altered the behavior of retail investors, leading to increased speculative trading and attention-driven investment patterns. Existing literature highlights the unique characteristics of these investors, the role of fintech innovations in shaping trading behavior, and the broader market impact of retail-driven trading.

3.1 Retail Investors and Attention-Induced Trading

One of the defining characteristics of Robinhood users is their tendency to engage in attention-induced trading. [Barber and Odean, 2011] document that retail investors are heavily influenced by limited attention and past return performance, often buying stocks that experience large daily price movements. This behavioral bias is further reinforced by Robinhood's design, which prominently features "Top Movers" lists and real-time price changes [Barber et al., 2021].

Ardia et al. [Ardia et al., 2023] provide a high-frequency analysis of Robinhood traders, showing that they react aggressively within an hour of extreme negative price movements. This suggests that Robinhood users are particularly prone to rapid contrarian trading, potentially due to real-time engagement with financial markets through mobile notifications and social media.

3.2 The Influence of Robinhood on Retail Trading Behavior

The trading behavior of Robinhood users differs from that of traditional retail investors. The simplicity and gamification of the app influence decision-making, leading to increased speculative trading. [Barber et al., 2021] find that Robinhood's "Top Mover" feature encourages users to buy both extreme gainers and losers at higher rates than traditional retail investors, diverging from the broader retail trading pattern of favoring winners.

Moreover, Robinhood traders exhibit a strong preference for stocks experiencing extreme price movements. [Ardia et al., 2023] report that users react particularly fast to negative price shocks, displaying contrarian tendencies not typically observed in traditional retail trading.

3.3 Market Impact of Robinhood Users

Several studies have examined the broader market impact of Robinhood-driven trading, particularly during periods of heightened retail activity. Concentrated retail buying can lead to short-term price pressure and subsequent reversals. [Barber et al., 2021] show that Robinhood herding episodes-where a large number of users buy the same stock-result in negative abnormal returns of approximately 4.7% over the following 20 days.

Additionally, retail-driven trading was amplified during the COVID-19 pandemic, with a surge in new users engaging in speculative trading. [Ardia et al., 2023] find that Robinhood users intensified their trading activity post-pandemic announcement, particularly in small-cap and volatile stocks.

Appendix A Tables

Table 1: Descriptive Statistics for Daily and Rolling Returns

	Mean	Std	Min	25%	50%	75%	Max
rh_portfolio	0.000719	0.018809	-0.132368	-0.006164	0.001141	0.009484	0.072851
mc	0.000396	0.015470	-0.125496	-0.003944	0.001012	0.006481	0.086673
voo	0.001081	0.016707	-0.124870	-0.003804	0.001254	0.007132	0.091087
mc_5_return	0.001940	0.031094	-0.207508	-0.008577	0.004961	0.016049	0.151511
rh_portfolio_5_return	0.003309	0.041768	-0.224427	-0.012162	0.004643	0.022078	0.153755
voo_5_return	0.005368	0.032968	-0.204425	-0.007355	0.007621	0.019069	0.162820
mc_15_return	0.005387	0.058358	-0.329924	-0.007216	0.014780	0.030739	0.232340
rh_portfolio_15_return	0.009097	0.084568	-0.370666	-0.016691	0.013130	0.036470	0.329388
voo_15_return	0.015704	0.058314	-0.329074	-0.001770	0.021466	0.039643	0.316891
mc_30_return	0.009890	0.080443	-0.401515	-0.015223	0.025072	0.046527	0.246006
rh_portfolio_30_return	0.015173	0.133421	-0.482722	-0.041325	0.020205	0.051931	0.408751
voo_30_return	0.030710	0.071248	-0.349674	0.005831	0.039493	0.067139	0.303026
mc_60_return	0.016554	0.099118	-0.356261	-0.013618	0.029848	0.062050	0.338109
rh_portfolio_60_return	0.010727	0.177731	-0.377518	-0.066831	0.001284	0.070271	0.641470
voo_60_return	0.056506	0.081384	-0.284299	0.029484	0.060684	0.084047	0.389273
mc_120_return	0.017214	0.075478	-0.307022	-0.031108	0.030972	0.070401	0.218409
rh_portfolio_120_return	-0.014462	0.114741	-0.310504	-0.098157	-0.014206	0.053267	0.370780
voo_120_return	0.093966	0.065267	-0.238772	0.059052	0.100453	0.130918	0.283035
mc_564_return	0.071788	0.069591	-0.200798	0.033823	0.070623	0.106613	0.224508
rh_portfolio_564_return	-0.011075	0.123340	-0.382891	-0.055196	-0.014086	0.045442	0.405724
voo_564_return	0.261809	0.153261	-0.002648	0.132996	0.253737	0.379423	0.611416

References

[Ardia et al., 2023] Ardia, D., Aymard, C., and Cenesizoglu, T. (2023). Fast and furious: A high-frequency analysis of robinhood users' trading behavior. *arXiv preprint arXiv:2307.11012*.

[Barber et al., 2021] Barber, B. M., Huang, X., Odean, T., and Schwarz, C. (2021). Attention induced trading and returns: Evidence from robinhood users. *SSRN Electronic Journal*.

[Barber and Odean, 2011] Barber, B. M. and Odean, T. (2011). The behavior of individual investors. *SSRN Electronic Journal*.