W200 Project 2 -- Invest in SAND?

Name: Chun Him Cheung, Dominic Lim, Ethan Nguonly, Sarah Zheng
GitHub repository: https://github.com/UC-Berkeley-ISchool/Project2_Cheung_Lim_Nguonly_Zheng

INTRODUCTION

As cryptocurrencies continue to shake the financial world, an area it has proven extraordinarily useful is that the adoption of these digital tokens in virtual worlds continues to bring us closer to the Metaverse. One particularly successful company that paves the way to this digital future is The Sandbox, bringing to the market an open Metaverse that gives players the ability to create and monetize their user generated worlds on Ethereum blockchain. In August 2020, The Sandbox launched their own cryptocurrency, SAND, with a market cap of $27M that has now grown to a market cap of $6.3B.[1] SAND can be used in many settings such as traded as a currency in a virtual economy and buy and sell NFTs. Given the increased popularity of the Metaverse and this token's explosive growth, we chose to analyze SAND for our research project.

For this paper, we'd like to investigate 200 Ethereum wallets that hold the most SAND and analyze its investment strategies and performance. As many people know, actively traded wallets, where traders may have a deeper knowledge of Ethereum-based coins, may have a significantly higher rate of return vs. passively holding SAND. *Therefore through our analysis, we hope to analyze the top performing wallets to understand their investment strategy and potentially uncover investment insights to buying-and-holding SAND.*

FOCUS QUESTIONS

1. What standard financial metrics can we modify and use to measure performance of crypto wallets?
2. What was the transfer movement or pattern of SAND tokens?
3. Which of the top holders of SAND has been performing the best due to their investment in SAND?
4. How did the top holders of SAND respond to significant price increases in terms of buying or selling tokens?

CRYPTOCURRENCY CONTEXT

The advent of blockchain technology, and by extensions, cryptocurrencies has immensely widened the universe of available financial data to analyze. In contrast to traditional financial markets, the transaction cryptocurrencies are entirely transparent and publicly available to see on blockchain explorers such as Etherscan or Ethplorer. However, unlike traditional asset classes, the technological component of cryptocurrency markets immensely complicates the analysis of buy and sell activities.

In traditional asset classes, investors, especially retail investors, would largely buy to seek price appreciation and sell to avoid financial losses. However, in the cryptocurrency world, transfers into wallets are not necessarily 'buy' transactions. Due to the programmable nature of cryptocurrencies, coins may enter a wallet as a result of an 'air drop' (the process of creators to coins use to reward and encourage early adoption), the rewards of staking (contributing to maintaining a proof-of-stake blockchain), transfer between individual with multiple wallets, or finally, a transfer that represents a traditional 'buy' order.

---

[1] https://coinmarketcap.com/currencies/the-sandbox/

Moreover, while blockchain brings the potential to allow for decentralized transaction (i.e. economic transaction without an intermediary), at the current stage of market development, the vast majority of transfers still occur on centralized platforms (i.e. on a centralized exchange such as Coinbase) as liquidity for buying/ selling remain best on those platforms. For our analysis, we began with the top 200 wallets holding SAND and filtered out wallets of smart contracts and centralized exchanges. This reduced the number of wallets of interest from 200 to 147.

We excluded the transaction of centralized exchanges and smart contracts because they are larger holders of SAND that do not transact on economic reasons. Centralized exchanges are akin to a broker and would execute transfers of coins to and from their wallet to fulfill client orders whereas smart contracts transact based on their pre-built logic that may have no relationship to the movement of price. As such, from the top 200 wallets, we try to isolate a list of wallets that would most likely behave like a traditional investor seeking monetary return.

SOURCE DATA

In order to compile a json file for the top 200 SAND wallets and their transactions, we used two sources to get our required information, namely the wallets, their transactions, the SAND USD price at each transaction's timestamp, and the tag names/whether each wallet is a smart contract for each wallet address.
1. Using a Python script and the Ethplorer API, we pulled a list of the top 200 largest holders for SAND.
2. We then needed to get the transaction history for each wallet address and once again, used an Ethplorer endpoint to each wallet individually.
3. Once we had each wallet and its list of transactions, we needed to determine the USD price of SAND at each transaction timestamp. However, we were unable to find an API that gave that exact price of SAND for a given unix timestamp. Accordingly, we called another Ethplorer endpoint that queried by transactionHash of each SAND transaction, and returned the transaction's individual information, including the current price at SAND at its transaction timestamp. Using this information, we added the SAND USD value for each transaction to our original json file.
4. In order to get the public tag names for each wallet address, we use the website etherscan, which showed this. By writing a simple page scraper, we were able to generate a JSON of wallet addresses to public tag name, and use this JSON dictionary to add the tag names to the final wallet json file.

Computational Methods:
It may be apparent from the description above that a large number of APIS calls were required to compile this data. In order to prevent failure of JSON generation due to API response errors, as well as speed up the required time needed, we used multiprocessing to be able to compute wallet information in chunks of 10. We can do this because each wallet's information is separate from others and by creating sub files of 10 wallet transaction and price information, a simple script to combine the results of these JSON files drastically reduces how long our script takes to run.

1. https://api.ethplorer.io/getTopTokenHolders/
   a. Holders: Holder information Metadata
      i. Address: Unique wallet address per each holder
      ii. Balance: Number of SAND tokens in the transaction as of the API call
      iii. Share: The percentage the balance is compared to total market cap as of the API call
   b. Transactions: Transaction information

       i.   Timestamp: Time each transaction was completed
     ii.   TransactionHash: Unique identifier for each transaction
   iii.   Type: transfers only (see Assumption 1 below for description)
   iv.   Value: Number of SAND tokens transferred in one transaction
    v.   From: Address the SAND tokens were transferred from
   vi.   To: Address the SAND tokens were transferred to

2. https://api.ethplorer.io/getTxInfo/
   a. USD price of SAND per each wallet transaction
3. https://www.coingecko.com/en/coins/the-sandbox/historical_data/usd?end_date=2021-11-26&start_date=2020-08-01#panel
   a. The historical price, market cap and total volume data for SAND

INITIAL EXPLORATION AND DATA PREPARATION

Assumptions
1) **Transaction Types:**
   a) "Transfers" were pulled from the API with the assumption that these transaction types were most representative of "Investment" activity: Buying and Selling. Notably, transaction types such as "issuance", "mint", and "burn" were excluded from the API calls.
   b. These other transaction types occur for mining rewards.
2) **Tagged Wallets:**
   a) We are assuming that all wallet holders that are either "smart contracts" or represent exchanges are being properly tagged.
      i) This provides us a basis for excluding such wallets from our dataset.
3) **Net Cash Flow:**
   a) Every crypto transaction has a sender and receiver denoted in our data set with "address_from" and "address_to", respectively.
   b) We are assuming that transactions in which the walletholder's address is different from the "address_from" (I.e the wallet holder is receiving tokens) represents a purchase of tokens.
   c) We are assuming that transactions in which the walletholder's address is the same as from the "address_from" (I.e the wallet holder is sending tokens) represents a sale of tokens.
4) **SAND Price Data:**
   a) While the transactions include the price of SAND at each timestamp, the daily price data from the CoinGeckoAPI is provided at 12:00 am GMT.

Sanity Checks / Data Cleaning

As we are building a dataset using API calls, rigorous sanity checks were performed before running any data analysis.

1) Total Number of transactions in dataset:
   a) Initial 18,139 transactions in line with estimates from API Calls.
   b) After excluding the "tagged" wallets, our transaction level dataset was filtered down to 12,277 transactions.
2) Only "transfer" type transactions should be used.
   a) "tx_type" for all transactions returned as "transfer".
   b) A transfer means the coins were transferred from one wallet to another. It is confirmed that this is the case for all transactions in the dataset
3) Our dataset should account for less than 100% of the total # of circulating tokens:

        a)   Our initial dataset with 200 wallet holders accounted for 97.4%.

        b)   After excluding "tagged" wallets, our remaining sample of 147 wallets accounted for 10.3% of the total supply of coins in circulation.

            i)   The excluded "smart contract" wallets made up 73.0% of total SAND balance.

           ii)   The excluded exchange wallets made up 13.7% of total SAND balance.

4)   Initial transaction for each wallet holder:

        a)   Should be an initial "purchase" or inflow of SAND.

        b)   Date stamp of the initial transaction should NOT come earlier than August 2020 (coin launch).

FOCUS QUESTION 1: What standard financial metrics can we modify and use to measure performance of crypto and wallet holders?

In identifying and assessing the best performing wallet holders, it is crucial to understand the underlying price movement in SAND. We wanted to apply traditional investment analysis methods from Modern Portfolio Theory to calculating the average daily returns as well as the volatility of the price movement. The total returns for a given date range effectively serves as the "market" or "beta" return to which we can compare our wallets' returns.

We can calculate a given wallet's "alpha" return interpreted as a return in excess of the underlying SAND return in the same time period. This would imply that a wallet holder's skill in timing of purchase and selling SAND earned the wallet holder above-market returns. We believe that in an emerging asset class such as SAND, inefficiencies in the market may be leveraged to achieve outsized returns.

In addition, we utilized Sharpe Ratios in calculating the wallet holders' investment returns relative to the volatility or risk of a given date frame. A Sharpe ratio is simply the excess return of an asset over volatility (or standard deviation) of the returns of that asset[2]. As one would expect higher returns for additional units of risk, an investor that had outsized returns relative to their risk profile might be characterized as being efficient with their allocation of capital (risk).
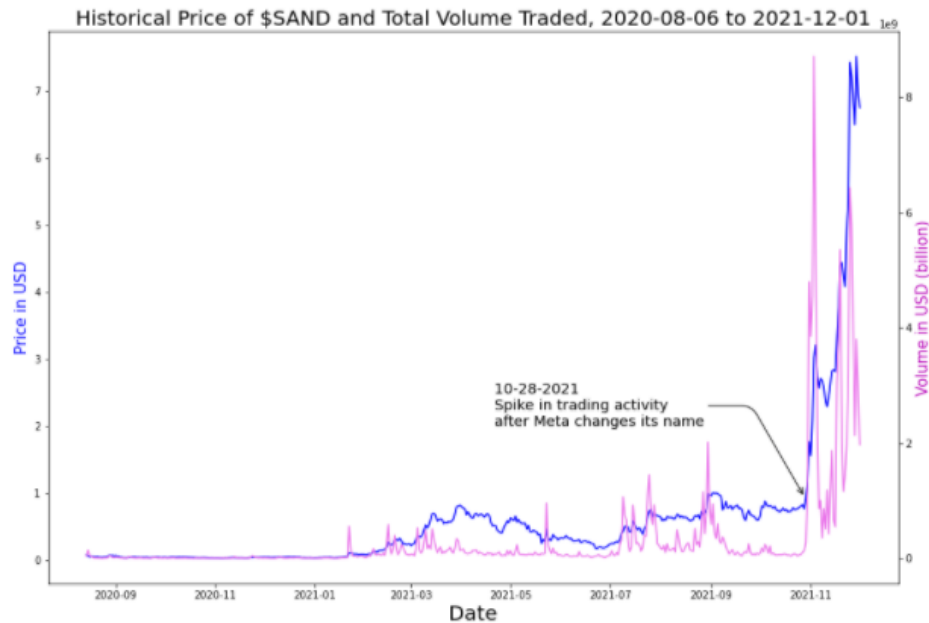
Holding period, or the duration that an asset is invested in, is a key component in contextualizing investment returns. If we are trying to run a comparative analysis on the returns and volatility of wallets with different holding periods, it is necessary to annualize both the total returns and volatility of daily price data. However, this assumes that the returns exhibit a normal distribution.

FOCUS QUESTION 2: What was the transfer movement or pattern of SAND tokens?

To understand the context behind SAND astronomical rise, we graphed its daily price since the inception of the coin to see if there were any anomalies (see below plot). After calculating the local maximum and minimum to identify major inflection points, it was not a coincidence to see that the lowest point before the price surge was also the same day Facebook changed its name to Meta, 10-28-2021. Meta's rebranding sent strong signals to investors of the growing ambition of large tech companies towards the Metaverse.[3] This is a strong basis for the incredible growth SAND has realized in just the last month, as this cryptocurrency, unlike other well known coins such as bitcoin, was created mainly to be used as currency in a virtual economy.

---

[2] https://www.investopedia.com/terms/s/sharperatio.asp

[3] https://www.cnbc.com/2021/10/28/facebook-changes-company-name-to-meta.html

Historical Price of $SAND and Total Volume Traded, 2020-08-06 to 2021-12-01

Additionally, in the plot above, to track the movement of SAND since inception, we see the largest spike in trading activity (shown in pink) for the token is also in the short timeframe beginning 10-28-2021 and into early November, before briefly dropping down to pre-Meta levels. However, throughout the rest of November, daily volumes continue to rise, which unsurprisingly perfectly lined up with the publication of major recent headline news such as well-known brands like Adidas tweeting a message in support of The Sandbox Game on 11/22/2021.[4]

FOCUS QUESTION 3: Which of the top 200 holders of SAND have been performing the best due to their investments in SAND?

In the ascent of SAND's value, we wanted to analyze the top performing wallet invested in SAND since the inception of the token. In order to determine which of the top 200 wallet holders of SAND have been performing the best, we decided to identify the top performing wallets by using the four financial metrics defined earlier: total returns, beta, Sharpe ratio, and alpha. In terms of the total returns, an assumption we made here is that we only want to calculate the returns generated by SAND transactions. That is why in our initial pulling of each wallet's transaction data, we filtered for only SAND transfers.

That said, from the below dataframes, we sorted our wallet dataframe by the highest alpha, returns, and Sharpe ratio. After analyzing the sorted data, we chose Sharpe ratio as the top metric of wallet performance because compared to alpha, the Sharpe ratio also accounts for the returns based on SAND's risk profile, the volatility of the asset, that we calculated in our analysis.

---

[4] https://www.coindesk.com/markets/2021/12/01/sand-mana-tokens-surged-in-november-as-crypto-traders-bet-on-metaverse-potential/

Top 5 Wallet Returns

	holders_address	return_by_wallet
41	0x4db1578cf389b1cda470c08e61c406f3e8bc2e7d	705.67
72	0x896078a63a1878b7fdc8dba468c9a59b94fd7a92	195.76
107	0xb701dec4186c8c7f1a7047b33f9f9a39febb1db1	140.76
56	0x6bcbe6c086cc66806b0b7b4a53409058d85f61c8	80.56
108	0xb8700c17bb966ab854a343377ddf6a2ef06630a3	75.30

Top 5 Wallet Sharpe Ratio

	holders_address	sharpe_ratio
41	0x4db1578cf389b1cda470c08e61c406f3e8bc2e7d	352.11
72	0x896078a63a1878b7fdc8dba468c9a59b94fd7a92	89.28
107	0xb701dec4186c8c7f1a7047b33f9f9a39febb1db1	69.54
56	0x6bcbe6c086cc66806b0b7b4a53409058d85f61c8	37.15
108	0xb8700c17bb966ab854a343377ddf6a2ef06630a3	35.56

Top 5 Wallet Alpha

	holders_address	wallet_alpha
41	0x4db1578cf389b1cda470c08e61c406f3e8bc2e7d	579.55
56	0x6bcbe6c086cc66806b0b7b4a53409058d85f61c8	63.17
72	0x896078a63a1878b7fdc8dba468c9a59b94fd7a92	27.96
117	0xc50174c3cdb60fb4b685891d5812743141ba557a	14.38
85	0x964affa16e70c1477e52670b97471f8db2f976f3	13.67

Coupled with the tables above, we can see from the plot below that five wallets in particular had a significantly higher return than the rest as measured by their Sharpe ratio - wallets 41, 72, 56, 107, and 108. Moving forward, we can now use these top performing wallets for further analysis as to what actions these wallet addresses took in order to generate such a high return.


Sand Volatility vs Wallet returns
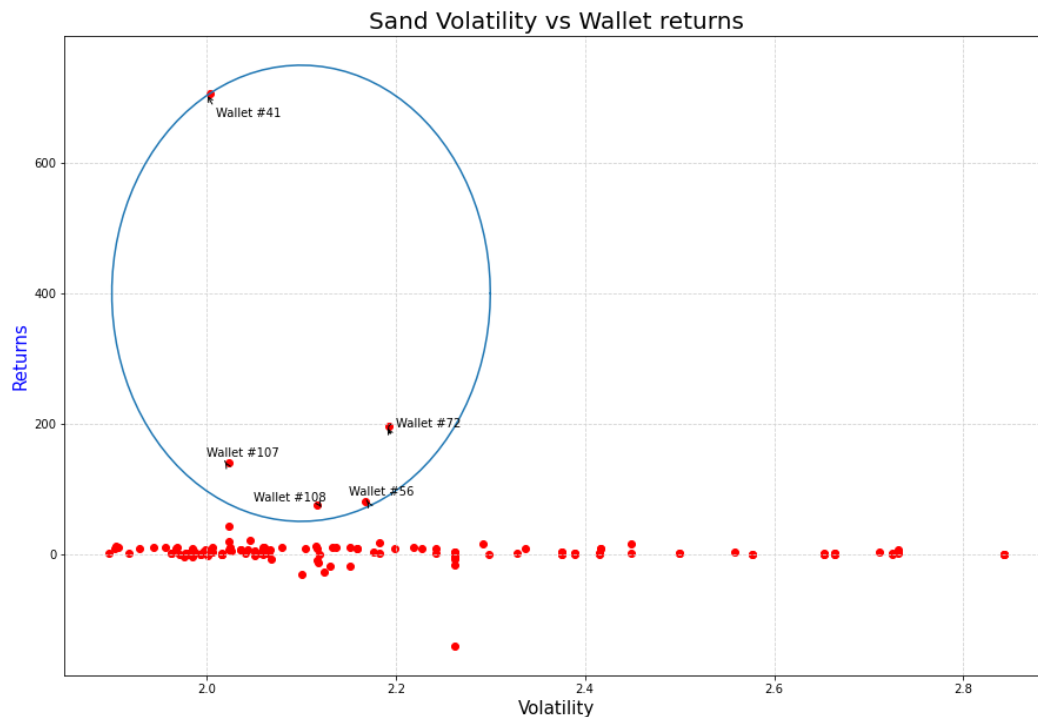
FOCUS QUESTION 4: How did the top performing wallets of SAND respond AFTER significant price increases in terms of buying or selling tokens?

After filtering down to our top 5 wallets, we found two wallets that seemed to exhibit market intuition and may warrant further observation and study in the waterfall charts below.[5]

---

[5] https://pbpython.com/waterfall-chart.html

- Wallet #72 was an early adopter of SAND and continued to be a net purchaser of SAND as represented by the green columns (see below chart).
- When the price of SAND increased dramatically as shown in the inflection point or "elbow" of the blue price plot,Wallet #72 sold a modest portion of its holdings suggesting it remained BULLISH in the price of SAND.
- Wallet 72 was handsomely rewarded as the SAND price would increase 24-fold from the $0.30 to $7.23 per SAND.

Trading activity over time for Wallet 72
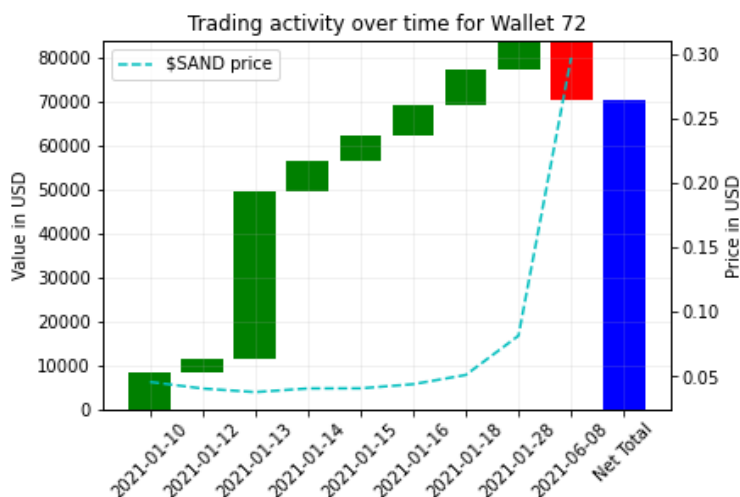
- Wallet holder #56 was also an early adopter of SAND (see below chart).
- When the price of SAND dramatically increased again in a separate breakthrough rally, Wallet #56 sold a substantial portion of its holdings suggesting a BEARISH sentiment.
- Wallet 56 took substantial profits as the price of SAND reached all time highs at $4.50 in this wallet holder's investment timeframe.


Trading activity over time for Wallet 56
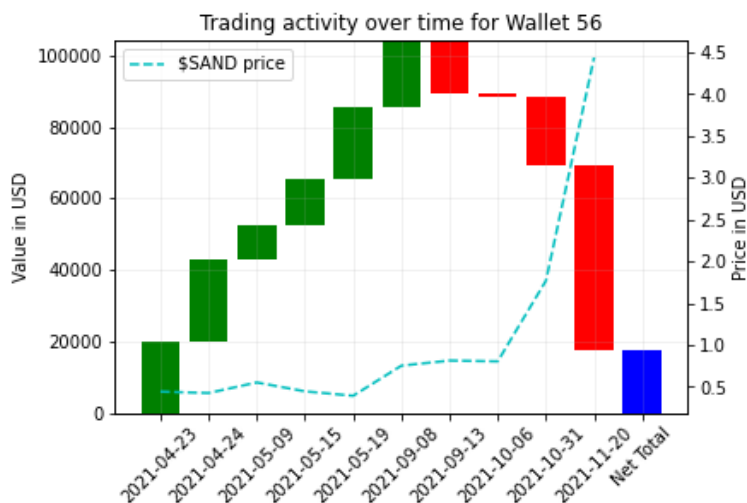
FINAL INSIGHTS

As we reflect on our project, it became clear just how important having a robust dataset is in order to conduct rigorous analysis. Even from the beginning, constructing our dataframes and pulling data from the API required planning and thinking ahead to ensure we have the right information and also that we had enough. From 18,000 rows of data, we manipulated and filtered it down to less than 200 after our meticulous sanity checks, leaving us with little data to be able to work with. This is also likely a consequence of us choosing SAND to analyze because the inception of the coin was only last year. This meant that the timeframe of the transactions collected is relatively short compared to what industry professionals typically analyze for regular financial assets, which have several more years of historical data. For an asset with a short time span that also displayed astronomical growth, it was hard to truly measure top performers as there were a lot of outliers. That said, we were inspired by our previous analysis and deep dive into our case studies and after further analysis, came across the below insights.

1. **Many wallets only began acquiring SAND in October 2021, after Facebook changed its strategic direction towards the metaverse.**

We looked at the distribution of the holding periods by wallet (see below chart). The median holding period of each wallet in SAND is 61 days but for a high concentration of the wallets, the holding period is less than 30 days. This makes sense because our data sample lasts from August 2020 until November 24th 2021. The biggest and latest price pump in the price of SAND was in the period after October 28th, when Facebook changed its name to Meta. Many of these wallets only began to hold SAND coins after attention to the metaverse increased after this development.
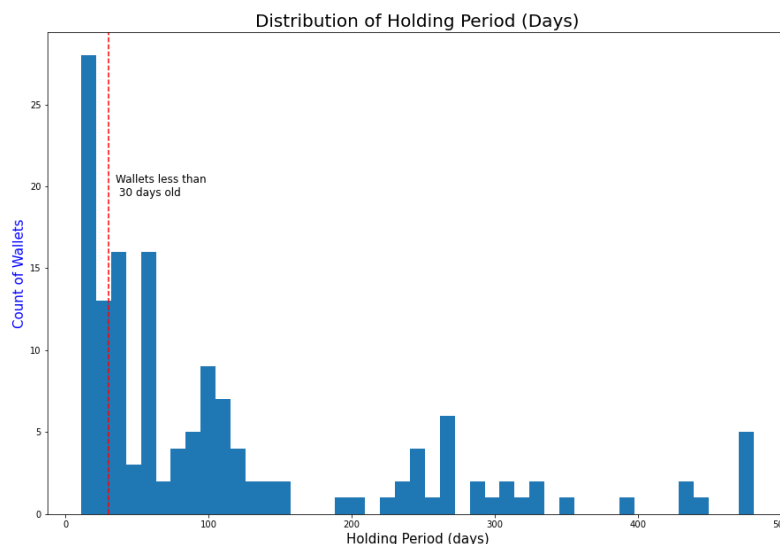

2. **Cryptocurrencies remain an emerging asset, with SAND among the latest additions**

Another angle of analysis we did is to compare the skewness of the SAND coin versus more mature cryptocurrencies such as BTC and ETH, in addition to the S&P 500. Skewness is a very important concept in finance because it provides the probability of infrequent but highly consequential moves.

The plot below shows, the conventional financial asset, the S&P500 is the least fat-tailed, followed by BTC, then ETH and finally SAND has an extremely fat tail, especially the right tail you see. This means, there were a small number of observations in the movement of SAND that made huge upward jumps,

which reinforces the front slides Sarah showed you regarding the price action.
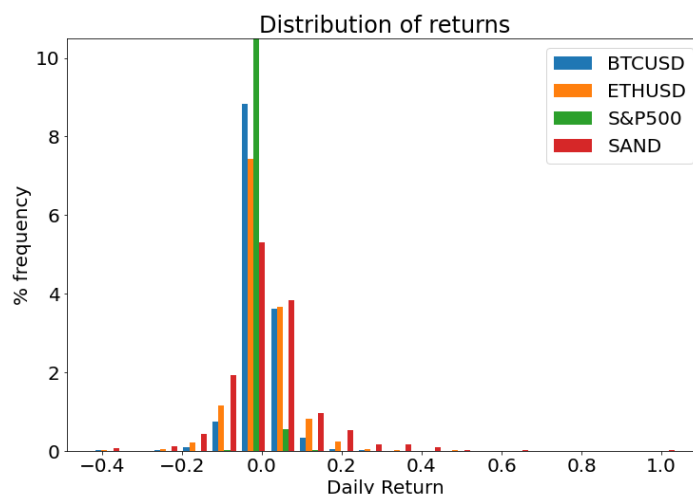
Furthermore, when we graphically show the skewness of our four assets in the below bar graph, according to our sample, the S&P500 and BTC are slightly negatively skewed. This is what you would typically expect for most financial assets, positive gains on most days but sharp sell-offs on selective days (think September 15, 2008 when Lehman Brothers declared bankruptcy or March 2020 when COVID globally). However, we can also see ETH is positively skewed and SAND is very positively skewed. This shows the relative market maturity of the assets. ETH and SAND are still relatively new assets which have generally been on an uptrend since inception, and hence explains why their right tail is larger than their left.

We also see the properties of these assets on the below table on the right. The relative size of the standard deviations also reflect their relative risk and level of market development. Once again, we see SAND being the most volatile, then followed by ETH, BTC and then the S&P 500. The row with the max daily fluctuation shows, the largest 1-day upside move for SAND was 104% while the biggest downside move was only 37%. The vast discrepancy between these two extremes once again shows the property of why is right skewed.
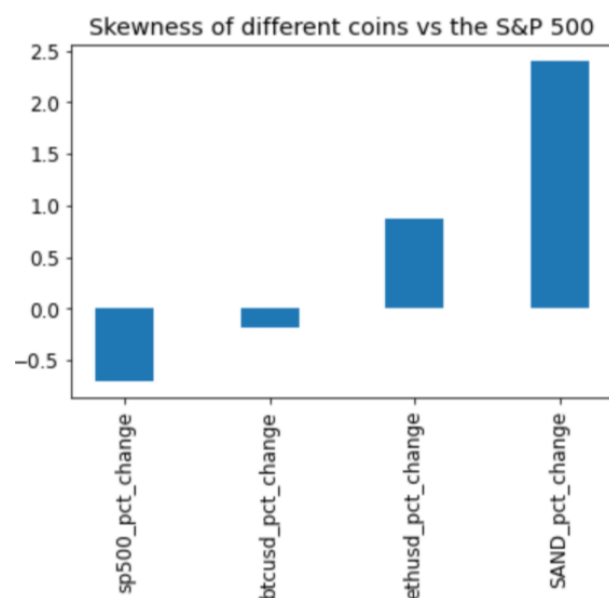
	btcusd_pct_change	ethusd_pct_change	sp500_pct_change	SAND_pct_change
count	1,818.00	1,594.00	1,819.00	473.00
mean	0.00	0.01	0.00	0.02
std	0.04	0.07	0.01	0.12
min	-0.37	-0.42	-0.11	-0.37
25%	-0.01	-0.02	-0.00	-0.04
50%	0.00	0.00	0.00	-0.00
75%	0.02	0.03	0.01	0.05
max	0.25	0.51	0.09	1.04

3. Traditional financial frameworks are still useful in understanding cryptocurrencies

And finally, the last analytic we looked at in the below chart is at the distribution of wallets who achieved "alpha" or essentially beat the passive strategy of buy-and-hold. From our sample, only about 20% of wallets achieved alpha through active management.


```
Number of wallets with POSITIVE alpha: 22
Number of wallets with NEGATIVE alpha: 111
```
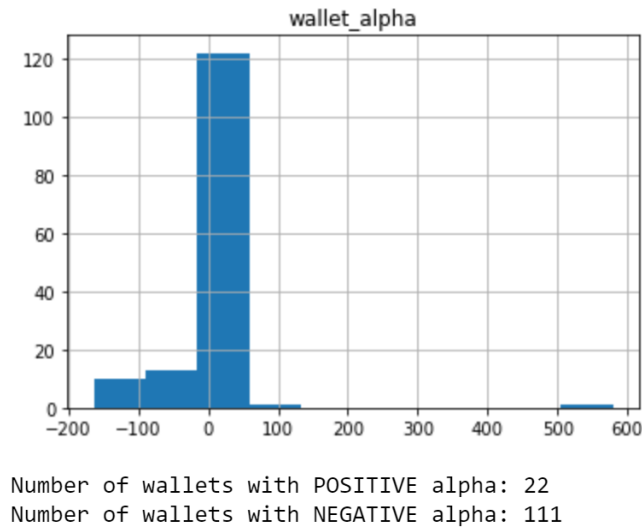
Surprisingly, this isn't too different from the world of traditional finance where about 20% of managers actually deliver alpha. When we look at the median holding period of those wallets that achieved alpha versus those that didn't, the traditional financial wisdom still applies: 1) Get in early; 2) Buy-low-sell-high!

REFERENCES

1. "The Sandbox Price Today, Sand to USD Live, Marketcap and Chart." *CoinMarketCap*, https://coinmarketcap.com/currencies/the-sandbox/.
2. sal19. "Facebook Changes Company Name to Meta." *CNBC*, CNBC, 29 Oct. 2021, https://www.cnbc.com/2021/10/28/facebook-changes-company-name-to-meta.html.
3. Wang, Tracy. "Sand, Mana Tokens Surged in November as Crypto Traders Bet on 'Metaverse' Potential." *CoinDesk Latest Headlines RSS*, CoinDesk, 1 Dec. 2021, https://www.coindesk.com/markets/2021/12/01/sand-mana-tokens-surged-in-november-as-crypto-traders-bet-on-metaverse-potential/.
4. Moffitt, Chris. "Creating a Waterfall Chart in Python." *Practical Business Python Atom*, 18 Nov. 2014, https://pbpython.com/waterfall-chart.html.
5. Fernando, Jason. "Sharpe Ratio Definition." *Investopedia*, Investopedia, 7 Dec. 2021, https://www.investopedia.com/terms/s/sharperatio.asp.