

Comparing Natural
Language Processing and
trends based algorithmic
trading strategies with a
basic mean reversion
strategy.



1. NLP based algorithmic trading strategy: trading cryptocurrencies with Twitter.

- A) Idea
- B) Code
- C) Pros and cons
- D) Areas for improvement

2. Trends based algorithmic trading strategy: trading stocks with Reddit

- A) Idea
- B) Code
- C) Pros and cons
- D) Areas for improvement

3. Mean reversion algorithmic strategy: Trading stocks with the Simple Moving Average

- A) Idea
- B) Code
- C) Pros and cons
- D) Areas for improvement

4. Comparing results

5. Conclusion

Note from the author:

These algorithms have not been tested with real money, nor backtested (because there is not such data for the first 2 algorithms) and do not take in count fees and liquidity issues. The first bot has been running on a free server (PythonAnywhere) during several days. This service provider has terminated several times the bot (because it took too many CPU) and the bot could run at full speed only 100 seconds each 24 hours (because of CPU management). The 2 other algorithms have been tested on Google Colab. All 3 programs can be executed at a specific time on a server on a daily basis, for instance on PythonAnywhere, for free (create 3 accounts). The tests have been execute on a really small amount of days, but it can be done once again on a larger scale quite easily, especially with a good server.

1. NLP based algorithmic trading strategy: trading cryptocurrencies with Twitter.

A) Idea

Let's imagine a class of 100 students. Let's imagine the professor asks the class if they like Tesla at 10:00. 60% will say yes. Then the professor asks again at 10:30. 70% say yes. That means that there is an increase in the sentiment about Tesla. Now he asks at 11:00 and only 20% say yes, that's a decrease.

This trading bot says: increase in sentiment → Buy, decrease → Sell.

NLP stands for Natural Language Processing. In Python there is some good libraries that provide NLP, like TextBlob or Vader. I will use TextBlob.

The idea is to select 1 crypto (I chose Bitcoin, but it can be almost anything, details in 1.C), to identify its underlying sentiment, by using sentiment analysis with NLP. Then to check once again the sentiments trends. If it increases, we buy X shares / coins. Once the position is open, we check again the sentiments movements. If it decreases, we close the position.

B) Code

```
1  import tweepy
2  import time
3  from textblob import TextBlob
4  import yfinance as yf
5  #####
6  ckey = ""
7  csecret = ""
8  atoken = ""
9  atsecret = ""
10 nb = 500 # Choose how many tweets you want the bot to analyze
11 keywords = ["BTC", "#BTC", "Bitcoin"] # Choose keywords related to the asset
    you want to trade
12 #####
13
14
15 auth = tweepy.OAuthHandler(ckey, csecret)
16 auth.set_access_token(atoken, atsecret)
17 api2 = tweepy.API(auth, wait_on_rate_limit=True,
18 wait_on_rate_limit_notify=True)
19
20
21 def perc(a,b):
22     temp = 100 * float(a) / float(b)
23     return format(temp, '.2f')
```

```

24 def get_current_price(symbol):
25     ticker = yf.Ticker(symbol)
26     todays_data = ticker.history(period='1d')
27     return todays_data['Close'][0]
28
29 def get_twitter_BTC():
30     ratios = 0
31
32     for keyword in keywords:
33         tweets = tweepy.Cursor(api2.search, q=keyword, lang="en").items(nb)
34
35         pos = 0
36         neg = 0
37
38         for tweet in tweets:
39             analysis = TextBlob(tweet.text)
40
41             if 0 <= analysis.sentiment.polarity <= 1:
42                 pos += 1
43             elif -1 <= analysis.sentiment.polarity < 0:
44                 neg += 1
45
46         pos = perc(pos, nb)
47         neg = perc(neg, nb)
48
49         if float(neg) > 0:
50             ratio = float(pos) / float(neg)
51         else:
52             ratio = float(pos)
53         ratios += ratio
54
55     return ratios
56
57
58 if __name__ == "__main__":
59
60     for k in range(1000):
61         score = get_twitter_BTC()
62         min1 = score + (score * 30 / 100)
63         time.sleep(60*5)
64         new_score = get_twitter_BTC()
65
66         if new_score > min1:
67             btc_price = get_current_price("BTC-USD")
68             buy = "\nBUY : " + str(btc_price)
69             f = open("output.txt", "a")
70             f.write(buy)
71             f.close()
72
73             time.sleep(60*5)
74             new_new_score = get_twitter_BTC()

```

```

73         min2 = new_score - (new_score * 30 / 100)
74
75         if new_new_score < min2:
76             new_btc_price = get_current_price("BTC-USD")
77             sell_at = "SELL : " + str(new_btc_price)
78             trade_profit = new_btc_price - btc_price
79             perc_profit = trade_profit / btc_price * 100
80             perc_profit_round = round(perc_profit, 3)
81             sell_message = sell_at + " | " + " Profit = " +
82 str(perc_profit_round) + " %"
83             f = open("output.txt", "a")
84             f.write(sell_message)
85             f.close()
86             time.sleep(60*5)
87
88         else:
89             time.sleep(60*5)

```

Comment:

L20-22: getting percentages

L24-27: getting current price of a ticker

L28-52: returns a number that corresponds to the current sentiment concerning all the keywords entered before.

L58-87: first check the current sentiment, sleeps, check once again. If new sentiment > previous one: buy. Sleep, check once again, if new sentiment < previous one: sell.

C) Pros and cons

What I like about that algorithm is that we can modulate it how we want. We can choose how "strict" the conditions must be to open or close a position. The "strictness" corresponds to how much the new sentiment analysis must be different from the previous one. For instance, we can say that a 5% increase is not an increase, and so we don't buy anything.

The main problem is clearly related to NLP. How NLP deals with noise, ironic or sarcastic ect. content? I have no idea, so I am quite pessimistic with that approach. Also, this approach must be proven by a correlation between prices and sentiment. Which brings me to the last part.

D) Areas for improvement

One could try to get a large dataset of sentiments upon an asset within a fixed time scope. Once it is done, one could compare it with the market price of that asset, and check for correlation. What would be really interesting is

that if the sentiments trends **precede** price movements. If not, most of the profit and losses from this bot were random. I wasn't able to do this because there is query rate limit which doesn't allow me to get the data each second.

To make the bot fully automated, one must take in count the fact that when the sentiment only increases, positions are held open, so assets are stacked. So, one must find the right moment to close the remaining positions. One moment could be to close when the price hits the higher position held + 0.1 % for instance.

2. Trends based algorithmic trading strategy: trading stocks with Reddit

A) Idea

Here, we will get the 20 stocks that are the most spoken about in Reddit. Once it is done, we will buy 1000 shares of them, and then close all the positions once the balance hits about 0.1% more from the previous one. The first part of the code is from asad70 on github, who wanted to use Vader for sentiment analysis. I used it and saw a lack of correlation, which makes no sense to buy only those with good compound scores.

B) Code

The part from line 9 to line 41 is taken from Asad70's Github (<https://github.com/asad70/wallstreetbets-sentiment-analysis/blob/master/Wallstreetbets.py>), this part gives me the 20 most mentioned stocks from Reddit. I modified it and replaced his sentiment analysis part (which I consider useless) with a direct connection to Alpaca's API.

```
import praw
1 import alpaca_trade_api as tradeapi
import time
2 import pandas as pd

3 APCA_API_KEY_ID = ""
  APCA_API_SECRET_KEY = ""
4 APCA_API_BASE_URL = "https://paper-api.alpaca.markets"
  api = tradeapi.REST(APCA_API_KEY_ID, APCA_API_SECRET_KEY,
5 APCA_API_BASE_URL, api_version='v2')
  account = api.get_account()
6

7 reddit = praw.Reddit(client_id='6cq9yDMVtyRcLA',
                        client_secret='cpwIv0_xvj-3rXILTQeU0KjNwh-7ew',
8                        user_agent='ksjdhsaq',
```

```

                                check_for_async = False)
9
us = {this list contains about 10 pages of us stocks}
1
0 # includes common words and words used on wsb that are also stock
  names
1 blacklist = {This list contains common words not to mess with stocks
1 names}

1 '''#####
2 #####'''
  # set the program parameters
1 subs = ['wallstreetbets', 'stocks', 'investing', 'stockmarket']      #
3 sub-reddit to search
post_flairs = {'Daily Discussion', 'Weekend Discussion', 'Discussion'}
1 # posts flairs to search || None flair is automatically considered
4 goodAuth = {'AutoModerator'}    # authors whom comments are allowed
  more than once
1 uniqueCmt = True                # allow one comment per author per
5 symbol
  ignoreAuthP = {'example'}        # authors to ignore for posts
1 ignoreAuthC = {'example'}        # authors to ignore for comment
6 upvoteRatio = 0.70              # upvote ratio for post to be considered,
  0.70 = 70%
1 ups = 20                        # define # of upvotes, post is considered if upvotes
7 exceed this #
  limit = 10                       # define the limit, comments 'replace more' limit
1 upvotes = 2                     # define # of upvotes, comment is considered if
8 upvotes exceed this #
  picks = 20                       # define # of picks here, prints as "Top ## picks are:"
1 '''#####
9 #####'''

2 posts, count, c_analyzed, tickers, titles, a_comments = 0, 0, 0, {},
0 [], {}
  cmt_auth = {}
2
1 for sub in subs:
    subreddit = reddit.subreddit(sub)
2    hot_python = subreddit.hot()    # sorting posts by hot
2    # Extracting comments, symbols from subreddit
    for submission in hot_python:
2        flair = submission.link_flair_text
3        author = submission.author.name

2        # checking: post upvote ratio # of upvotes, post flair, and
4 author

```



```

    if submission.upvote_ratio >= upvoteRatio and submission.ups >
2 ups and (flair in post_flairs or flair is None) and author not in
5 ignoreAuthP:
        submission.comment_sort = 'new'
2        comments = submission.comments
6        titles.append(submission.title)
        posts += 1
2        submission.comments.replace_more(limit=limit)
7        for comment in comments:
            # try except for deleted account?
2            try: auth = comment.author.name
8            except: pass
            c_analyzed += 1
2
9            # checking: comment upvotes and author
            if comment.score > upvotes and auth not in
3 ignoreAuthC:
0                split = comment.body.split(" ")
                for word in split:
3                    word = word.replace("$", "")
1                    # upper = ticker, length of ticker <= 5,
excluded words,
3                    if word.isupper() and len(word) <= 5 and word
2 not in blacklist and word in us:

3                    # unique comments, try/except for key
3 errors
                    if uniqueCmt and auth not in goodAuth:
3                        try:
4                            if auth in cmt_auth[word]: break
                        except: pass
3
5                    # counting tickers
                    if word in tickers:
3                        tickers[word] += 1
6                        a_comments[word].append(comment.body)
                        cmt_auth[word].append(auth)
3                        count += 1
7                    else:
3                        tickers[word] = 1
                        cmt_auth[word] = [auth]
8                        a_comments[word] = [comment.body]
                        count += 1
3
9 # sorts the dictionary
    symbols = dict(sorted(tickers.items(), key=lambda item: item[1],
4 reverse = True))
0 top_picks = list(symbols.keys())[0:picks]

```



```

times = []
4 top = []
1 for i in top_picks:
    times.append(symbols[i])
4     top.append(i)
2
    print("Stocks loaded successfully")
4
3 for t in top:
    try:
4         api.submit_order(
4             symbol= t,
4             qty=1000,
4             side='buy',
5             time_in_force='gtc',
4             type='market')
4     except:
6         print(t, "can't buy / or sell")
4         continue
4
7 print("Positions opened successfully")

4 while(True):
8     time.sleep(1)
    account = api.get_account()
4     min = float(account.last_equity) + float(account.last_equity) *
9     0.1/100
    if (float(account.equity) > min):
5         for t in top:
0             try:
                api.submit_order(
5                    symbol= t,
1                    qty=1000,
                    side='sell',
5                    time_in_force='gtc',
2                    type='market')
            except:
5                print(t, "can't close position")
3                continue
            break
5
4 print("Positions closed successfully")
    account = api.get_account()
    print("Profit : $", (float(account.equity) -
        float(account.last_equity)))

```

Comment:

The initial algorithm from Asad70 outputs something like:

```
10 most mentioned picks:
PLTR: 54
TSLA: 33
AAPL: 31
TLRY: 23
GME: 23
AMD: 22
NOK: 19
RKT: 19
NET: 17
MSFT: 16

Sentiment analysis of top 5 picks:
      Bearish Neutral Bullish Total/Compound
PLTR  0.061   0.790   0.150             0.269
TSLA  0.123   0.752   0.125             0.001
AAPL  0.098   0.764   0.138             0.176
TLRY  0.102   0.824   0.074             0.057
GME   0.078   0.804   0.118             0.195
```

As I told, I chose not to use the sentiment analysis part, but only to pick the most mentioned US stocks.

L43-47: Buys all the top stocks (1000 shares, GTC "Good till canceled")

L48-53: Checks every second if the balance has increased by 0.1%. If yes : close the positions

L54: prints the profit

C) Pros and cons

Which is important with that strategy is to understand how correlated to the trend it is. Actually, sometimes Reddit **is** the trend. So, one must switch from buy-then-sell to sell-then-buy according to the current trend. That's why one must first check with both directions (buy-then-sell then sell-then-buy) to invest more money on the most profitable direction. Beyond all that, that strategy is simple, and seems to work quite well.

D) Areas for improvement

To nuance what I just said, sometimes one could do buy-then-sell and then sell-then-buy, and for both, lose money (getting for both sides a negative profit). It happened sometimes to me. For this I don't have ideas, but I think that one would need to create a program which tells the stocks that are making us losing money (whether we are in short or long positions). Once we get them, we just launch the script once again making sure the "bad" stocks are not traded the same way that others.

3. Mean reversion algorithmic strategy: Trading stocks with the Simple Moving Average

A) Idea

This strategy is widely spread and often opposed to the momentum strategy. Mean reversion means that every asset has an average price. If the current market price is lower than the current average, then it should increase. On the other hand, if it is higher than the average, then it should decrease. So, when it should increase, we buy, and when it should decrease, we sell. I made the following algorithm simple in order to compare with the first 2 bots.

B) Code

Note that I chose to trade 10 financial stocks during/post-Covid crisis.

```
1  import pytz
2  import datetime
3  import alpaca_trade_api as tradeapi
4  import investpy
5
6  APCA_API_KEY_ID = ""
7  APCA_API_SECRET_KEY = ""
8  APCA_API_BASE_URL = "https://paper-api.alpaca.markets"
9  api = tradeapi.REST(APCA_API_KEY_ID, APCA_API_SECRET_KEY,
10 APCA_API_BASE_URL, api_version='v2')
11 account = api.get_account()
12
13 def get_price(stock):
14     full_stock_price = api.get_barset(stock, '1Min', limit=1)
15     close_price = full_stock_price.df.iat[0,3]
16     return close_price
17
18 mean_r_universe=["JPM", "HSBC", "C", "WFC", "GS", "MS", "PNC", "USB", "SCHW", "
19 BCS"]
20 portfolio = api.list_positions()
21 open_positions = []
22
23 for position in portfolio:
24     open_positions.append(position.symbol)
25
26 for i in mean_r_universe:
27     utc = datetime.datetime.now(pytz.timezone("UTC"))
28     try:
29         data = investpy.moving_averages(name=i, country='United States',
30 product_type='stock', interval='daily')
```

```

1  except:
6      print("Error with :", i)
1      continue
7  sma = data.iloc[0, 1]
1  price = get_price(i)
8  if price < sma :
1      if i not in open_positions:
9          print("BUY", i, "at", price, utc)
2          try:
0              api.submit_order(
2                  symbol= i,
1                  qty=1000,
2                  side='buy',
2                  time_in_force='gtc',
2                  type='market')
3          except:
2              print("Can't buy", i)
4              continue
2  else:
5      print("SELL", i, "at", price, utc)
2      try:
6          api.submit_order(
2              symbol= i,
7              qty=1000,
2              side='sell',
8              time_in_force='gtc',
2              type='market')
9      except:
3          print("Can't sell", i)
0          continue

```

Comment:

L7-9 : returns the current price of a stock

L11-13: stacking current opened positions (tickers)

L14-30: getting SMA value and market price value for each stock. If $SMA > price$ and position not opened yet : buy. Else if $SMA < price$: sell.

Investpy's moving averages function looks like this :

	period	sma_value	sma_signal	ema_value	ema_signal
0	5	5.279	buy	5.274	buy
1	10	5.211	buy	5.223	buy
2	20	5.120	buy	5.097	buy
3	50	4.767	buy	4.867	buy
4	100	4.584	buy	4.529	buy
5	200	3.856	buy	4.225	buy

C) Pros and cons

This strategy is extremely simple. However, we must adapt our strategy between momentum and mean reversion. I have not done it, but this strategy can be backtested, because it relies on stock prices and movements only. I consider it more of a long-term strategy and would not advise to use it within a market relying on trends.

D) Areas for improvement

In my opinion, one could use all the SMAs (5 days, 10 days...) combined with the EMAs (Exponential Moving Averages) to take decisions. One could also combine all that with other technical indicators (a lot of libraries provide some, like TA-lib)

4. Comparing results

1st Algorithm)

First a decided to buy when there was 40% of sentiment increase:

1 BUY : 34527.50390625	SELL : 34934.4375	Profit = 1.179 %
2 BUY : 34705.78125		
3 BUY : 35875.6328125	SELL : 35901.76953125	Profit = 0.073 %
4 BUY : 35814.38671875	SELL : 35834.828125	Profit = 0.057 %
5 BUY : 35404.38671875	SELL : 35323.16796875	Profit = -0.229 %
6 BUY : 35450.9140625		

40% is quite strict, so I decided to lower it to 30%:

1 BUY : 36145.83984375	SELL : 36124.16796875	Profit = -0.06 %
2 BUY : 36213.40625	SELL : 36168.55859375	Profit = -0.124 %
3 BUY : 36256.453125		
4 BUY : 36699.03515625	SELL : 36729.83984375	Profit = 0.084 %
5 BUY : 36773.92578125		
6 BUY : 36675.89453125	SELL : 36688.19921875	Profit = 0.034 %
7 BUY : 36170.921875	SELL : 36191.08984375	Profit = 0.056 %
8 BUY : 36364.03515625	SELL : 36417.52734375	Profit = 0.147 %
9 BUY : 36865.4296875	SELL : 36680.76171875	Profit = -0.501 %
10 BUY : 36423.2421875	SELL : 36358.07421875	Profit = -0.179 %
11 BUY : 36115.24609375		
12 BUY : 36187.515625		
13 BUY : 36492.41796875		
14 BUY : 36789.21875	SELL : 36842.26953125	Profit = 0.144 %
15 BUY : 36884.6484375	SELL : 37128.74609375	Profit = 0.662 %
16 BUY : 37318.54296875		
17 BUY : 37270.625	SELL : 37278.3046875	Profit = 0.021 %

```

18 BUY : 38088.01953125 SELL : 38038.94921875 | Profit = -0.129 %
19 BUY : 37287.16796875 SELL : 37261.2109375 | Profit = -0.07 %
20 BUY : 38840.984375
21 BUY : 38673.09375 SELL : 38657.91015625 | Profit = -0.039 %
22 BUY : 38840.0859375 SELL : 38811.48046875 | Profit = -0.074 %
23 BUY : 38527.58984375 SELL : 38453.19921875 | Profit = -0.193 %
24 BUY : 37744.546875
25 BUY : 36894.6015625 SELL : 36772.21875 | Profit = -0.332 %
26 BUY : 35931.7734375 SELL : 35859.265625 | Profit = -0.202 %
27 BUY : 36071.19921875
28 BUY : 36106.11328125 SELL : 35970.46484375 | Profit = -0.376 %
29 BUY : 36142.52734375 SELL : 36132.40625 | Profit = -0.028 %
30 BUY : 35119.19140625
31 BUY : 35625.203125 SELL : 35562.3359375 | Profit = -0.176 %
32 BUY : 36173.3984375 SELL : 36164.9375 | Profit = -0.023 %
33 BUY : 36123.421875
34 BUY : 36123.828125 SELL : 36191.2109375 | Profit = 0.187 %
35 BUY : 35903.81640625 SELL : 35875.41796875 | Profit = -0.079 %
36 BUY : 36118.94140625 SELL : 36143.51171875 | Profit = 0.068 %
37 BUY : 35928.0546875 SELL : 35939.76171875 | Profit = 0.033 %
38 BUY : 35620.83984375 SELL : 35634.2890625 | Profit = 0.038 %
39 BUY : 35800.88671875 SELL : 36131.80078125 | Profit = 0.924 %

```

As we can see, there is much more trades when conditions to buy are less strict.

Then I decided to make conditions to sell stricter (only when current sentiment is < previous one – 30%):

```

1 BUY : 36325.25
2 BUY : 36284.109375
3 BUY : 36572.5625
4 BUY : 33764.7265625
5 BUY : 32961.4296875
6 BUY : 32732.67578125
7 BUY : 32850.5859375
8 BUY : 33498.90234375 SELL : 33593.84375 | Profit = 0.283 %
9 BUY : 36810.0625
10 BUY : 37874.4609375
11 BUY : 37934.12109375 SELL : 37778.3359375 | Profit = -0.411 %
12 BUY : 36995.49609375
13 BUY : 36732.31640625
14 BUY : 36306.640625
15 BUY : 37199.54296875 SELL : 37070.07421875 | Profit = -0.348 %
16 BUY : 37016.49609375
17 BUY : 37228.3359375 SELL : 37345.73046875 | Profit = 0.315 %
18 BUY : 37391.3203125

```

As we can see, there is less position closing.

There were 63 opened positions and 36 closed, so about 57% closed positions. Among this 53%, 17 were winning trades, and 19 losing, so about 47% winning trades. So, it looks like randomness was high (which is bad). We can't say yet if the whole algorithm is profitable because we have to sell the 27 remaining Bitcoins.

2nd Algorithm)

Daily program outputs:

Day 1	Day 2	Day 3	Day 4	Day 5
Profit : \$ 640.0799999999581	PE can't buy / or sell Profit : \$ 2228.710000000079	Profit : \$ 383.71999999997206	Algorithm interrupted (always losing money)	PE can't buy / or sell PE can't close position Profit : \$ 600.0

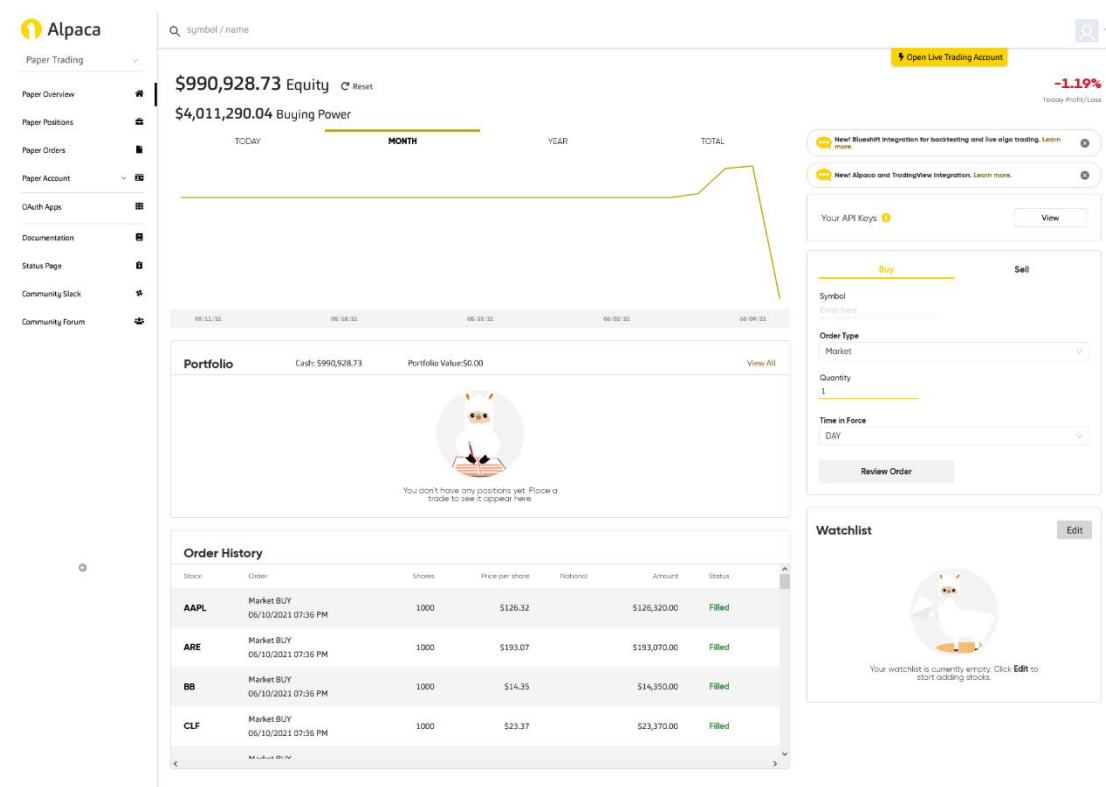
List of open / closed positions (screenshots):

Day 1: <https://ibb.co/tz3pVFD>, Day 2: <https://ibb.co/VmSHJBQ>

Day 3: <https://ibb.co/6R3PZH4>, Day 4: <https://ibb.co/3F7mQrZ>

Day 5: <https://ibb.co/Lnk4KxG>

Day 4 was tough:



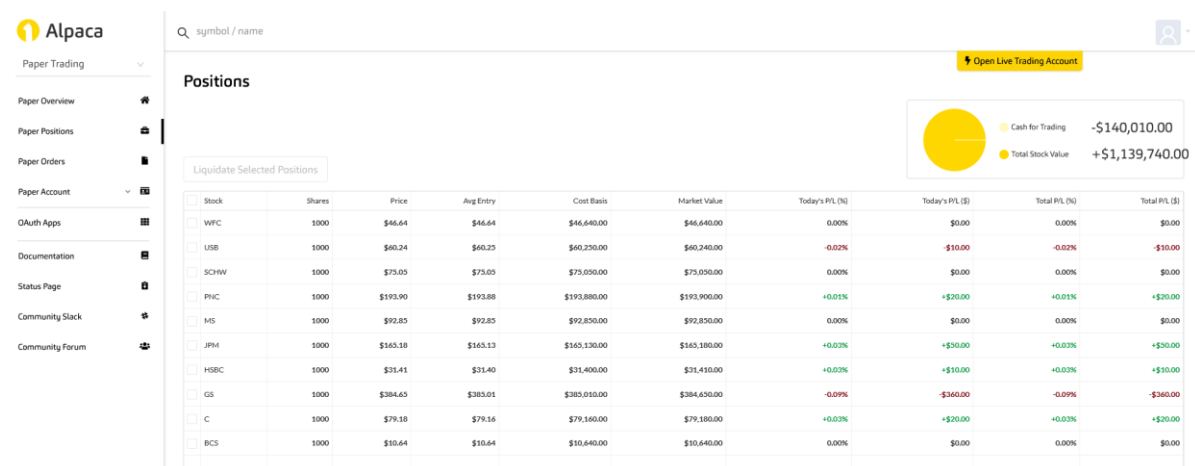
There is more profit on day 2 because I ran the bot several times on this day. I could make the bot even more profitable if the condition to close positions was not so strict (i.e. a higher take profit).

3rd Algorithm)

First (and only) output:

```
1 BUY JPM at 165.15 2021-06-08 19:49:37.499971+00:00
2 BUY HSBC at 31.405 2021-06-08 19:49:38.283278+00:00
3 BUY C at 79.15 2021-06-08 19:49:38.811455+00:00
4 BUY WFC at 46.635 2021-06-08 19:49:39.293030+00:00
5 BUY GS at 385.08 2021-06-08 19:49:39.934558+00:00
6 BUY MS at 92.875 2021-06-08 19:49:40.430601+00:00
7 BUY PNC at 193.94 2021-06-08 19:49:41.108389+00:00
8 BUY USB at 60.25 2021-06-08 19:49:41.676054+00:00
9 BUY SCHW at 75.06 2021-06-08 19:49:42.223928+00:00
10 BUY BCS at 10.635 2021-06-08 19:49:42.849609+00:00
```

Orders on Alpaca:



PNL Dashboards:

Day 2: <https://ibb.co/BP2LrHy>, Day 3: <https://ibb.co/KLkQ3Mf>

Day 4: <https://ibb.co/NsSbH9q>, Global PNL: <https://ibb.co/gRmx9mM>

5. Conclusion

These results have been extracted from a very few days of experimentation, so it must not be taken as a proof of anything. On the overall, all three algorithms have made me lose money, even though, as I said, the first one is not finished yet, as it required some manual actions, and the second one could have been perfect if there was not the problem I encountered the 4th day. The mean reversion bot results were quite bad but seemed to get greater while time passed.

