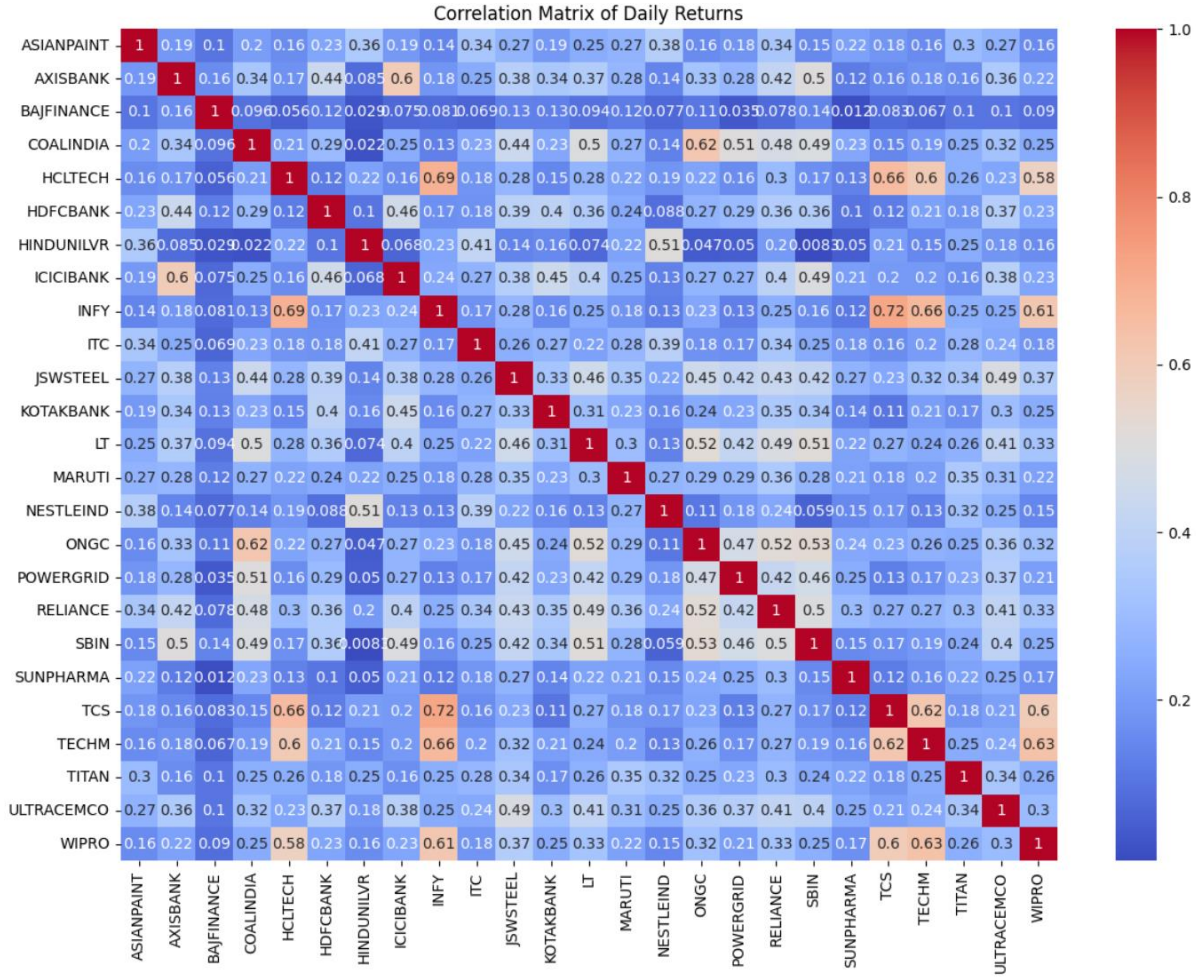
// Pair Trading Project by Ayush Goel

1. Fetch the data of common names in the Indian Stock market.

Name	Date modified	Type	Size
ASIANPAINT.csv	14-06-2025 01:14 PM	Microsoft Excel C...	44 KB
AXISBANK.csv	14-06-2025 01:14 PM	Microsoft Excel C...	42 KB
BAJFINANCE.csv	14-06-2025 01:14 PM	Microsoft Excel C...	43 KB
COALINDIA.csv	14-06-2025 01:14 PM	Microsoft Excel C...	44 KB
HCLTECH.csv	14-06-2025 01:14 PM	Microsoft Excel C...	45 KB
HDFCBANK.csv	14-06-2025 01:14 PM	Microsoft Excel C...	42 KB
HINDUNILVR.csv	14-06-2025 01:14 PM	Microsoft Excel C...	42 KB
ICICIBANK.csv	14-06-2025 01:14 PM	Microsoft Excel C...	43 KB
INFY.csv	14-06-2025 01:14 PM	Microsoft Excel C...	45 KB
ITC.csv	14-06-2025 01:14 PM	Microsoft Excel C...	45 KB
JSWSTEEL.csv	14-06-2025 01:14 PM	Microsoft Excel C...	41 KB
KOTAKBANK.csv	14-06-2025 01:14 PM	Microsoft Excel C...	42 KB
LT.csv	14-06-2025 01:14 PM	Microsoft Excel C...	45 KB
MARUTI.csv	14-06-2025 01:14 PM	Microsoft Excel C...	39 KB
NESTLEIND.csv	14-06-2025 01:14 PM	Microsoft Excel C...	43 KB
ONGC.csv	14-06-2025 01:14 PM	Microsoft Excel C...	45 KB
POWERGRID.csv	14-06-2025 01:14 PM	Microsoft Excel C...	44 KB

OHLC Daily Data from June 2023 to June 2025. (Source: Yahoo Finance)

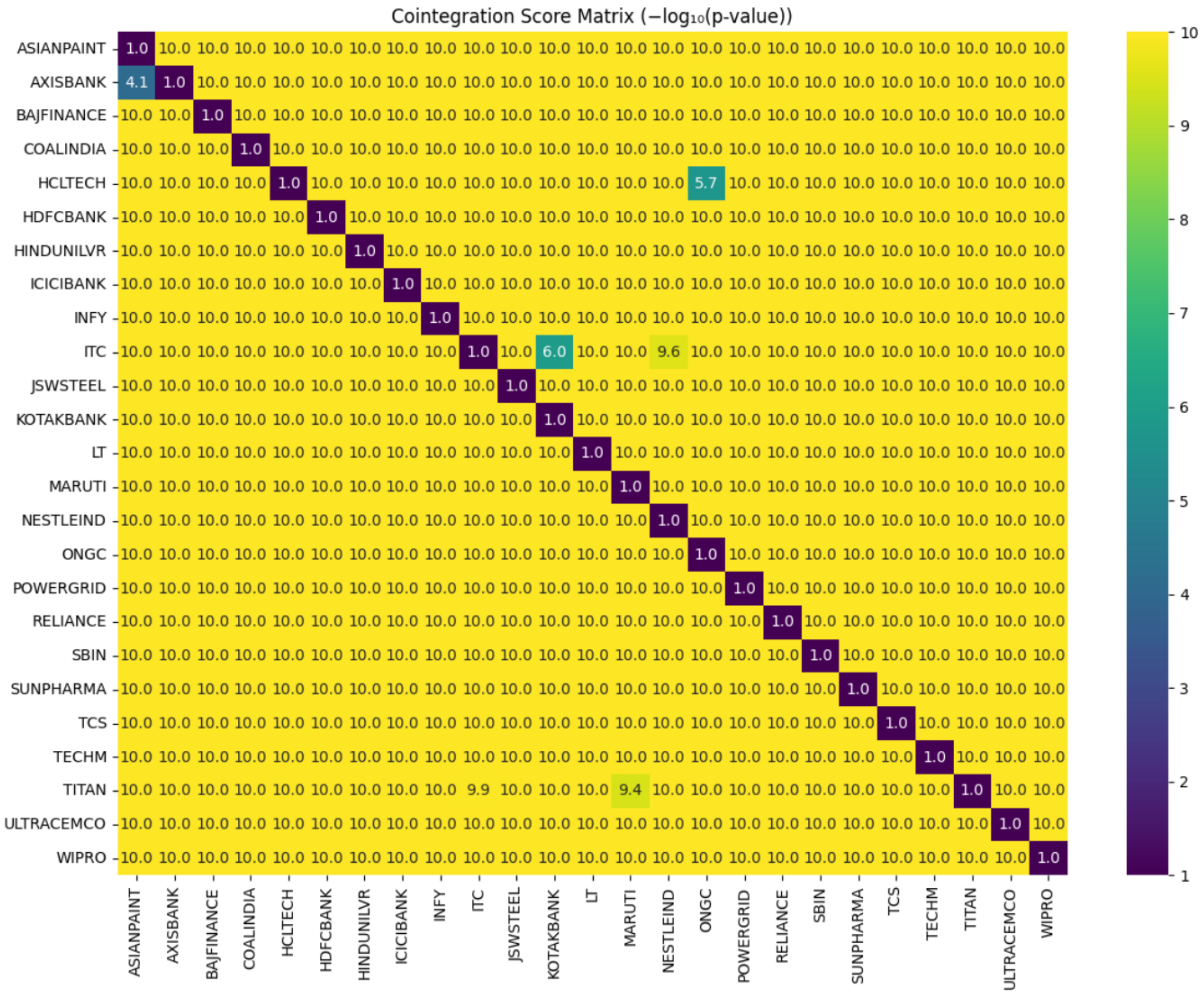
2. Let's see if we can find any kind of correlation in the stocks maybe after that we can go into the co-integration test (engle-granger) – The idea behind correlation is maybe there exists stocks which give similar returns over time (btw all this is done on daily returns of the stock)



Interpretation:

Stock A	Stock B	Correlation
INFY	**HCLTECH**	0.69
INFY	**TCS**	0.66
TECHM	**TCS**	0.62
TECHM	**INFY**	0.6
ONGC	**COALINDIA**	0.62
HDFCBANK	**AXISBANK**	0.5
RELIANCE	**ONGC**	0.53
LT	**ULTRACEMCO**	0.51

We found some nicely co-related stocks now, 69% doesn't sound a lot but judging by the sentiment driven Indian market, over a 2-year period it seems pretty song. But, before any conclusion we will now perform the co-integration test which is a much better metric than correlation.



The negative log values of p-values are taken from the engle granger test as they are very small values and hard to visualize on the heatmap.

This heatmap makes our hypothesis very strong.

We have found some beautifully cointegrated pairs.

Pair	Score ($-\log_{10}(p)$)
INFY – TCS	9.6
INFY – TECHM	9.4
INFY – HCLTECH	5.7
ITC – INFY	6
AXISBANK – HDFCBANK	4.1

Infosys and TCS feels correct as they both are IT companies and work in basically the same sector.

Infosys and Tech Mahindra Again very expected to give such good results.

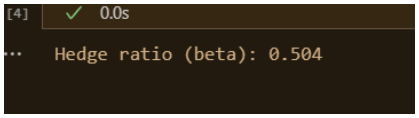
Infosys and HCL Tech same.

ITC and Infosys this is weird and maybe just a fluke or a false positive because they work in different sectors and getting even a decent co-integration feels absurd but hey you never know before you check deeper.

Axis Bank and HDFC well expected.

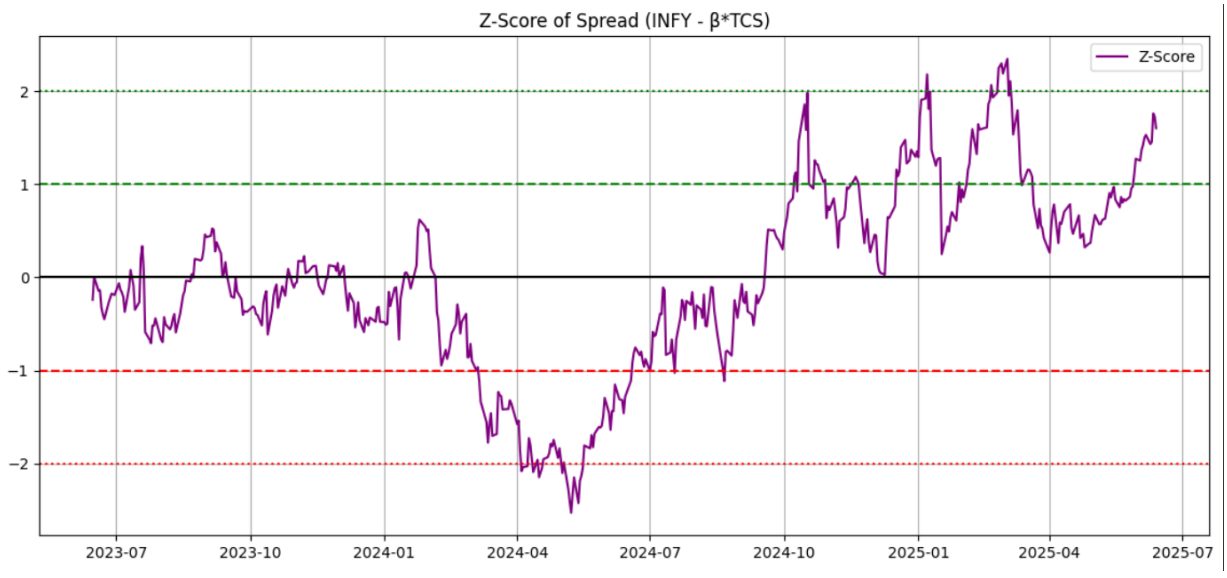
Now let's dive into each pair and see what's happening.

Pair 1 : Infosys and TCS



Hedge Ratio calculated using OLS Regression. (Industry Standard)

Not the absolute best but lets see if we can try something.



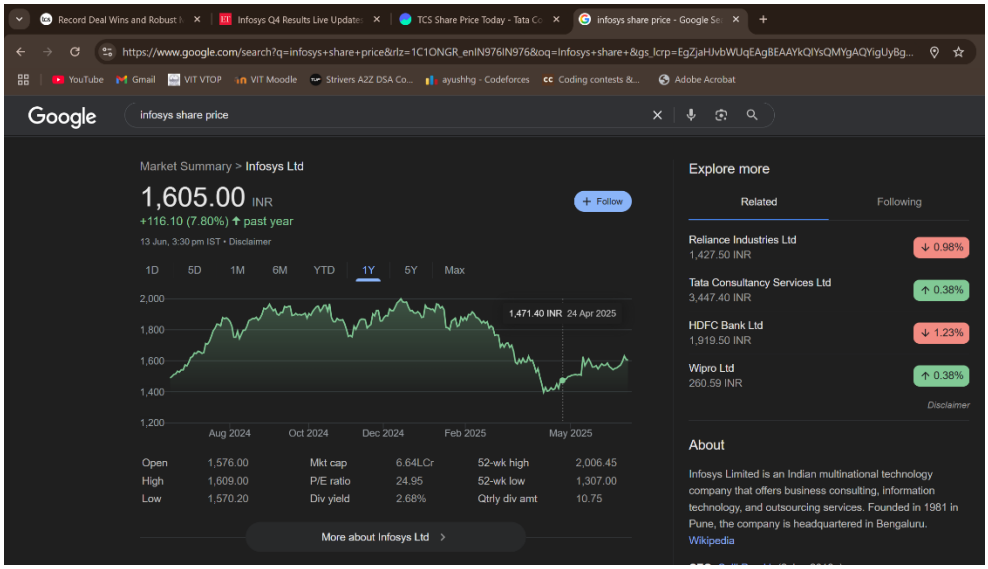
Okay we can clearly see there is some Rhyme and Rhythm in this spread.

2023 Q3 and 2024 Q1 almost gave a very nice mean reversion which continued until the exile of Feb – 2024.

The spread closed around October 2024 and has been nicely mean reverted till now.

(Quick Research Time)

Well, No need to research Sep/Oct 2024 onwards the Indian market has been on the biggest bullish trend ever (Sarcasm if not obvious),



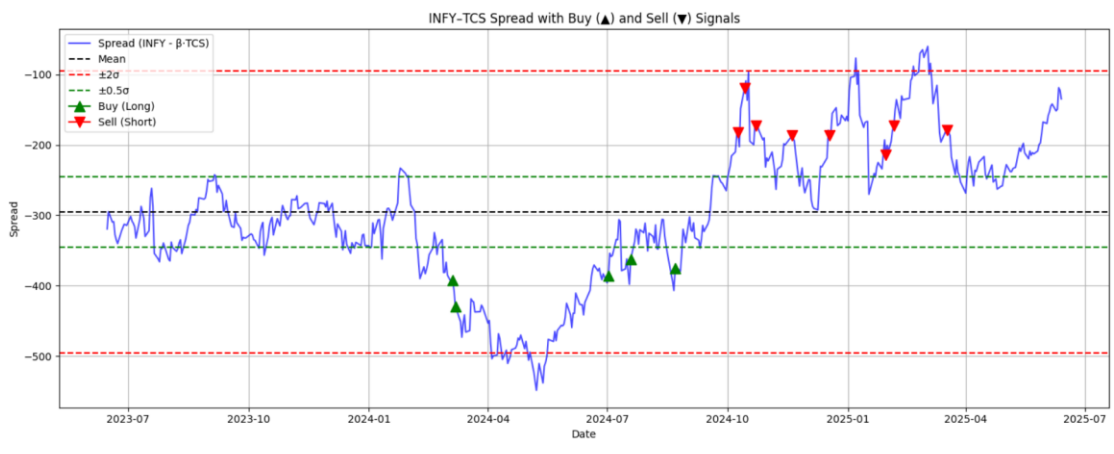
(Due to Time constrain I'll have to limit my research here)

We should now Backtest the strategy

	EntryDate	ExitDate	Direction	PnL	HoldingDays
0	2024-04-05	2024-04-10	Long	30.176987	5
1	2024-04-15	2024-04-16	Long	20.178678	1
2	2024-04-18	2024-04-22	Long	18.174221	4
3	2024-05-03	2024-05-06	Long	16.014473	3
4	2024-05-07	2024-05-21	Long	40.053538	14
5	2025-01-08	2025-01-09	Short	28.510140	1
6	2025-02-21	2025-02-24	Short	4.986939	3
7	2025-02-27	2025-03-05	Short	16.492511	6
8	2025-03-06	2025-03-07	Short	33.216726	1

Total PnL: 207.8042137508686
Total Trades: 9

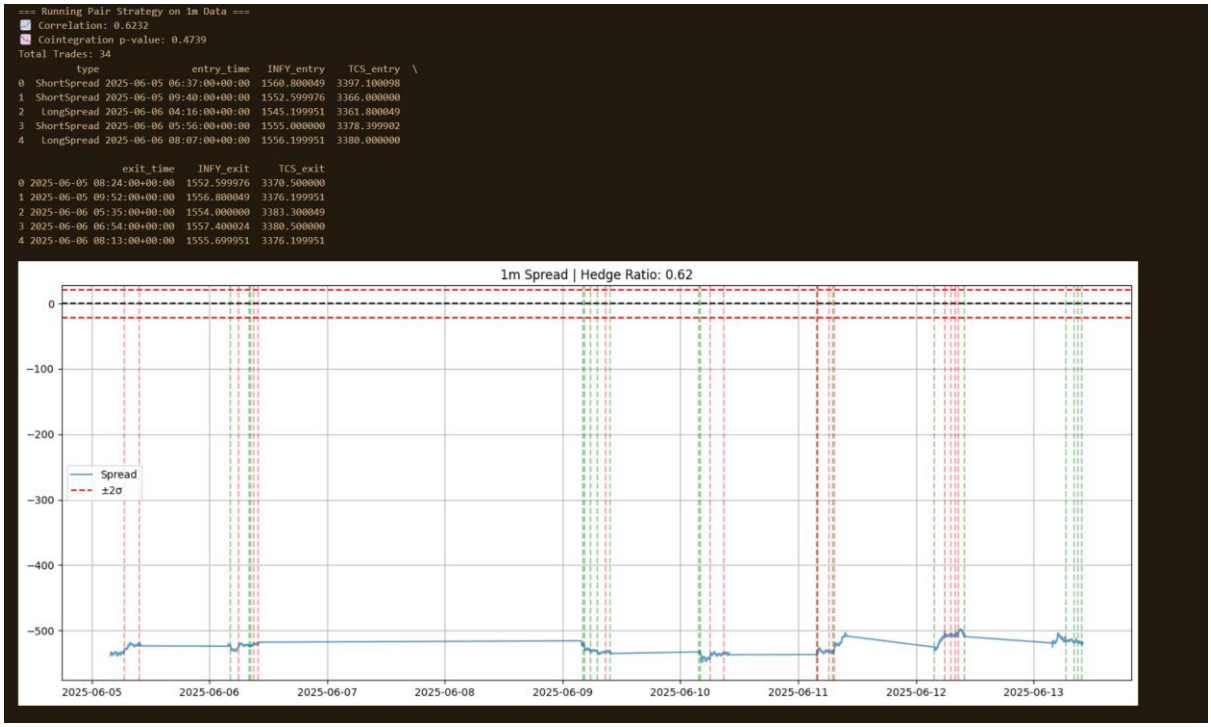
Since the strategy is working on daily data the results seem Underwhelming. But the fact that we did not encounter any negative trade means the co-integration is real.



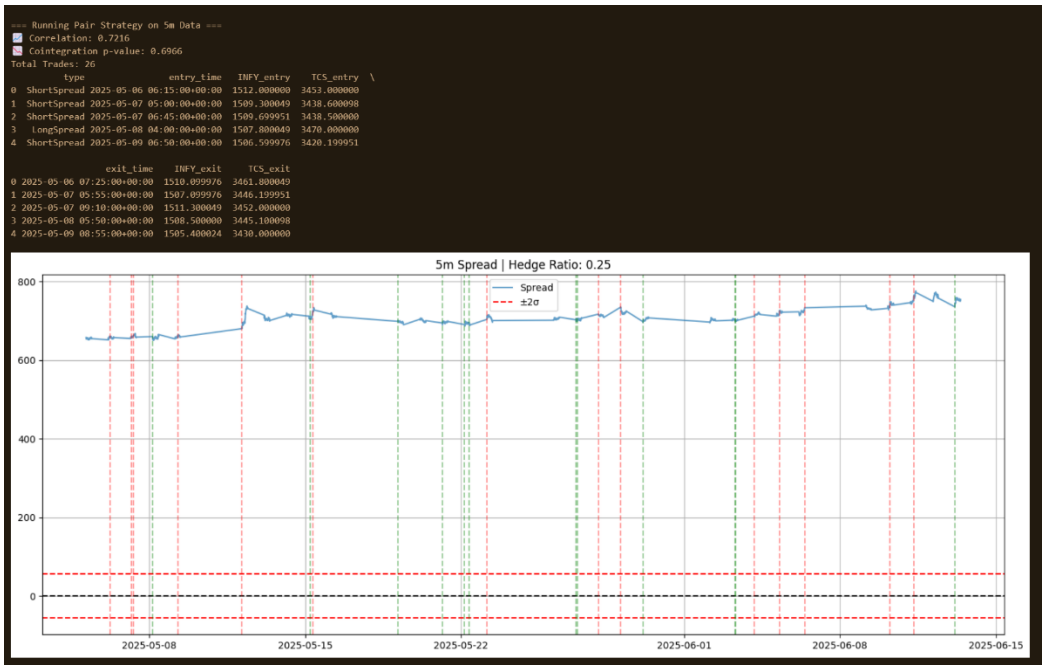
Now what can we do further?

- The strategy is far from ready we have to go into smaller data frames and find if there is any rhyme and rhythm
- We need to test for Filtering logics to find when the strategy can and cannot run
- Also, We need to find the different backtesting benchmarks to see if the strategy is any good.

Now, we will go into smaller time frames and see for co-integration.



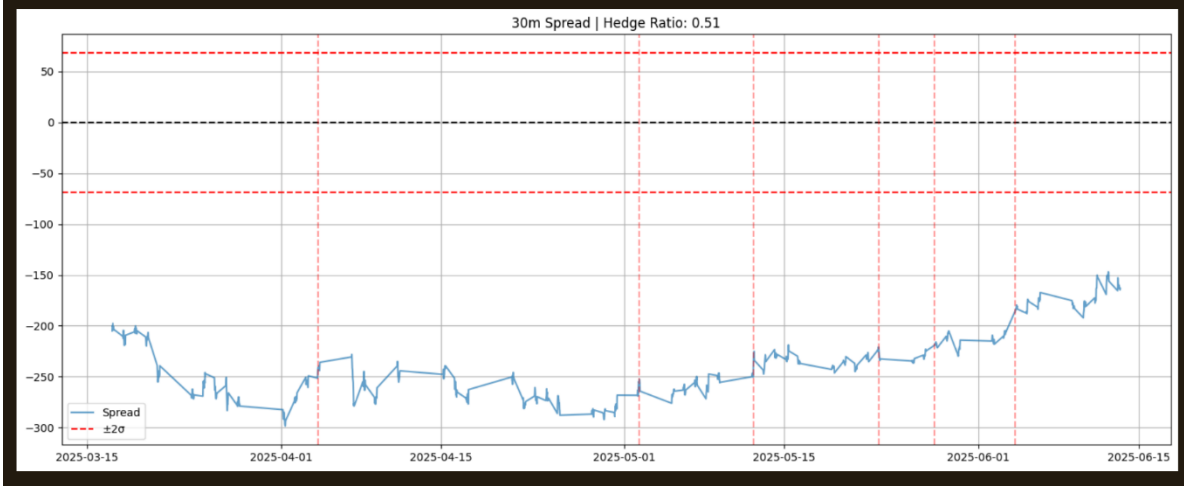
So, 1 Minute is a mess and we can't trade it.



5 minute is again underwhelming

```
=== Running Pair Strategy on 30m Data ===
📈 Correlation: 0.8154
📉 Cointegration p-value: 0.7923
Total Trades: 6
         type               entry_time   INFY_entry    TCS_entry  \
0  ShortSpread 2025-04-04 05:30:00+00:00  1457.000000  3313.300049
1  ShortSpread 2025-05-02 08:00:00+00:00  1510.900024  3439.100098
2  ShortSpread 2025-05-12 07:30:00+00:00  1608.900024  3598.199951
3  ShortSpread 2025-05-23 06:00:00+00:00  1577.000000  3506.199951
4  ShortSpread 2025-05-28 04:00:00+00:00  1577.500000  3503.600098

                  exit_time    INFY_exit     TCS_exit
0 2025-04-07 08:00:00+00:00  1379.699951  3208.050049
1 2025-05-05 04:00:00+00:00  1517.099976  3497.399902
2 2025-05-16 08:30:00+00:00  1586.800049  3558.100098
3 2025-05-23 08:00:00+00:00  1573.199951  3518.600098
4 2025-05-30 04:00:00+00:00  1563.000000  3492.199951
```



30 Minute is a nicer timeframe, it his highly correlated and can be further used to calculate mispricing and take smarter trades. (One strategy could be if a EMA crossover on 30m time frames gives a leading indication for the lagging stock to move behind – but again speculation)

```
=== Running Pair Strategy on 60m Data ===
📈 Correlation: 0.8200
📉 Cointegration p-value: 0.8053
Total Trades: 3
         type               entry_time   INFY_entry    TCS_entry  \
0   LongSpread 2025-04-25 08:45:00+00:00  1480.900024  3451.899902
1  ShortSpread 2025-05-12 06:45:00+00:00  1608.000000  3602.100098
2  ShortSpread 2025-05-28 05:45:00+00:00  1578.300049  3497.000000

                  exit_time    INFY_exit     TCS_exit
0 2025-04-30 08:45:00+00:00  1500.300049  3449.800049
1 2025-05-19 09:45:00+00:00  1557.900024  3518.800049
2 2025-05-30 03:45:00+00:00  1561.099976  3488.899902
```

Similar results on 60m Timeframe.

Future Prospect.

- Breaking both pairs into Microstructure to find any rhythm
- Find if any parts of the two pairs are leading lagging
- How does news and events affect the price of two.
- Does spikes in one spread to the other
- Looking at the cointegration between there constant volume candle
- How does 2Sigma moves correlates.