



COINTEGRATION AND CLUSTERING BASED APPROACH TO PAIR TRADING OF STOCKS OF THE INDIAN STOCK MARKET



Aayush Sahu
Anirudh Pratap Singh
Ashutosh Pradhan
Gopad Shukla
Saif Rizvi
Umesh Bugata

PRESENTATION FRAMEWORK

Key talking points



PAIRS TRADING

1. Introduction

2. Business Problem

3. Methodology

4. Results

5. Demo

6. Future Scope

FINANCIAL MARKET AND PAIRS TRADING

NSE and BSE - The Facilitators

- Financial instruments are exchanged and prices are determined
- Act as a common ground for buyers and sellers



STOCK EXCHANGE



Pairs Trading of Stocks

- Involves undertaking long and short positions in stocks having high correlation





WHAT WE AIM TO ACHIEVE

Aid investors in taking the right decision

We aim to enable users take the right decisions and make profit from virtually any market condition - uptrend, downtrend, or sideways movement.





Get the stock pairs in which you can invest and become the owner of your dream car!!! Or get bag full of money to stash in your basement.

Just browse the website and get Highly recommended stock pairs. Want to dare, then visit our analysis tab for more risk associated stock pairs. You a market nerd, then explore the explanation tab for detailed explanation



Highly recommended stocks

Analysis

Sectoral analysis with Cluster assisted Cointegration approach

Choose the sector:

IT

Testing Period(Months): 12

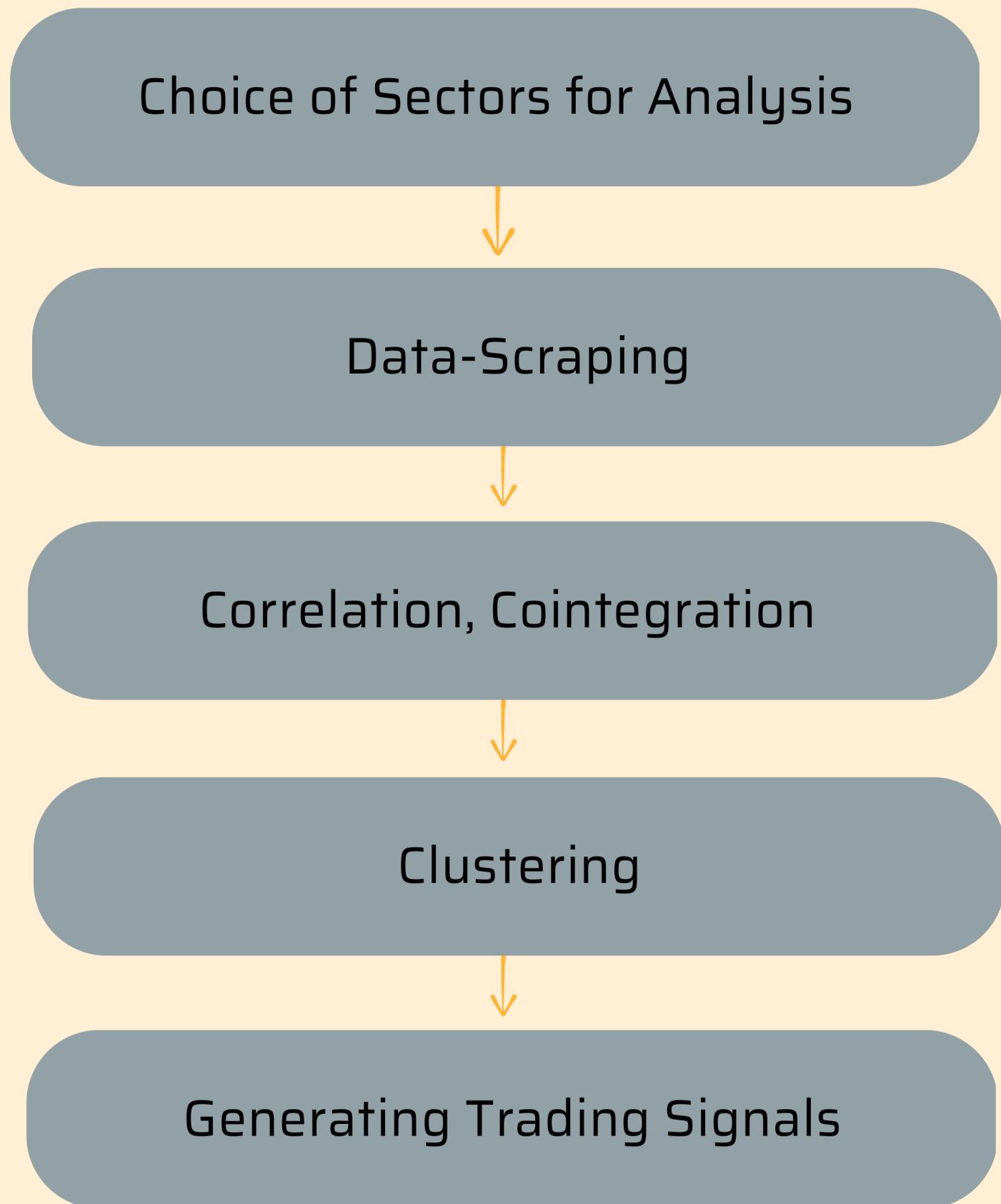
Initial investment amount: 200000

SUBMIT TO FIND THE BEST STOCK PAIRS TO INVEST

	Stock Pairs	Initial Investment	Final Value	Profit	Return(%)
1	(HCLTECH, INFY)	200000	440399.162476	400399.162476	100.00%
2	(INFY, TCS)	200000	425684.038520	25684.038520	12.84%

Our website helps investors select the best stock pairs for different sectors of the Indian Stock Market

METHODOLOGY



Points of Investment Opportunities



COINTEGRATION

Top 10 stocks from different sectors



Data from Yahoo Finance



Coint() - Python



OLS Regression

Generating Trading Signals



Visualizing Signals and Positions

CLUSTERING

Top 10 stocks from different sectors



Data from Yahoo Finance



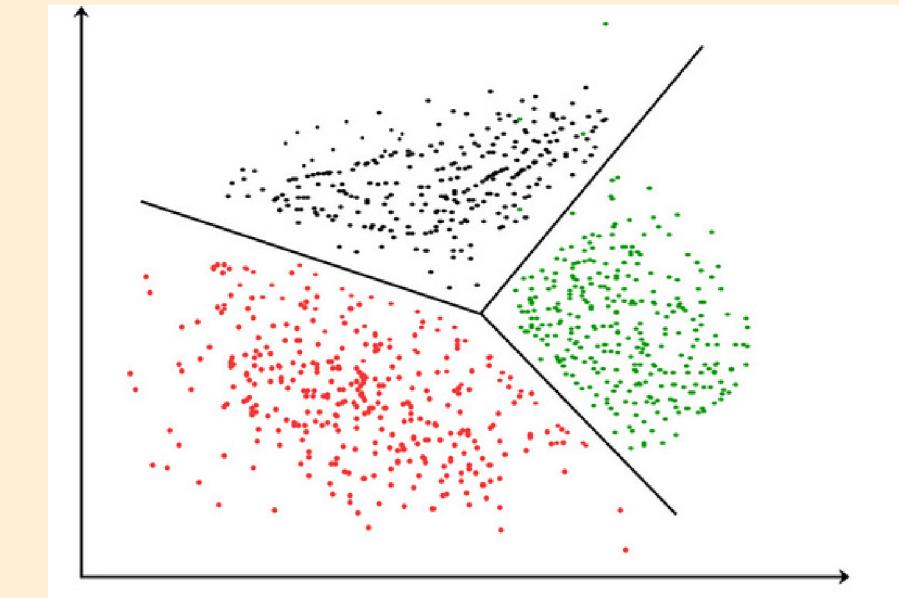
Return and Volatility as variables



K-Means, Hierarchical, Affinity
Propagation



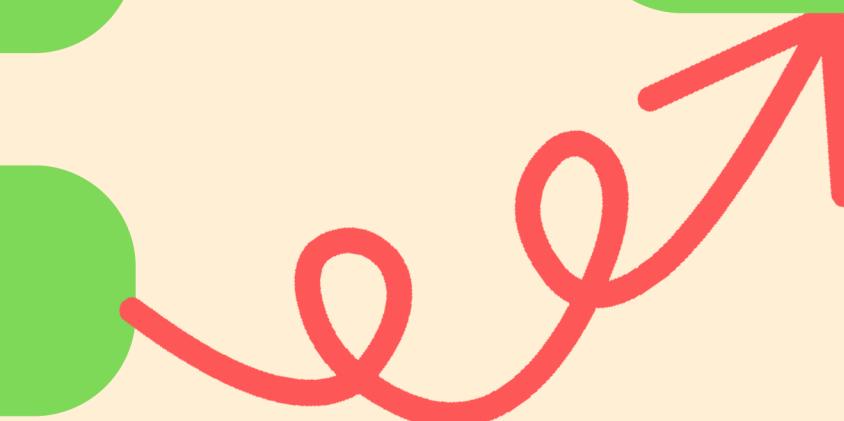
Cluster Evaluation



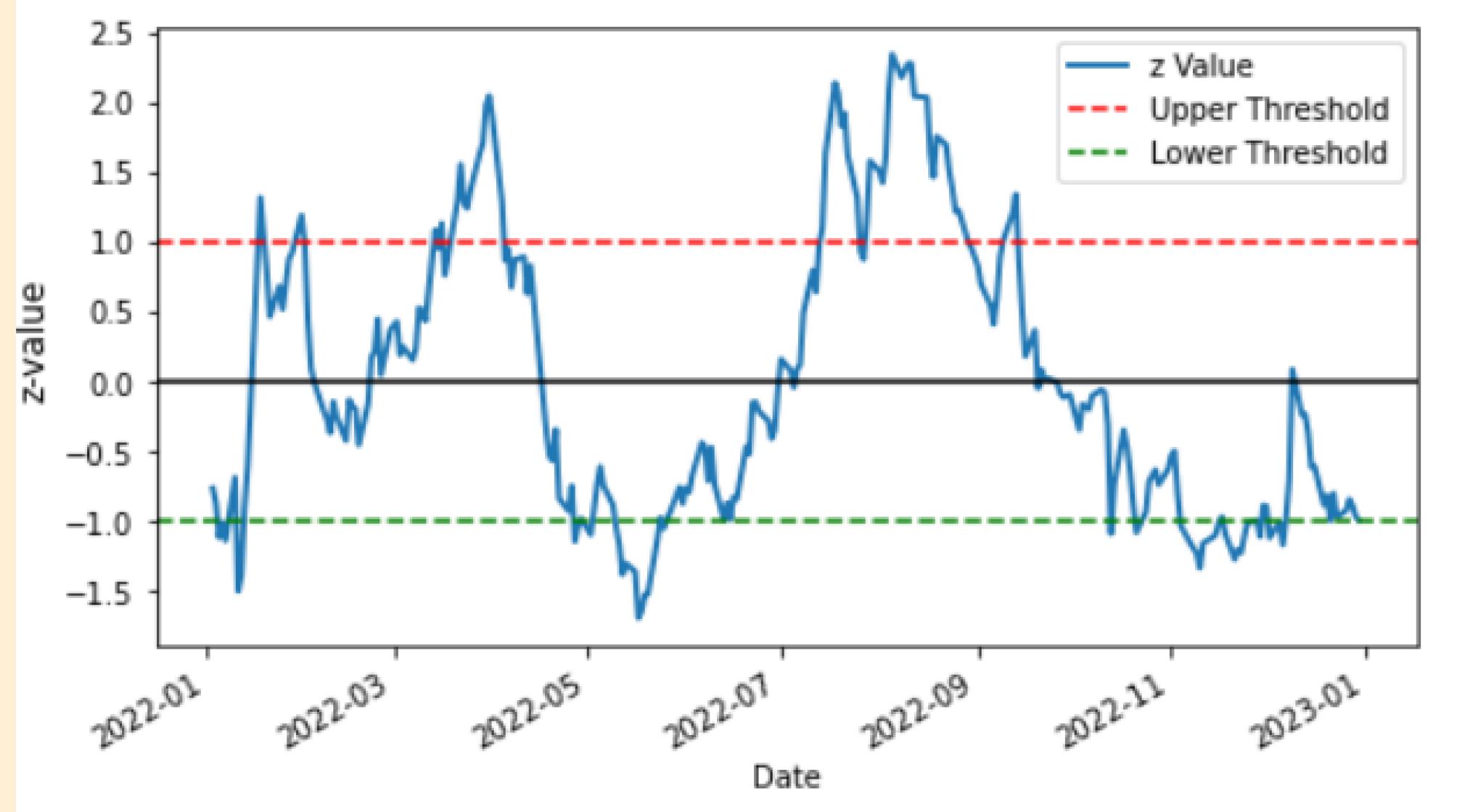
Obtaining the common pairs of stocks



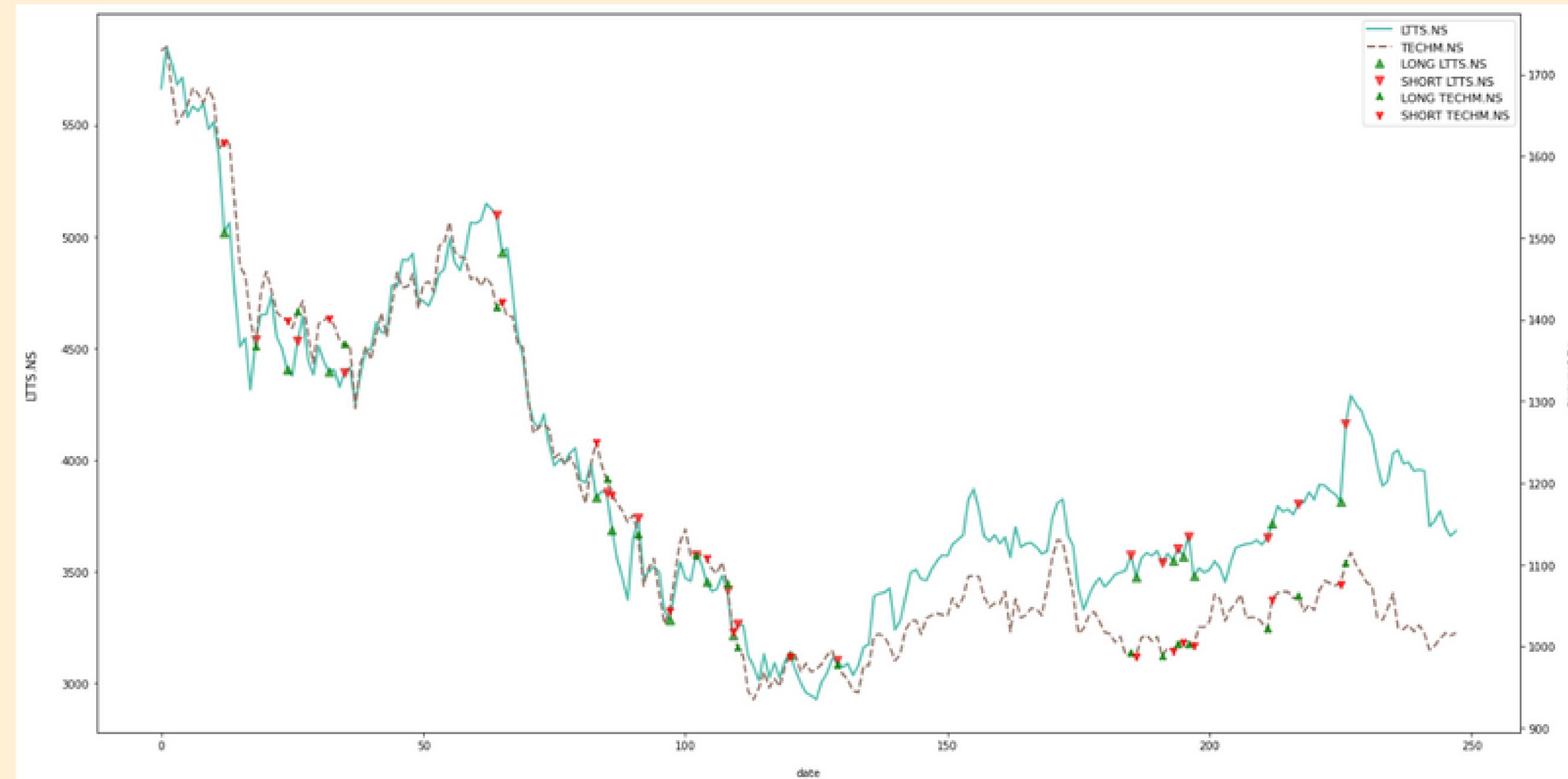
Clustering technique chosen



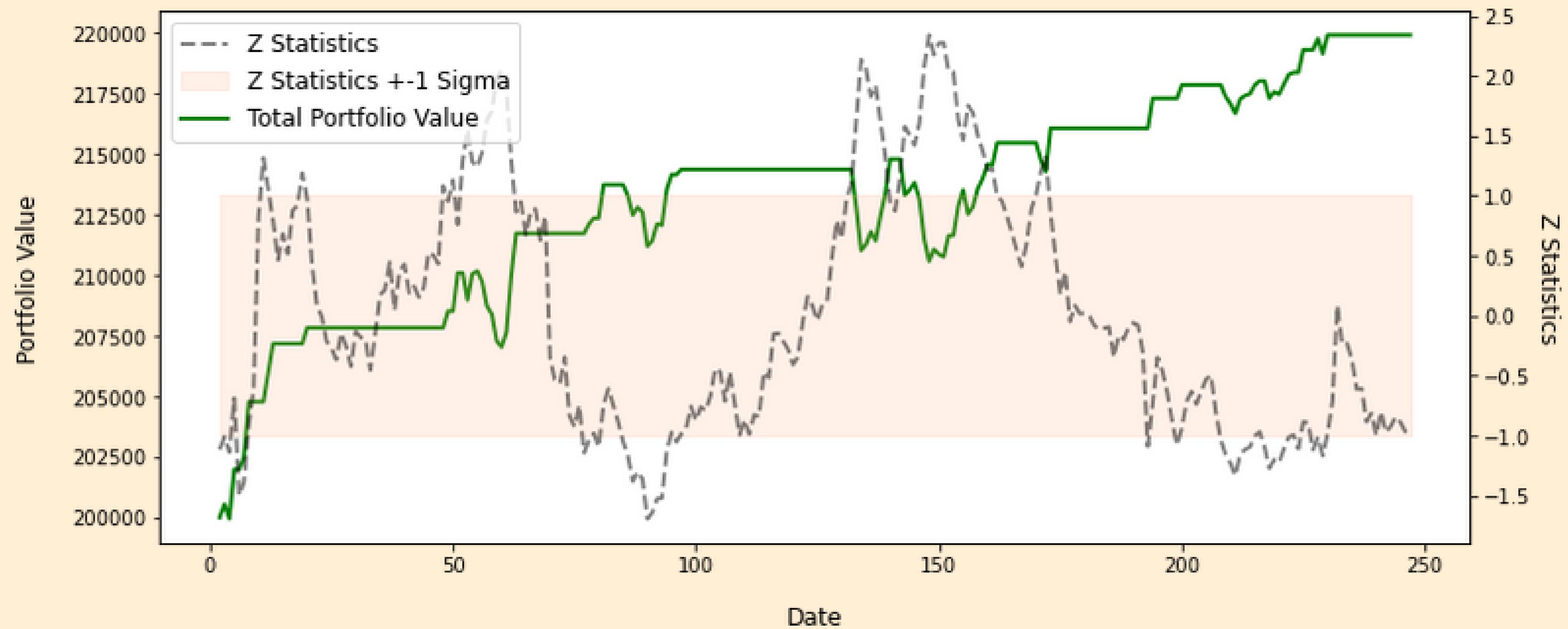
RESULTS



IT Sector Signal Generation



IT Sector Trading signals and Positions



IT Sector Portfolio Valuation

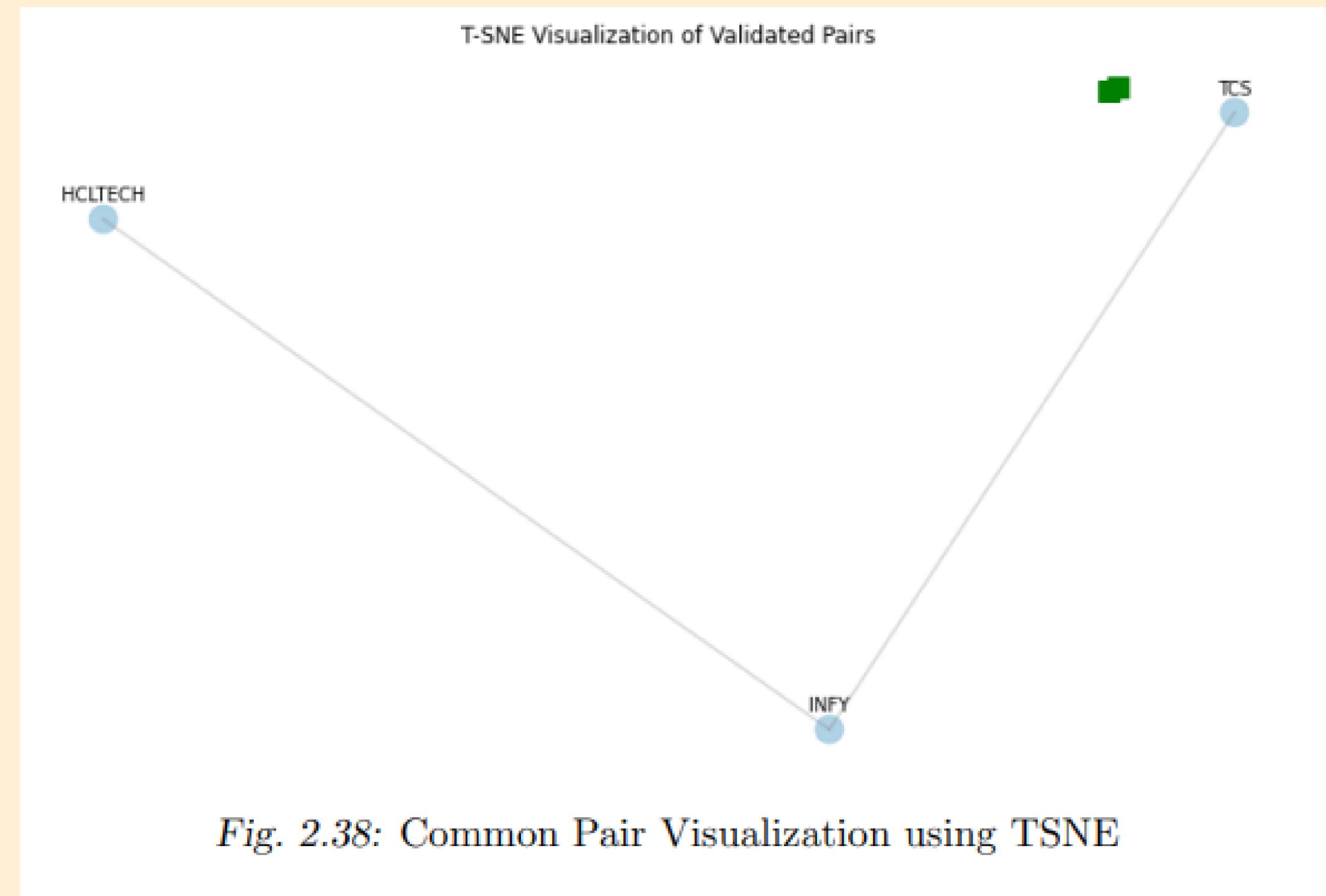
**Tab. 2.4: THE ANNUAL RETURNS OF THE PAIR-TRADING PORTFOLIOS
OF THE IT SECTOR STOCKS (PERIOD: JANUARY 1, 2022 - DE-
CEMBER 31, 2022).**

Stock pairs	Initial Investment	Final value	Profit	Return
HCLTECH INFY	200000	440109.19	40109.19	10.03
INFY TCS	200000	425684.03	25684.03	6.42
LTTS TECHMAHINDRA	200000	470378.45	70378.45	17.59

Stock Time Series for Cluster 1



No. of clusters - 2
Common stock pairs - [(HCLTECH,INFY) and (INFY,TCS)]



DIGITAL

Demo

CONCLUSION

- **Sectors with remarkably Positive Returns**
Oil & Gas, Pharma, Media, IT
- **Sectors with Negative Returns**
Metal Sector(1), FMCG(1), NIFTY50(11)
- **12 Trigger points on an Average**
Window of investment opportunities - 1 month

FUTURE SCOPE OF WORK

- **In-depth analysis of stock pairs with negative returns**
- **Extension to inter-sectorial analysis as well as to Global Markets**
- **Integrating Predictive Algorithms**

THANK YOU

```
from sklearn import metrics
print("km", metrics.silhouette_score(X, kmeans.labels_, metric='euclidean'))
print("hc", metrics.silhouette_score(X, hc.fit_predict(X), metric='euclidean'))
print("ap", metrics.silhouette_score(X, ap.labels_, metric='euclidean'))
```

km 0.6151468263219049

hc 0.571871877599359

ap 0.571871877599359

IT Sector Silhouette Scores

```
# function to find cointegrated pairs
def find_cointegrated_pairs(data):
    n = data.shape[1]
    pvalue_matrix = np.ones((n, n))
    keys = data.keys()
    pairs = []
    for i in range(n):
        for j in range(i+1, n):
            result = coint(data[keys[i]], data[keys[j]])
            pvalue_matrix[i, j] = result[1]
            if result[1] < 0.05:
                pairs.append((keys[i], keys[j]))
    return pvalue_matrix, pairs
```

```
pvalues, pairs = find_cointegrated_pairs(train_data)
```

```
# Cointegrating p
print(f"There are {len(pairs)} cointegrated pairs and they are:")
for i in range(len(pairs)):
    print(pairs[i])
```

There are 3 cointegrated pairs and they are:

```
('HCLTECH.NS', 'INFY.NS')
('INFY.NS', 'TCS.NS')
('LTTS.NS', 'TECHM.NS')
```

IT Sector Cointegrated Pairs

```
In [41]: # conduct Augmented Dickey-Fuller test  
adf = adfuller(residual, maxlag = 1)  
  
# Test statistics  
print('Test Statistic = ', adf[0])  
  
# probability critical values  
print(adf[4])
```

```
Test Statistic = -3.094412142160525  
{'1%': -3.4369994990319355, '5%': -2.8644757356011743, '10%': -2.5683331327427803}
```

IT Sector ADF

OLS Regression Results

```
=====
```

Dep. Variable:	asset2	R-squared (uncentered):	0.994
Model:	OLS	Adj. R-squared (uncentered):	0.994
Method:	Least Squares	F-statistic:	1.546e+05
Date:	Wed, 19 Apr 2023	Prob (F-statistic):	0.00
Time:	16:46:43	Log-Likelihood:	-5688.2
No. Observations:	987	AIC:	1.138e+04
Df Residuals:	986	BIC:	1.138e+04
Df Model:	1		
Covariance Type:	nonrobust		

```
=====
```

	coef	std err	t	P> t	[0.025	0.975]
asset1	1.4341	0.004	393.136	0.000	1.427	1.441

```
=====
```

Omnibus:	160.359	Durbin-Watson:	0.036
Prob(Omnibus):	0.000	Jarque-Bera (JB):	252.373
Skew:	1.079	Prob(JB):	1.58e-55
Kurtosis:	4.218	Cond. No.	1.00

```
=====
```

IT Sector OLS Regression