



Citi Code Challenge 2022

Approach Paper

College Name: MKSSS's Cummins College of Engineering for Women, Pune.

Team Name: Hash_Paav

Overview

➤ **Describe your understanding on problem statement**

To identify the pairs that have maximum correlation of one instrument with other, so that the Important parameter which have influence on the stock price are not missed using some indicators, based upon whose values we can decide which stock among the pair to sell or buy and thus increase our cumulative profit.

➤ **Team members**

Komal Vhanmane

Mrunal Jambenal

Shreya Uttarwar

Mukta Joshi

High Level Solution Approach

➤ **Describe Solution**

1. Data Pre-processing

Data Pre-processing is needed to organise and clean the data.

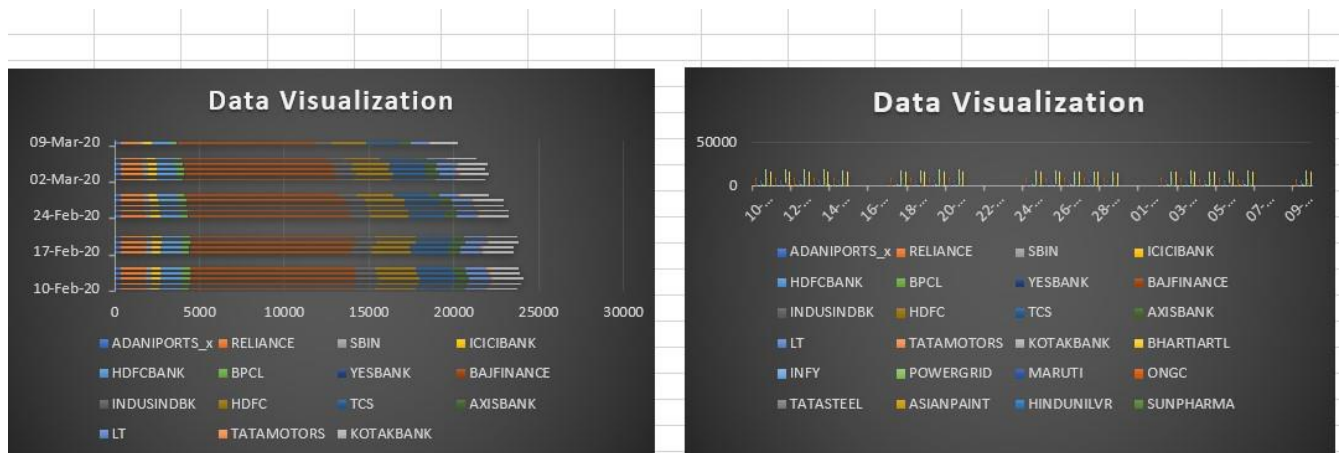
After data pre-processing

| | Date | ADANI PORTS_x | RELIANCE | SBIN | ICICIBANK | HDFCBANK | BPCL |
|----|-----------|---------------|----------|--------|-----------|----------|--------|
| 0 | 09-Mar-20 | 321.90 | 1114.15 | 253.45 | 457.75 | 1107.30 | 423.85 |
| 1 | 06-Mar-20 | 341.40 | 1271.00 | 270.50 | 486.35 | 1134.90 | 403.10 |
| 2 | 05-Mar-20 | 349.15 | 1311.15 | 288.50 | 504.50 | 1151.35 | 418.40 |
| 3 | 04-Mar-20 | 345.95 | 1339.70 | 285.30 | 508.35 | 1148.85 | 418.70 |
| 4 | 03-Mar-20 | 347.30 | 1342.85 | 289.85 | 514.75 | 1181.80 | 425.00 |
| 5 | 02-Mar-20 | 340.65 | 1316.15 | 287.40 | 506.10 | 1179.60 | 412.35 |
| 6 | 28-Feb-20 | 342.20 | 1328.65 | 303.00 | 497.25 | 1177.65 | 426.35 |
| 7 | 27-Feb-20 | 353.70 | 1386.25 | 321.95 | 515.35 | 1199.45 | 444.05 |
| 8 | 26-Feb-20 | 358.40 | 1392.00 | 328.20 | 523.70 | 1199.25 | 448.80 |
| 9 | 25-Feb-20 | 358.40 | 1416.40 | 326.80 | 530.95 | 1200.30 | 456.75 |
| 10 | 24-Feb-20 | 362.25 | 1444.95 | 322.95 | 529.85 | 1209.95 | 464.65 |
| 11 | 20-Feb-20 | 370.85 | 1485.95 | 327.65 | 547.00 | 1217.10 | 471.75 |
| 12 | 19-Feb-20 | 371.30 | 1503.80 | 320.35 | 544.80 | 1227.20 | 475.55 |
| 13 | 18-Feb-20 | 366.60 | 1467.40 | 317.55 | 541.20 | 1213.25 | 470.75 |
| 14 | 17-Feb-20 | 362.80 | 1478.25 | 314.20 | 541.60 | 1217.15 | 461.65 |
| 15 | 14-Feb-20 | 366.85 | 1487.60 | 319.40 | 545.80 | 1219.35 | 476.45 |
| 16 | 13-Feb-20 | 369.10 | 1474.15 | 327.45 | 541.00 | 1241.40 | 469.00 |
| 17 | 12-Feb-20 | 374.10 | 1470.30 | 320.20 | 549.30 | 1249.00 | 474.75 |
| 18 | 11-Feb-20 | 374.20 | 1452.75 | 324.35 | 539.75 | 1240.60 | 481.00 |
| 19 | 10-Feb-20 | 368.25 | 1435.60 | 318.50 | 533.95 | 1240.30 | 485.45 |

2. Data Visualization

Data visualization is to make it easier to identify patterns, trends and outliers in large data sets.

In the 2nd graph X-axis is representing date and Y-axis representing as closing prices of stocks on that date.



In the 1st graph X-axis is represented as closing prices of stocks and Y-axis is represented as date. We can interpret fluctuations of closing prices of particular company stocks.

3. Finding Correlation coefficient Matrix/Cointegration Matrix

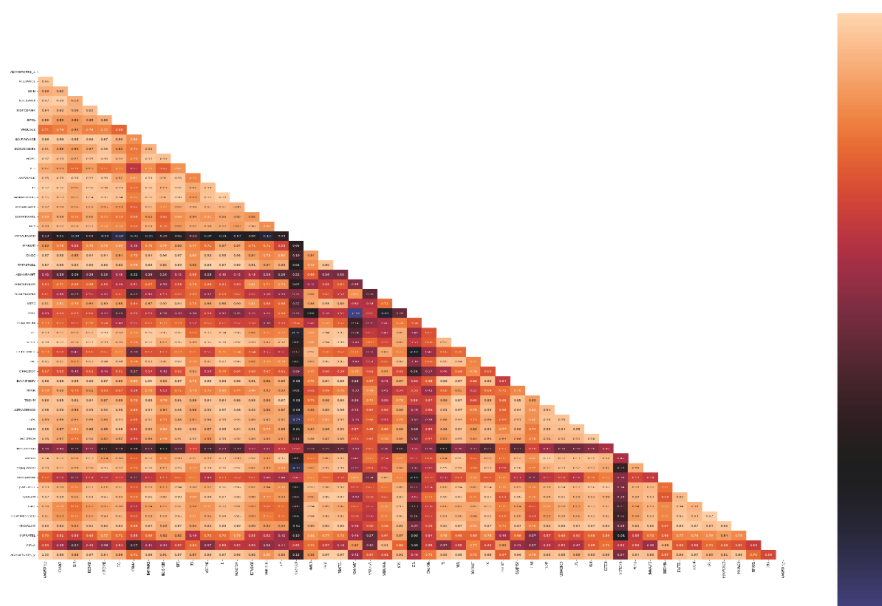
In general terms, the Pearson Correlation Coefficient is a measure of how two random variables X & Y are linearly representative. The Pearson Correlation

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

The formula above is the measure of the normalized covariance

Cointegration:

Cointegration is a statistical test to understand if two time series behave similarly to one another in the long term. Formally, cointegration occurs when two-time series





In the below given graphs,

First graph represents the line plot of 'CLOSE' prices of AXISBANK and

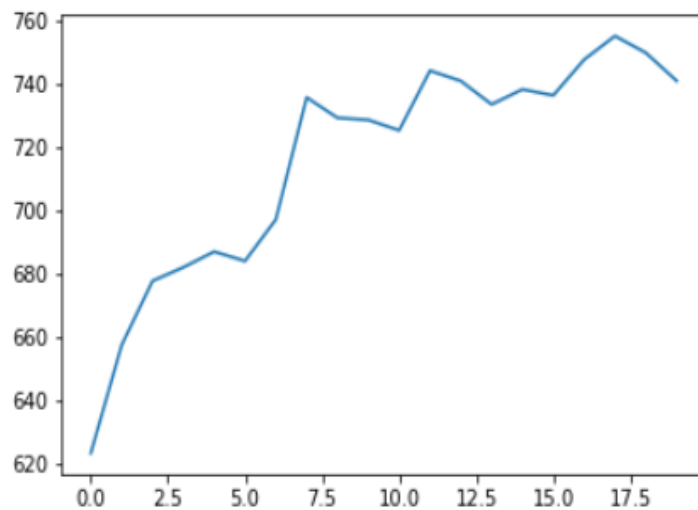
The second graph represents line plot of 'CLOSE' prices of BAJFINANCE

Correlation coefficient between these two is 0.98 and From the Below plots We can clearly observe they are highly correlated.



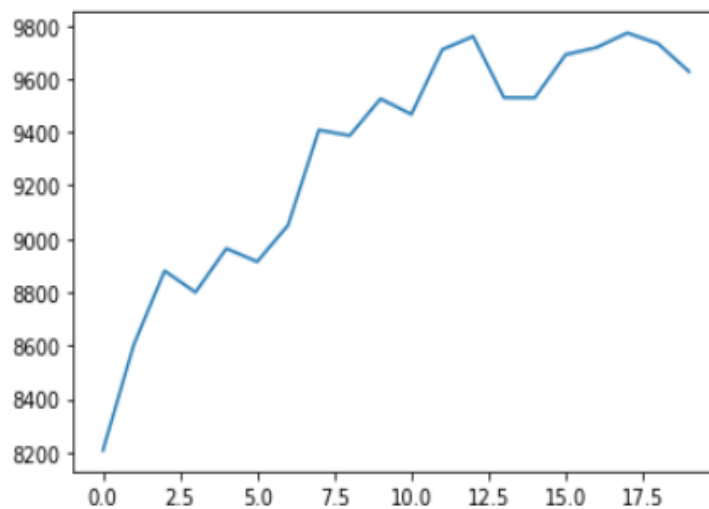
```
data_mat['AXISBANK'].plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc5461602d0>
```



```
[53] data_mat['BAJFINANCE'].plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc5460d6d90>
```



4. Finding highly correlated Pairs of stocks.

```
sorted_mat.head(50)
```

| | | |
|------------|------------|----------|
| ADANIPTS_y | ADANIPTS_x | 1.000000 |
| BAJAJFINSV | BAJFINANCE | 1.000000 |
| ITC | HDFCBANK | 0.989458 |
| AXISBANK | BAJFINANCE | 0.988376 |
| BAJAJFINSV | AXISBANK | 0.988376 |
| VEDL | TATASTEEL | 0.986121 |
| ICICIBANK | RELIANCE | 0.984324 |
| TATASTEEL | BAJFINANCE | 0.983677 |
| BAJAJFINSV | TATASTEEL | 0.983677 |
| HINDALCO | TATAMOTORS | 0.982710 |
| TATASTEEL | AXISBANK | 0.982402 |
| HINDALCO | LT | 0.981983 |
| TATAMOTORS | LT | 0.979312 |
| GRASIM | HDFC | 0.978296 |
| TATAMOTORS | HDFC | 0.977515 |
| ADANIPTS_y | VEDL | 0.977212 |
| VEDL | ADANIPTS_x | 0.977212 |
| | LT | 0.977126 |
| HINDALCO | JSWSTEEL | 0.976717 |
| TATASTEEL | TATAMOTORS | 0.975622 |
| | HDFC | 0.975186 |
| IOC | HDFC | 0.974439 |
| HEROMOTOCO | GAIL | 0.974256 |
| TECHM | VEDL | 0.974096 |
| VEDL | RELIANCE | 0.973849 |
| GRASIM | TATASTEEL | 0.973295 |
| VEDL | ICICIBANK | 0.972694 |
| ADANIPTS_y | HDFC | 0.972433 |
| HDFC | ADANIPTS_x | 0.972433 |
| TATASTEEL | ADANIPTS_x | 0.972123 |
| ADANIPTS_y | TATASTEEL | 0.972123 |
| WIPRO | TECHM | 0.972074 |



Finding RSI values.

data_mat

1 to 20 of 20 entries ?

| index | Date | AXISBANK | BAJFINANCE | Ratio | change | Daily_gain | Daily_loss |
|-------|-----------|----------|------------|---------------------|-------------------------|------------------------|------------------------|
| 0 | 09-Mar-20 | 623.55 | 8207.1 | 0.07597689805168695 | 0.0 | 0.0 | 0.0 |
| 1 | 06-Mar-20 | 657.65 | 8603.2 | 0.07644248651664497 | 0.0004655884649580194 | 0.0004655884649580194 | 0.0 |
| 2 | 05-Mar-20 | 677.9 | 8879.9 | 0.07634094978547055 | -0.00010153673117441198 | 0.0 | 0.00010153673117441198 |
| 3 | 04-Mar-20 | 682.2 | 8799.7 | 0.07752537018307443 | 0.001184420397603872 | 0.001184420397603872 | 0.0 |
| 4 | 03-Mar-20 | 687.15 | 8963.65 | 0.07665961968617695 | -0.00086575049689748 | 0.0 | 0.00086575049689748 |
| 5 | 02-Mar-20 | 684.2 | 8914.5 | 0.0767513601435863 | 9.174045740935333e-05 | 9.174045740935333e-05 | 0.0 |
| 6 | 28-Feb-20 | 697.3 | 9050.75 | 0.07704333894981079 | 0.00029197880622448824 | 0.00029197880622448824 | 0.0 |
| 7 | 27-Feb-20 | 735.85 | 9408.0 | 0.07821534863945578 | 0.0011720096896449955 | 0.0011720096896449955 | 0.0 |
| 8 | 26-Feb-20 | 729.4 | 9387.0 | 0.07770320656226697 | -0.000512142077188818 | 0.0 | 0.000512142077188818 |
| 9 | 25-Feb-20 | 728.75 | 9525.0 | 0.07650918635170603 | -0.0011940202105609338 | 0.0 | 0.0011940202105609338 |
| 10 | 24-Feb-20 | 725.45 | 9466.9 | 0.0766301534821325 | 0.0001209671304264609 | 0.0001209671304264609 | 0.0 |
| 11 | 20-Feb-20 | 744.3 | 9709.7 | 0.07665530345942716 | 2.514997729466395e-05 | 2.514997729466395e-05 | 0.0 |
| 12 | 19-Feb-20 | 741.05 | 9758.65 | 0.07593775778411972 | -0.000717545675307435 | 0.0 | 0.000717545675307435 |
| 13 | 18-Feb-20 | 733.65 | 9529.5 | 0.07698725011805446 | 0.00104949233393474 | 0.00104949233393474 | 0.0 |
| 14 | 17-Feb-20 | 738.4 | 9528.85 | 0.07749098789465675 | 0.0005037377766022866 | 0.0005037377766022866 | 0.0 |
| 15 | 14-Feb-20 | 736.5 | 9690.5 | 0.07600227026469222 | -0.0014887176299645255 | 0.0 | 0.0014887176299645255 |
| 16 | 13-Feb-20 | 747.8 | 9717.2 | 0.0769563248672457 | 0.0009540546025534752 | 0.0009540546025534752 | 0.0 |
| 17 | 12-Feb-20 | 755.25 | 9771.85 | 0.0772883332480543 | 0.00033200845755973485 | 0.00033200845755973485 | 0.0 |
| 18 | 11-Feb-20 | 749.95 | 9731.6 | 0.07706338115006782 | -0.00022495217473761375 | 0.0 | 0.00022495217473761375 |
| 19 | 10-Feb-20 | 741.1 | 9627.0 | 0.07698140646099512 | -8.197468907270022e-05 | 0.0 | 8.197468907270022e-05 |

Show 25 per page

Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

```
[ ] print(avg_gain)
    print(avg_loss)

0.00032057378158407285
0.0002673449515821644

[ ] rs = avg_gain/avg_loss
    rs

1.1991016837493915

rsi = 100 - (100/(1+rs))
rsi

54.526886710621085
```

5. Analysing RSI indicators Result

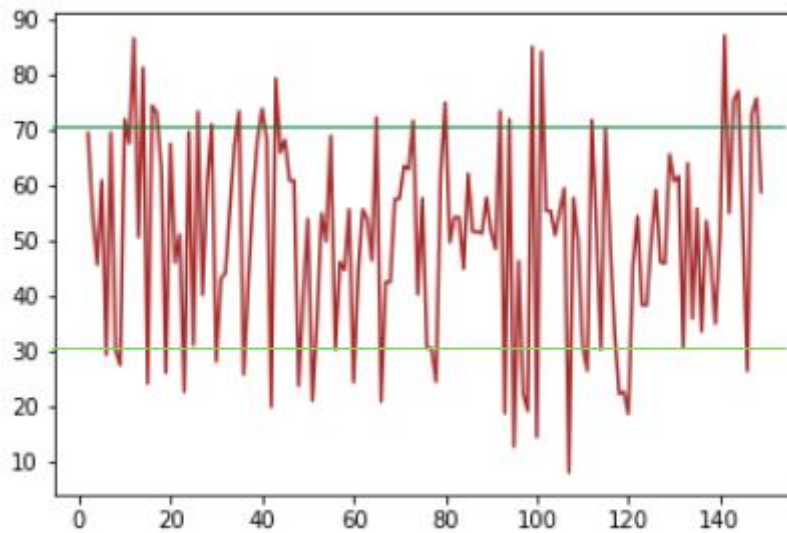
Pair Trading with RSI

| PARAMETERS | PAIR RATIO LONG | PAIR RATIO SHORT |
|---|--|--|
| DAILY PRICE RATIO RATIO = STOCKA / STOCKB | ENTRY <ul style="list-style-type: none"> • RSI(14) < 30 • LONG STOCK A • SHORT STOCK B | ENTRY <ul style="list-style-type: none"> • RSI(14) > 70 • SHORT STOCK A • LONG STOCK B |
| APPLY RSI INDICATOR FOR RATIO, USE 14 PERIOD RSI THAT IS RSI(14) | EXIT <ul style="list-style-type: none"> • 5% PROFIT TARGET OR • 5 PERIOD - TIME EXIT OR • 5% STOP LOSS | EXIT <ul style="list-style-type: none"> • 5% PROFIT TARGET OR • 5 PERIOD - TIME EXIT OR • 5% STOP LOSS |

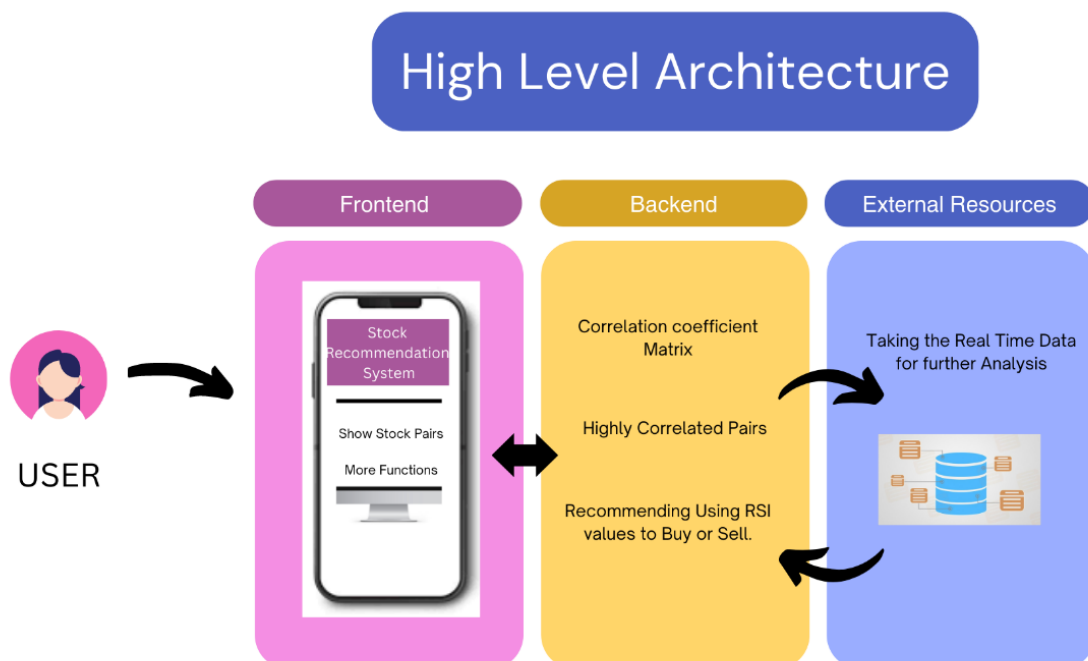
| Stock A | Stock B | RSI values |
|-----------|------------|-------------------|
| HINDALCO | TATAMOTORS | 27.30389452138685 |
| TATASTEEL | AXISBANK | 71.87647810717914 |

- In the first record stock A = HINDALCO and stock B = TATAMOTORS with RSI values = 27.303894521 where it is < 30 so Long Stock A and Short stock B.
- In the second record stock A= TATASTEEL and stock B = AXISBANK with RSI values = 71.87647810717914 where it is > 70 so Short Stock A and Long Stock B.

- The following graph is of Relative Strength Index where x axis is representing Sr. No of stocks and y axis represents RSI value ranging 0-100
- RSI above 70 “overbought” and RSI below 30 “oversold”



➤ High Level Architecture





Impact

➤ Product Approach

Our end product will be a fully functional web app which not only finds the correlated pairs among stocks but also suggest which stock to buy and sell thus help us in minimizing ambiguity and chances of encountering loss.

➤ Business Impact

Our prime intention is to use statistical and economical trading strategies to maximize the profit for traders under market neutral conditions. The historical trend shows that stock market returns are relatively lower during periods of rising inflation, but with our pair trading approach trader can sustain profit.

Non-functional Requirements

- **Scalability** Now the application works on historical data. Also it will be working on the real time data, which will be the scalability of the application.
- **Throughput** Building another server for the operations and using API we can fetch the data to the application. which leads to improving speed of the application
- **Security** User authentication (google authentication, OTP)
Data sharing & role-based access control: - Only admin can do changes, no access to the users for downloading/changing the data.
Project management software data encryption: - when user uploads the sheet encrypt that data and while performing operations decrypt data.



- **Entitlement** Entitlement is a right to access a benefit or service to the user. In our application a valid user has access to get recommended by the application based on trade pairs to buy or sell the stock.
- **Cloud Deployment** As we have a web app we will go ahead with Platform as a service (PaaS), which offers active environment except application code, users, data giving us time to focus on application code.
- **Test Automation** Code Analysis, Unit Testing is also done on each function of application. As we are using API to fetch data from a different server, we will be doing Integration Testing on our application.