STATISTICAL ARBITRAGE – Machine Learning PROJECT

Source – NSE data provided combined 2016 and 2017

Job – to predict Arbitrage Opportunities in 2017 data

Industry selected – Banking & Finance

Working DETAILS of the Project

First Step➔ Create 2016 Dataset

Separate 2016 data of stocks from 2017

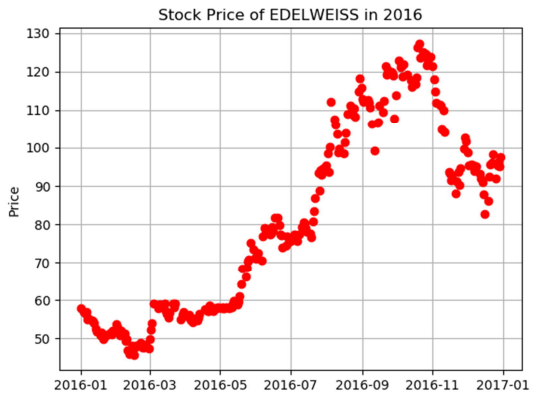
Specify Target Stock for which arbitrage opportunity will be explored – in this case, I have identified EDELWEISS as Target stock

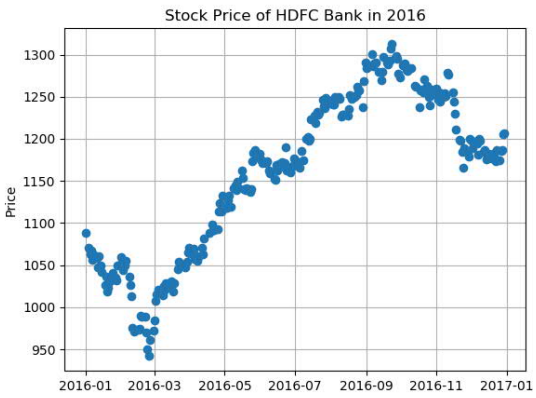
Apart from that I have taken following stocks against which we will compare target

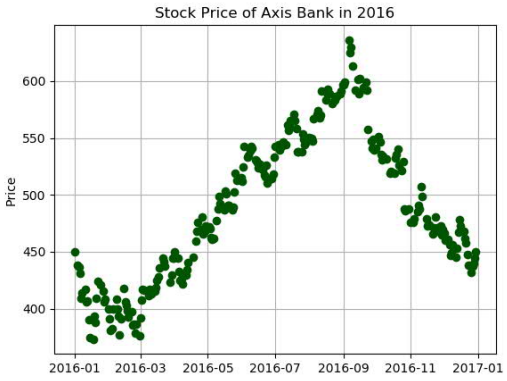
- HDFC Bank
- ICICI Bank
- Kotak Mahindra Bank
- Max Financial Service Ltd (MFSL)
- Axis Bank
- India Bull

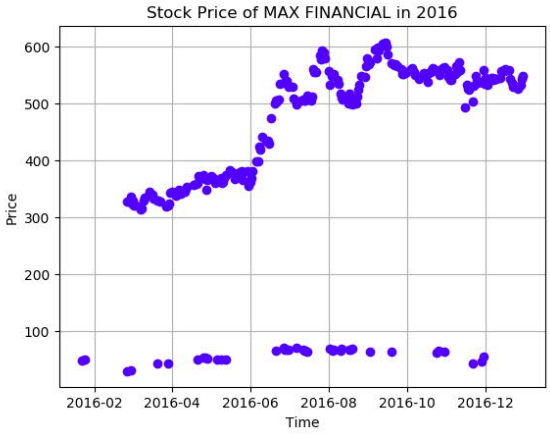
Second Step➔ Compare price movement of different stocks

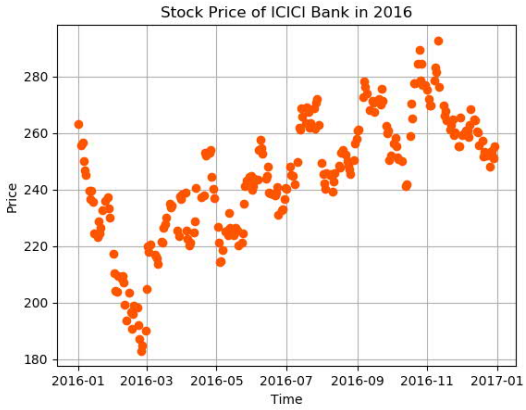
Let us pay attention to individual stock price movement with time over the year 2016:

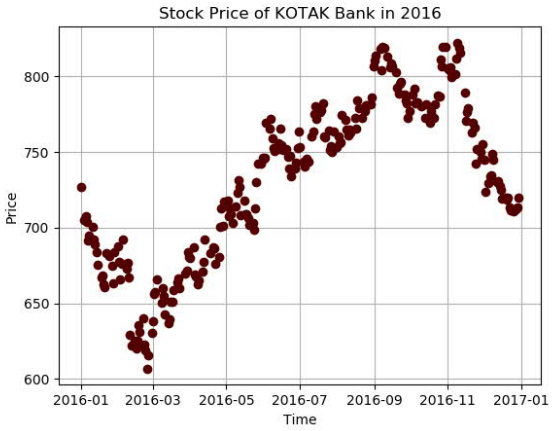


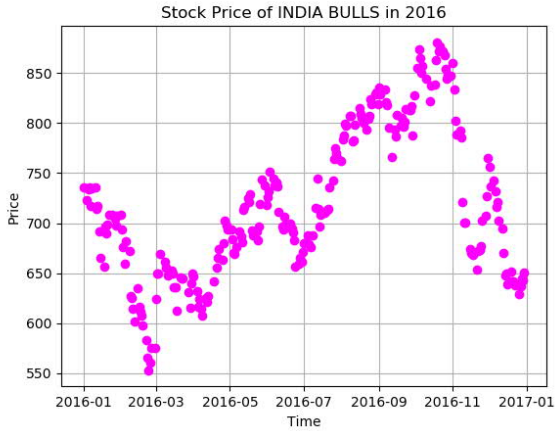



Stock Price of Axis Bank in 2016
Price
600
550
500
450
400
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01


Stock Price of MAX FINANCIAL in 2016
Price
600
500
400
300
200
100
2016-02
2016-04
2016-06
2016-08
2016-10
2016-12
Time


Stock Price of ICICI Bank in 2016
Price
280
260
240
220
200
180
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
Time


Stock Price of KOTAK Bank in 2016
Price
800
750
700
650
600
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
Time

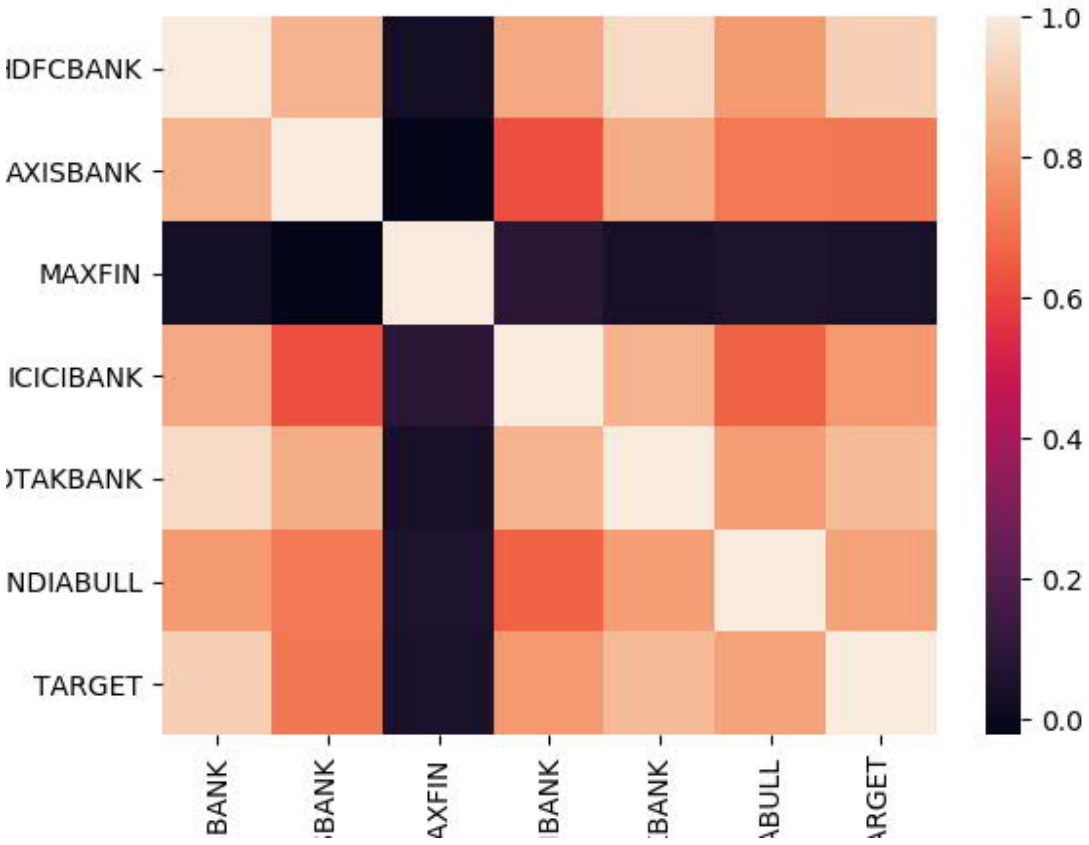

Stock Price of INDIA BULLS in 2016
Price
850
800
750
700
650
600
550
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
Time

Observation➔

All the stocks under selection is moving similarly with our target stock (EDELWEISS) except Max Financial (MFSL). We might need to take this stock out of our analysis after looking at Coefficient Matrix. .

Third Step➔Correlation Matrix

Above plots arel not definitive answer for best and distinct pair of stock selections. To get clear picture, we will now plot correlation matrix



Correlation Coefficients from highest to lowest with Target:

TARGET 1.000000

HDFCBANK 0.920602

KOTAKBANK 0.871357

INDIABULL 0.810035

ICICIBANK 0.787410

AXISBANK 0.711601

MAXFIN 0.051432

It is clear that our training set should exclude Max Financial data

Fourth Step➔ Data Frame & Train/Test Split

Sample data of new dataset:

[8 rows x 7 columns]

	HDFCBANK	AXISBANK	MAXFIN	ICICIBANK	KOTAKBANK	INDIABULL	TARGET
0	1091.15	467.95	50.55	253.05	686.70	665.60	57.15
1	1140.90	490.65	360.60	226.50	708.25	716.10	59.85
2	1119.65	461.45	50.90	218.60	702.95	676.45	58.05
3	1053.80	429.45	373.75	223.40	671.50	614.80	56.15
4	1234.90	589.50	50.45	247.60	765.65	809.20	101.45

Train and Test Data Split (MaxFin excluded basis Correlation Matrix)

Shape of Feature: (247, 6)

Training Feature Shape: (197, 6)

Training Labels Shape: (197,)

Testing Feature Shape: (50, 6)

Testing Labels Shape: (50,)

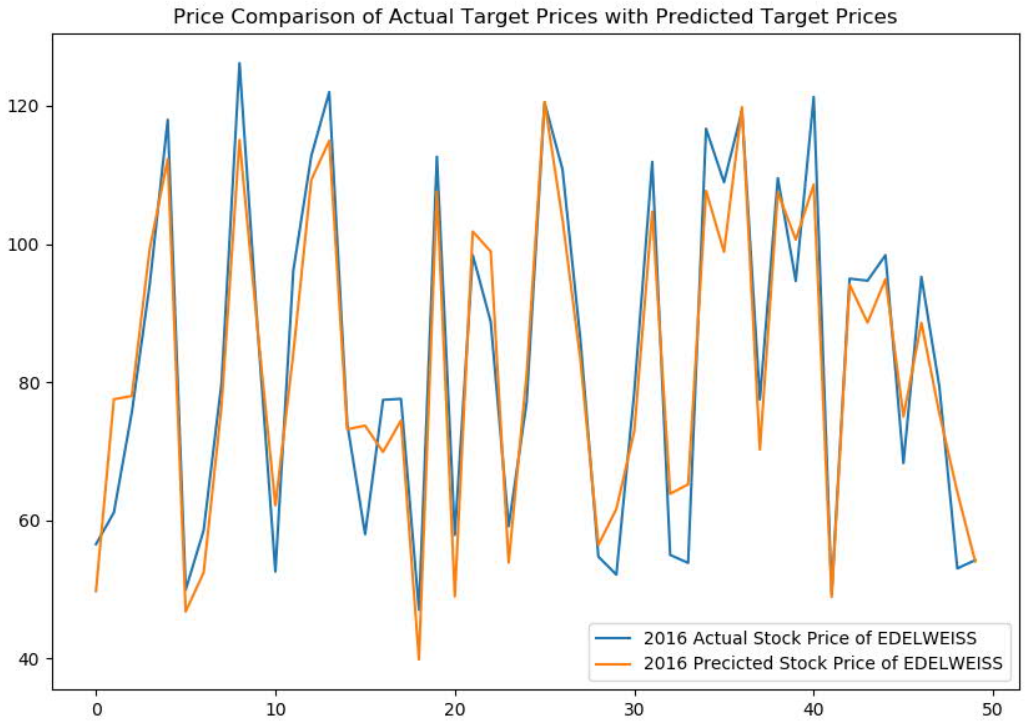
Evaluation of Linear Regression Model

Explained Variance Score: 0.9134675004622271

Mean Abosolute Error: 6.102948565730823

Mean Squared Error: 52.91514488059111

R Squared Error: 0.9110470555423941



Average Price of testing sample (2016): 83.586

Average Price of predicted stock (2016): 82.386

Evaluation of KNN Regression Model

k=1, accuracy=99.33%

k=3, accuracy=98.94%

k=5, accuracy=99.14%

k=7, accuracy=98.08%

k=9, accuracy=97.92%

k=1, achieved highest accuracy of 99.33%

This KNN regression output is much better than Linear Regression

Sample Predicted Price of the Target:

0 95.20

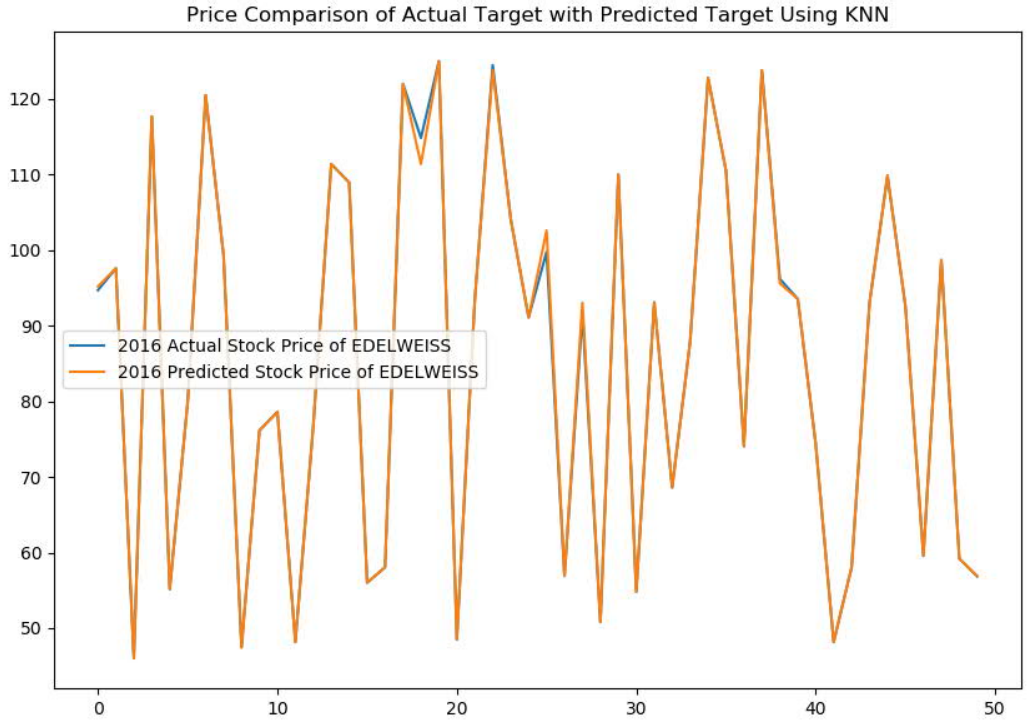
1 97.60

2 46.00

3 117.70

4 55.15

Plot of Actual Target Price vs. Predicted Price:



Predicted value is just same as target value at every point due to high accuracy of the model.

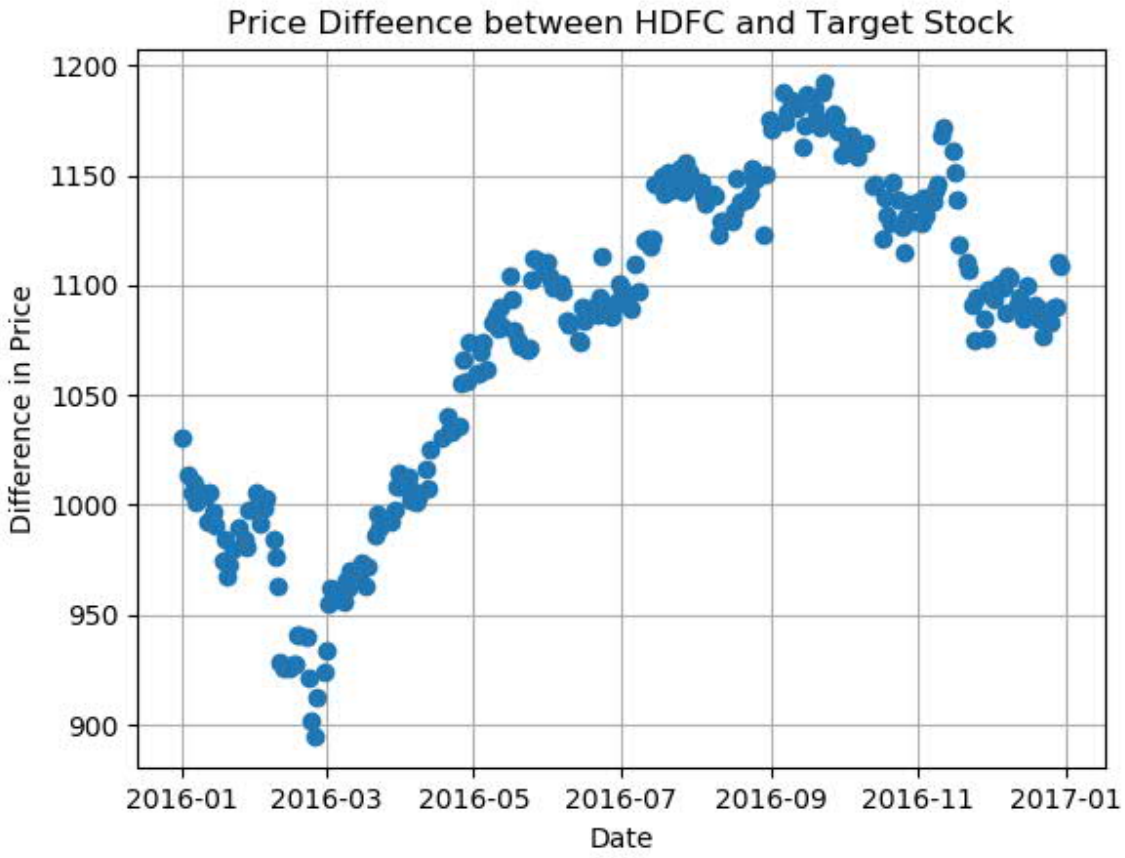
Determination of Co-integration to analyze long-short positions of a Pair of Stocks (one of them being Target Stock):

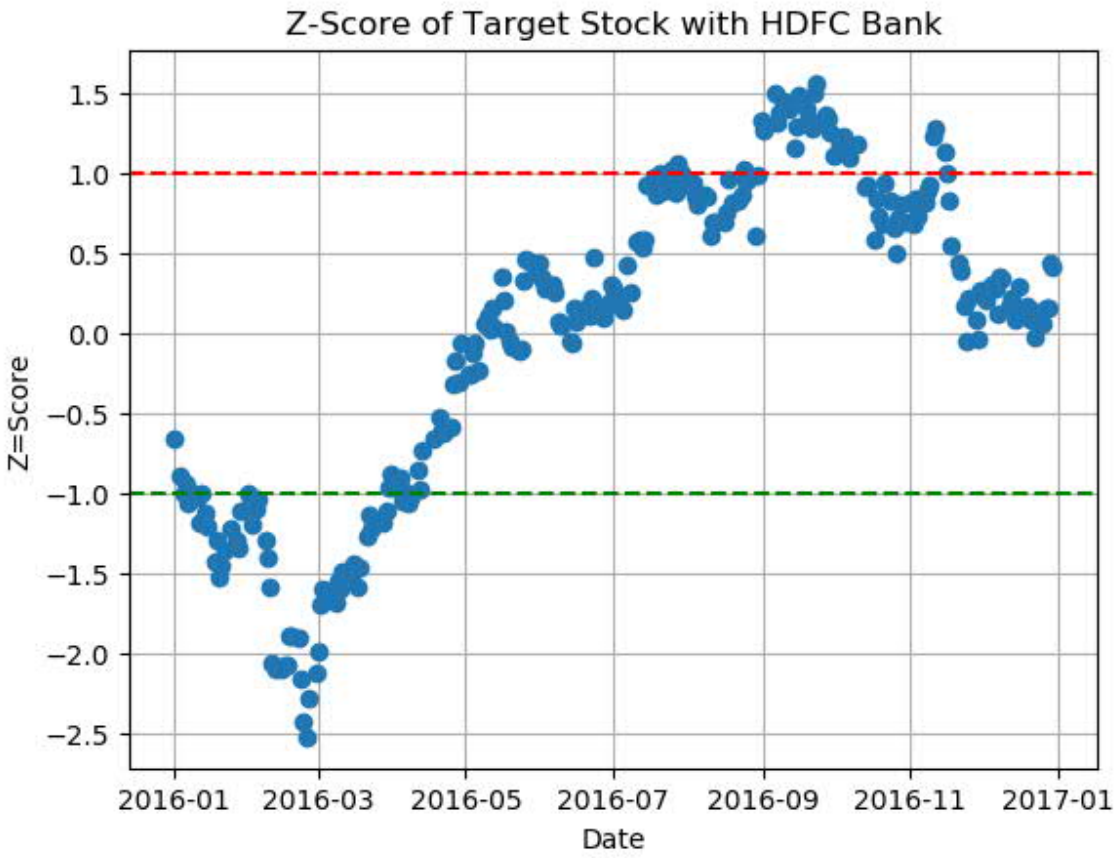
Fifth Step➔ Determination Z-Score (Arbitrage Indicator) for each pair of stocks

For Pair of HDFC and Target (EDELWEISS):

Co-integration Score: -14.937919539814336

pValue: 1.1441238589224678e-26





Arbitrage Indicator:

Go “Long” the spread whenever the z-score is below -1.0

Go “Short” the spread whenever the z-score is above 1.0

Exit positions when z-score approaches zero

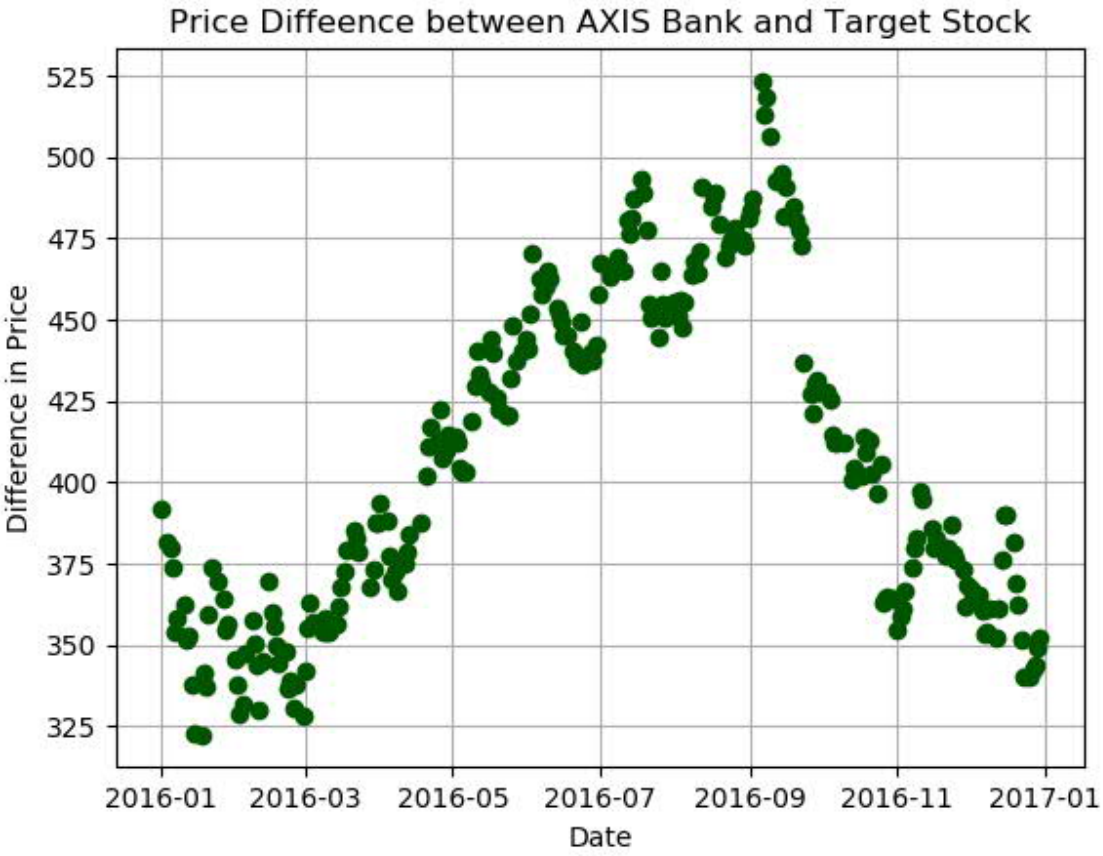
Since we originally defined the "spread" as HDFC - TARGET, "Long" the spread would mean "Buy 1 share of HDFC, and Sell Short 1 share of EDELWEISS" and if you were going "Short" the spread, “Sell 1 share of HDFC and Buy Short 1 share of EDELWEISS).

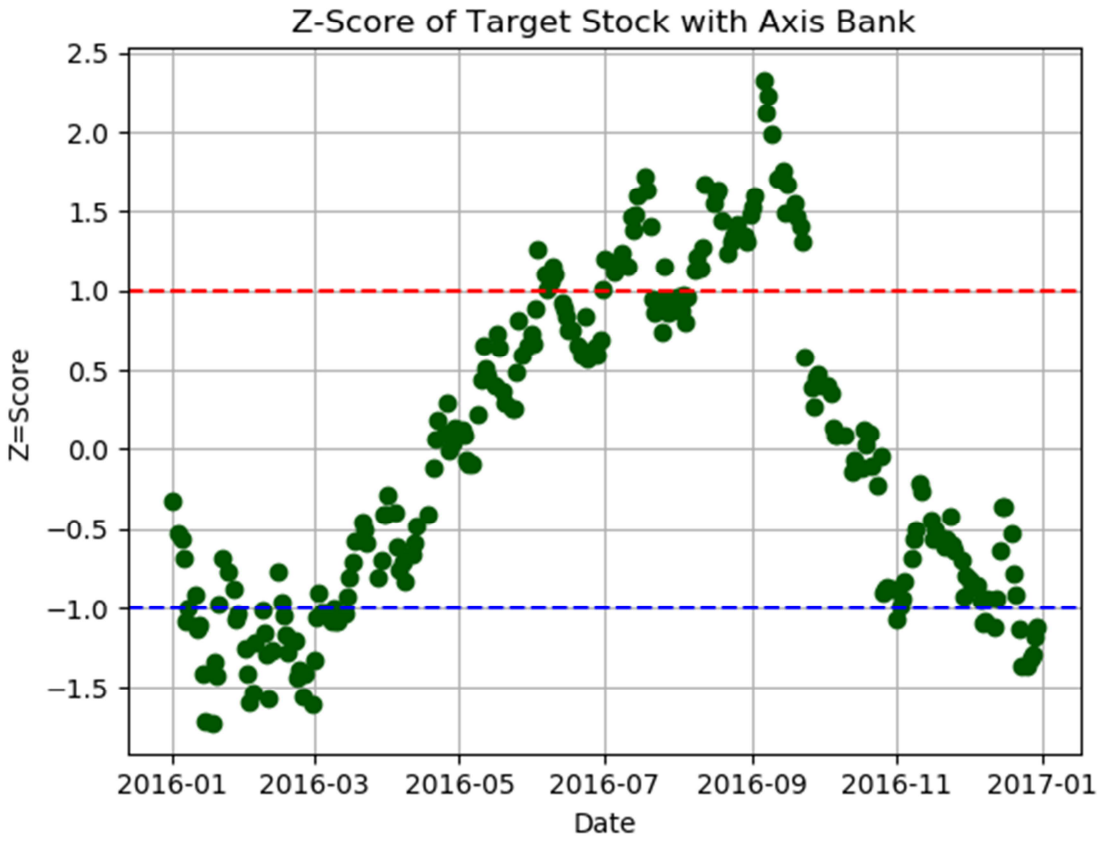
Another observation is in the month of March’16, there is a huge opportunity of arbitrage by buying 1 HDFC share and selling 1 EDELWEISS share while in Sep’16, there is an arbitrage opportunity by selling 1 HDFC share and buying 1 EDELWEISS share.

For Pair of Axis Bank and Target (EDELWEISS):

Co-integration Score: -15.312194880782178

pValue: 3.4497169042875244e-27


Price Diffeence between AXIS Bank and Target Stock
Difference in Price
525
500
475
450
425
400
375
350
325
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
Date

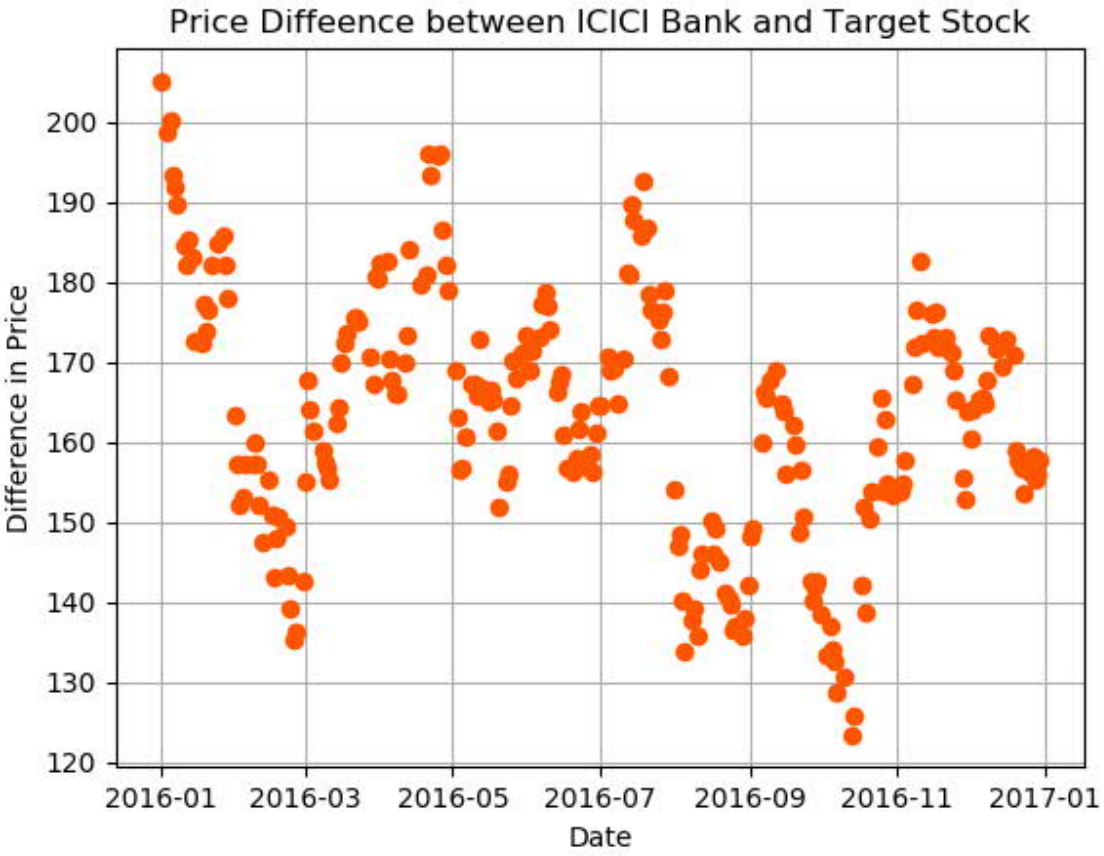


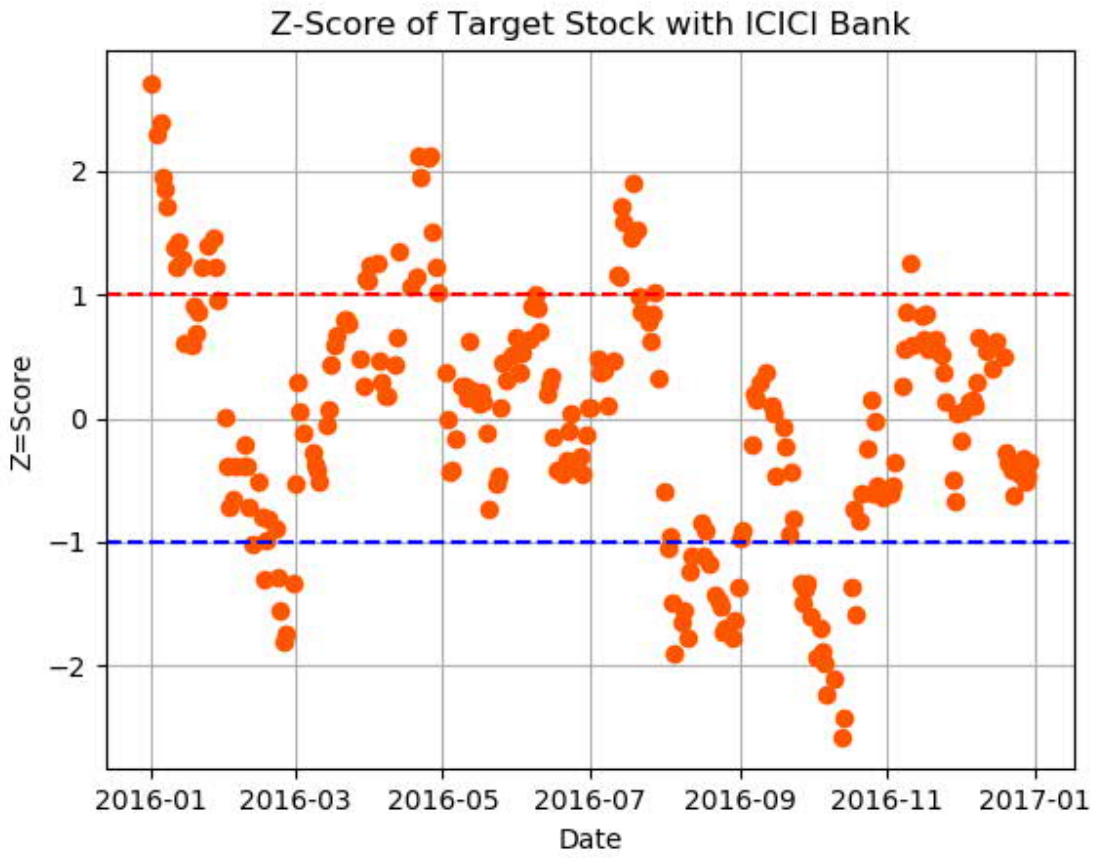
Arbitrage Indicator:
Go “Long” the spread whenever the z-score is below -1.0
Go “Short” the spread whenever the z-score is above 1.0
Exit positions when z-score approaches zero
Since we originally defined the "spread" as AXIS - TARGET, "Long" the spread would mean "Buy 1 share of Axis, and Sell Short 1 share of EDELWEISS" and if you were going "Short" the spread, “Sell 1 share of Axis and Buy Short 1 share of EDELWEISS).
Another observation is in the month of Feb-March’16 and again in Dec’16, there is an opportunity of arbitrage by buying 1 Axis share and selling 1 EDELWEISS share while in July-Sep’16, there is an arbitrage opportunity by selling 1 Axis share and buying 1 EDELWEISS share.

For Pair of ICICI Bank and Target (EDELWEISS):

Co-integration Score: -12.405978421037375

pValue: 4.903245598626143e-22

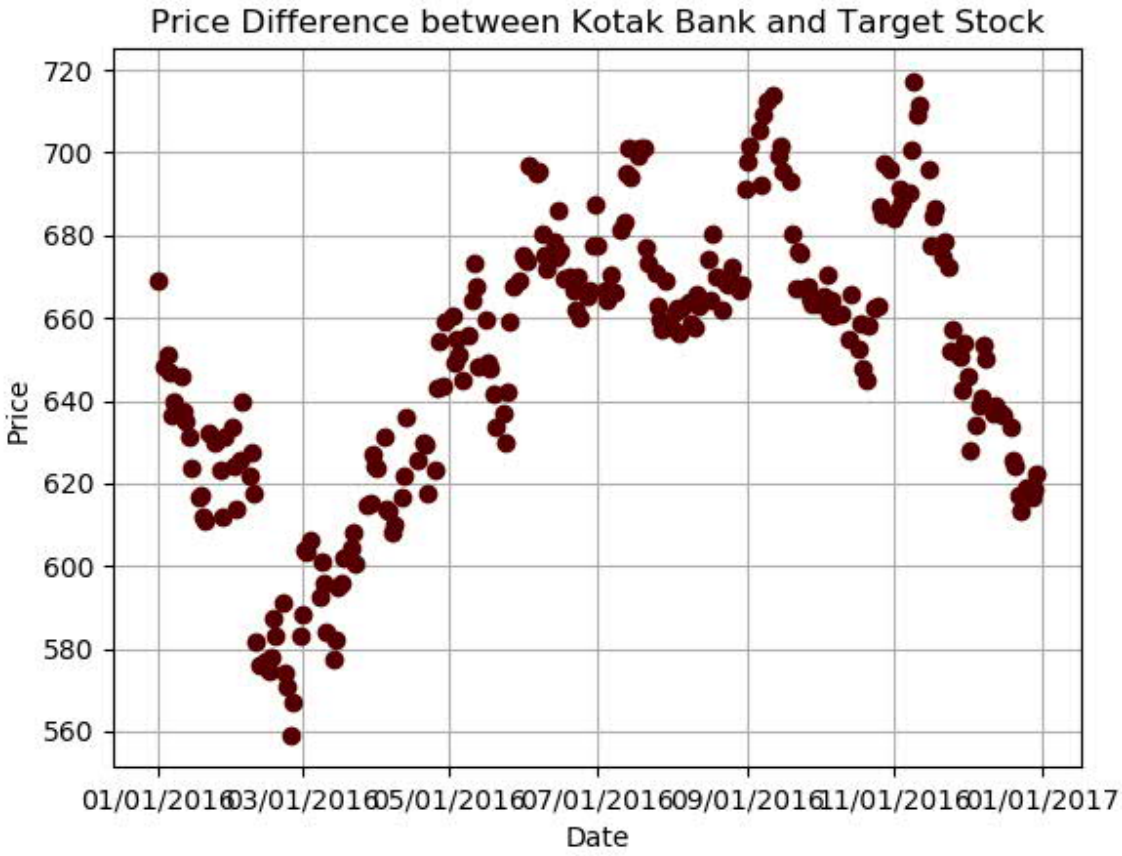

Price Diffeence between ICICI Bank and Target Stock
Difference in Price
200
190
180
170
160
150
140
130
120
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
Date

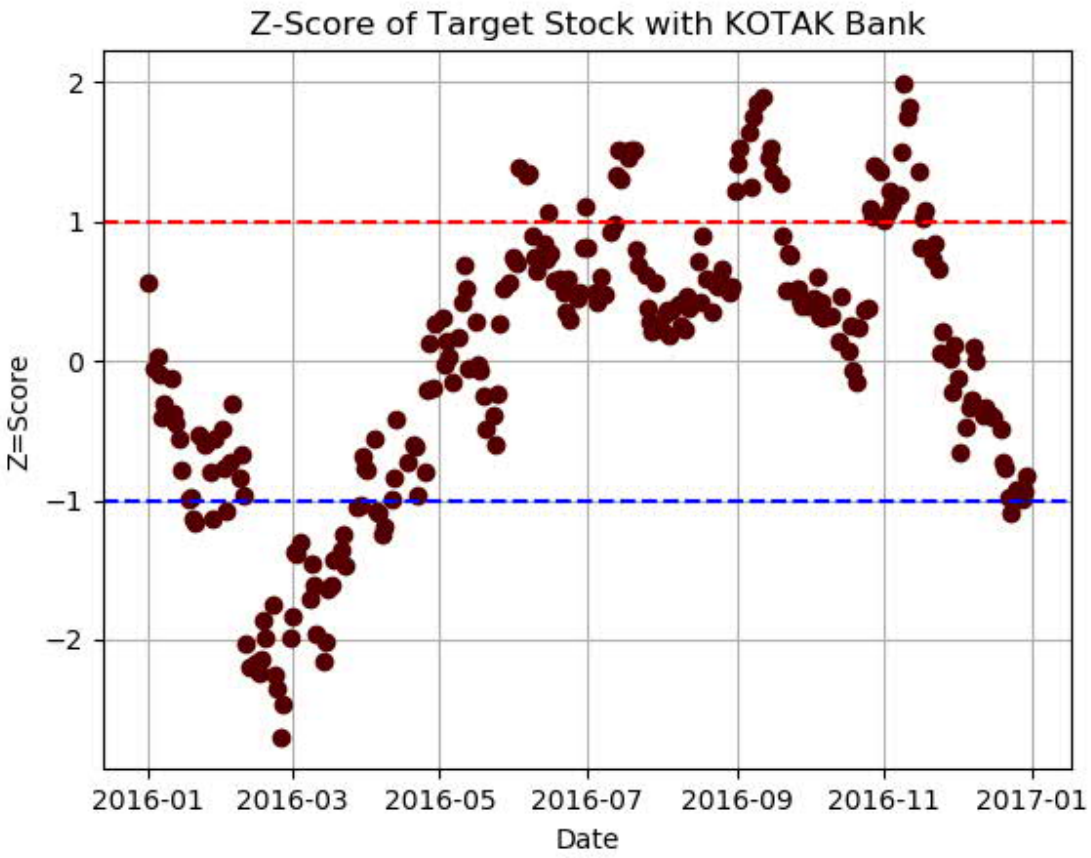


Arbitrage Indicator:
Go “Long” the spread whenever the z-score is below -1.0
Go “Short” the spread whenever the z-score is above 1.0
Exit positions when z-score approaches zero
Since we originally defined the "spread" as ICICI - TARGET, "Long" the spread would mean "Buy 1 share of ICICI, and Sell Short 1 share of EDELWEISS" and if you were going "Short" the spread, “Sell 1 share of ICICI and Buy Short 1 share of EDELWEISS).
Another observation is in the month of Mar’16, Aug’16 and Oct16, there is an opportunity of arbitrage by buying 1 ICICI share and selling 1 EDELWEISS share while in Jan’16, April’16 and July’16, there is an arbitrage opportunity by selling 1 ICICI share and buying 1 EDELWEISS share.

For Pair of Kotak Bank and Target (EDELWEISS):

Co-integration: Score: -7.779281952920594
pValue: 1.0771733343825084e-10


Price Difference between Kotak Bank and Target Stock
720
700
680
660
640
620
600
580
560
Price
01/01/2016
03/01/2016
05/01/2016
07/01/2016
09/01/2016
11/01/2016
01/01/2017
Date

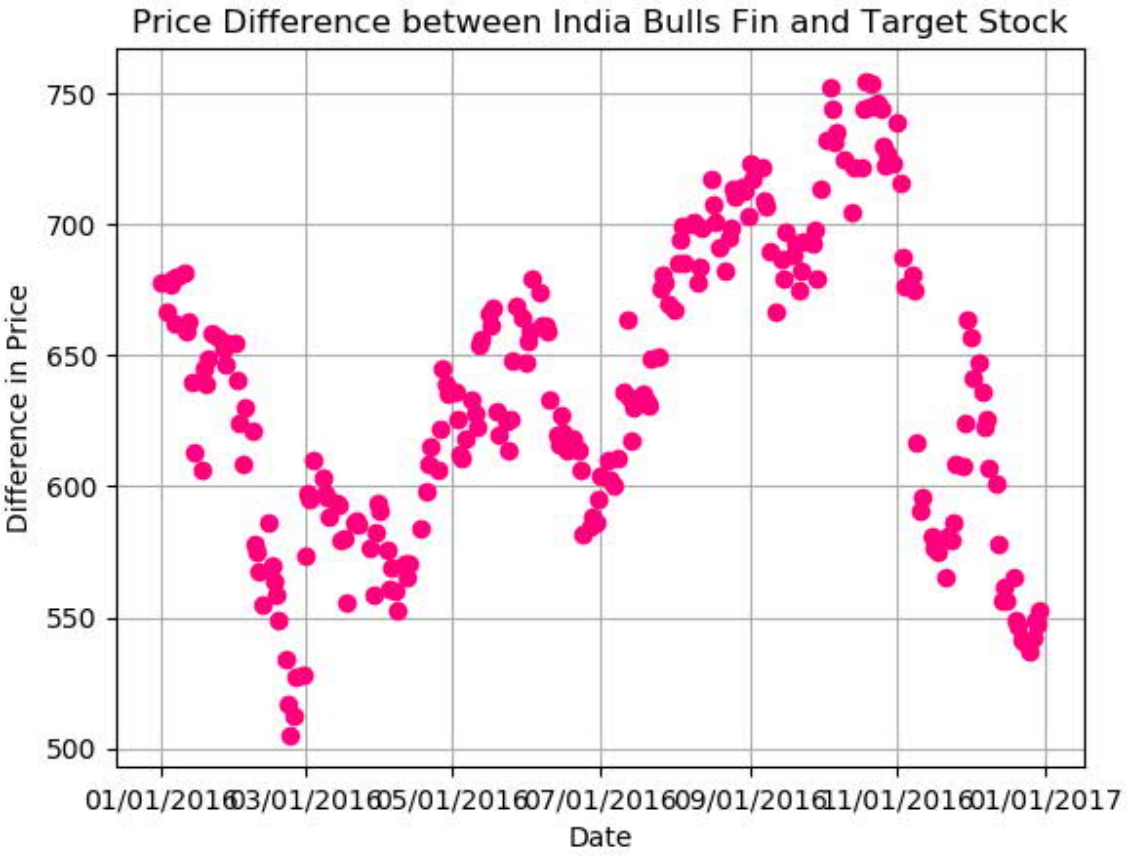


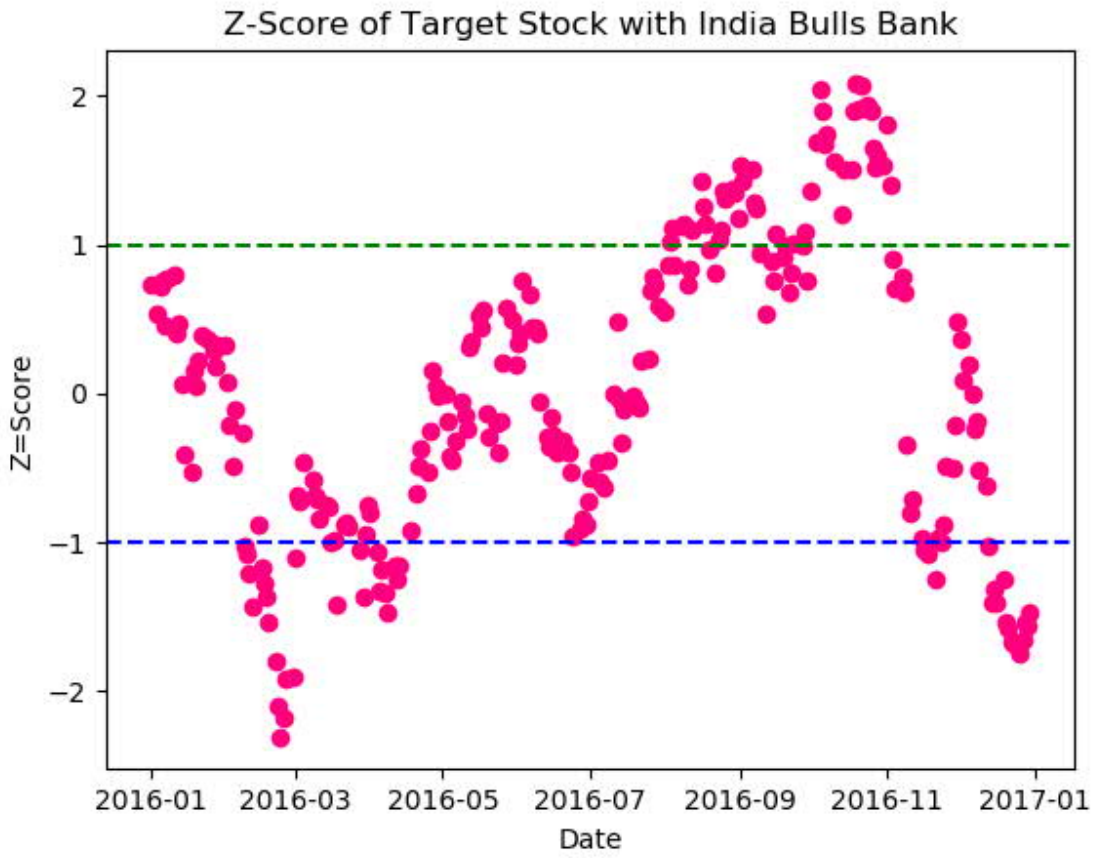
Arbitrage Indicator:

Go “Long” the spread whenever the z-score is below -1.0
Go “Short” the spread whenever the z-score is above 1.0
Exit positions when z-score approaches zero
Since we originally defined the "spread" as KOTAK - TARGET, "Long" the spread would mean "Buy 1 share of KOTAK, and Sell Short 1 share of EDELWEISS" and if you were going "Short" the spread, “Sell 1 share of KOTAK and Buy Short 1 share of EDELWEISS).
Another observation is in the month of Mar’16 huge opportunity of arbitrage by buying 1 KOTAK share and selling 1 EDELWEISS share while in July’16, Sep’16 and Nov’16, there is an arbitrage opportunity by selling 1 KOTAK share and buying 1 EDELWEISS share.

For Pair of India Bulls and Target (EDELWEISS):

Co-integration Score: -14.934572847560133
pValue: 1.157014872837643e-26


Price Difference between India Bulls Fin and Target Stock
Difference in Price
750
700
650
600
550
500
01/01/2016
03/01/2016
05/01/2016
07/01/2016
09/01/2016
11/01/2016
01/01/2017
Date



Arbitrage Indicator:

Go “Long” the spread whenever the z-score is below -1.0

Go “Short” the spread whenever the z-score is above 1.0

Exit positions when z-score approaches zero

Since we originally defined the "spread" as INDIA BULLS - TARGET, "Long" the spread would mean "Buy 1 share of INDIA BULLS, and Sell Short 1 share of EDELWEISS" and if you were going "Short" the spread, “Sell 1 share of INDIA BULLS and Buy Short 1 share of EDELWEISS).

Another observation is in the month of Mar’16 and Dec’16 there is huge opportunity of arbitrage by buying 1 INDIA BULLS share and selling 1 EDELWEISS share while in Oct’16, there is an arbitrage opportunity by selling 1 INDIA BULLS share and buying 1 EDELWEISS share.

Co-Integration Matrix of Pair of Component Stocks & EDELWEISS

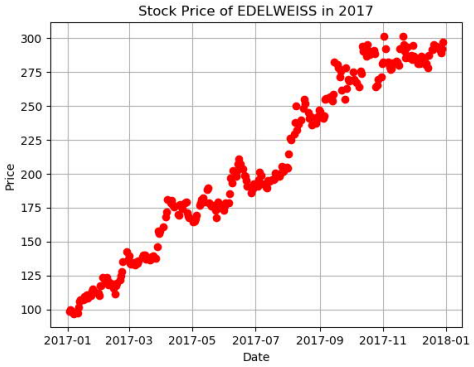
	Co-Integration Score	P-Value
HDFC Bank	-14.93791954	1.14E-26
Axis Bank	-15.31219488	3.45E-27
ICICI Bank	-12.40597842	4.90E-22
Kotak Bank	-7.779281953	1.08E-10
India Bulls	-14.93457285	1.16E-26

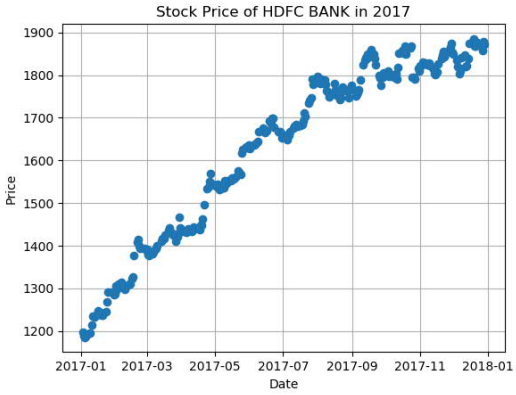
2017 Data Analysis, Prediction, Arbitrage Opportunities:

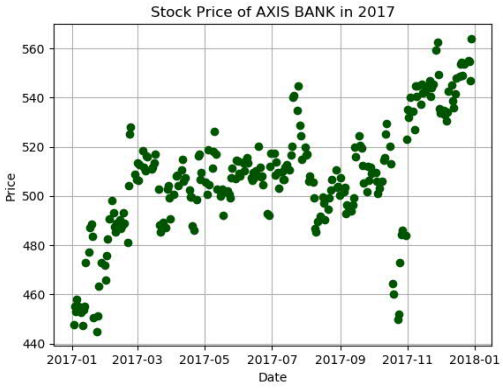
Data Size:

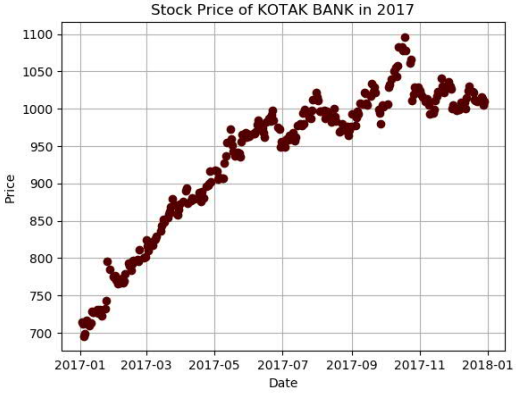
Size of Target: 248
Size of HDFC Bank Data: 248
Size of Axis Bank Data: 248
Size of ICICI Bank Data: 248
Size of Kotak Bank Data: 248
Size of India Bull Data: 248

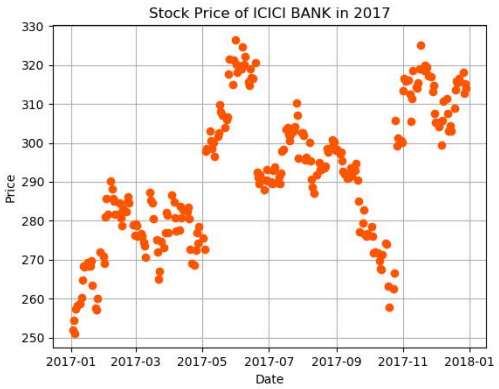
Plot 2017 Price of Individual Stocks including Target:

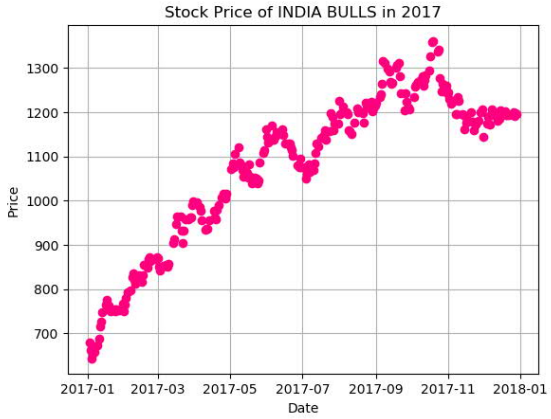










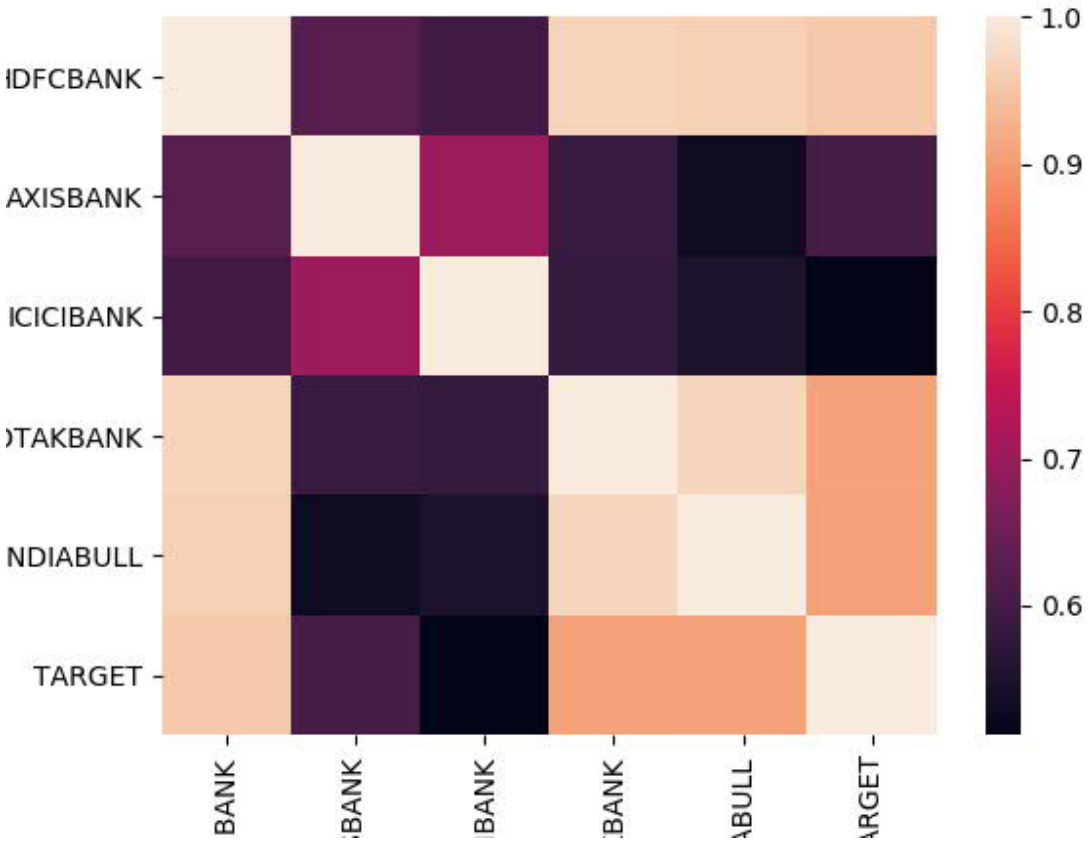


Observations:
From above plots, it seems price movement of ICICI Bank is different from rest of stocks. We will check correlation matrix

It will be more evident while we plot heat map of correlation matrix as detailed below:
Here is the 2017 co-relation compared to 2016 co-relation:

	Correlation Coefficient	
Stocks	2016	2017
TARGET	1	1
HDFC	0.920602	0.95477
KOTAK	0.871357	0.908872
INDIABULL	0.810035	0.907921
AXISBANK	0.711601	0.603131
ICICIBANK	0.78741	0.511893

This supports our plots above where we could see less correlation of target with ICICI stocks, but in 2016, correlation coefficients were better for this stock.
The below heat map is display of correlation map

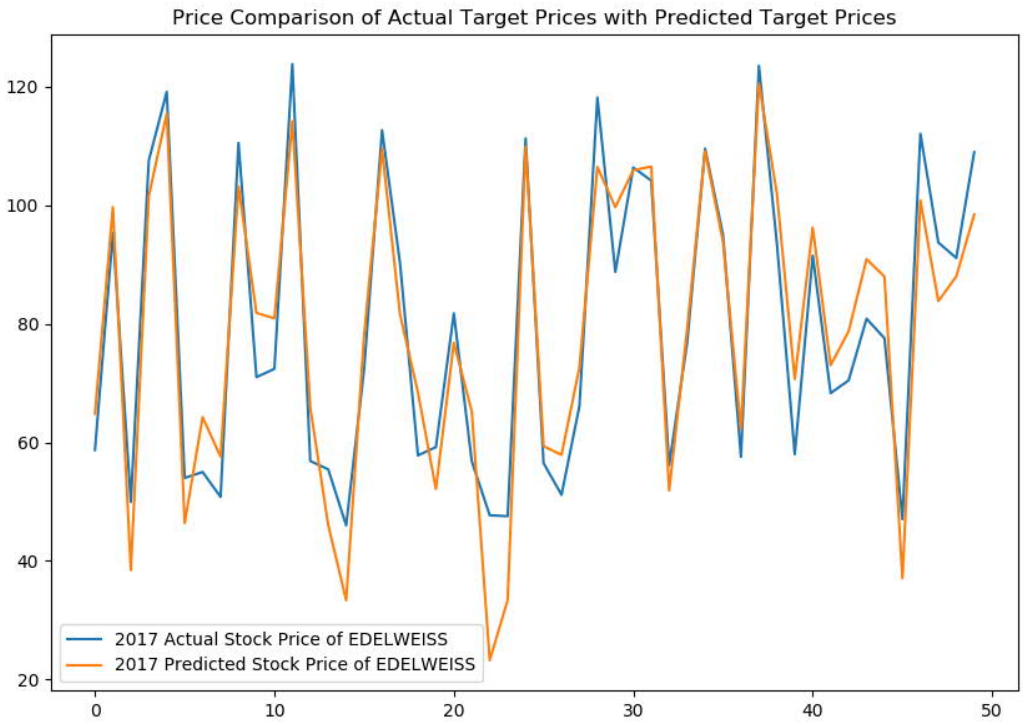


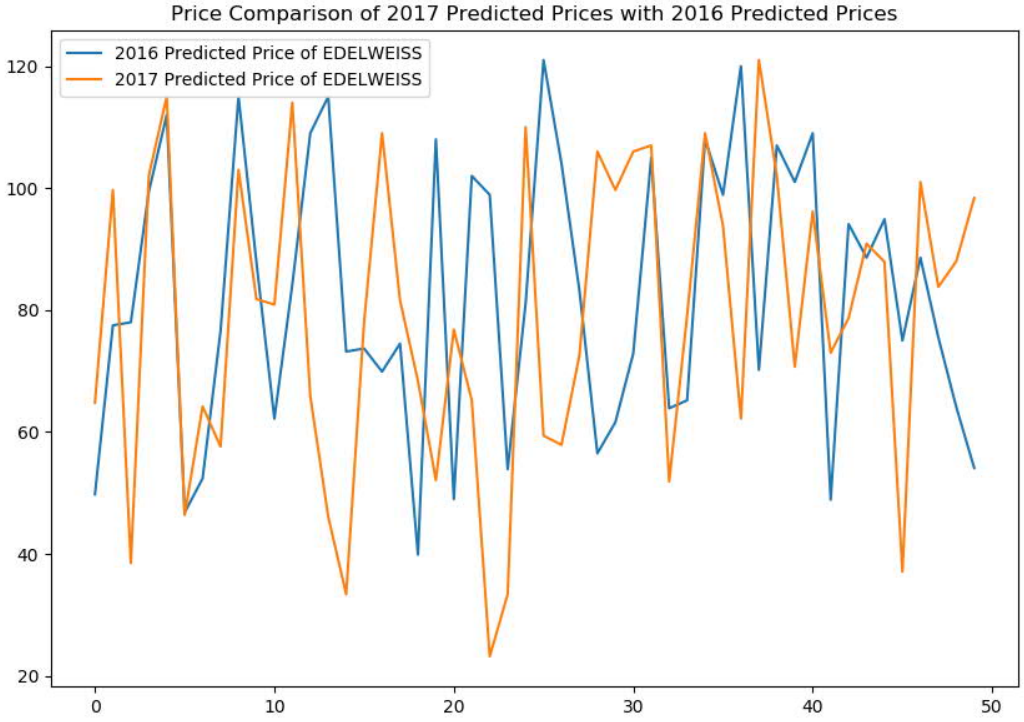
Now, we will set up Data Frame for train-test split and estimate prediction of target stocks:

	HDFCBANK	AXISBANK	ICICIBANK	KOTAKBANK	INDIABULL	TARGET
0	1666.45	492.00	290.35	972.20	1075.15	188.85
1	1865.35	562.55	313.20	1029.85	1200.05	287.75
2	1390.10	506.65	276.35	802.10	871.30	137.20
3	1546.50	509.65	278.50	901.95	1016.95	167.65
4	1436.10	502.85	275.05	854.60	963.70	138.00

We have got predicted price based on 2017 actual data. Now two interesting plots:

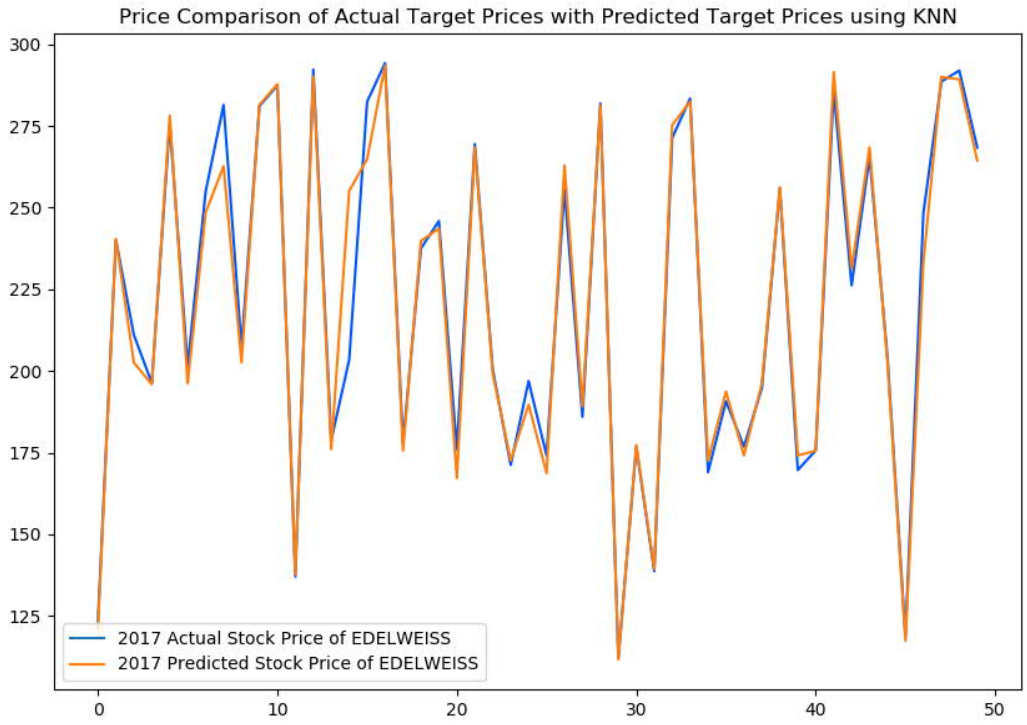
1. Compare actual price with predicted price
2. Compare predicted 2016 price with 2017 predicted price


Price Comparison of Actual Target Prices with Predicted Target Prices
2017 Actual Stock Price of EDELWEISS
2017 Predicted Stock Price of EDELWEISS
120
100
80
60
40
20
0
10
20
30
40
50


Price Comparison of 2017 Predicted Prices with 2016 Predicted Prices
2016 Predicted Price of EDELWEISS
2017 Predicted Price of EDELWEISS
120
100
80
60
40
20
0
10
20
30
40
50

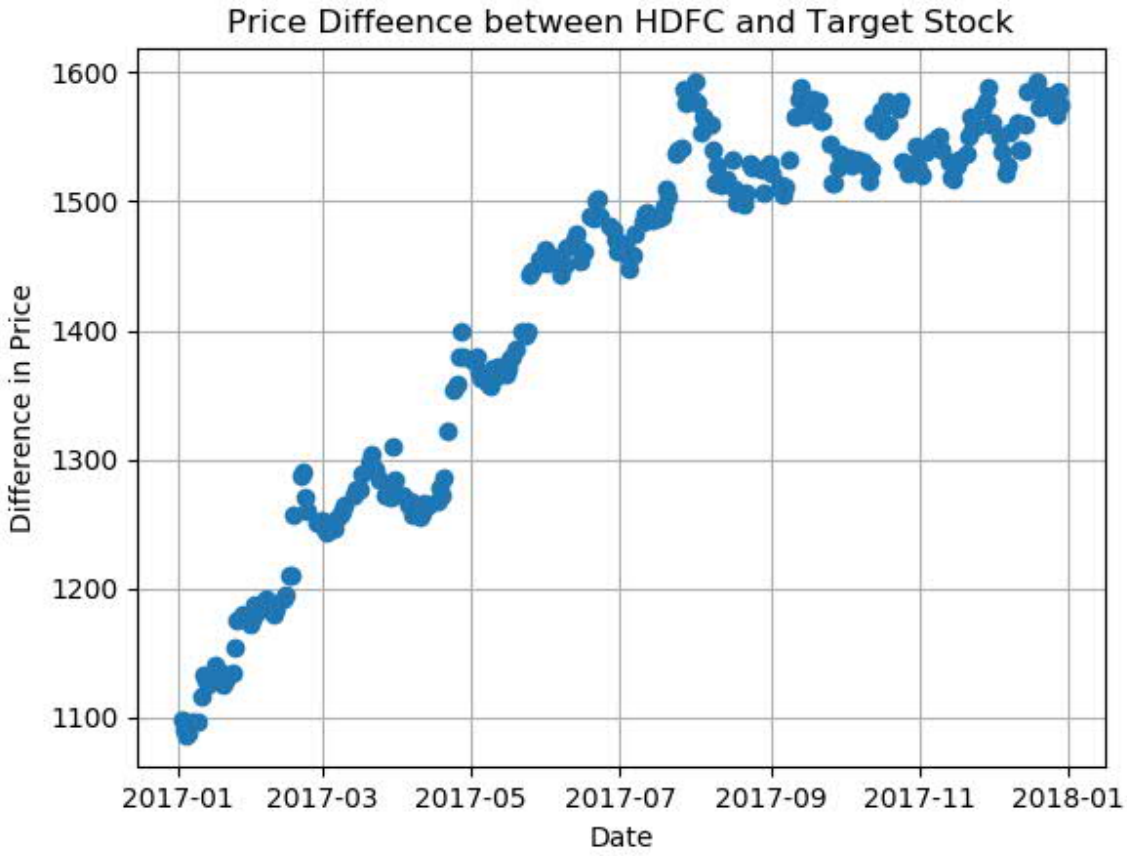
Output of KNN Regression:
k=1, accuracy=92.96%
k=3, accuracy=96.38%
k=5, accuracy=96.16%
k=7, accuracy=95.88%
k=9, accuracy=94.93%
k=3, achieved highest accuracy of 96.38%

Comparison of Actual vs. Predicted value of Target (using KNN):

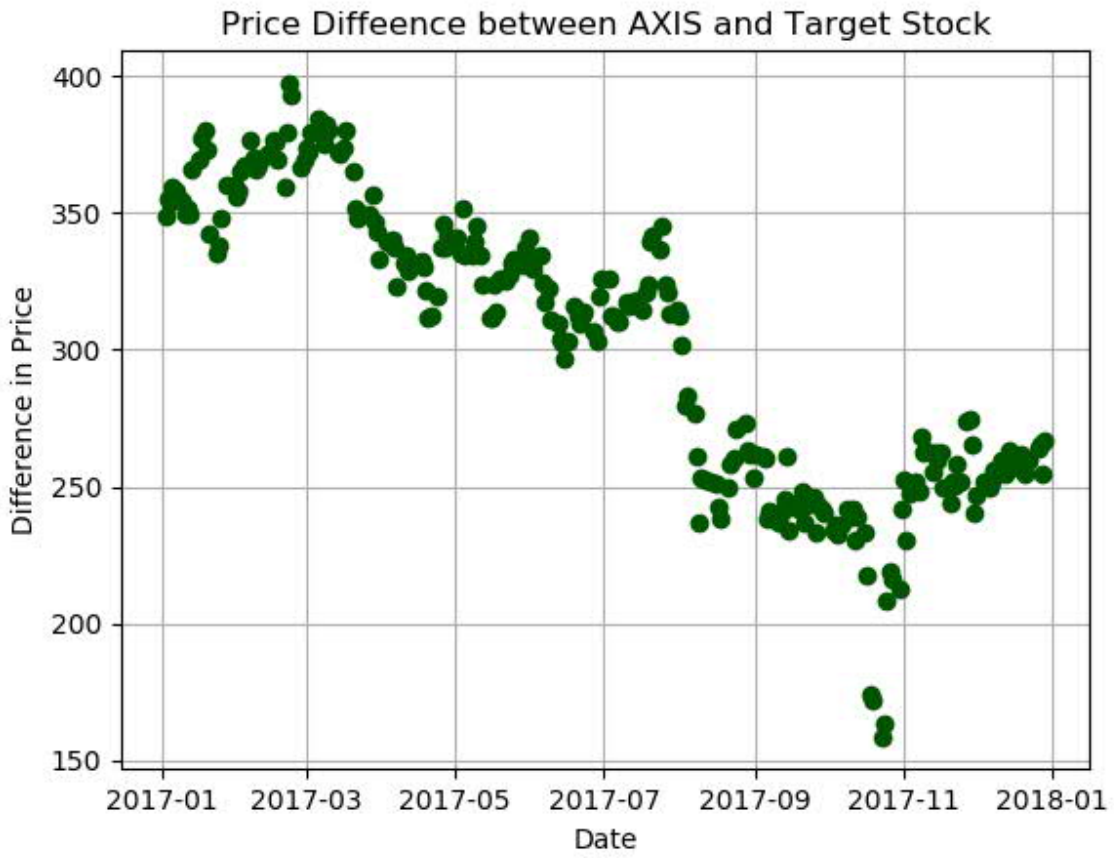


For Pair of HDFC and Target (EDELWEISS):

Co-integration Score: -11.817967718387173
pValue: 9.933693044347266e-21

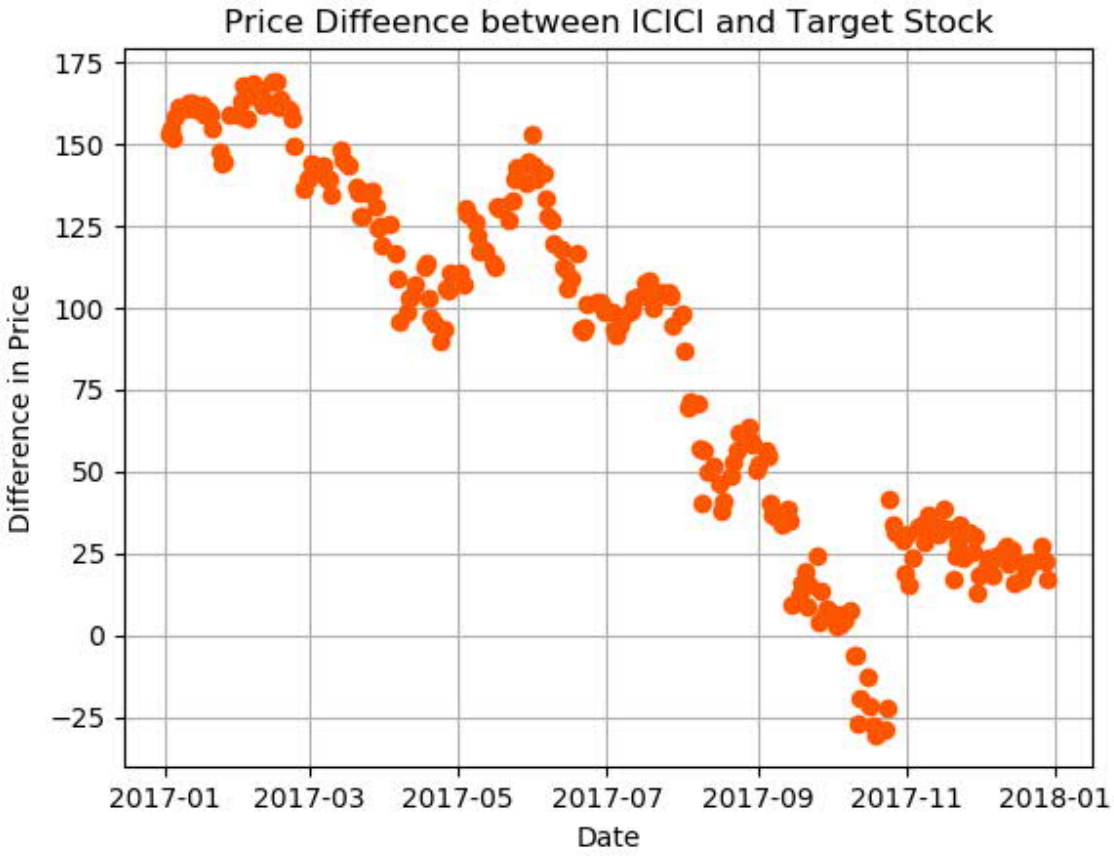


For Pair of AXIS and Target (EDELWEISS):
Co-integration Score: -13.116841680675433
pValue: 1.6361851461453943e-23



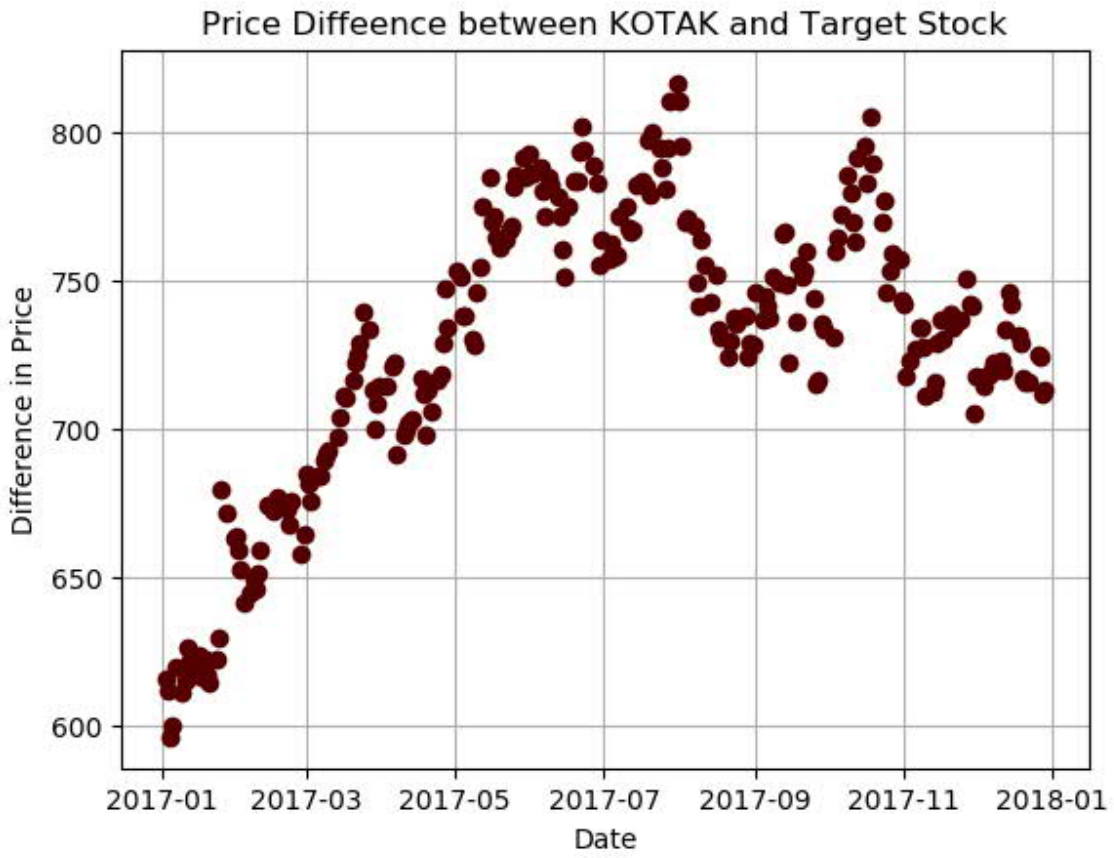
- ⇨ Price difference between Axis and EDELWEISS are actually decreasing over the year which means prices are merging towards each other till Nov'17 and then starts moving upward.

For Pair of ICICI and Target (EDELWEISS):
Co-integration Score: -12.572601010142705
pValue: 2.1560463268945154e-22

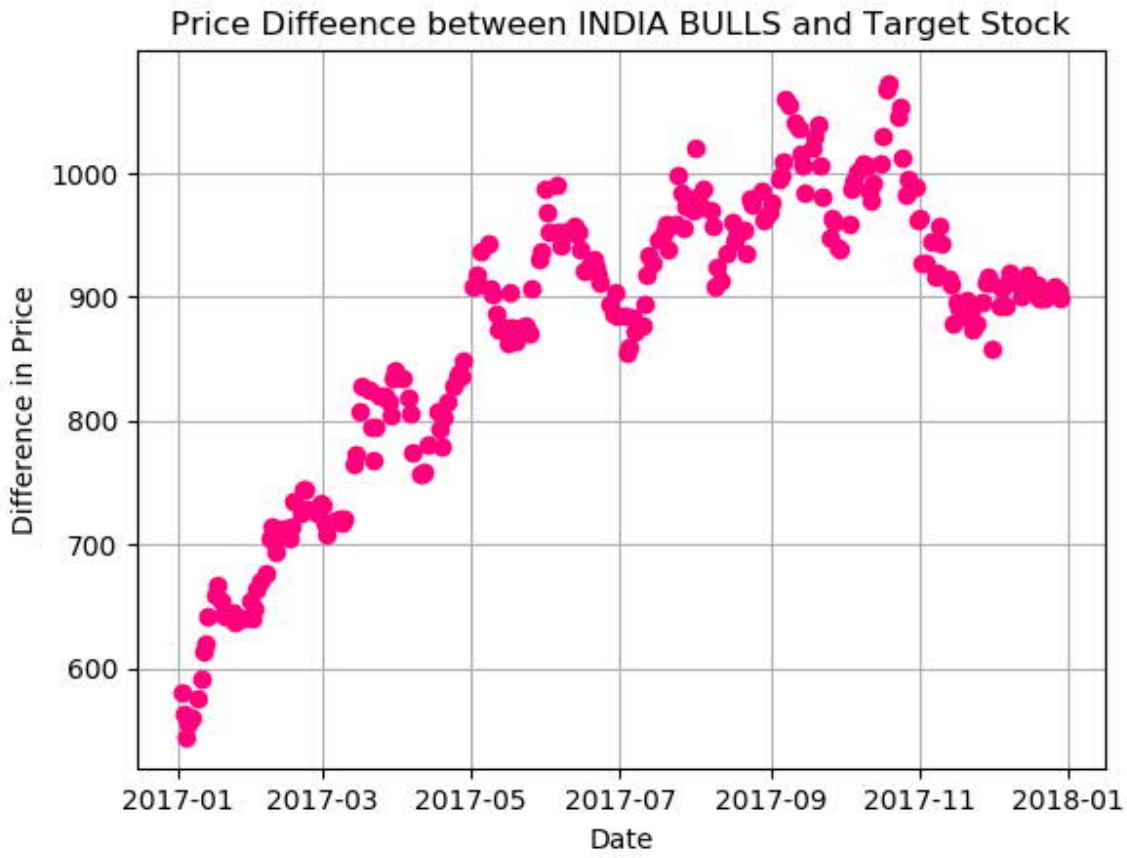


⇨ Same trend as above in case of Axis-Target

For Pair of KOTAK and Target (EDELWEISS):
Co-integration Score: -11.878567908553526
pValue: 7.231664733178348e-21



For Pair of INDIA BULLS and Target (EDELWEISS):
Co-integration Score: -11.800744164676091
pValue: 1.0874854046033709e-20



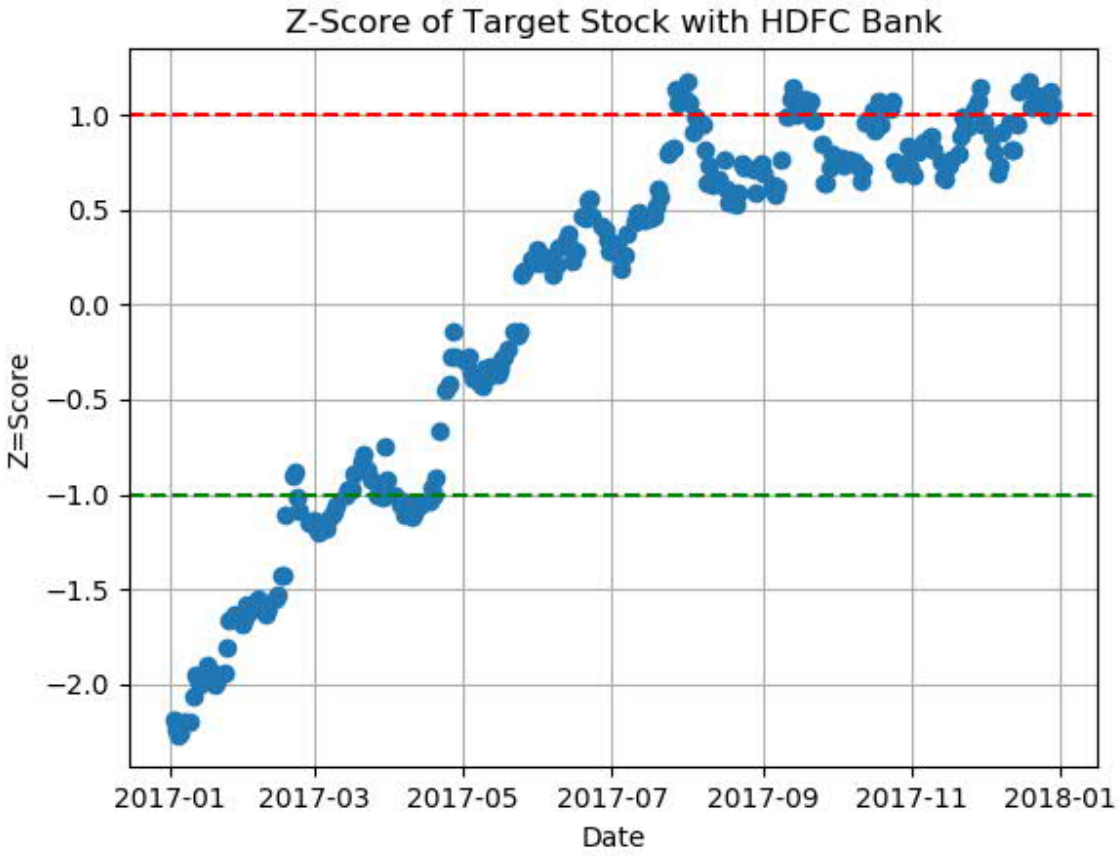
Co-Integration Matrix of Pair of Component Stocks & EDELWEISS

	Co-Integration Score	P-Value
HDFC Bank	-11.81796772	9.93E-21
Axis Bank	-13.11684168	1.64E-23
ICICI Bank	-12.57260101014270	2.16E-22
Kotak Bank	-11.87856791	7.23E-21
India Bulls	-11.80074416	1.09E-20

Prediction of Arbitrage Strategy for 2017:

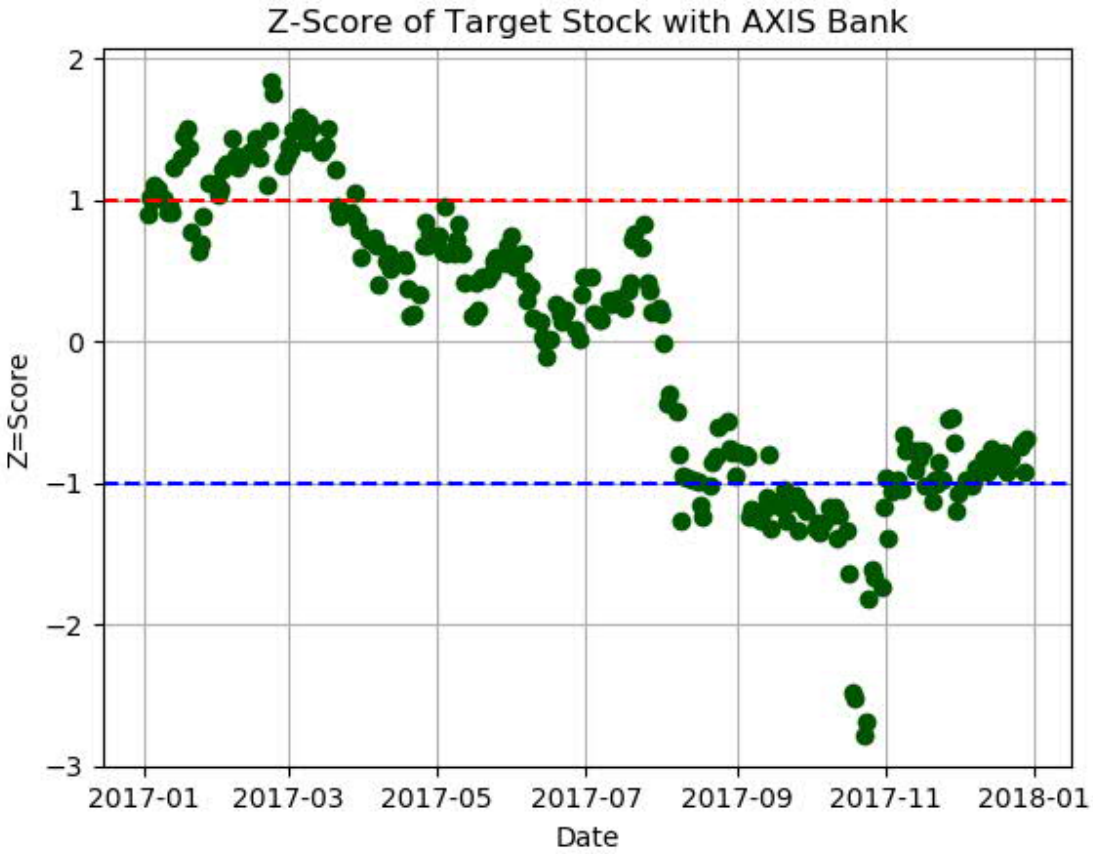
In this section, we will find out arbitrage opportunity for each pair of stocks one of the stock in the pair will be our target i.e. **EDELWEISS**

For Pair of HDFC and Target (EDELWEISS):



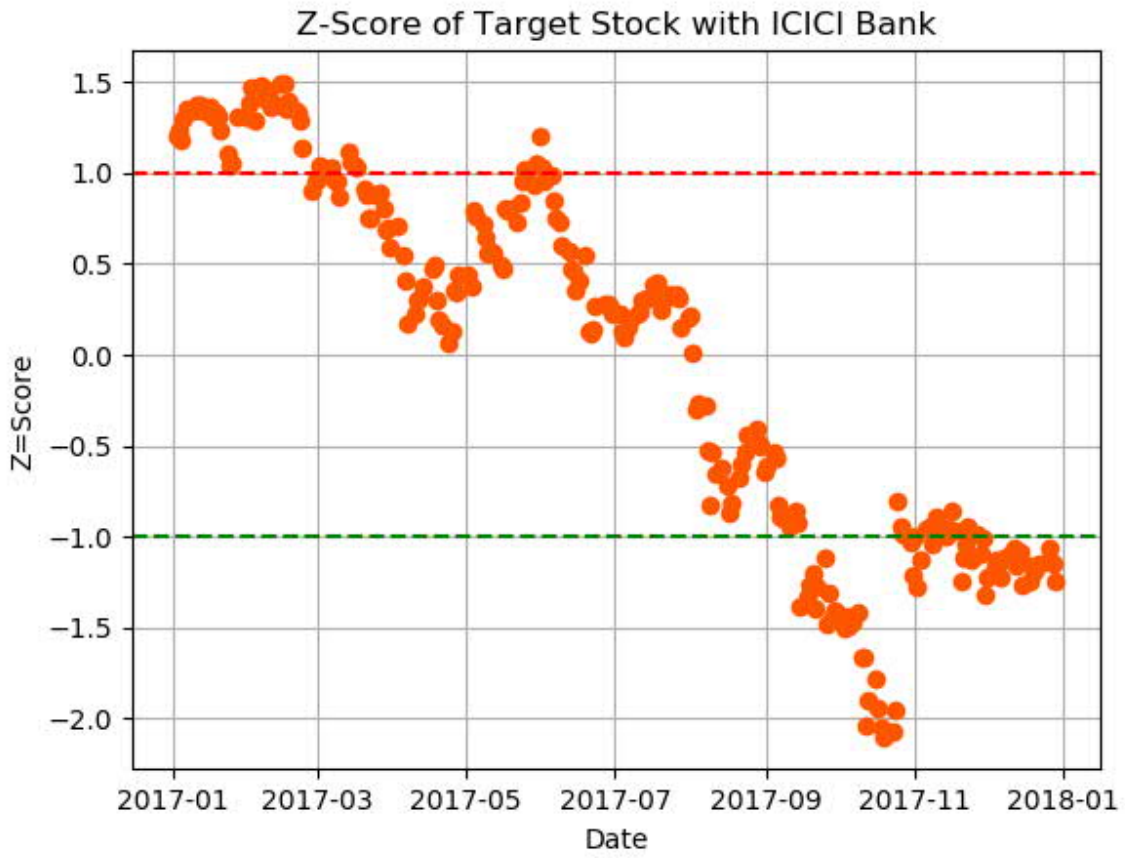
For pair of HDFC Bank with EDELWEISS, clearly in the first quarter of 2017, 'long' position is dominant which means investor should buy 1 share of HDFC and sell 1 share of EDELWEISS. However, there is no such material 'short' position during the year.

For Pair of AXIS and Target (EDELWEISS):



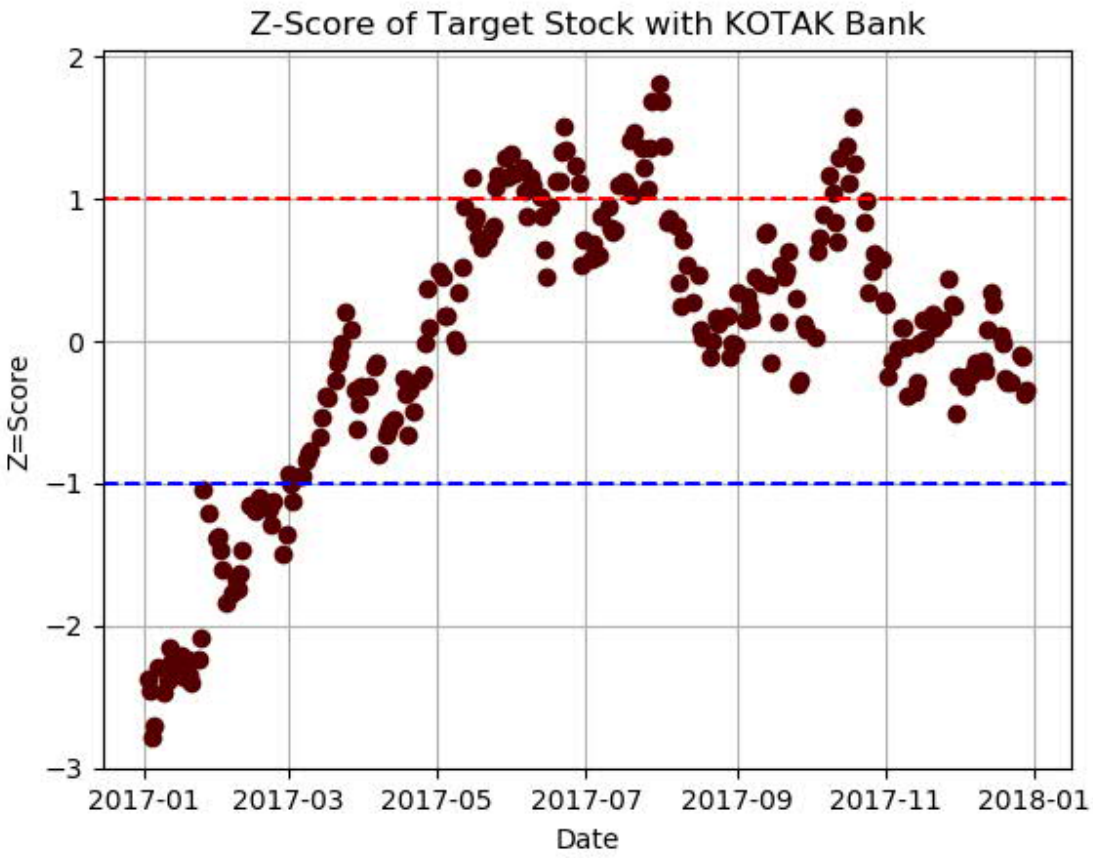
For pair of Axis Bank with EDELWEISS, in March'17, there is a 'short' position and in Nov'17, it is 'long' position, but arbitrage scope is not that significant, which is supported by the difference in price plot above indicating axis price is getting very close to target price over the year.

For Pair of ICICI and Target (EDELWEISS):



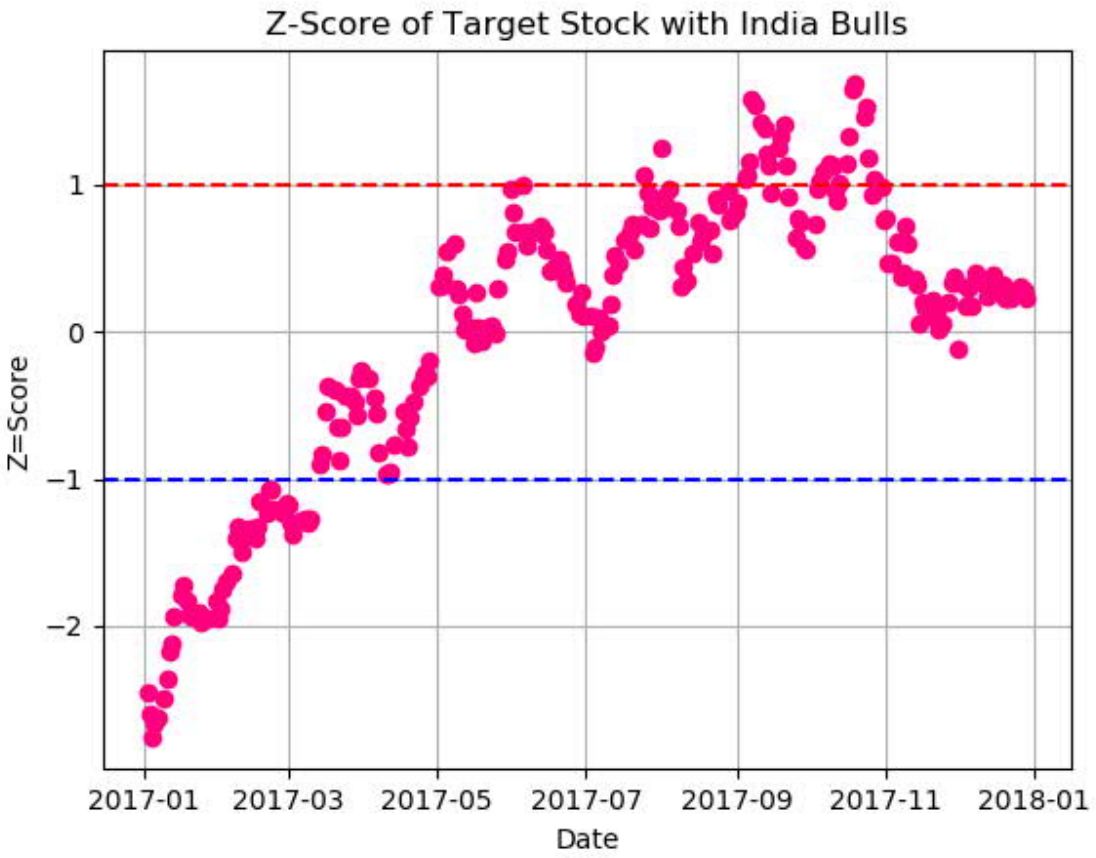
We can see there is huge arbitrage opportunity in Oct-Nov'17 being in "long" position while little opportunity in Jan-Feb'17 being in "short" position.

For Pair of KOTAK and Target (EDELWEISS):



Clearly investor should buy Kotak and sell EDELWEISS during Jan-Mar'17, but should sell Kotak and buy EDELWEISS mainly during July'17

For Pair of INDIA BULLS and Target (EDELWEISS):



Exactly same 'Long" and 'Short" position as between Kotak and Edelweiss as plotted above

BEYOND THE PROJECT➔

How about predicting 2018 Arbitrage?

Now, based on historical data of stocks for 2016 and 2017, I am planning to project stock price of our sample stocks in 2018 and find out future arbitrage opportunities. To do that we need to create 2018 dataset and then we need to predict the price of our target stock – EDELWEISS. Also in this exercise, test size has been modified to 40% instead of 20% to increase the 2017 predicted data points, otherwise in predicting 2018 price 20% of 50 is very less i.e.10 which doesn't produce any significant statistical results.

Hence KNN model is built with 60:40 ratio of train/test data set for 2017 below.
We will create 2018 dataset using predicted price already determined using 2017 data
This 2018 dataset will be having 100 data points for each stock as size of our predicted vector 40

KNN results using EDELWEISS as target:
k=1, accuracy=95.03%
k=3, accuracy=96.78%

k=5, accuracy=96.16%
k=7, accuracy=95.93%
k=9, accuracy=95.30%
k=3, achieved highest accuracy of 96.78%
Results would have been much better if we would have taken 80/20 ratio, but ultimate target vector length would have been very small of only 10. Want to avoid that

KNN results using HDFCBANK as target:
k=1, accuracy=98.12%
k=3, accuracy=98.43%
k=5, accuracy=97.88%
k=7, accuracy=97.37%
k=9, accuracy=97.21%
k=3, achieved highest accuracy of 98.43%
Size of HDFC Bank's 2017 predicted Stock Price: 100

KNN results using AXISBANK as Target:
k=1, accuracy=84.01%
k=3, accuracy=80.64%
k=5, accuracy=80.04%
k=7, accuracy=72.97%
k=9, accuracy=68.97%
k=1, achieved highest accuracy of 84.01%
Size of Axis Bank's 2017 Predicted Stock Price: 100

KNN results using ICICIBANK as Target:
k=1, accuracy=80.50%
k=3, accuracy=79.27%
k=5, accuracy=71.16%
k=7, accuracy=65.33%
k=9, accuracy=61.17%
k=1, achieved highest accuracy of 80.50%
Size of ICICI Bank's 2017 Predicted Stock Price: 100

KNN results using KOTAKBANK as Target:
k=1, accuracy=98.10%
k=3, accuracy=97.86%
k=5, accuracy=97.70%
k=7, accuracy=97.56%
k=9, accuracy=97.25%
k=1, achieved highest accuracy of 98.10%
Size of Kotak Bank's 2017 Predicted Stock Price: 100

KNN results using INDIABULL as Target:
k=1, accuracy=96.71%
k=3, accuracy=97.58%
k=5, accuracy=96.82%
k=7, accuracy=96.10%

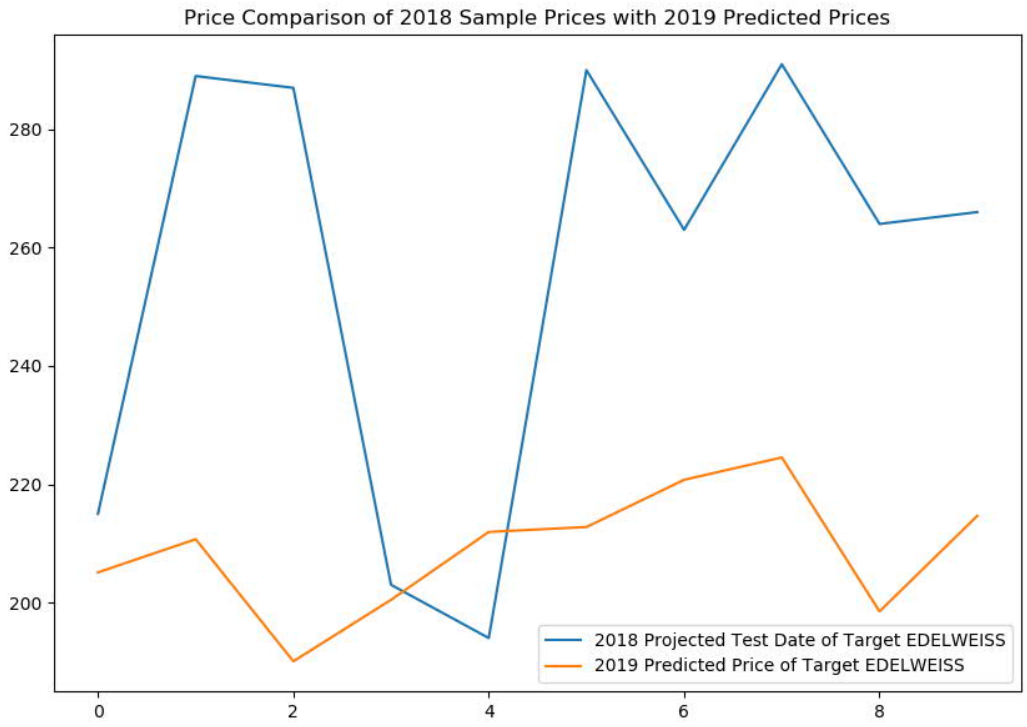
k=9, accuracy=95.86%
k=3, achieved highest accuracy of 97.58%
Size of India Bull's 2017 Predicted Stock Price: 100

2018 Sample Dataset (based on 2017 predicted stock price):

	HDFCBANK	ICICIBANK	AXISBANK	KOTAKBANK	INDIABULLS	EDELWEISS
0	1810.0	300.0	506.0	1030.0	1190.0	110.0
1	1750.0	294.0	512.0	907.0	670.0	194.0
2	1760.0	281.0	510.0	979.0	757.0	266.0
3	1780.0	287.0	489.0	970.0	731.0	289.0
4	1200.0	285.0	512.0	996.0	1220.0	147.0

I adopted Linear Regression and got predicted values as compared to test Label data as:

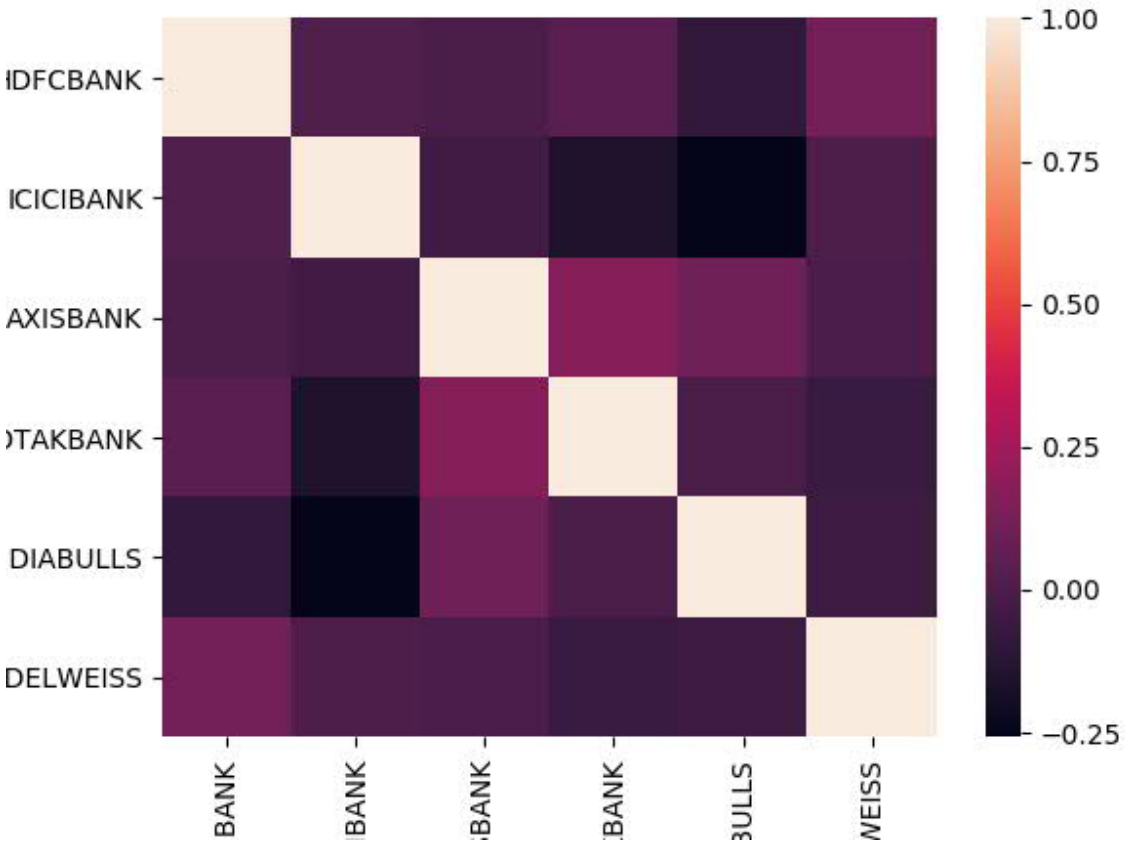
2018_predicted	2017_test_data
205.09	215
210.71	289
190.07	287
200.5	203
211.93	194
212.77	290
220.75	263
224.53	291
198.49	264
214.65	266



The prediction for 2018 indicates lower stock price compared to actual 2017 sample data. It will be interesting to see correlation of target stock with other sample stocks. Here is the map of correlation coefficient between stocks:

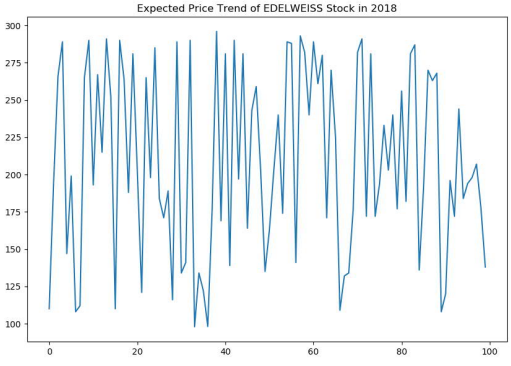
Correlation Coefficients from highest to lowest with Target:
EDELWEISS 1.000000
HDFCBANK 0.111480
ICICIBANK -0.001037
AXISBANK -0.005935
INDIABULLS -0.050536
KOTAKBANK -0.062468

The above indicates no correlation between individual stocks. Heat-map says it all

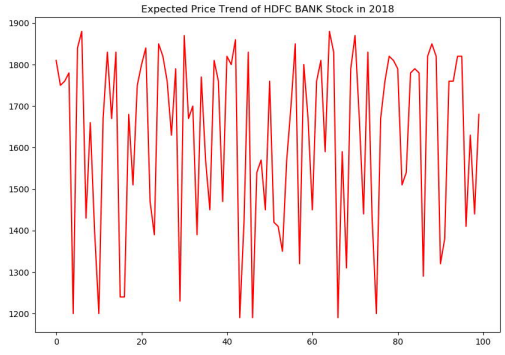


The above display prompts one important fact that prediction of predicted values does not make sense. We were trying to split predicted values to predict future.

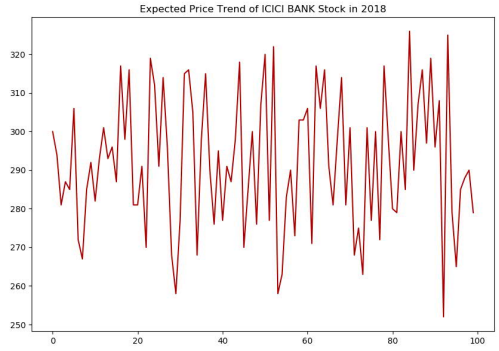
Plotting individual stocks' 2018 price



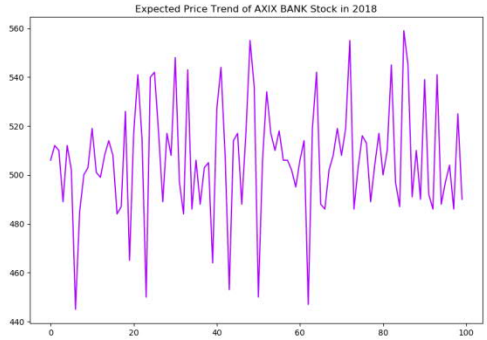
EDELWEISS (2018)



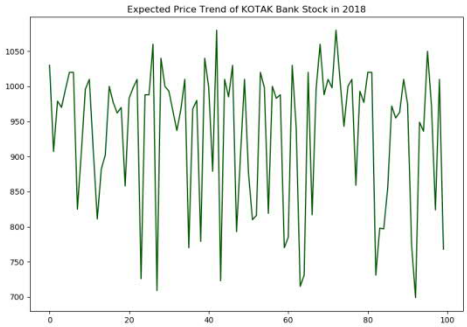
HDFC (2018)



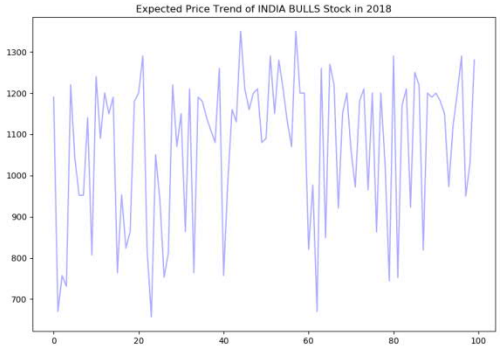
ICICI Bank (2018)



AXIS Bank (2018)



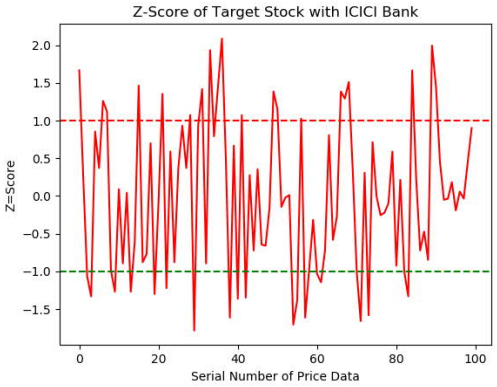
KOTAK BANK (2018)



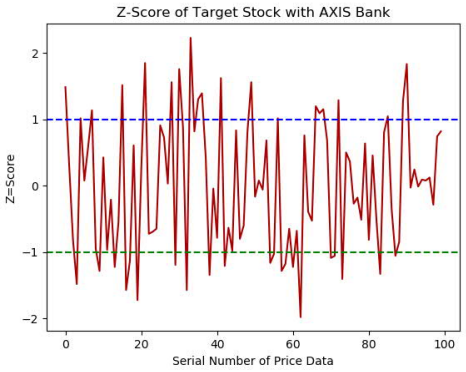
INDIA BULLS (2018)

From the above plots, it is clear that predicted values are comparable; hence we can proceed further with prediction of arbitrage in 2018. In the z-score plots below, we could not predict arbitrage by time period as we don't have predicted data on date time. The X-axis indicates price at each data position just like serial number of data entries.

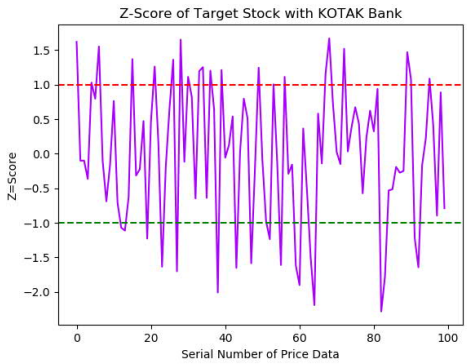
HDFC-EDELWEISS Pair (2018)



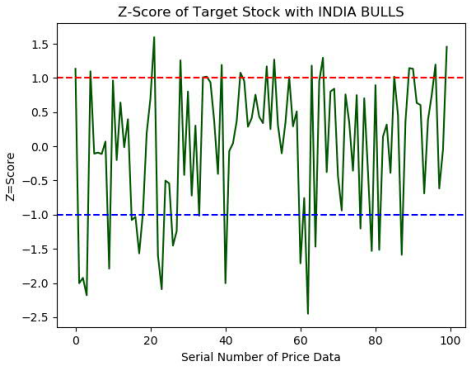
ICICI-EDELWEISS Pair (2018)



AXIS-EDELWEISS Pair (2018)



KOTAK-EDELWEISS Pair (2018)



INDIABULLS-EDELWEISS Pair (2018)

"Long" the spread whenever the z-score is below -1.0
Go "Short" the spread when the z-score is above 1.0
Exit positions when the z-score approaches zero

"Long" the spread i.e. z-score below -1.0 means 1 share of EDELWEISS to sell and 1 share of paired stock to buy. “Short” the spread i.e. z-score above 1.0 means 1 share of EDELWEISS to buy and 1 share of paired stock to sell.
In the above z-score plot of each pair of individual stock with target stock, we could predict Arbitrage Opportunities in 2018

**********END**********