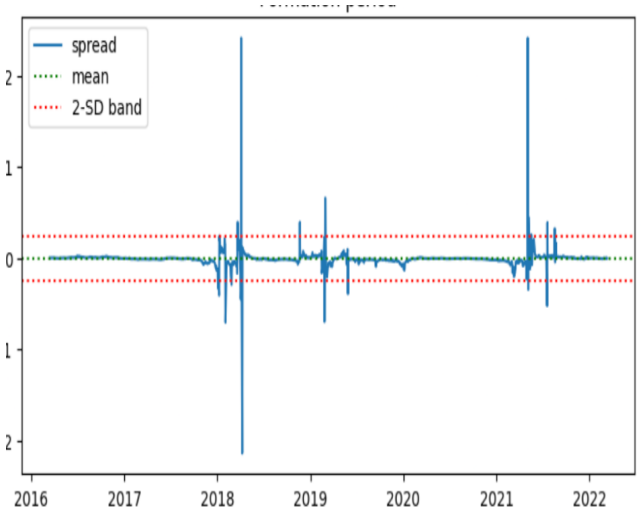
PAIR SELECTION

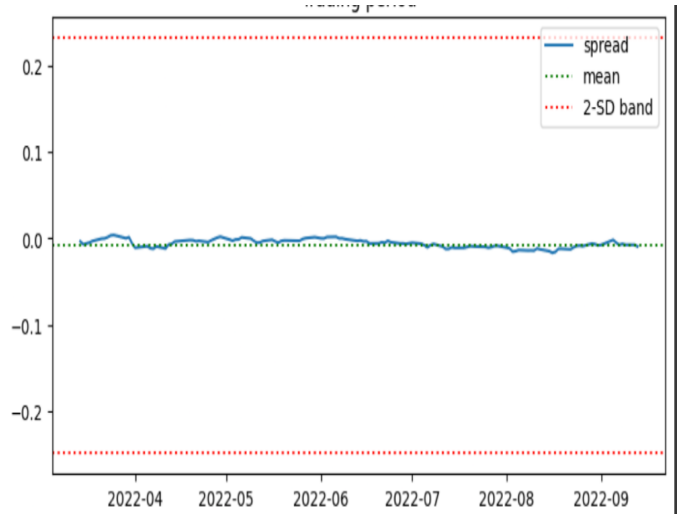
1.Pairs didn't show mean reversion behavior using shortest Euclidean Distance as parameter. it might not capture direction and magnitude of of relationship between the assets as effective as other parameters

CADF P-value optimal range is not satisfied (all p-value<0.01)
ADF P-value optimal range is not satisfied (all p-value <0.05)
Hurst Exponent optimal range is not satisfied (0<H<0.5)

```
pairs48_36.sort_values(by='Distance')
```

	Distance	Num zero-crossings	Pearson r	Spread mean	Spread SD	Hurst Exponent	Half-life of mean reversion	Hedge ratio
Instr13-Instr15	5.207439	59.0	0.980251	-0.011598	0.164921	0.377067	1.046612	0.958129
Instr13-Instr16	5.990257	55.0	0.973781	-0.015984	0.189713	0.342856	1.059581	0.957957
Instr12-Instr15	7.079409	55.0	0.963564	-0.020289	0.224207	0.391095	1.157321	0.946647
Instr2-Instr3	7.224833	49.0	0.973940	-0.019601	0.228813	0.398039	1.026466	0.986419
Instr5-Instr3	7.723215	68.0	0.966340	0.015500	0.244597	0.331505	1.086225	0.922349
...	...	...	...	...	...	...	...	...





2.Number of zero crossings

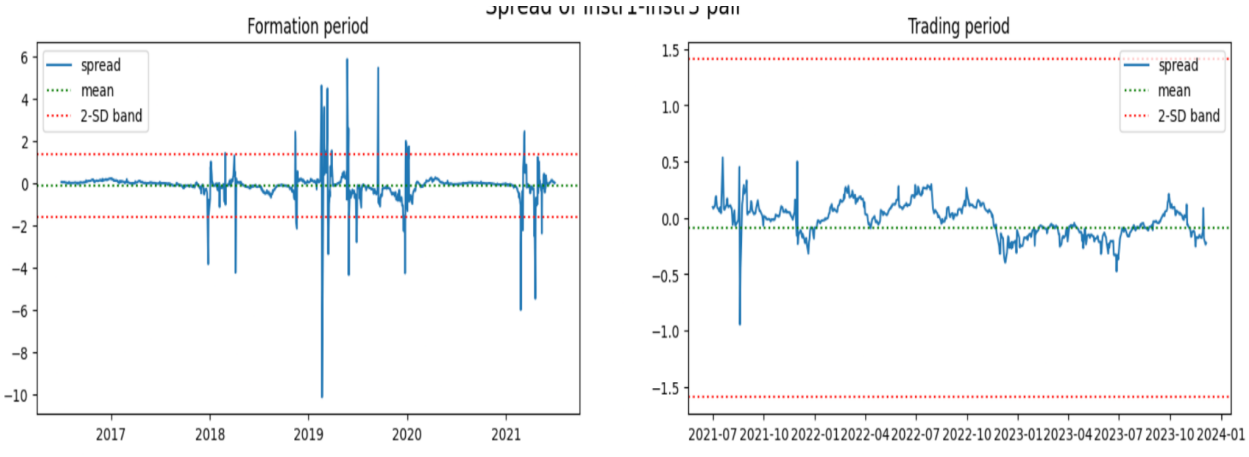
Many pairs have high number of zero crossings but may not optimal in Trading period because of other parameters such as CADF ‘p’ ,value ADF p value ,hurst exponent etc
These pairs have high number of zero crossings but pearson r and hurst exponent are out of optimal range

CADF P-value optimal range is not satisfied (all p-value<0.01)
ADF P-value optimal range is not satisfied (all p-value <0.05)
Hurst Exponent optimal range is not satisfied (0<H<0.5)

```
[ ] pairs48_36.sort_values(by='Num zero-crossings', ascending=False)
```

	Distance	Num zero-crossings	Pearson r	Spread mean	Spread SD	Hurst Exponent	Half-life of mean reversion	Hedge ratio
Instr12-Instr1	15.740888	88.0	0.803925	0.077197	0.498519	0.336020	1.651213	0.651888
Instr10-Instr5	9.444148	86.0	0.936933	0.048451	0.299099	0.328421	1.089814	0.843343
Instr13-Instr3	14.035674	86.0	0.846101	0.074251	0.444514	0.295572	1.576632	0.708378
Instr13-Instr1	16.272322	86.0	0.786208	0.078857	0.515350	0.323201	1.904827	0.634271
Instr10-Instr6	8.877180	82.0	0.944495	0.037577	0.281143	0.339951	1.009428	0.837666
...	...	...	...	...	...	...	...	...
Instr4-Instr10	13.243300	49.0	0.905784	-0.101287	0.419420	0.329735	1.070602	1.047647
Instr6-Instr16	16.508910	49.0	0.840472	-0.123823	0.522842	0.328791	1.643807	0.956653
Instr3-Instr18	22.211365	49.0	0.708008	-0.186741	0.703441	0.324403	1.982595	0.752896
Instr4-Instr16	18.696044	49.0	0.801340	-0.159621	0.592110	0.323374	1.802752	0.935643
Instr4-Instr13	16.103788	49.0	0.857030	-0.138938	0.510012	0.325629	1.326280	1.017197

130 rows × 8 columns



3.Pearson r

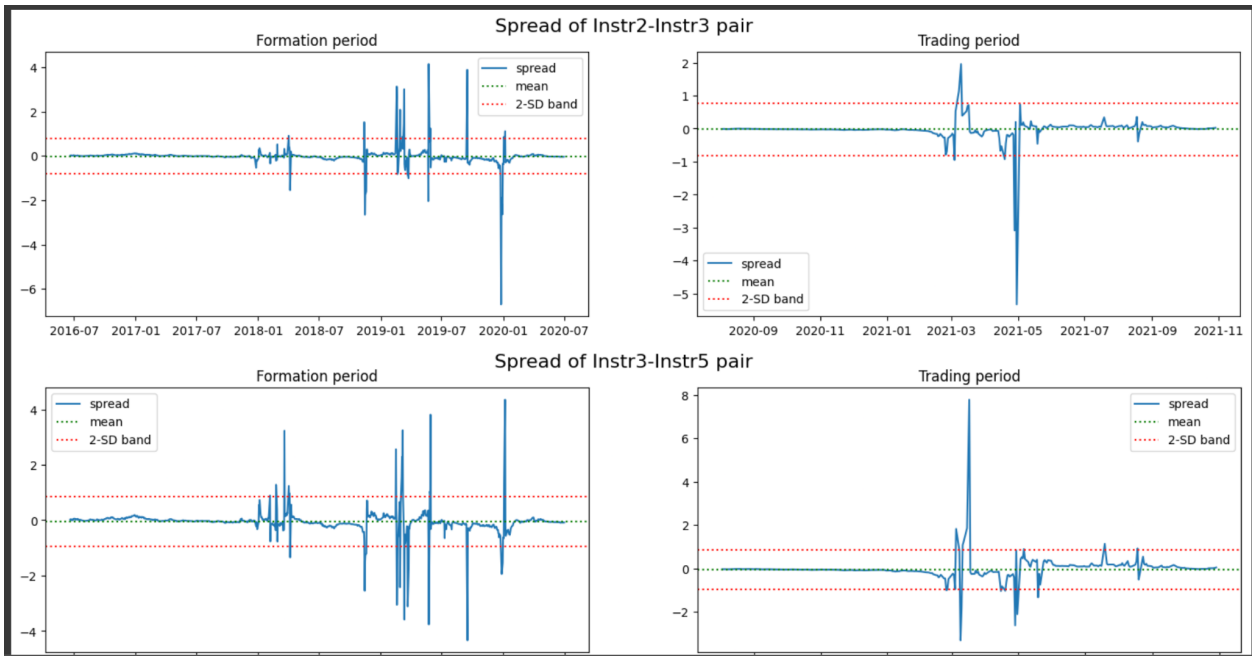
Pearson r identifies pairs that have strong historical relationships.it gives strength and direction of collinearity between the pairs.we can establish correlation threshold for finding potential trading opportunity

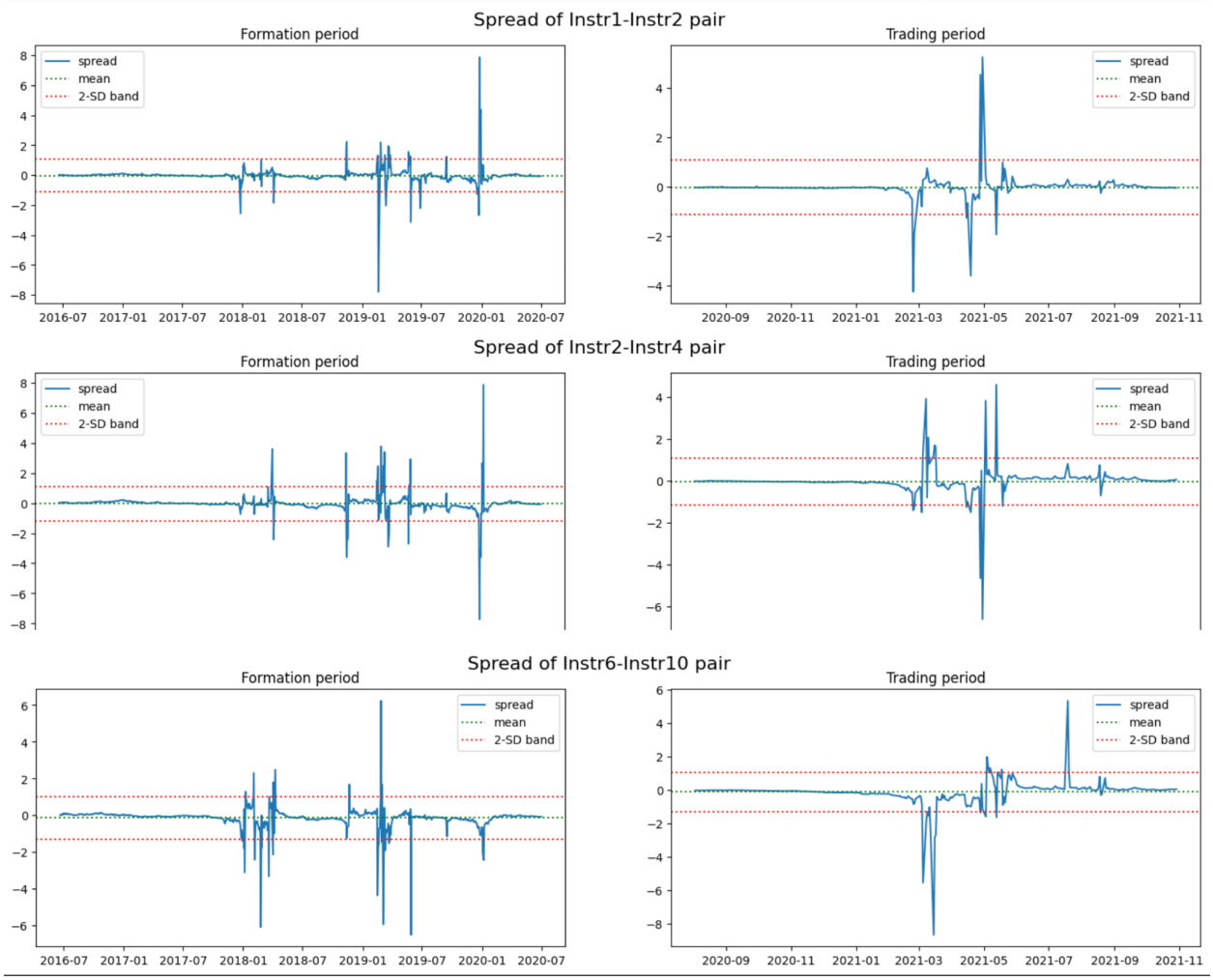
CADF P-value optimal range is satisfied (all p-value<0.01)
ADF P-value optimal range is satisfied (all p-value <0.05)
Hurst Exponent optimal range is satisfied (0<H<0.5)

These are the TOP 5 Pairs which have best potential among others pairs to Trade.

```
calculate_metrics(top5_pearson, test12_6, pairs12_6)
```

	Euclidean distance	CADF p-value	ADF p-value	Spread SD	Pearson r	Num zero-crossings	Hurst Exponent	Half-life of mean reversion	% days within historical 2-SD band
Instr2-Instr3	7.353046	0.000068	0.00001	0.399058	0.979823	13	0.369644	0.776569	97.826087
Instr3-Instr5	11.815422	0.000124	0.000018	0.450056	0.947659	13	0.409262	1.188314	94.409938
Instr1-Instr2	9.91908	0.000298	0.000011	0.552283	0.962969	26	0.44488	0.936453	97.515528
Instr2-Instr4	12.398662	0.000011	0.000001	0.566409	0.942354	13	0.305336	0.667357	95.031056
Instr6-Instr10	14.507469	0.018441	0.003679	0.581266	0.921979	23	0.421666	1.372063	94.409938




Spread of Instr1-Instr2 pair
Formation period
Trading period
spread
mean
2-SD band
Spread of Instr2-Instr4 pair
Formation period
Trading period
spread
mean
2-SD band
Spread of Instr6-Instr10 pair
Formation period
Trading period
spread
mean
2-SD band