

Multivariate Analysis-Oil Price Prediction Using LSTM & GRU

Data Description: -

Source of Data Collection: Individual Data series collected from (<https://in.investing.com/>) and (<https://finance.yahoo.com/>) between 4th Jan 2000 to 10th June 2019.

Data Dimensions: -

- 4947 Attributes
- 7 Features

Date: Date of Trading

WTI : Includes spot values of West Texas Intermediate which is mix of crude oil traded on NYMEX.

GOLD : With Increase in Oil price ,inflation increases and in turn GOLD futures also increases. Gold futures is a good indicator of economic situations.

SP 500 : Overall market stability can be ascertained by S&P 500 as it contains 500 large companies. Can indicate market sentiments and economic conditions.

US DOLLAR INDEX: Represents US Dollar Index Futures . Foreign exchange market has quite an impact on oil prices since the US dollar is international metric to interpret the global economic conditions.

US 10YR BOND: Represents US 10-Year Bond yield . Is a certificate of loan with federal government that matures in 10 years. Great indicator of investor sentiment about the economy.

DJU: Represents Dow Jones Utility Average. Performance of 15 important utility stocks, these being sensitive to changes in the price of commodities such as natural gas or coal.

Quick Summary Table

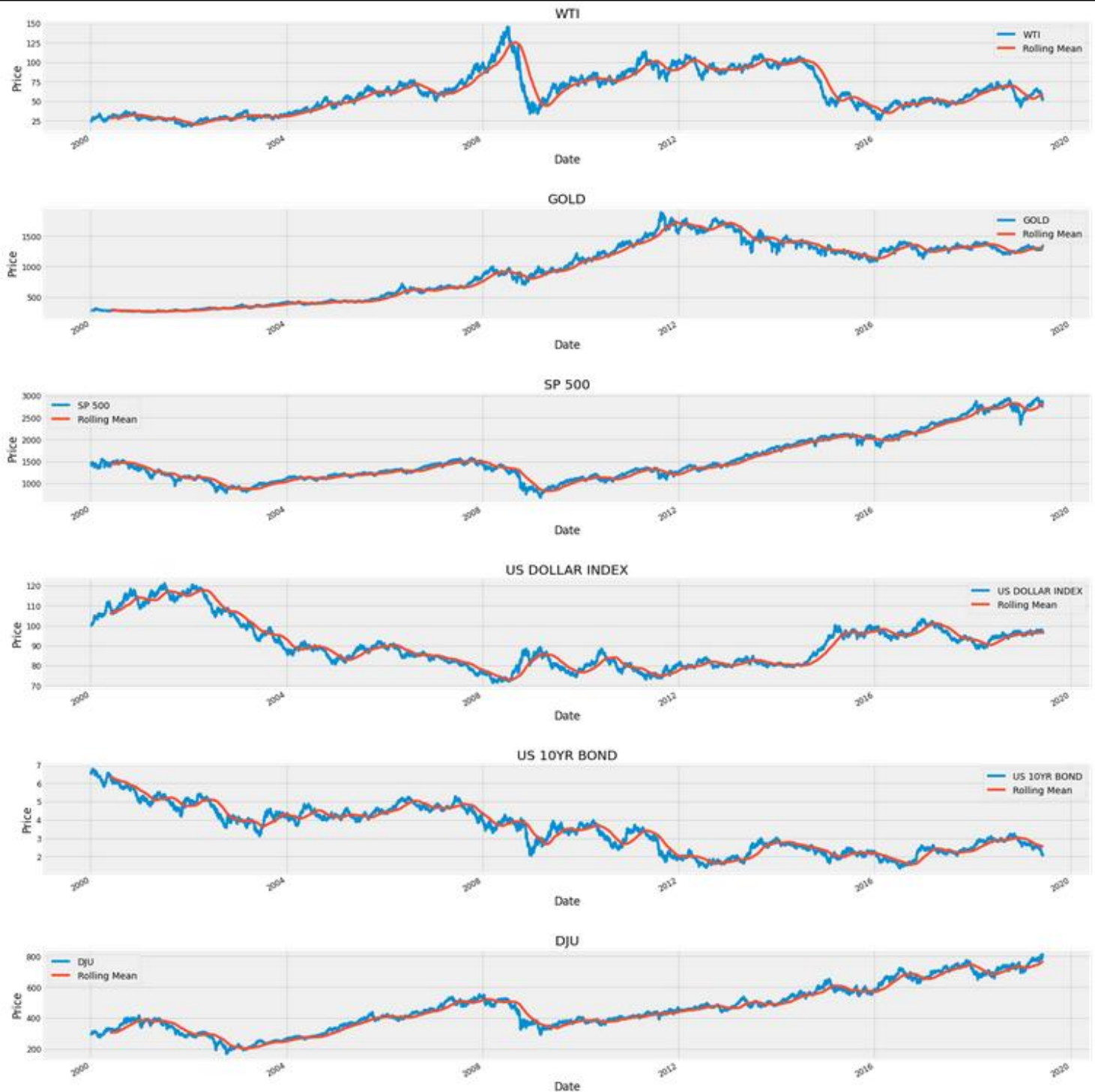
	WTI	GOLD	SP 500	US DOLLAR INDEX	US 10YR BOND	DJU
count	4947.000000	4941.000000	4946.000000	4917.000000	4920.000000	4859.000000
mean	62.068803	943.852311	1544.704848	90.503182	3.453490	459.169801
std	26.470270	473.656860	544.716725	11.691564	1.222532	147.334229
min	17.450000	255.100000	676.000000	71.304000	1.358000	167.570007
25%	40.115000	424.000000	1150.250000	80.946000	2.382000	359.664994
50%	59.300000	1088.500000	1356.250000	88.350000	3.389500	438.119995
75%	83.725000	1329.800000	1945.687500	96.940000	4.413000	553.309998
max	145.290000	1888.700000	2948.500000	121.210000	6.790000	812.909973

Target Variable- (WTI)

Features -(GOLD), (SP 500), (US DOLLAR INDEX), (US 10YR BOND), (DJU)

Exploratory Data Analysis

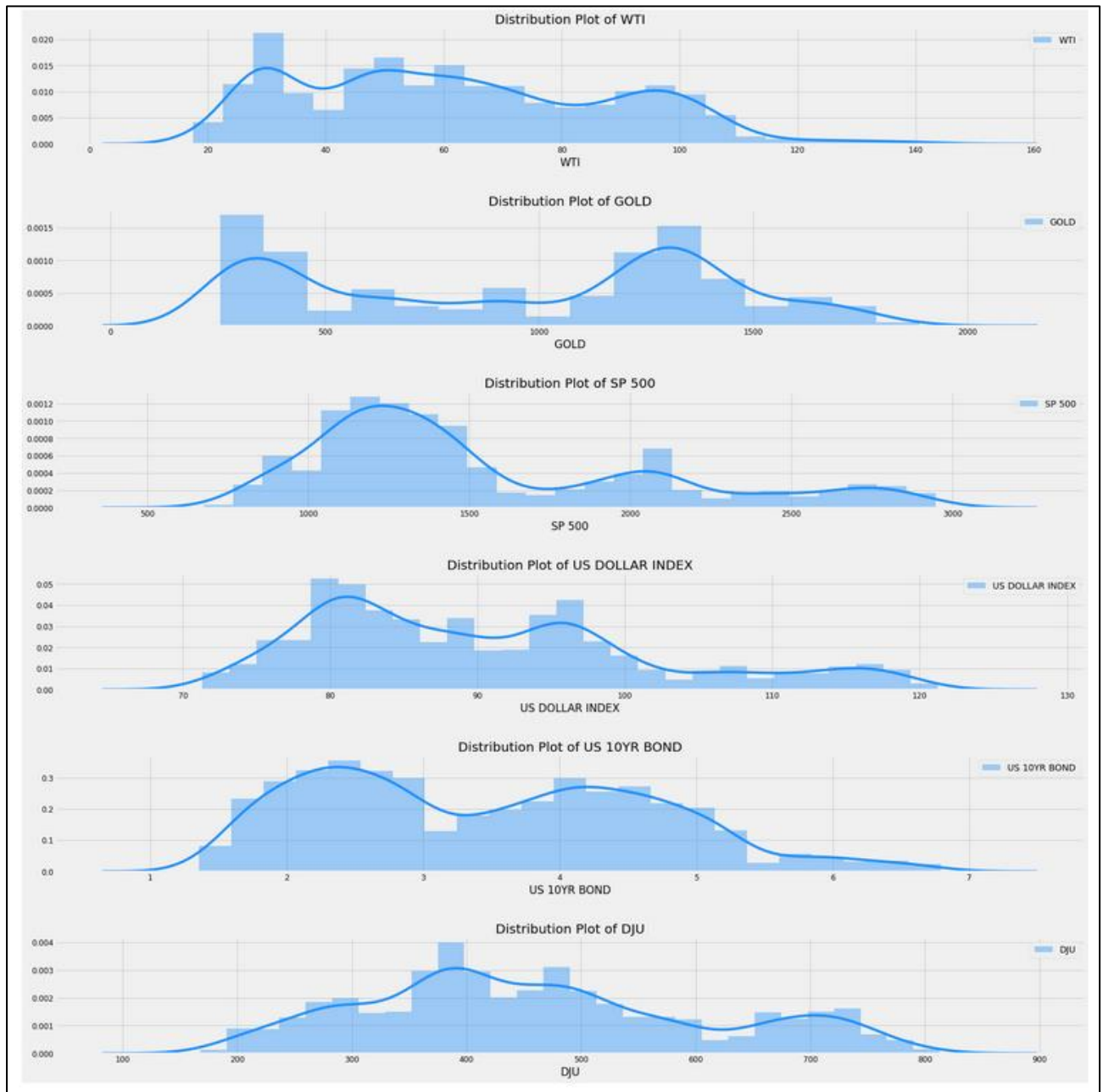
Univariate Analysis



Inferences: -

- WTI has mix of both UP and DOWN trends.
- Gold, DJU has UP Trend.
- US Dollar Index, US 10YR Bond has downward trend

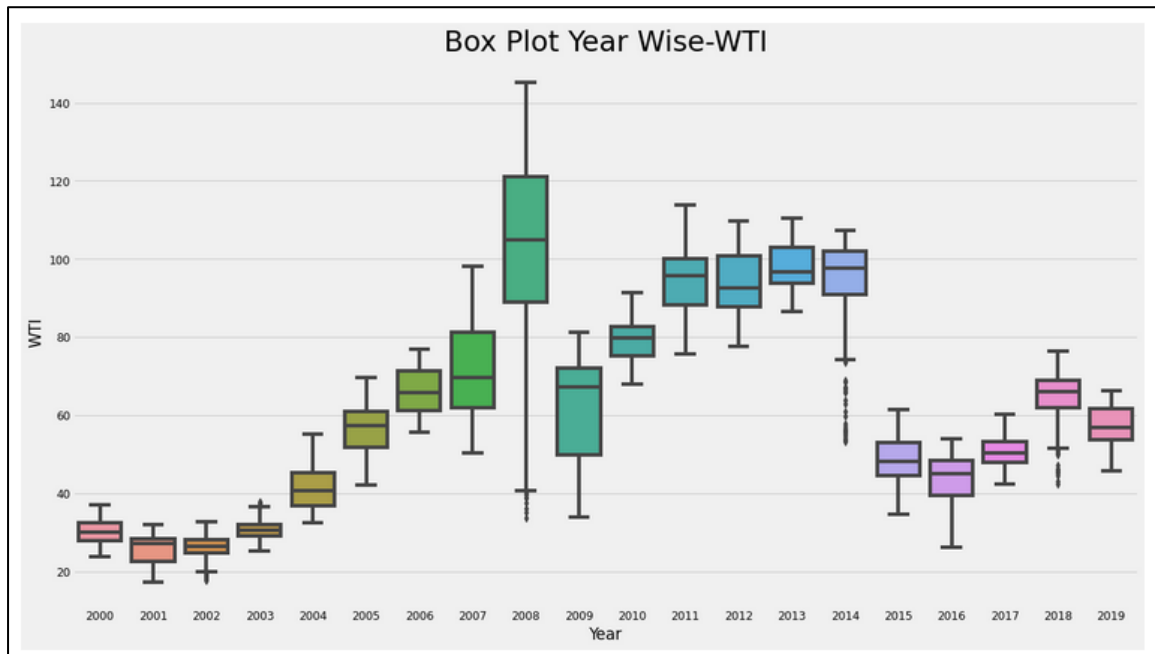
Distribution Plots



Inference from above Distribution plots: -

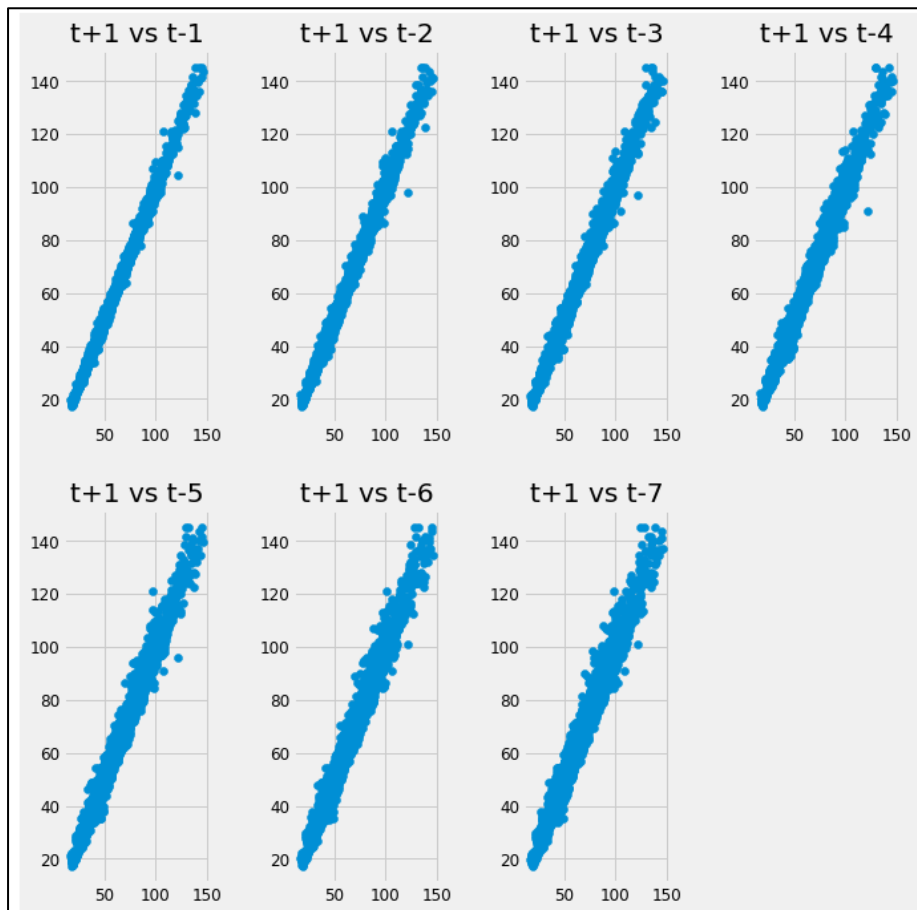
- Most frequent values of WTI are in the range of 25 to 35.
- The WTI oil price values above 120 are less frequent.
- All features & WTI don't follow Normal Distribution.

Box Plots



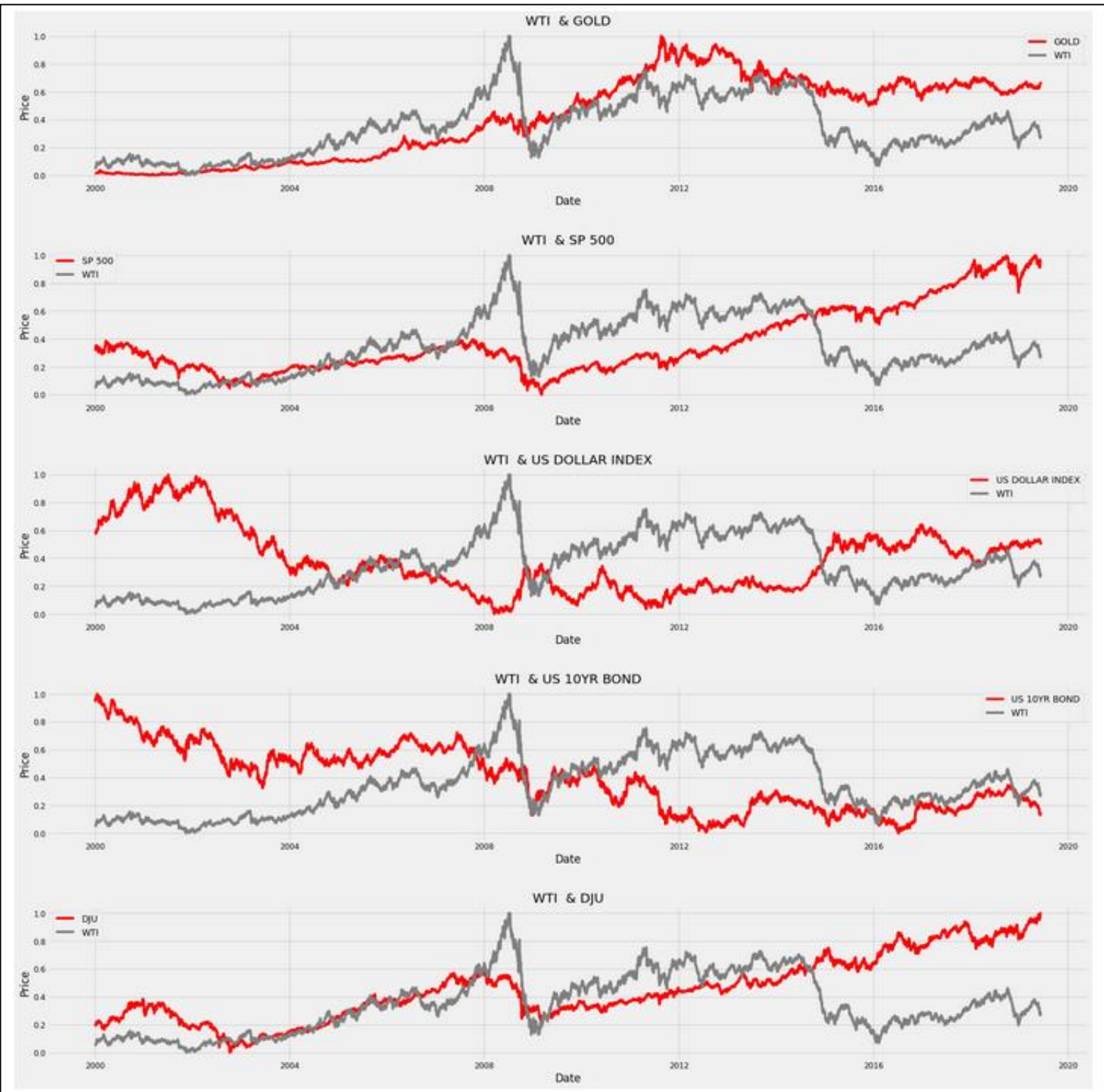
- During the years 2007, 2008, and 2009 wide variations in the values of WTI have been observed.
- In the year 2008, which is the period of the financial crisis and huge volatility, the WTI oil price varies between 30-140\$.
- Years 2002, 2008, 2014 & 2018 show outlier.

Checking Lag Order



- The above Lag plot indicates that WTI has good positive correlation with each of its lagging orders.
- This kind of series is good for time series as its showing auto-regressive nature.

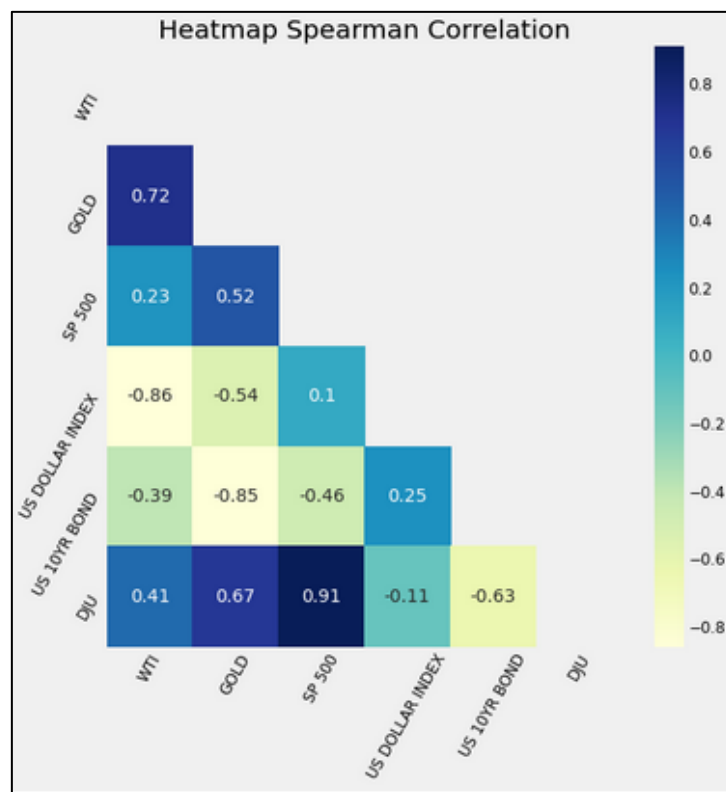
Bivariate Analysis



Inferences from Bivariate Feature Analysis: -

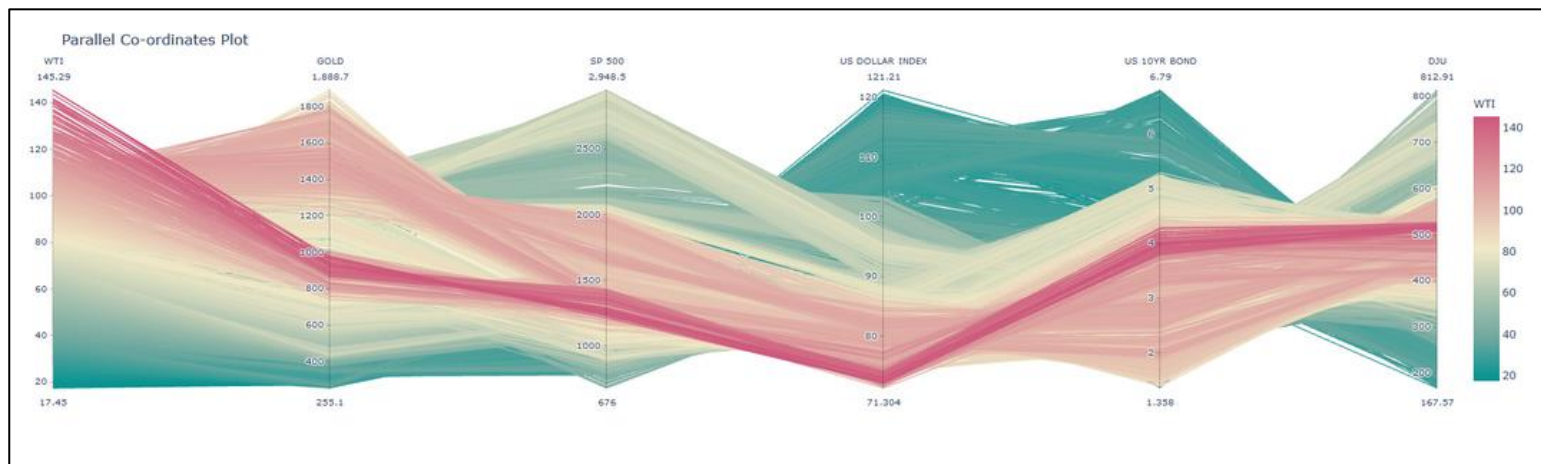
- During 2009 the value for both WTI and SP 500 drop sharply indicating that both these series are affected by same external factors. Move together.
- US Dollar index coinciding at several points with WTI, meaning that both the variables are strongly correlated. Apart from some common points the relationship seems to be inverse specifically when observing periods of 2000, 2007- 09 and 2010-16. Don't Move together.
- Both WTI and gold are fluctuating together. The gold seems to be lagging to the WTI in some ways especially if we observe years 2007 wherein gold has not had much difference as WTI. Move together.
- US 10-year bond interest rates seem to be a leading indicator to volatility of WTI price, as we can see specifically from 2000-09 US bond progresses downward and then the WTI follows it. The US interest rate basically gave us a leading indication to financial crisis is what we have observed. Don't Move together.
- DJU & WTI move together. Very High correlation between 2003-2009.

Multivariate Analysis: -



- The highest positive correlation with target variable WTI is with Gold($\rho:0.72$)
- We also observe that SP500 and WTI have less correlation. ($\rho : 0.23$)
- US 10 YR Bond and WTI are moderately inversely correlated. ($\rho : -0.39$)
- DJU & WTI positive moderate correlation. ($\rho : 0.41$)
- US Dollar Index & WTI have high correlation.($\rho : -0.86$)
- We observe high correlation between DJU & SP 500 ($\rho : 0.91$) observed the SP 500 and DJU also have very high positive correlation. This is kind of Expected since both are indexes.

Parallel Co-ordinates Plots

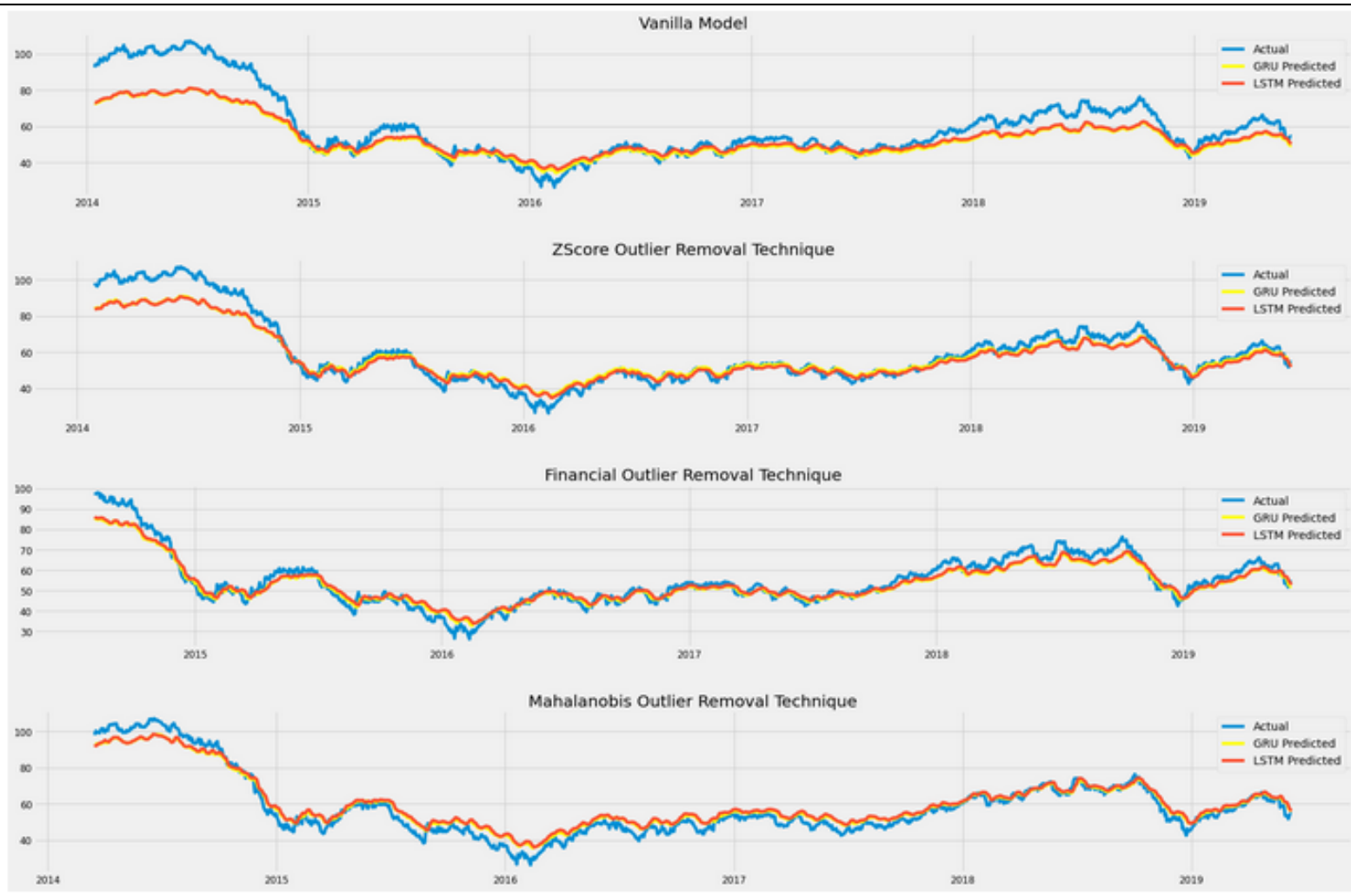


Inferences from Parallel Co-ordinates Plot: -

- The higher values of WTI correspond to average values for GOLD, US 10 YR BOND, DJU. (Red shaded Path)
- Low Values of WTI correspond to High Values of US Dollar Index (Dark Green Shaded Path).

Since Target & Feature don't follow normal distribution, we will use **Spearman Correlation**.

Forecasting Using LSTM & GRU



General Issues with Traditional RNN's

- Suffer from short memory due to vanishing gradients they forget initial information.
- LSTM & GRU tackle short term memory problem. They are able to regulate flow of information.

Comparing LSTM & GRU

Long Term Short Memory (LSTM): -

- Have three gates (input, output & forget gate)
- More memory consumption and more training time in comparison to GRU.
- With Bigger datasets LSTM is better suited.

Modelling Results LSTM: -

- Mahalanobis Distance Outlier Removal (R2 Score): 0.934
- Financial Outlier Removal (R2 Score): 0.918
- Z-Score Outlier Removal (R2 Score): 0.898
- Vanilla Model-No Outlier Removal (R2 Score): 0.728

Gated Re-current Unit (GRU): -

- Have two gates (reset & update gate)
- Less memory consumption, faster execution, less training time.
- More suited for smaller datasets.

Modelling Results GRU: -

- Mahalanobis Distance Outlier Removal (R2 Score): 0.946
- Financial Outlier Removal (R2 Score): 0.915
- Z-Score Outlier Removal (R2 Score): 0.901
- Vanilla Model-No Outlier Removal (R2 Score): 0.706

Best Outlier Elimination Method: -

- In both LSTM & GRU, the order of performance remains same.
- Mahalanobis Distance as outlier removal is the best technique among all we tested.

Best to worst performing model

Mahalanobis Outlier Removal > Financial Outlier Removal > Z-Score Outlier Removal > No Outlier Removal

Overall Verdict: -

- Both LSTM & GRU are very capable for long sequential prediction problems. Both can be handy according to the type of problem.
- In our forecasting we observed both being very close at predicting and there is no clear winner.