MDSC-102 Final Lab

Regd No - 23906
Date - 16/10/2023

National Stock Exchange Dataset

The Data set chosen for the current assignment is the Indian Banking stocks. The dataset has 5 years of banking stock data fetched from the National Stock Exchange. It consists of more than 25 banks with each bank consisting of more than 1300 rows. For the current exercise we will not be using the whole dataset but only a part of it which will be consisting of 1-2 banks.

Dataset link - National Stock Exchnage

Libraries used for the exercise - Seaborn, Plotly, Matplotlib, scipy, Pandas, Numpy

The dataset consist of 15 features with 41231 rows.

```
[76]: df.shape

[76]: (41231, 15)
```

First few rows of the dataset looks like.

```
df.head()
```

	DATE	SYMBOL	SERIES	PREV CLOSE	OPEN	HIGH	LOW	LAST	CLOSE	VWAP	VOLUME	TURNOVER	TRADES	DELIVERABLE VOLUME	%DELIVERBLE
0	2016-01-01	HDFC	EQ	1263.75	1261.00	1266.90	1250.65	1257.80	1258.45	1258.39	676161	8.508751e+13	13230	308262	0.4559
1	2016-01-04	HDFC	EQ	1258.45	1250.00	1253.90	1212.05	1217.15	1216.70	1227.55	1995329	2.449371e+14	78529	1360507	0.6818
2	2016-01-05	HDFC	EQ	1216.70	1229.90	1233.45	1206.50	1208.15	1209.40	1219.50	2325929	2.836464e+14	109820	1644980	0.7072
3	2016-01-06	HDFC	EQ	1209.40	1209.60	1220.75	1202.40	1207.55	1209.30	1210.81	2746330	3.325286e+14	96546	2001431	0.7288
4	2016-01-07	HDFC	EQ	1209.30	1198.85	1203.55	1175.00	1176.35	1179.45	1186.35	1780298	2.112055e+14	60151	1172564	0.6586

Features:

1. Date: date on which data is recorded
2. SYMBOL: Stock symbol according to NSE
3. SERIES: Different series in which a particular stock could be traded
4. PREV CLOSE: Previous day closing price
5. OPEN: Day's opening price
6. HIGH: Day's Highest Price
7. LOW: Day's Lowest Price

8. LAST: Day's Last Price
9. CLOSE: Day's Closing Price
10. VWAP: volume-weighted average price is the ratio of the value traded to total volume traded over a particular time horizon
11. Volume: the amount of a security that was traded during a given period of time. For every buyer, there is a seller, and each transaction contributes to the count of total volume.
12. Turnover: Total Turnover of the stock till that day
13. Trades: Number of buy or Sell of the stock.
14. Deliverable: Volume the quantity of shares which actually move from one set of people (who had those shares and are selling today) to another set of people (who have purchased those shares).
15. %Deliverble: percentage deliverables of that stock.

Also, the dataset does not consist of any null values.

```
df.isnull().sum()
```

```
DATE                 0
SYMBOL               0
SERIES               0
PREV CLOSE           0
OPEN                 0
HIGH                 0
LOW                  0
LAST                 0
CLOSE                0
VWAP                 0
VOLUME               0
TURNOVER             0
TRADES               0
DELIVERABLE VOLUME   0
%DELIVERBLE          0
dtype: int64
```

List of all the banks with their respective number of rows

```
df['SYMBOL'].value_counts() # Taking the count of values of different Banks.
```

```
SYMBOL
HDFC          1337
BANKINDIA     1337
KTKBANK       1337
J&KBANK       1337
SOUTHBANK     1337
DCBBANK       1337
KARURVYSYA    1337
PSB           1337
CENTRALBK     1337
CUB           1337
UCOBANK       1337
INDIANB       1337
MAHABANK      1337
ICICIBANK     1337
UNIONBANK     1337
FEDERALBNK    1337
CANBK         1337
IOB           1337
SBIN          1337
YESBANK       1337
KOTAKBANK     1337
AXISBANK      1337
IDBI          1337
INDUSINDBK    1337
BANKBARODA    1337
PNB           1337
DHANBANK      1297
RBLBANK       1173
AUBANK         962
BANDHANBNK     784
IDFCBANK       752
IDFCFIRSTB     585
CSBBANK        370
UJJIVANSFB     364
EQUITASBNK     141
SURYODAY        41
Name: count, dtype: int64
```

The basic information of the dataset tells us that their are no null values and their are 3 features with object and 12 with numeric (integer and float) datatype.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41231 entries, 0 to 41230
Data columns (total 15 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   DATE                41231 non-null  object 
 1   SYMBOL              41231 non-null  object 
 2   SERIES              41231 non-null  object 
 3   PREV CLOSE          41231 non-null  float64
 4   OPEN                41231 non-null  float64
 5   HIGH                41231 non-null  float64
 6   LOW                 41231 non-null  float64
 7   LAST                41231 non-null  float64
 8   CLOSE               41231 non-null  float64
 9   VWAP                41231 non-null  float64
 10  VOLUME              41231 non-null  int64  
 11  TURNOVER            41231 non-null  float64
 12  TRADES              41231 non-null  int64  
 13  DELIVERABLE VOLUME  41231 non-null  int64  
 14  %DELIVERBLE         41231 non-null  float64
dtypes: float64(9), int64(3), object(3)
memory usage: 4.7+ MB
```

Data subset Selection

Out of the full dataset only 2 banks are chosen for the current analysis which are SBI and HDFC.

```
SBI = df[(df.SYMBOL == 'SBIN') & (df.DATE >= '2021-01-01')]  # Taking only SBI's data (ONLY 2021)
HDFC = df[(df.SYMBOL == 'HDFC') & (df.DATE >= '2021-01-01')] # Taking only HDFC's data (ONLY 2021)
# print(d['HIGH'])
# print(SBI)
HDFC
# ,'INDUSINDBK','IDBI','AXISBANK','KOTAKBANK','YESBANK','SBIN','ICICIBANK','HDFC'
```

:

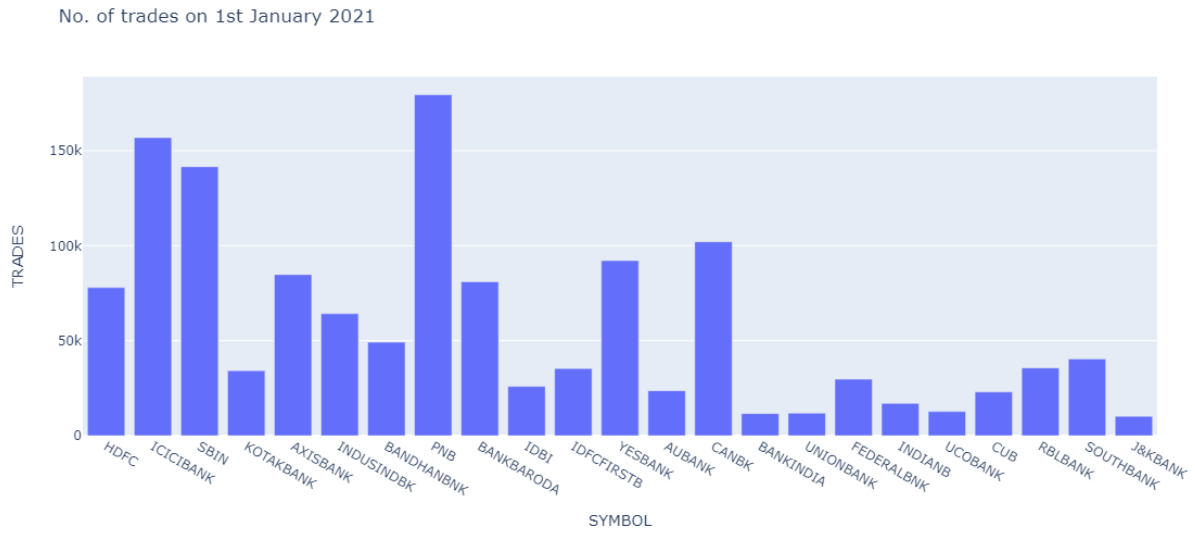
	DATE	SYMBOL	SERIES	PREV CLOSE	OPEN	HIGH	LOW	LAST	CLOSE	VWAP	VOLUME	TURNOVER	TRADES	DELIVERABLE VOLUME	%DELIVERBLE
1238	2021-01-01	HDFC	EQ	2558.65	2549.00	2593.30	2541.15	2565.00	2568.75	2574.95	2054108	5.289220e+14	77982	564766	0.2749
1239	2021-01-04	HDFC	EQ	2568.75	2576.90	2592.35	2533.15	2579.00	2578.50	2568.41	2795972	7.181188e+14	99023	1419319	0.5076
1240	2021-01-05	HDFC	EQ	2578.50	2589.00	2659.45	2572.40	2654.95	2651.85	2631.43	5870743	1.544844e+15	191487	3136561	0.5343
1241	2021-01-06	HDFC	EQ	2651.85	2664.70	2665.00	2619.00	2631.00	2638.85	2642.51	3401085	8.987386e+14	126726	1679728	0.4939
1242	2021-01-07	HDFC	EQ	2638.85	2670.00	2709.35	2647.40	2653.45	2661.35	2680.83	4106042	1.100758e+15	183933	2215445	0.5396
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1332	2021-05-24	HDFC	EQ	2518.05	2515.00	2558.80	2515.00	2531.70	2540.90	2545.15	2655396	6.758374e+14	132449	1356421	0.5108

Checking the daily close price of SBI stock in 2021.

We can easily infer from the below graph that their has been an increase in the closing price of the stock.
The SBI stock also had a exponential increase from 31st January till 10th February 2021 and after that it became constant. But from March till mid April it had a downfall but the price again started increasing from April ending.

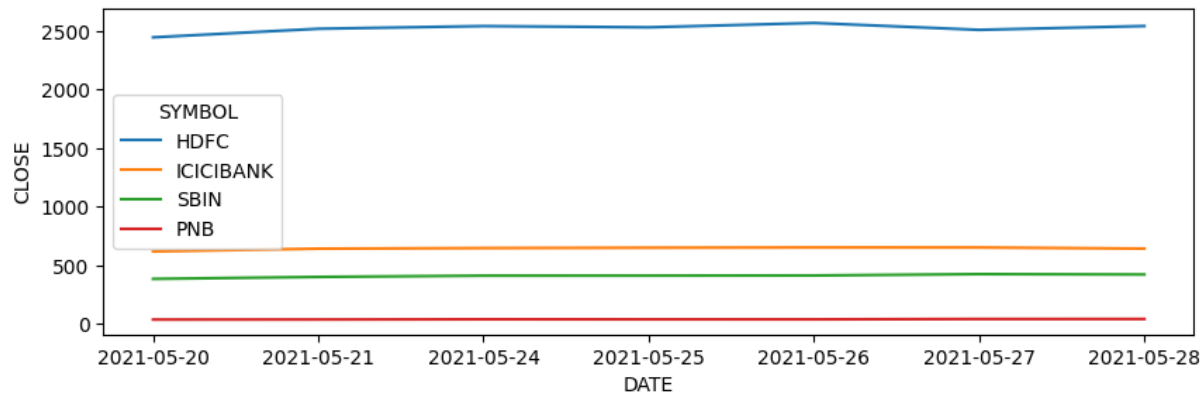
No. of trades on 1st January 2021

The below graph tells us that on the first day of the year all the Banks had atleast 10K trades, with lowest trades in J&K Bank (10.229K) and highest in PNB (179.492K). The average number of trades on 1st Jan 2021 was 58K.



Comparing the Close price of the day May in 2021

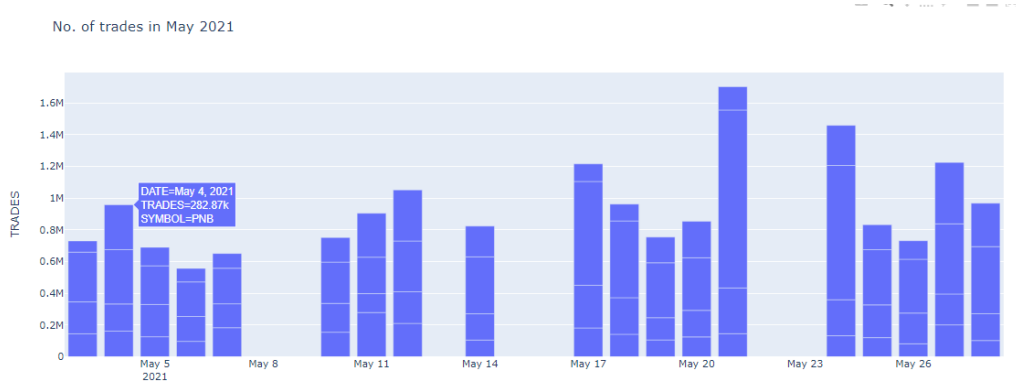
We can clearly see that the closing price for HDFC Bank has been significanlty higher than other banks like ICICI, SBI, PNB.



Checking the No. of trades in may for each bank

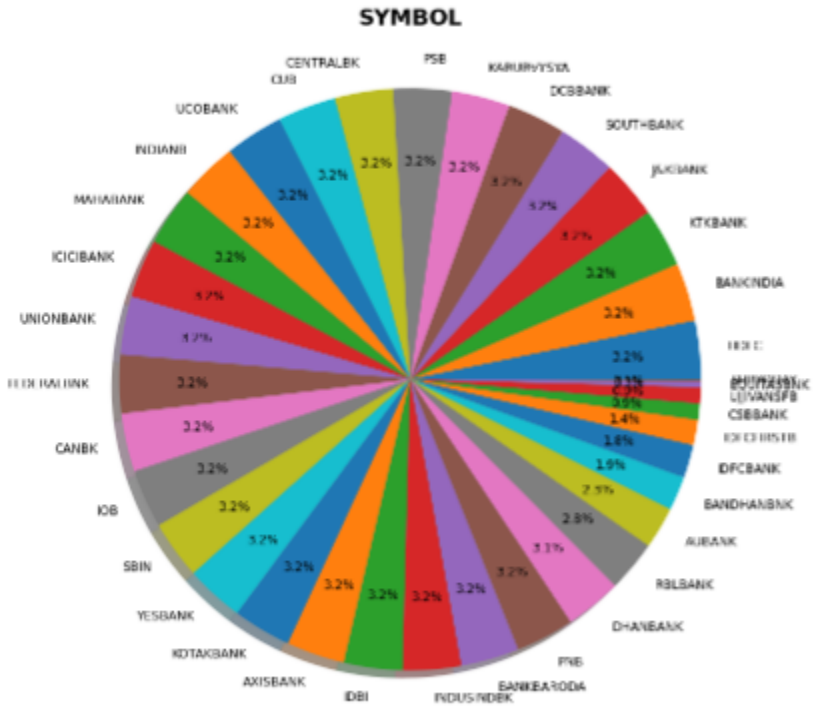
From the below graph we can infer that the no of trades for SBI was the highest among all the 4 banks (SBI, HDFC, ICICC, PNB). Also we can see from the above and the

below graph that even though the closing price of pnb is the least but the sbi is having the highest number of trades.



The % of data available from each bank

The below pie chart shows that % of the data available

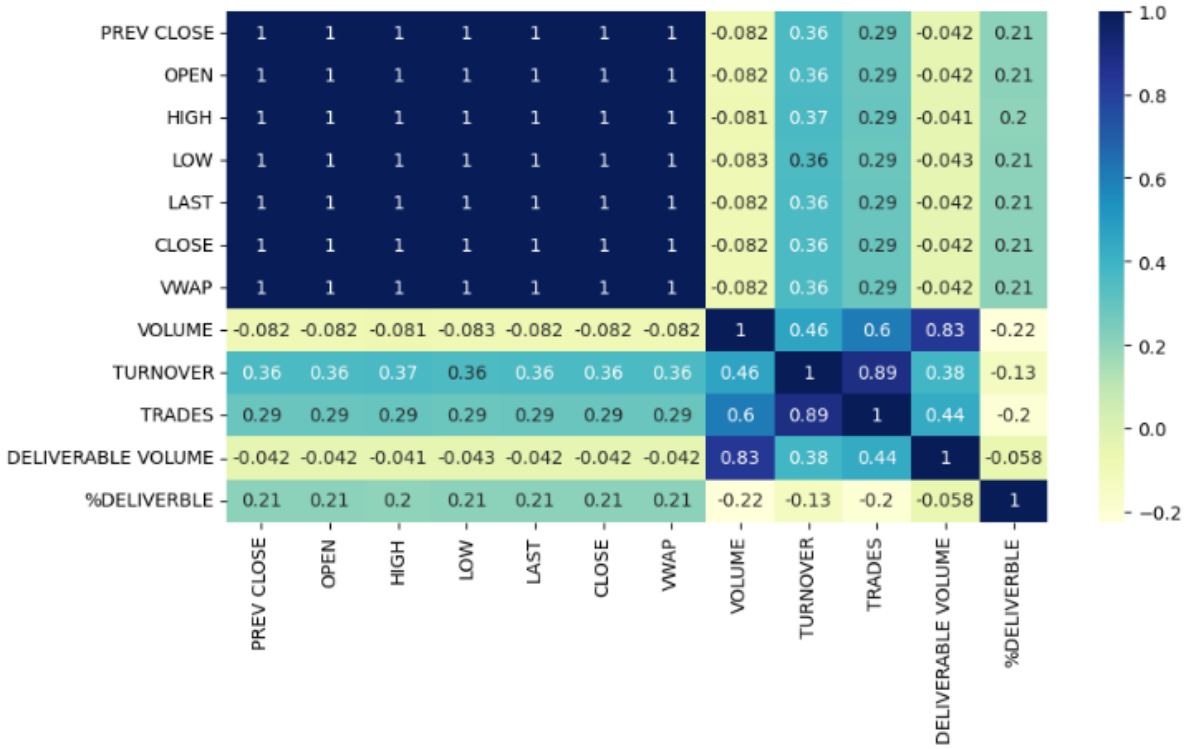


Checking the Skewness

The taken data set was highly negatively skewed which can be seen from the given table.

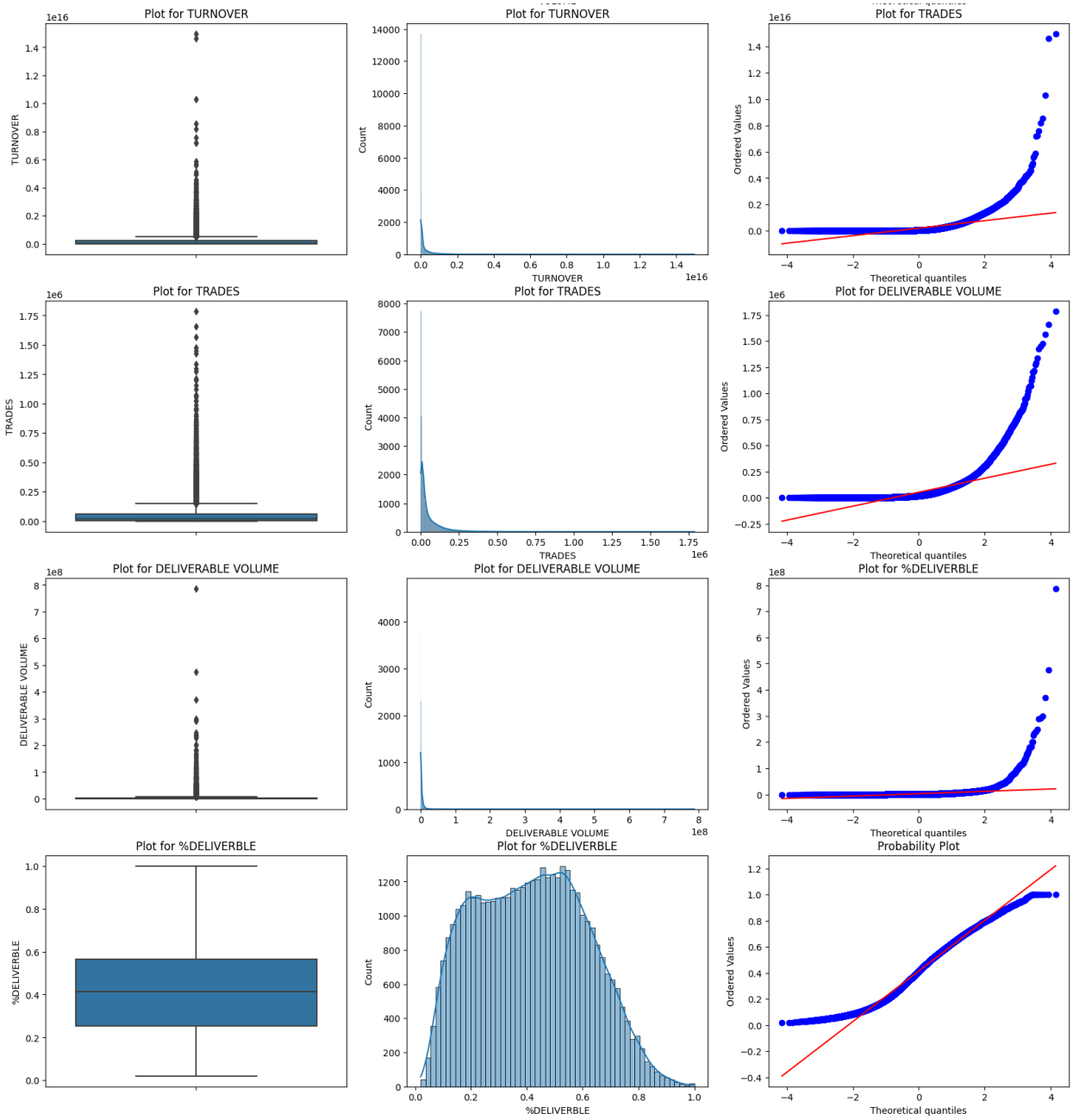
```
PREV CLOSE      -2.486035093902925
OPEN    -2.4862733615145975
HIGH    -2.4797932203375037
LOW     -2.4918352347566195
LAST    -2.486348093165274
CLOSE   -2.486259192454467
VWAP    -2.4849640772251127
VOLUME  -12.965941515012204
TURNOVER        -6.78818485847942
TRADES  -4.598922893752054
DELIVERABLE VOLUME      -27.443206566978784
%DELIVERBLE     -0.14606995077450433
```

Checking the correlation



The above correlation heatmap shows us that the first few features like prev close, open, high are highly correlated. Some of them are quiet clear as in the case of volume and deliverable percentage, Turnover and the number of trades

The same thing was confirmed by the boxplot, histplot and the probability plot.

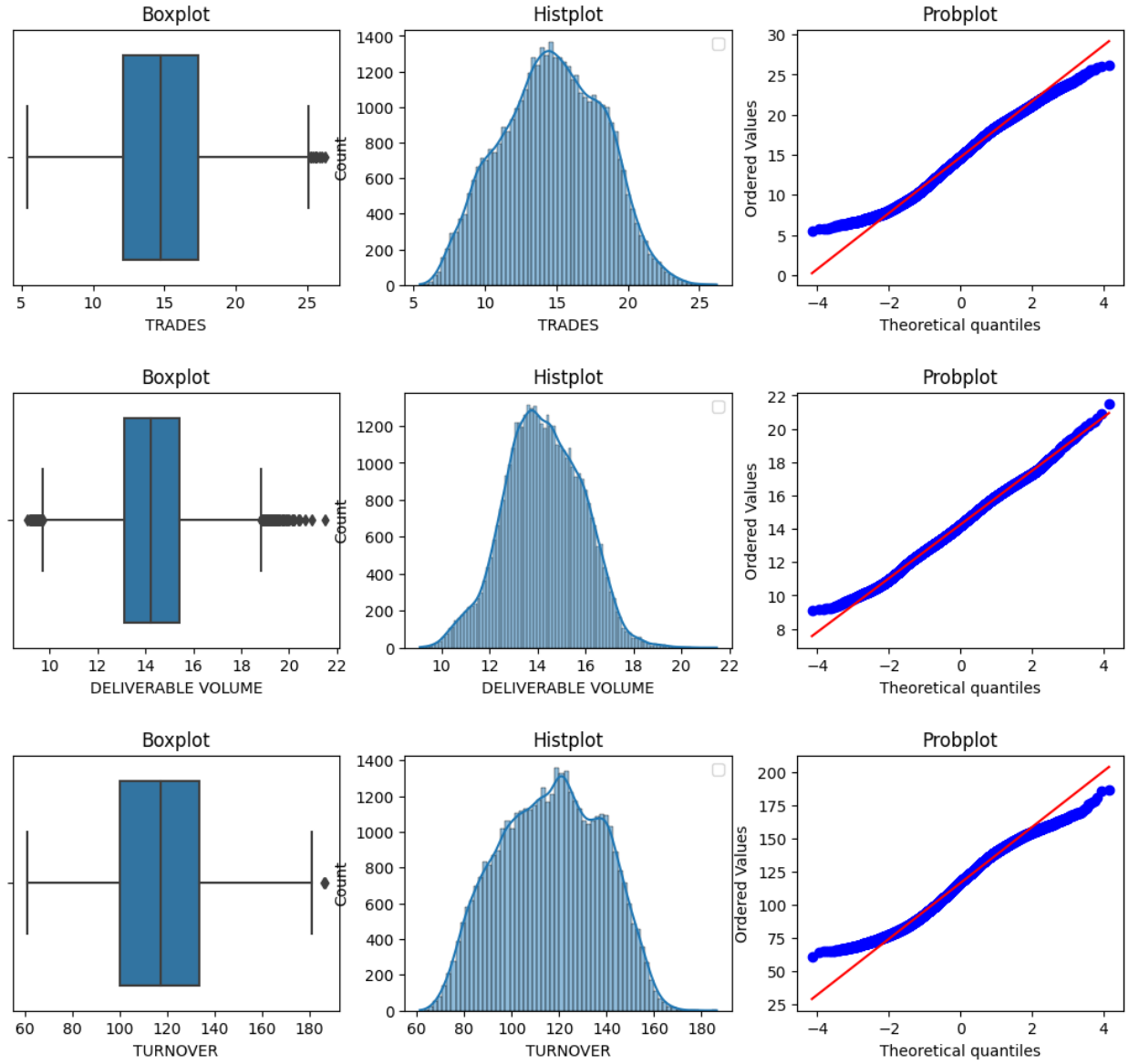


The above plots are for Turnover, Trades, Deliverable and %Deliverable.
The histplot shows that Turnover, trades and Deliverable are highly right skewed and %Deliverable is close to normal.

Normalizing

To normalize the data here we have used Box-Cox. It is a statistical technique that involves transforming the target variable so that the data follows a normal distribution. It helps us to improve the predictive power

Features after Normalization



It can been seen very clearly that after applying the boxcox that the features (Turnover, Deliverable Volume, Trades)are either normal or very close to normal.

Testing Hypothesis

As we have already seen that SBI is having the highest number of Trades in a day in the month of may , with (highest) 1M daily trades in may.
Now we test the hypothesis that the average trade in a day for SBI in 2021 is 3M.

$H_0 : \mu = 300000$ $H_1 : \mu \neq 300000$

```
alpha = 0.5   # Setting the level of significance
m0 = 300000   # Setting the null hypothesis

z_cal = (SBI['TRADES'].mean() - m0)/(np.var(SBI['TRADES'])/(len(SBI['TRADES']))**0.5) # Calculating the z_cal
print(z_cal)
```

```
2.0463249345351912e-05
```

```
p = 2 * (1- sp.stats.norm.cdf(abs(z_cal)))   # calculating the p-value
print("p value : ",p)
```

```
p value :  0.9999836726892846
```

```
if p < alpha:
    print("Reject: The Average number of trades in a day in 2021 is not 300000.")
else:
    print("Do Not Reject: The Average number of trades in a day in 2021 is 300000. ")
```

```
Do Not Reject: The Average number of trades in a day in 2021 is 300000.
```

From the above we can see that we have applied z-test from which we got our z_cal.
The z_cal value that we got is 2.0463259

After that we calculated the p value to check our hypothesis.
Our p value is 0.99998

Then we compare our alpha value with the p value.
As our p value is > alpha so we DO NOT REJECT our hypothesis. Which means that for SBI the Average number of trades in a day in 2021 is around 300000.