Analytathon-3

Energeia Group - Analysis, Forecasting and Strategy Development

1. Abstract

This report aims to represents the taks as follows

- PART 1: Identify the factors resulted in the rise of the I-SEM Day Ahead Price for month July 2021
- PART 2: Develop different Trade stratergies for maximum profit revenue

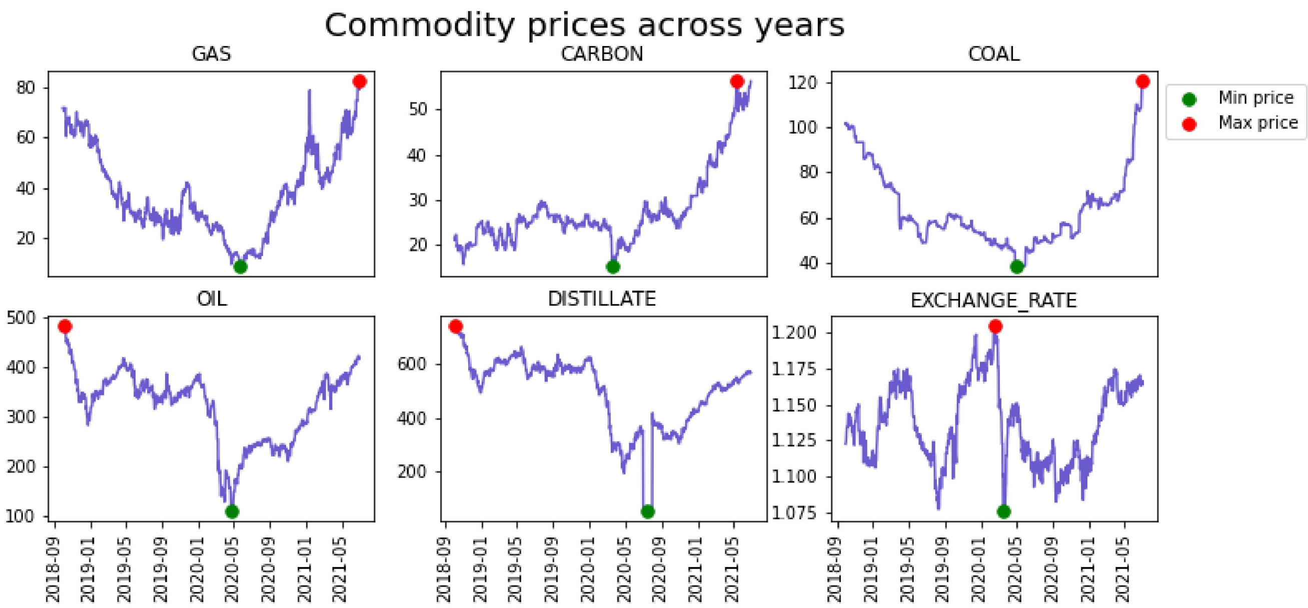
2. PART 1: Data Preprocessing

The Data is provided in an excel sheets and is spread accross different sheets. The data is imported sucessfully and below is quick glimpse on the dimension of all the data to be used for further Analysis

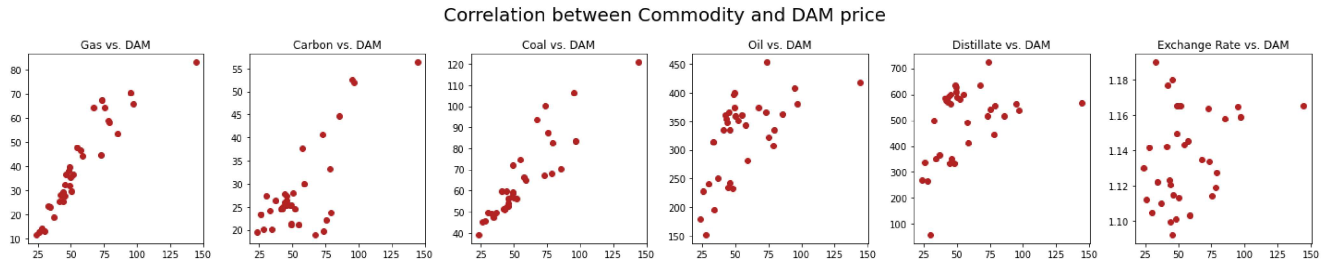
```
Monthy DAM :: (24790, 2)
Hourly DAM :: (24790, 6)
Commodity Prices :: (1004, 7)
Net Demand :: (49392, 10)
Generator Fuel :: (18, 2)
Key generator Availaibility :: (48138, 20)
Key generator Cost :: (18558, 24)
```

The main motive of part 1 is to identify the reason for increase in DAM price for July 2021, To understand this lets get an overview of price trends accross years for varisous commodities. Currently in market there is a high demand for Coal and Gas, used fuel to generate electricity. To Our surprise we see there us a spiked seen in year 2021 with the highest price recorded for both on June 30 2021.

```
Out[60]: <matplotlib.legend.Legend at 0x1b600267610>
```

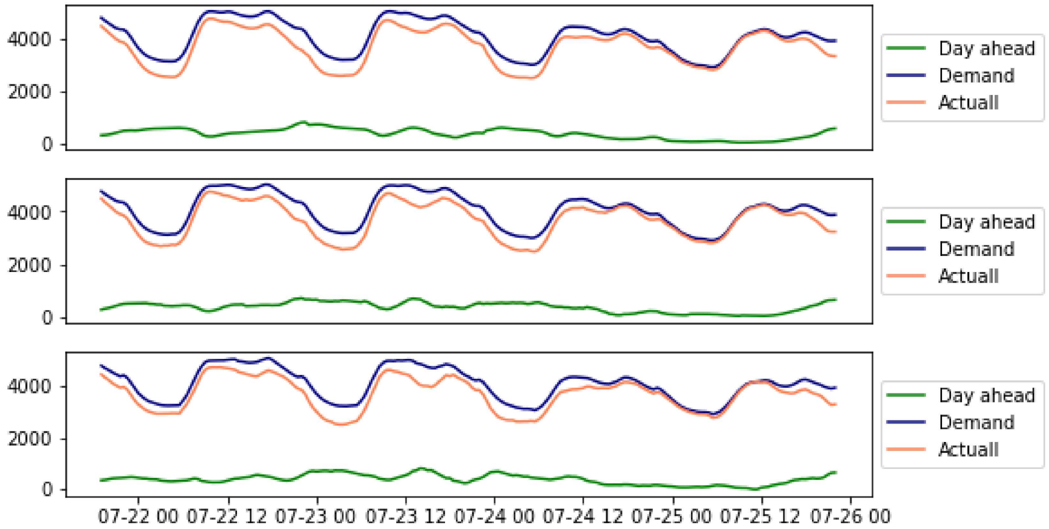


To get a better understand on the relationship between the commodity prices and DAM prices a correaltion plot is created. There is a good positive correlation between coal vs DAM and Gas vs DAM, which means increase in Coal or Gas price will defintiely increase the DAM price. Other commodities like oil and distillate has a moderate correlation but there is no recent increase seen in them and hence may not be the factors driving recent price change.

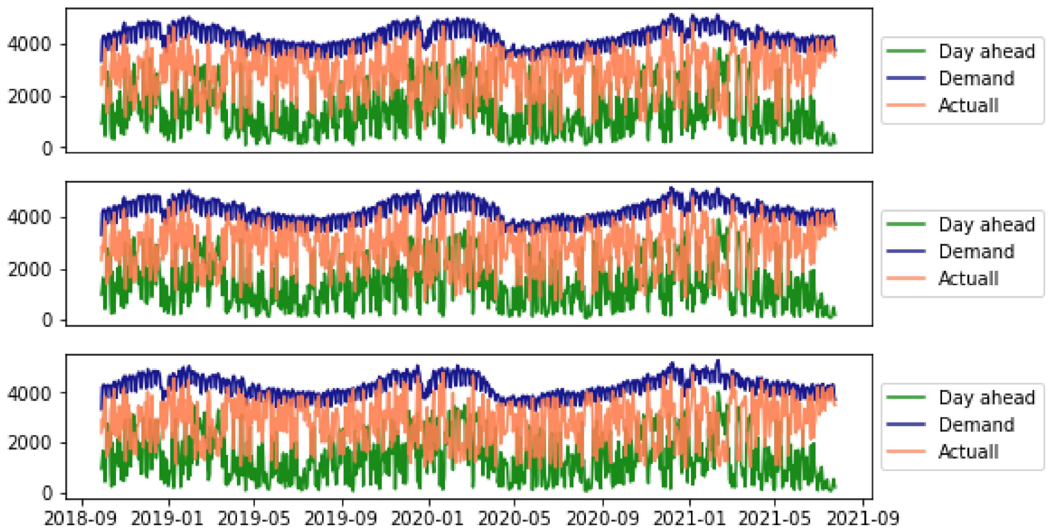


Next focus was to gain some insight on the wind data and see if there are any seasonal patterns occuring. The below plots is created to study hourly and seasonal patterns and below are some of the observations

Hourly: There are weak winds during midnight and strong winds observed during day time after 12PM



Seasonal: We do there are patterns occuring in second plot where the wind is stronger mostly during the winter seasons and wind strength drops during summer, as July and June are the mid of summer months and hence Electricity Generators cannot depends much on wind for electricity generation and look for other commodities.



The below table shows that the majority of Generators use Gas as there fuel to produce electricity

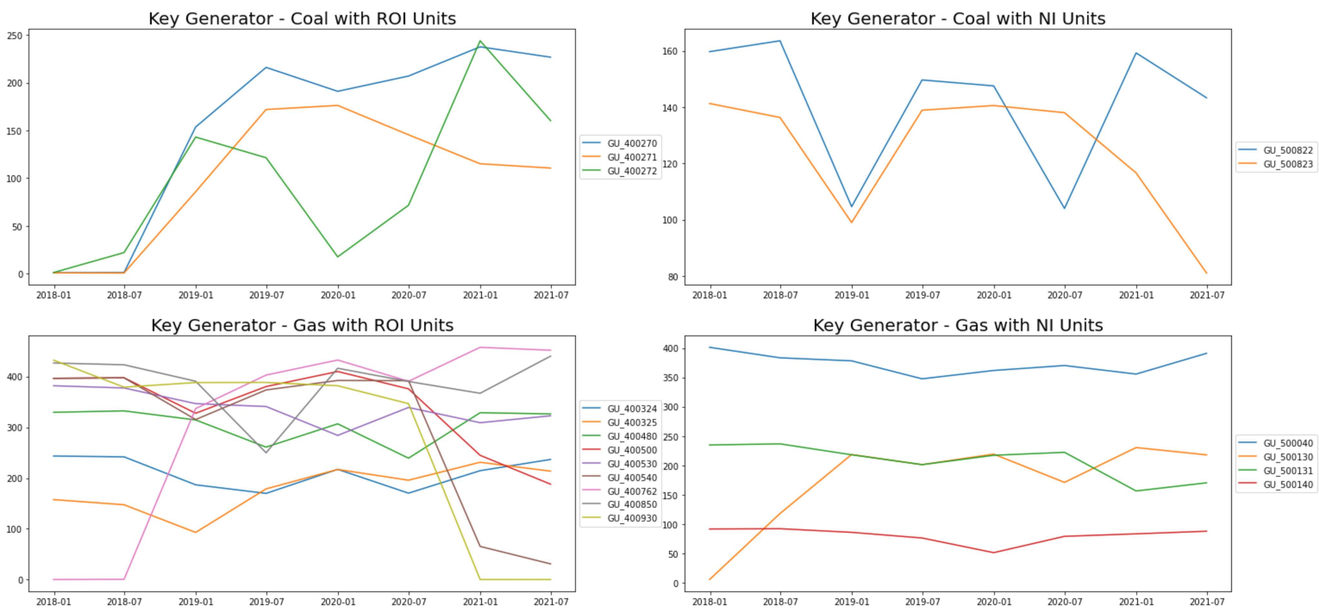
Out[86]:

fuel	Generator
Coal	5
Gas	13

Generators producing electricity based on the fuel used are as below

```
Coal Generators:: ['GU_400270', 'GU_400271', 'GU_400272', 'GU_500822', 'GU_500823']
Gas Generators:: ['GU_400324', 'GU_400325', 'GU_400480', 'GU_400500', 'GU_400530', 'GU_400540', 'GU_4
00762', 'GU_400850', 'GU_400930', 'GU_500040', 'GU_500130', 'GU_500131', 'GU_500140']
```

Here we are with the most important factor "Key Generators". The plots clearly shows that some of the Key generators are down during the month of July either due to maintainance or producing less units of electricity. This made the other active generating units to increase the price by a margin of 5% - 10% from June to July.



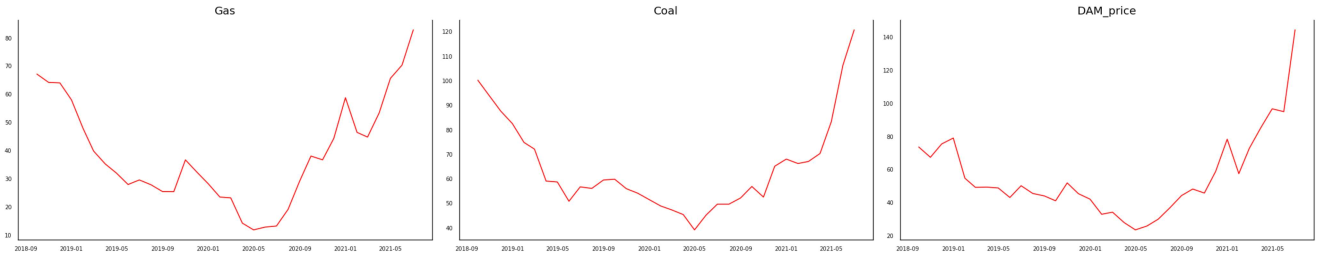
Out[127...

	June Price	July Price
0	69.04	72.97
1	66.01	71.32
2	65.09	68.93
3	69.04	72.97
4	78.68	85.00
5	78.68	85.00
6	84.48	88.87
7	50.07	54.30
8	179.87	207.10

3. PART 1: Machine Learning model (Vector autoregression)

To understand the increase in price rise to need to find the future values of Coal and Gas and use these new prices to forecase the DAM prices. A Vector autoregression is used along with a lag of 3.

Below plot is quicK look on how exactly the Coal, Gas and DAM relate. Each of the series have a fairly similar trend patterns over the years



Testing Causation using Granger's Causality Test Looking at the P-Values in the below table, we can pretty much observe that all the variables (time series) in the system are interchangeably causing each other.

Out[131...

	Gas_x	Coal_x	DAM_price_x
Gas_y	1.0000	0.0110	0.0126
Coal_y	0.1057	1.0000	0.0002
DAM_price_y	0.0311	0.0001	1.0000

Cointegration test helps to establish the presence of a statistically significant connection between two or more time series and below are the details

```
Name   ::  Test Stat > C(95%)    =>   Signif
 ----------------------------------------
Gas    ::  44.34     > 24.2761   =>   True
Coal   ::  12.88     > 12.3212   =>   True
DAM_price ::  1.85      > 4.1296    =>   False
```

Check for Stationarity and Make the Time Series Stationary. ADF Test. Augmented Dickey-Fuller Test (ADF Test)

```
    Augmented Dickey-Fuller Test on "Gas" 
    -----------------------------------------------
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -2.5331
 No. Lags Chosen       = 4
 Critical value 1%     = -3.724
 Critical value 5%     = -2.986
 Critical value 10%    = -2.633
 => P-Value = 0.1076. Weak evidence to reject the Null Hypothesis.
 => Series is Non-Stationary.


    Augmented Dickey-Fuller Test on "Coal" 
    -----------------------------------------------
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -3.2661
 No. Lags Chosen       = 0
 Critical value 1%     = -3.679
 Critical value 5%     = -2.968
 Critical value 10%    = -2.623
 => P-Value = 0.0165. Rejecting Null Hypothesis.
 => Series is Stationary.


    Augmented Dickey-Fuller Test on "DAM_price" 
    -----------------------------------------------
```

```
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -1.6874
 No. Lags Chosen       = 4
 Critical value 1%     = -3.724
 Critical value 5%     = -2.986
 Critical value 10%    = -2.633
 => P-Value = 0.4376. Weak evidence to reject the Null Hypothesis.
 => Series is Non-Stationary.
```

ADF test confirms none of the time series is stationary. Let's difference all of them once and check again.

```
    Augmented Dickey-Fuller Test on "Gas"
    -----------------------------------------------
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -3.5248
 No. Lags Chosen       = 8
 Critical value 1%     = -3.833
 Critical value 5%     = -3.031
 Critical value 10%    = -2.656
 => P-Value = 0.0074. Rejecting Null Hypothesis.
 => Series is Stationary.


    Augmented Dickey-Fuller Test on "Coal"
    -----------------------------------------------
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -2.9507
 No. Lags Chosen       = 8
 Critical value 1%     = -3.833
 Critical value 5%     = -3.031
 Critical value 10%    = -2.656
 => P-Value = 0.0398. Rejecting Null Hypothesis.
 => Series is Stationary.


    Augmented Dickey-Fuller Test on "DAM_price"
    -----------------------------------------------
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -6.5834
 No. Lags Chosen       = 2
 Critical value 1%     = -3.724
 Critical value 5%     = -2.986
 Critical value 10%    = -2.633
 => P-Value = 0.0. Rejecting Null Hypothesis.
 => Series is Stationary.
```

```
C:\Users\MYPC\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:524: ValueWarning: No fre
quency information was provided, so inferred frequency MS will be used.
  warnings.warn('No frequency information was'
```

Training the model with LAG 4 of selected order

Out[146…

```
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Sun, 08, Aug, 2021
Time:                     01:12:43
--------------------------------------------------------------------
No. of Equations:         3.00000    BIC:                    13.0237
Nobs:                     24.0000    HQIC:                   11.6172
Log likelihood:          -196.476    FPE:                    98559.0
AIC:                      11.1093    Det(Omega_mle):         26898.3
```

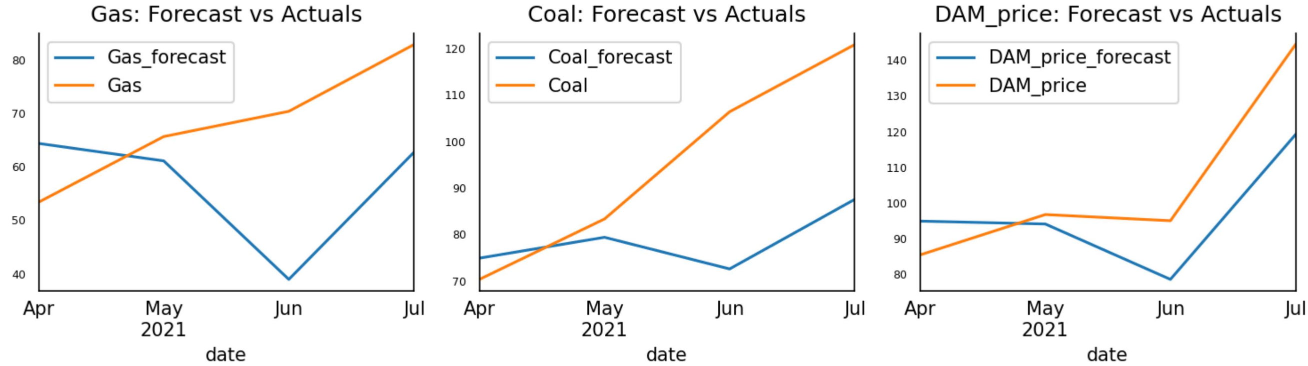
```
-----------------------------------------------------------------
Results for equation Gas
==============================================================================
                 coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const               0.122477         1.511723            0.081           0.935
L1.Gas             -0.683998         0.652393           -1.048           0.294
L1.Coal             0.177842         0.348584            0.510           0.610
L1.DAM_price        0.068586         0.541670            0.127           0.899
L2.Gas             -1.013498         0.571050           -1.775           0.076
L2.Coal             0.073626         0.458528            0.161           0.872
L2.DAM_price        0.264562         0.493914            0.536           0.592
L3.Gas              0.185863         0.524247            0.355           0.723
L3.Coal             0.711153         0.446166            1.594           0.111
L3.DAM_price       -0.113837         0.438142           -0.260           0.795
L4.Gas             -0.028781         0.473535           -0.061           0.952
L4.Coal             0.707350         0.386499            1.830           0.067
L4.DAM_price        0.258357         0.354492            0.729           0.466
==============================================================================

Results for equation Coal
==============================================================================
                 coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const               0.774424         1.556245            0.498           0.619
L1.Gas              0.040403         0.671607            0.060           0.952
L1.Coal            -0.975773         0.358851           -2.719           0.007
L1.DAM_price        0.080246         0.557623            0.144           0.886
L2.Gas             -0.395257         0.587869           -0.672           0.501
L2.Coal            -0.592691         0.472032           -1.256           0.209
L2.DAM_price        0.328142         0.508461            0.645           0.519
L3.Gas              0.056439         0.539687            0.105           0.917
L3.Coal            -0.195923         0.459306           -0.427           0.670
L3.DAM_price        0.146160         0.451046            0.324           0.746
L4.Gas             -0.052290         0.487482           -0.107           0.915
L4.Coal            -0.058910         0.397882           -0.148           0.882
L4.DAM_price        0.229531         0.364932            0.629           0.529
==============================================================================

Results for equation DAM_price
==============================================================================
                 coefficient       std. error           t-stat            prob
------------------------------------------------------------------------------
const               1.467736         2.171145            0.676           0.499
L1.Gas              0.053944         0.936970            0.058           0.954
L1.Coal             0.178357         0.500639            0.356           0.722
L1.DAM_price       -1.064627         0.777950           -1.369           0.171
L2.Gas             -0.767787         0.820146           -0.936           0.349
L2.Coal            -0.166212         0.658540           -0.252           0.801
L2.DAM_price       -0.440636         0.709363           -0.621           0.534
L3.Gas              0.259075         0.752927            0.344           0.731
L3.Coal             0.609861         0.640786            0.952           0.341
L3.DAM_price       -0.549093         0.629263           -0.873           0.383
L4.Gas             -0.473655         0.680094           -0.696           0.486
L4.Coal             0.634296         0.555092            1.143           0.253
L4.DAM_price        0.308696         0.509123            0.606           0.544
==============================================================================

Correlation matrix of residuals
                 Gas      Coal  DAM_price
Gas         1.000000  0.355854   0.911493
Coal        0.355854  1.000000   0.479120
DAM_price   0.911493  0.479120   1.000000
```

The data is forecasted with Vector Autoregression model and below is the forecast for the test data

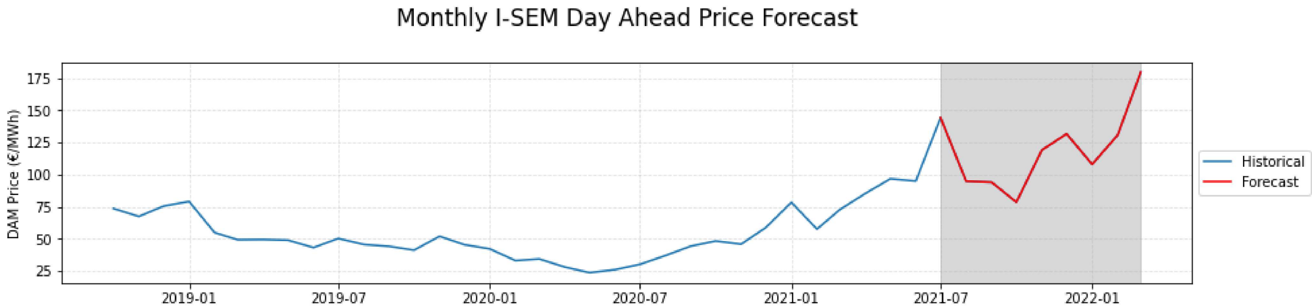


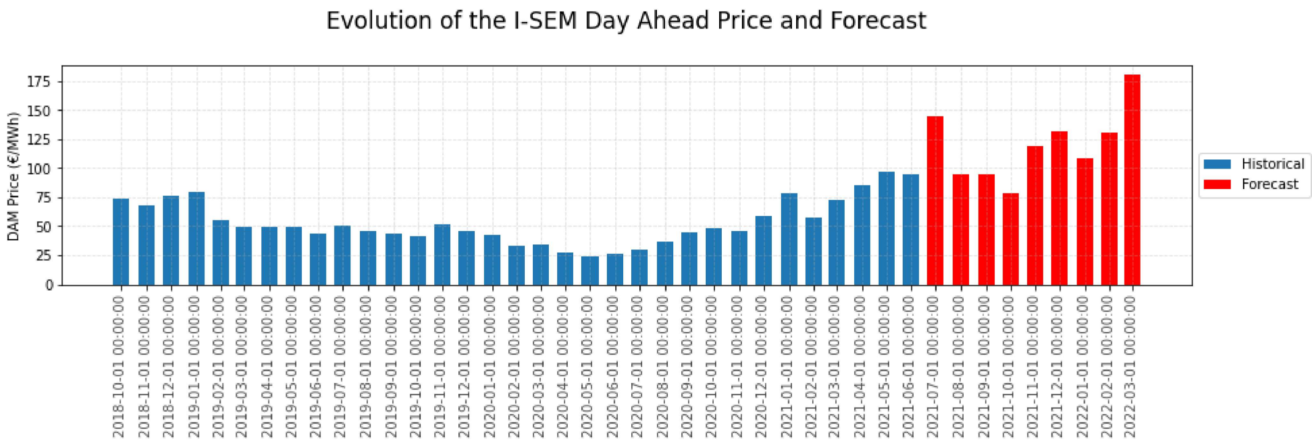
Out[188...

trade_month	Gas_2d	Coal_2d	DAM_price_2d	Gas_1d	Gas_forecast	Coal_1d	Coal_forecast	DAM_price_1d
2021-08-01	21.235461	6.988168	6.723502	19.570931	64.377382	7.818467	74.902338	22.027751
2021-09-01	-22.833692	-3.342785	-22.835533	-3.262761	61.114621	4.475683	79.378021	-0.807783
2021-10-01	-18.905778	-11.270343	-14.711078	-22.168540	38.946081	-6.794660	72.583361	-15.518861
2021-11-01	45.936827	21.729882	56.091741	23.768287	62.714368	14.935222	87.518584	40.572880
2021-12-01	-7.100162	-8.696956	-27.989896	16.668125	79.382493	6.238266	93.756850	12.582984

4. PART 2: CONLCUSION

By taking all the factors into consideration, the future prices were predictd and there new prices were uwed to forecast the DAM data. These Observation bring us to a conclusion that Coal, gas and key generator price resulted in a increase of DAM price and as per the forecast the prices will reduce in the next comming months and again spiked back in March 2022, which is something to be worried about.





5. PART 2: DATA PREPROCESSING

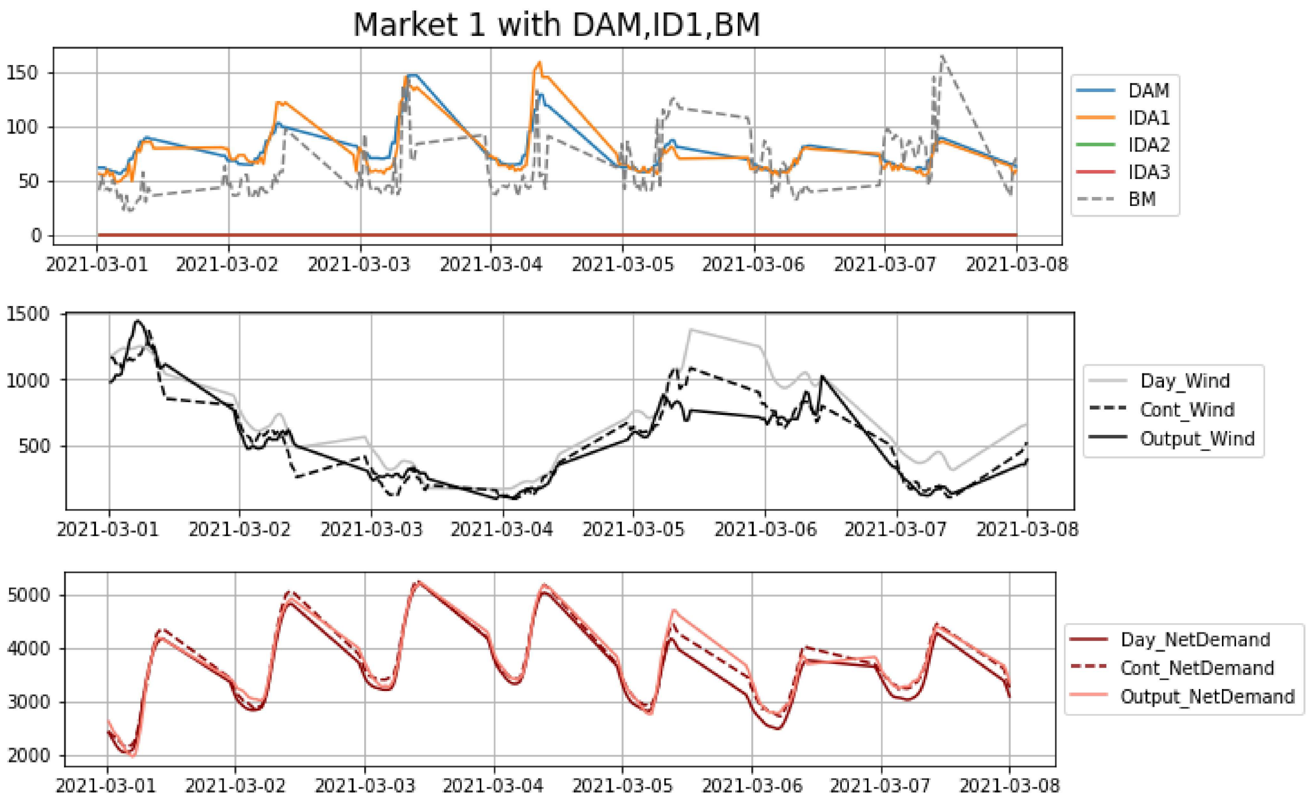
The Data is provided in an excel sheets and is spread accross different sheets. The data provided has information about the variuous prices seen in ISEM markets based on theere auction trade prices (DAM,IDA1,IDA2,IDA3,BM). The data is divided into 3 years 2019,2020,2021.

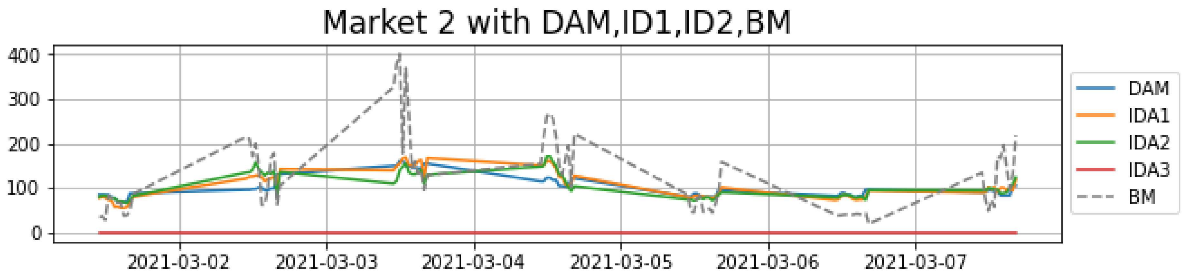
The data types are changes for all features and any NA's are converted to 0

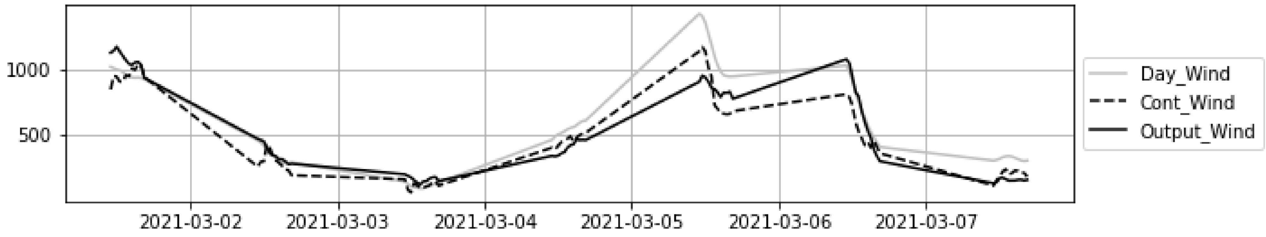
Verify if there is any NA data in the complete data frame

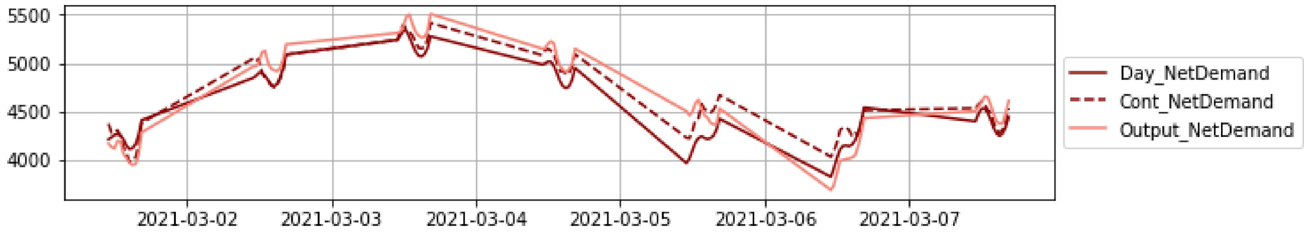
```
No NA found in year 2019
No NA found in year 2020
No NA found in year 2021
```

For the data given for 2021 patterns were captured between the different trade markt prices and pther factors like wind and net demand. The trends were captured for Yearly, houry, Weekend, Monthly and seasonal. Below is a snippet taken from a week report and split acrros 3 markets to understand the prices variations.

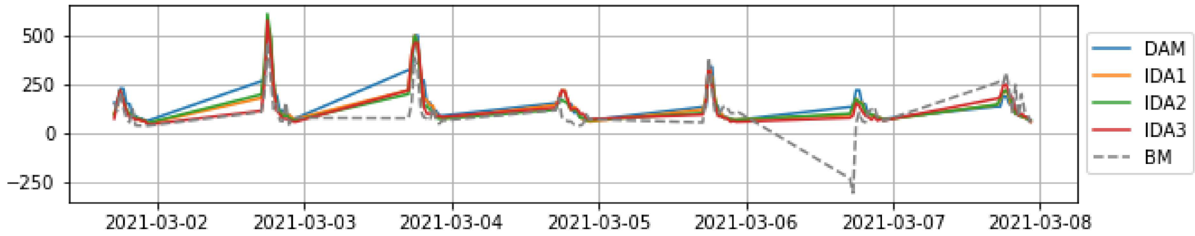


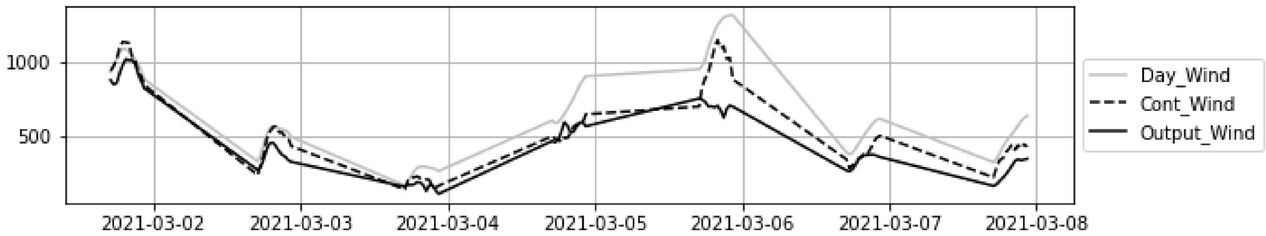


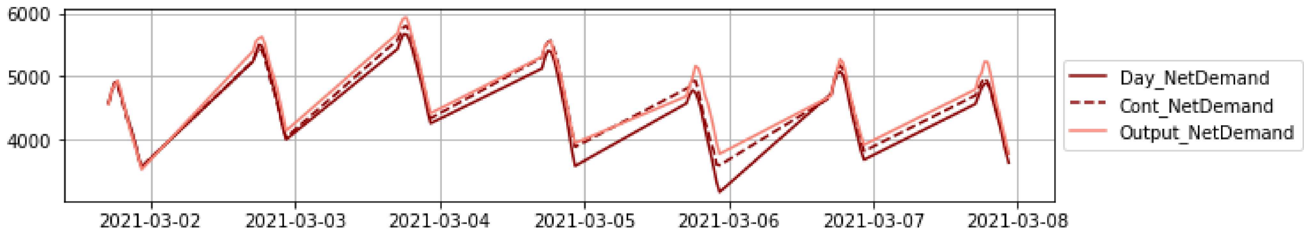




Market 3 with DAM,ID1,ID2,ID3,BM







From the trends and patterns examined, there are 3 trading stratergies finalised each for every trade market. There are 2 short trades and 1 long trade. The below are stratergies by keeping the wind and net demand patterns keeping same accross all the auction trades.

Trade in Market 1: Buy in DAM and Sell in BM when Output Wind < 1200 and Day net demand is > 1200

Trade in Market 2 Buy in IDA1 and Sell in BM when Output Wind < 1200 and Day net demand is > 1200

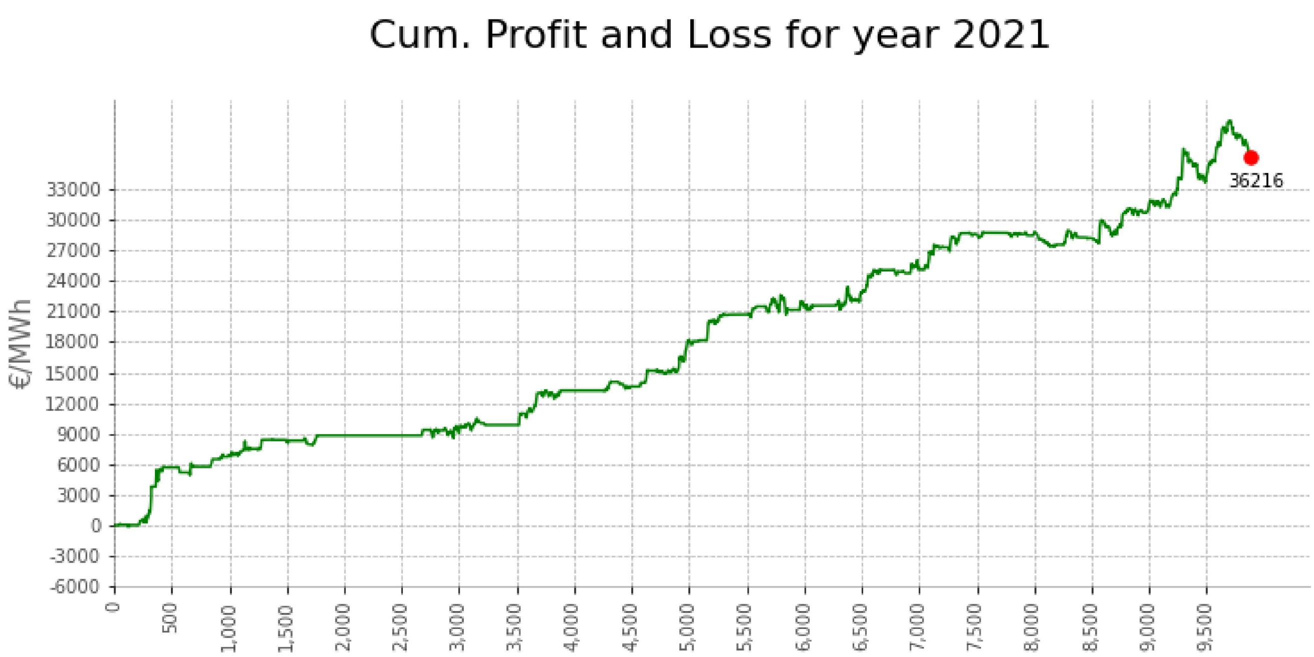
Trade in Market 3 Sell in IDA3 and Buy in DAM when Output Wind < 1200 and Day net demand is > 1200

The win and loss trades numbers and along with there profit and lose values are given below.

```
--------------------------------
Details of year 2019
--------------------------------
Market 1 with win trades 1329 and max profit is 368.85
Market 1 with Loss trades 2614 and min profit is -311.05
Market 2 with win trades 1989 and max profit is 368.85
Market 2 with Loss trades 3827 and min profit is -311.05
```

```
Market 3 with win trades 2864 and max profit is 368.85
Market 3 with Loss trades 4777 and min profit is -311.05
-------------------------------
Details of year 2020
-------------------------------
Market 1 with win trades 1442 and max profit is 349.31
Market 1 with Loss trades 2011 and min profit is -224.61
Market 2 with win trades 2136 and max profit is 421.84
Market 2 with Loss trades 2821 and min profit is -224.61
Market 3 with win trades 2915 and max profit is 421.84
Market 3 with Loss trades 3651 and min profit is -224.61
-------------------------------
Details of year 2021
-------------------------------
Market 1 with win trades 1083 and max profit is 348.74
Market 1 with Loss trades 1500 and min profit is -336.73
Market 2 with win trades 1732 and max profit is 1091.05
Market 2 with Loss trades 2061 and min profit is -336.73
Market 3 with win trades 2499 and max profit is 1091.05
Market 3 with Loss trades 2460 and min profit is -336.73
```

The graph below shows a Drmatic increase in the DAM values for every 1MWatt

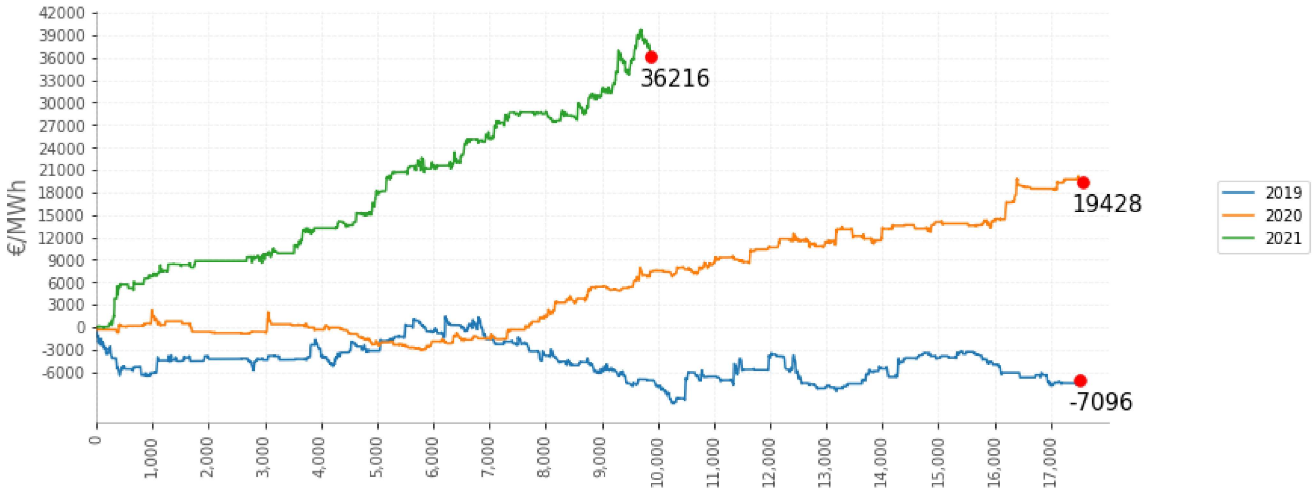


6. PART 2: CONCLUSION

Looking at the sucessfull trade startergies, the same were applied on historic data 2019 and 2020.The graph below clearly shows the stratergies failed for 2019 resulting in a loss with negative values, But gave a good trade profit returns for 2020. Where as for year 2021 within mid of year we have gained highest profit compared to other years and by continuing the same trade stratergies we can double the profit in comming months.

Out[182… `<matplotlib.legend.Legend at 0x1b600008af0>`

Cum. Profit and Loss



Below figure provides information on all the trade wins and lose.

```
***********************************
Profit Loss data for year 2019
***********************************
Total trades: 7641
Total Available trades: 17520
Percentage of trades: 43.61%
------------------------------------
Winning trades: 2864
Loss trades: 4777
------------------------------------
Winning Ratio: 37.48%
Loss Ratio: 62.52%
------------------------------------
Average Win trades: 33.69
Average Loss trades: -21.69
------------------------------------
Max Win trades: 368.853
Max Loss trades: -311.05

***********************************
Profit Loss data for year 2020
***********************************
Total trades: 6566
Total Available trades: 17568
Percentage of trades: 37.37%
------------------------------------
Winning trades: 2915
Loss trades: 3651
------------------------------------
Winning Ratio: 44.4%
Loss Ratio: 55.6%
------------------------------------
Average Win trades: 27.97
Average Loss trades: -17.01
------------------------------------
Max Win trades: 421.8399999999999
Max Loss trades: -224.61

***********************************
Profit Loss data for year 2021
***********************************
Total trades: 4959
Total Available trades: 9888
Percentage of trades: 50.15%
------------------------------------
Winning trades: 2499
Loss trades: 2460
```

```
-----------------------------------
Winning Ratio: 50.39%
Loss Ratio: 49.61%
-----------------------------------
Average Win trades: 43.05
Average Loss trades: -29.01
-----------------------------------
Max Win trades: 1091.05
Max Loss trades: -336.73
```

7. FUTURE SOLUTIONS

A Machine Learning model can be implemented to predict the right trade stratergies for the data given and all other factors and can be further implemented as reinforcement ML model to make a sequence of decisions based on the current market affairs and trading.

8. REFERENCES

- https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/
- https://info.ams-insights.com/machine-learning-market-research#additional_resources
- https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/
- https://core.ac.uk/download/pdf/326318195.pdf