Statistical analysis of electric energy production of Romania and trend forecasting

**Introduction**

**Parameters for statistical Analysis**

In order to gain a better understanding in the insight of the trends in our energy production data we use the following statistical measures.

*Mean*

The central tendency of our data is measured using the mean (or the expected value)

*Standard deviation (Std)*

The amount of dispersion of our data is measured using the standard deviation. A low value of the standard deviation shows that the data values tend to be close to the expected value of the set, a higher value indicates that the data values are spread out on a bigger interval

*Skew*

For short- and medium-term predictions the expected value is less likely to be achieved, so the take this into account we also want to analyse the extremes of the data set by calculating the skew.

The skewness can pe positive (right skewness) negative (left skewness) or zero (bell curve)

In general, the more skewed the data, either positive or negative, the less accurate the future predictions will be.

*Kurtosis*

We also consider how ‘fat’ the data distribution is by calculating the kurtosis. A positive kurtosis means that our data distribution has a higher peak than a normal curve with the same mean and std and thus a tinner tail. A negative kurtosis translates in a flatter than normal distribution and thus a ‘fatter’ tail.

*Range*

The minimum and maximum values of our data is also noted as to begin from a interval with our future predictions of energy production.

*Variation (Coefficient of variation)*

It is well known that big values of standard deviation in relation to the average value may lead to limitations in the quality of forecasts. So, we use the coefficient of variation to rank the reliability of our future predictions

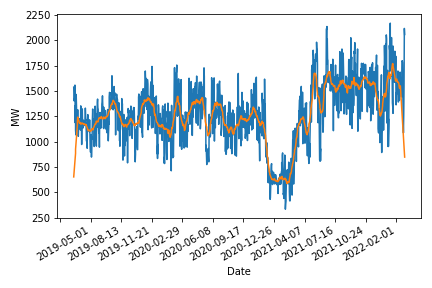
**Plots**

**Analiza e buna trebuie schiimbate luniile. in cod e trebuie np.flip si la data.valori**

The following section consists of plots of electrical energy production by method of production on the interval 7th March 2019 – 9th March 2022, together with a visual analysis of the plots.

Coal - constant value on the interval March 2019 - December 2020. 50% Decrease on the interval January 2021 - May 2021. 400% increase in the subsequent period June 2021 – March 2022

Hydro – We can notice a periodicity in the lows and highs of the data. Highs corresponding to months in which snow usually melts in Romania (February, March and April). Lows corresponding to summer months and autumn months. We notice an anomaly in 2021 where the low is period is longer than that of the previous year, this is due to less rain this year

Chart, histogram

Description automatically generated

Fig 1: Coal Fig 2: Hydro

Oild & Gas – We notice a periodicity in the lows and highs that is opposite to the one in Hydro. The highs are in the summer months where hydro production in not that great. The lows are in the winter and spring period

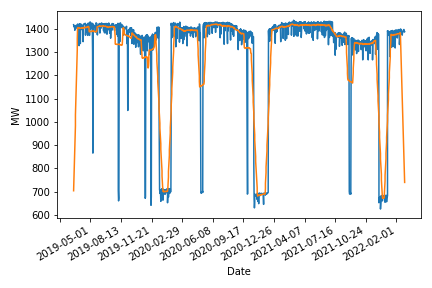
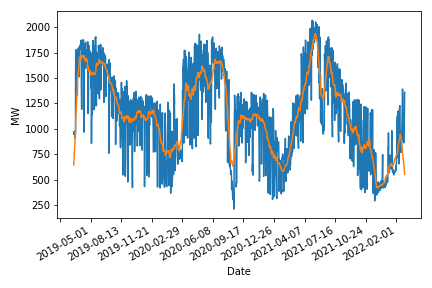


Fig 3: Oil&Gas Fig 4: Nuclear

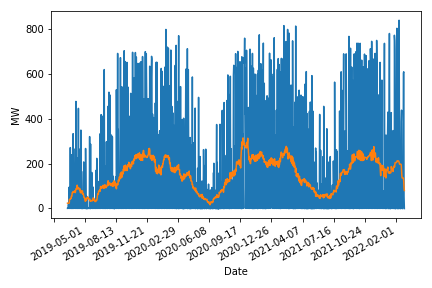
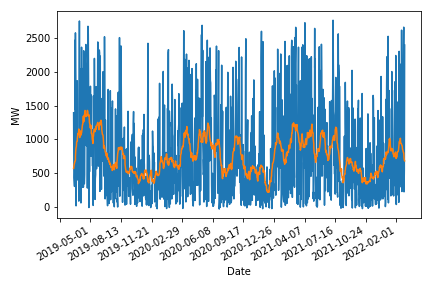


Fig 5: Wind Fig 6: Solar

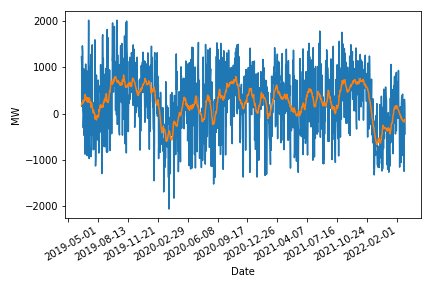
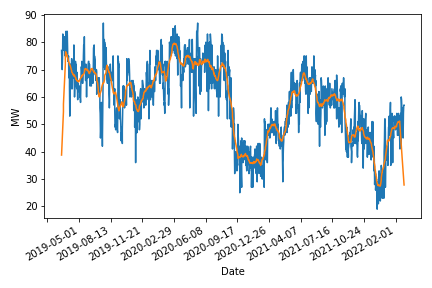


Fig 6: Biomass Fig 8: Sold

**Histograms**

The following section consists of histograms of electrical energy production by method of production on the interval 7th March 2019 – 9th March 2022, together with an visual analysis

Coal – Highest peak is at 1250 MW

Hydro – Highest peak is at 1750 MW

Oil& Gas Highest peak is at 1100 MW

Nuclear – highest peak is at 1400 MW but the histogram exhibits 2 peaks at 1400MW and 700MW majority of values are centred at 700 and 1400 MW

Wind – highest peak is at 0 Mw, but the majority of values are at 0-250MW

Solar – highest peak is at 0MW

Biomass – highest peak is at 65 MW

Sold – highest peak is at 250MW and extremely close the second one at 750MW

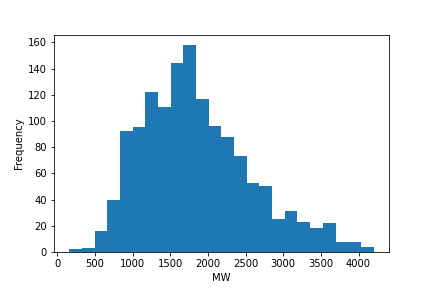
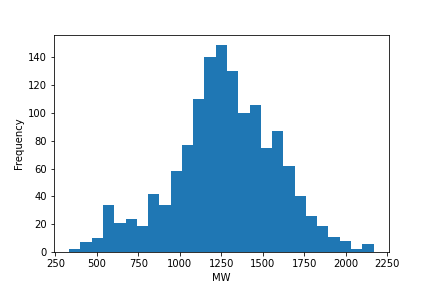
****

Fig 9: Coal Fig 10: Hydro

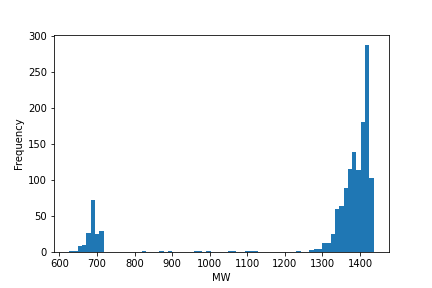
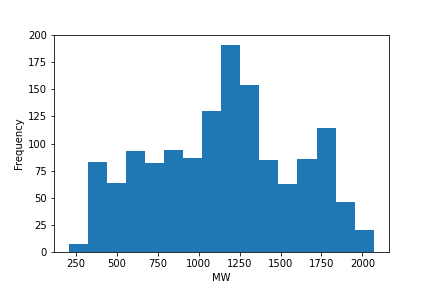


Fig 11: Oil&Gas Fig 12: Nuclear

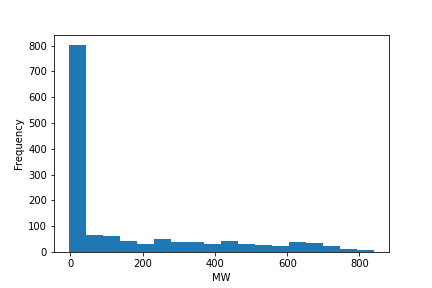
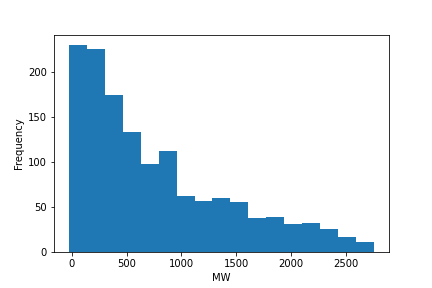


Fig 13: Wind Fig 14: Solar

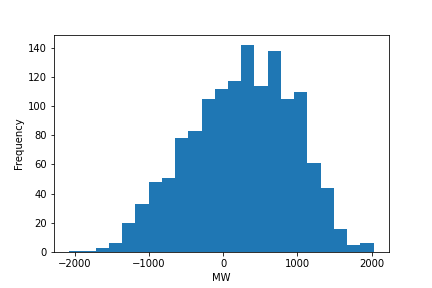
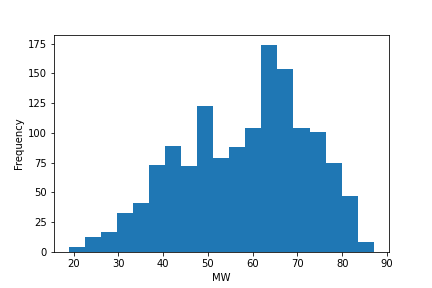


Fig 13: Biomass Fig 14: Sold

**Results of Statistical Analysis**

The selected results of the daily distribution of energy production grouped by method of production are presented in Table 1.

The highest 3 means are hydro, nuclear and coal, lowest mean is biomass, solar and sold, but if we exclude sold because it allows negative values, we get that the lowest 3 means are those of biomass.

This means that most of the energy in Romania is generated by means of hydro, nuclear and coal.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Std | Skew | Kurtosis | Variation | Range |
| Coal | 1270 | 315.2 | (-)0.231 | 0.103 | 0.248 | 333 - 2168 |
| Hydro | 1853 | 724.3 | 0.661 | 0.101 | 0.390 | 159 - 4200 |
| Oil & Gas | 1150.21 | 431.5 | (-)0.05 | (-)0.877 | 0.375 | 210 - 2068 |
| Nuclear | 1298.89 | 236.54 | (-)2.12 | 2.68 | 0.182 | 626 - 1436 |
| Wind | 764.81 | 672.96 | 0.972 | 0.013 | 0.879 | 0 - 2756 |
| Solar | 153.671 | 225.67 | 1.304 | 0.366 | 1.468 | 0 - 839 |
| Biomass | 58.25 | 14.02 | (-)0.345 | (-)0.676 | 0.240 | 19 - 87 |
| Sold\* | 248.417 | 700.098 | (-)0.248 | (-)0.458 | 2.818 | (-)2073 - 2025 |

Table 1: Results of the statistical analysis by method of production

\*Negative values of sold represent export of electrical energy and positive values represent import.

**Trend forecasting method**

**Conclusions**

**References**

**https://www.sistemulenergetic.ro**