Renewable energy world generation 1965-2018

Introduction

This dataset provides information about Renewable Energy generation (TWh) worldwide. The dataset consists of:

1. 5036 rows
2. 7 columns:
    a) countries of the world;
    b) country code;
    c) types of renewable energy sources:
        – Hydropower (terawatt-hours);
        – Solar (terawatt-hours);
        – Wind (terawatt-hours);
        – Other renewables (terawatt-hours).

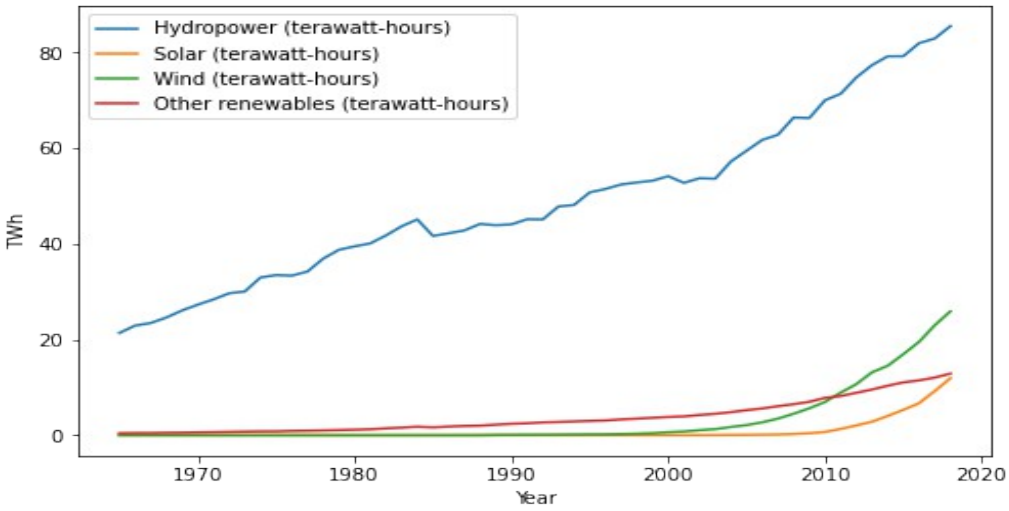
The aim of the analysis is to analyze the dataset according to their global distribution, their annual growth and to develop a more linear regression model.

Note: the data set consists of countries, regions and the data used for analysis as they stand in the dataset because regions bring together countries that are not listed separately.

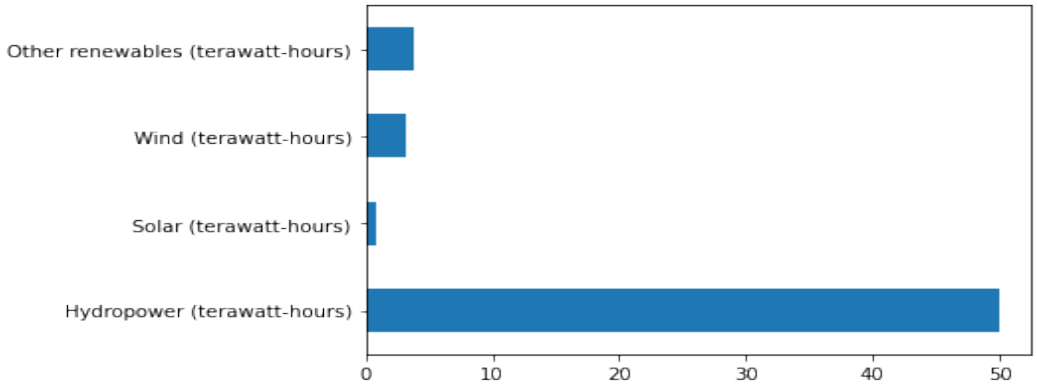
P.S.: TWh - commonly used to calculate the electricity consumed or produced by electrical appliances.

Data overview

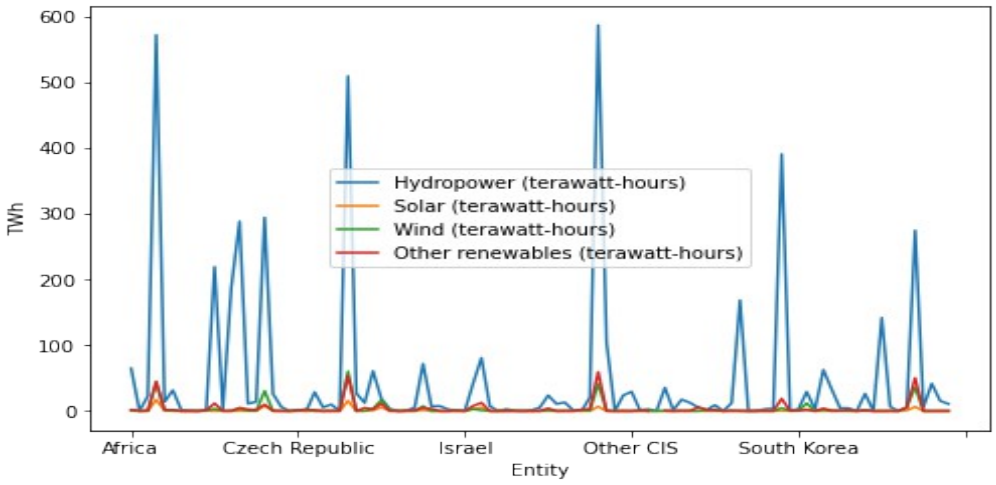
- Calculated WORLD average production of renewable energy sources (TWh) by year



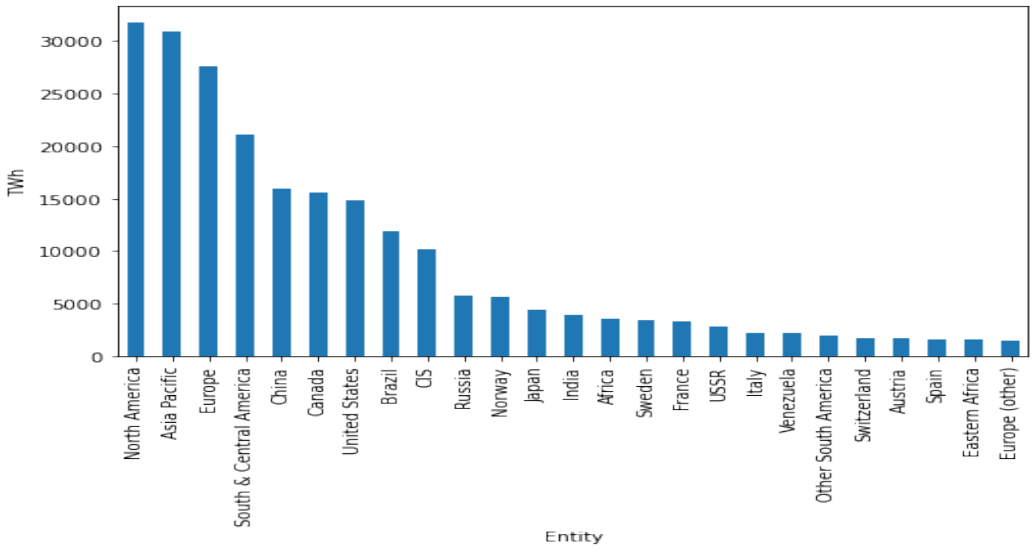
- Calculated WORLD average production of renewable energy sources (TWh) throughout the period



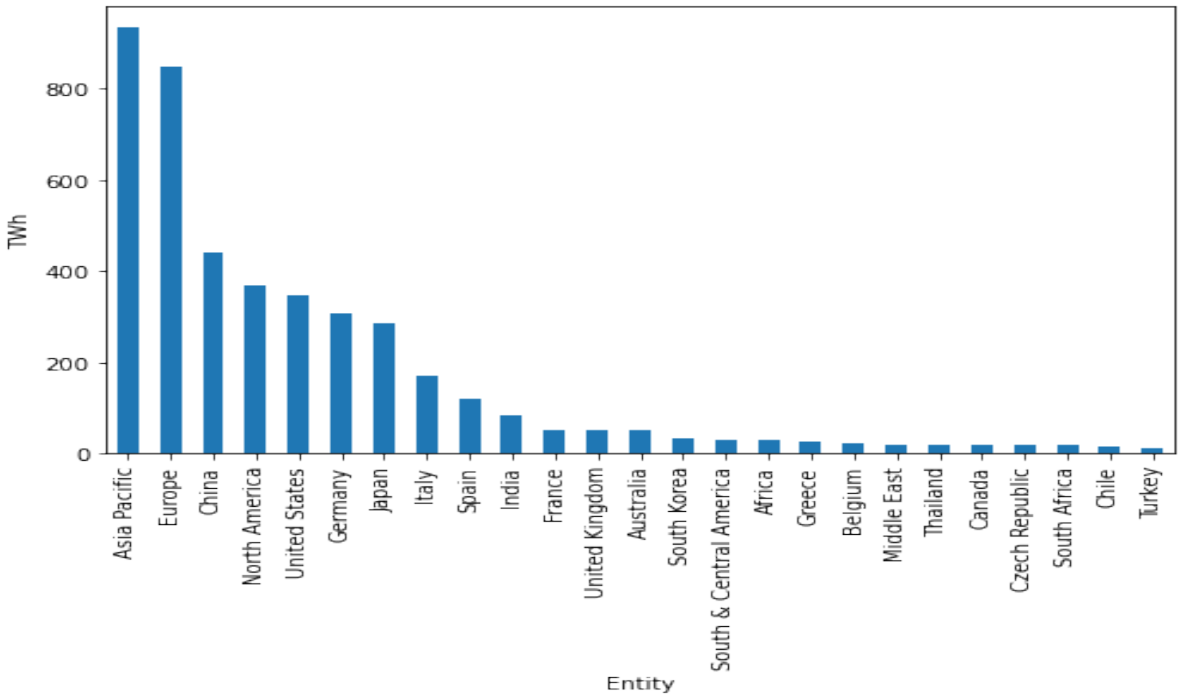
- Average world renewable energy production by category



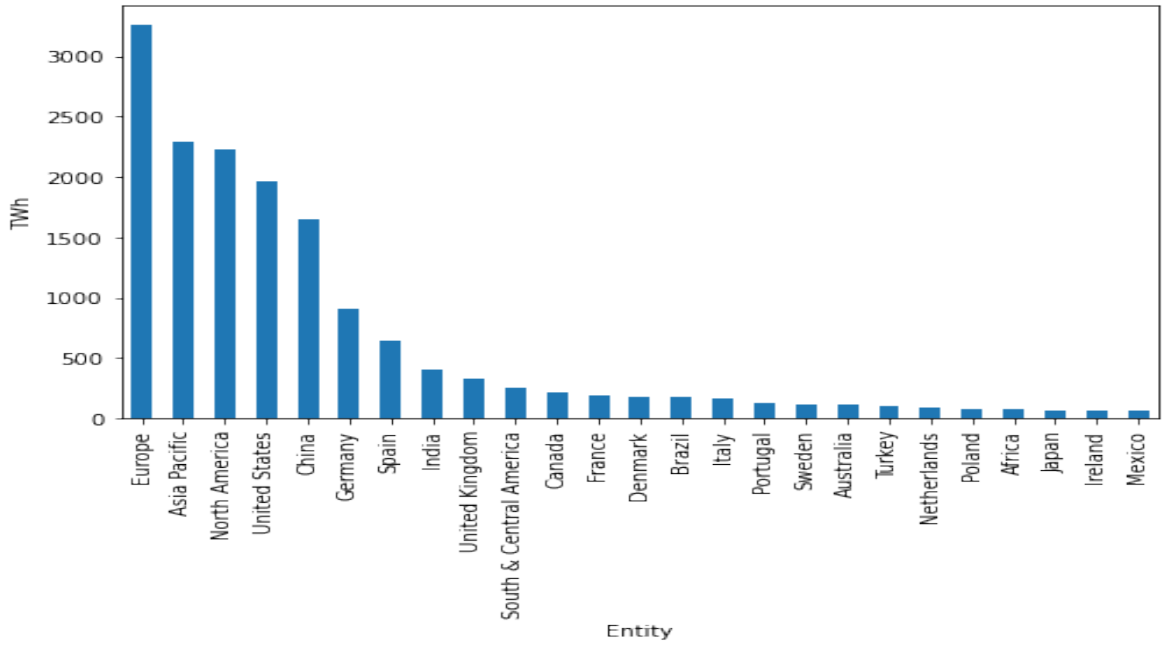
- TOP 25 countries with the largest HYDROPOWER production (TWh)



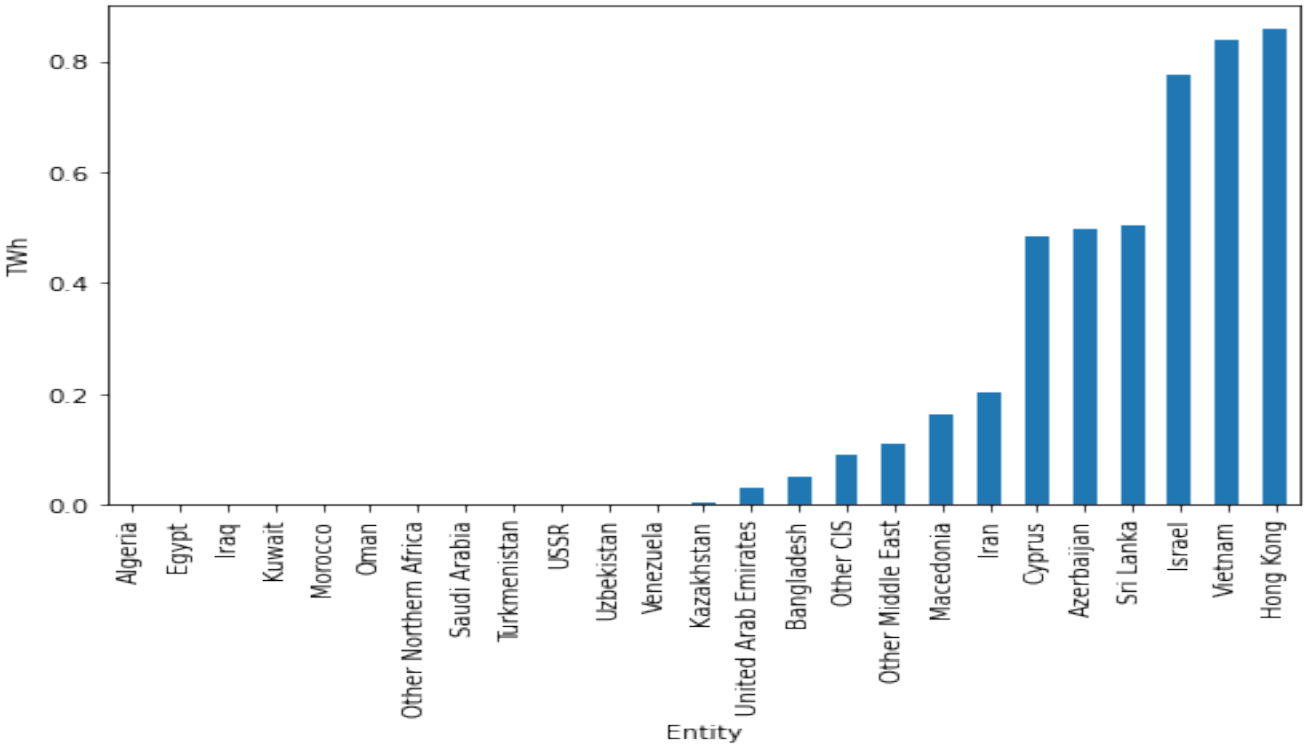
- TOP 25 countries with the largest SOLAR production (TWh)



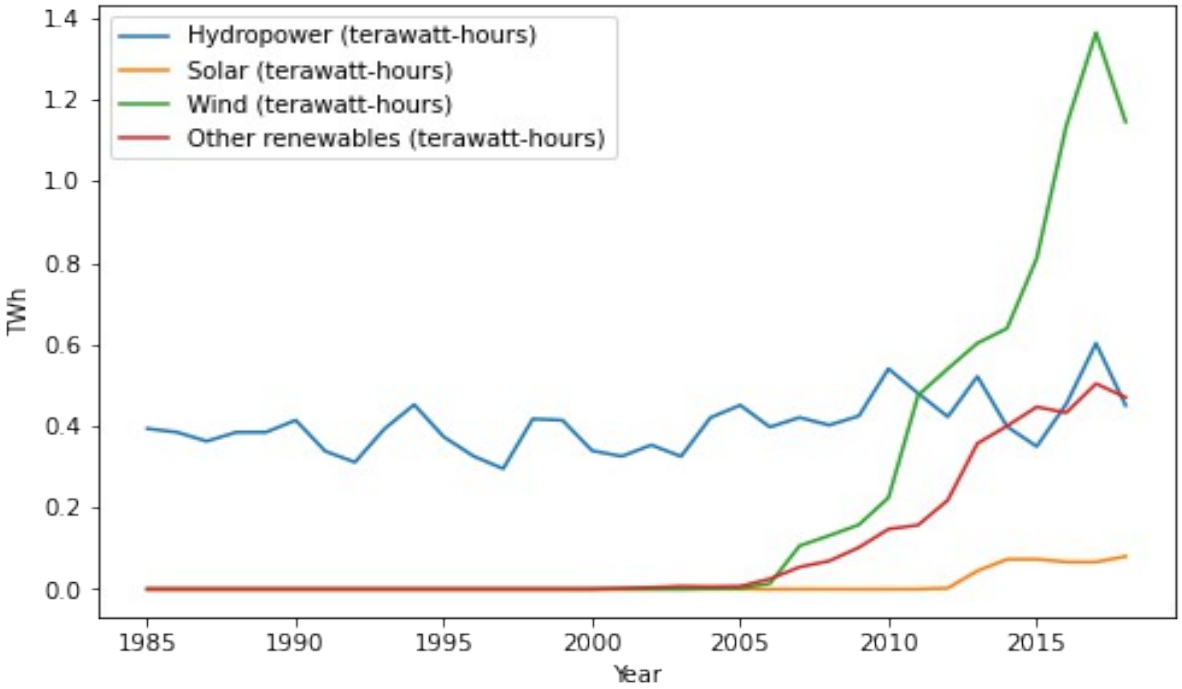
- TOP 25 countries with the largest WIND production (TWh)



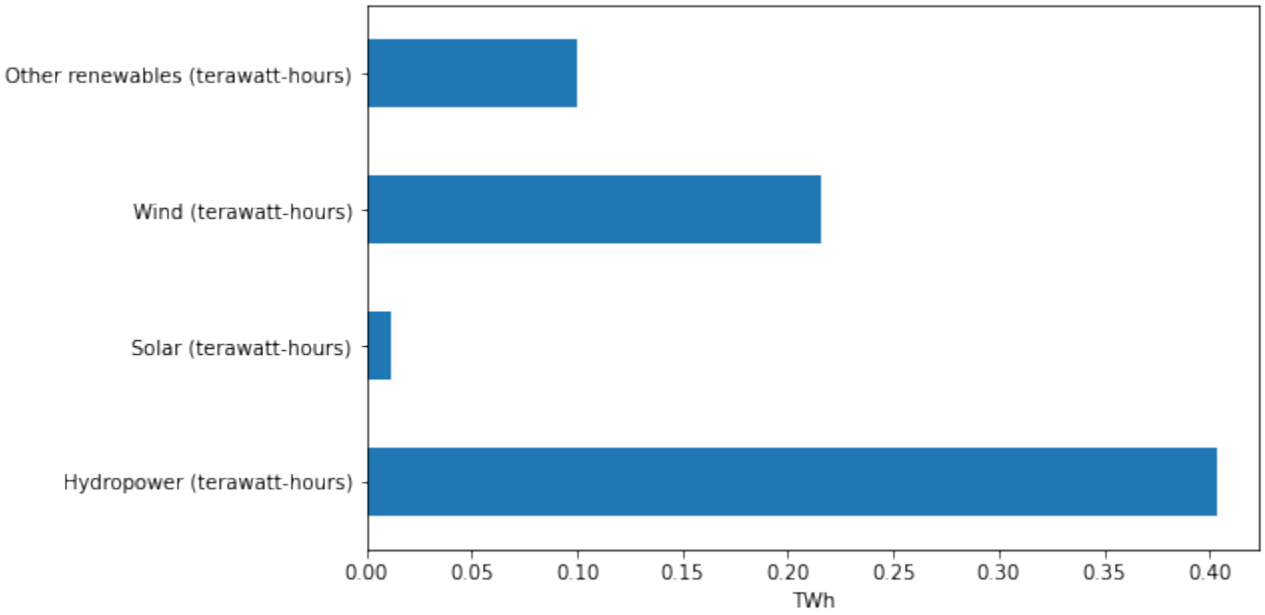
- TOP 25 countries with the largest OTHER RENEWABLES production (TWh)



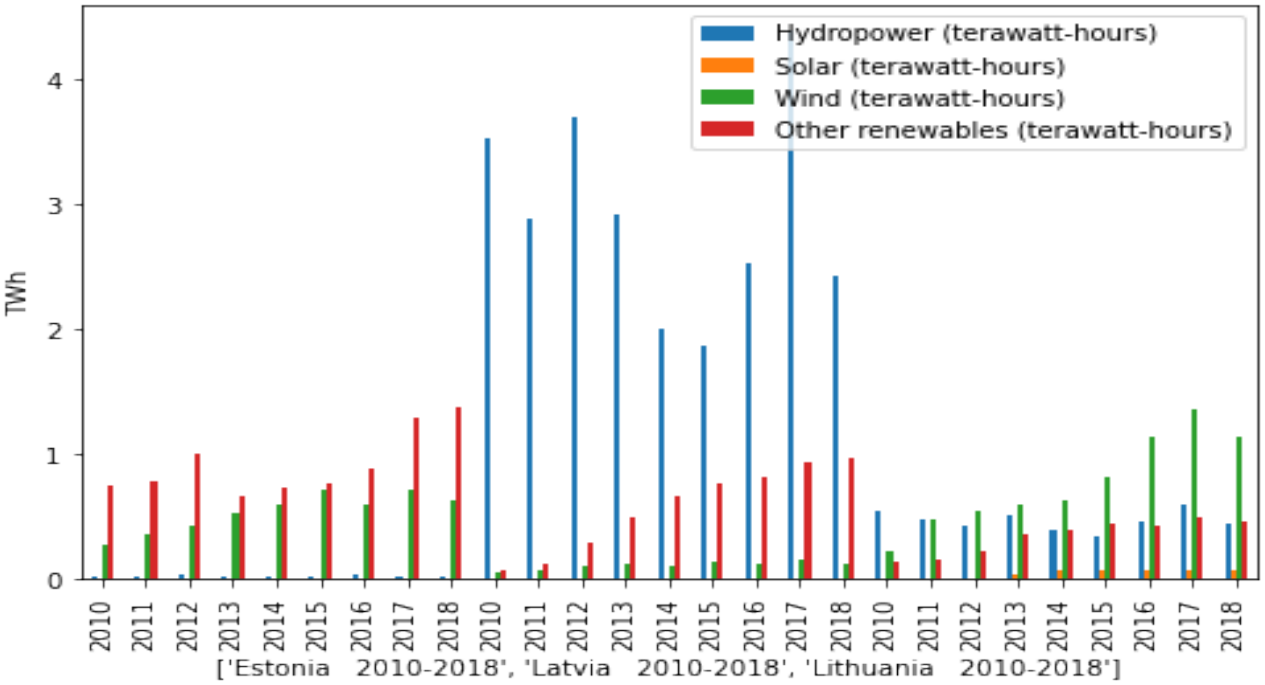
- Calculated LITHUANIA renewable energy sources (TWh) production by year



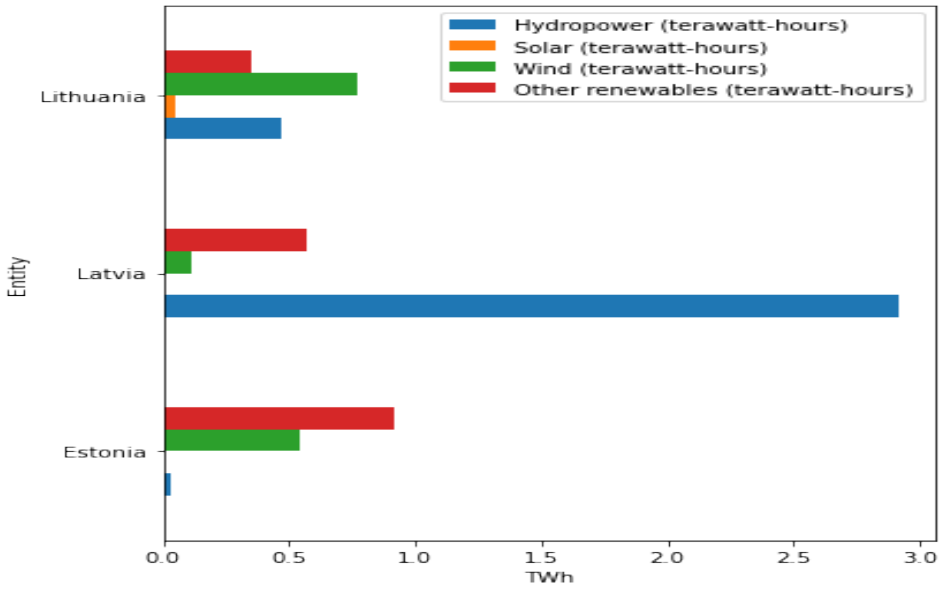
- LITHUANIAN average of renewable energy sources production (TWh) during the whole period



- Distribution of renewable energy sources in the Baltic Countries (Lithuania, Latvia, Estonia) since 2010 – 2018



- Average production of renewable energy sources in the Baltic Countries by categories



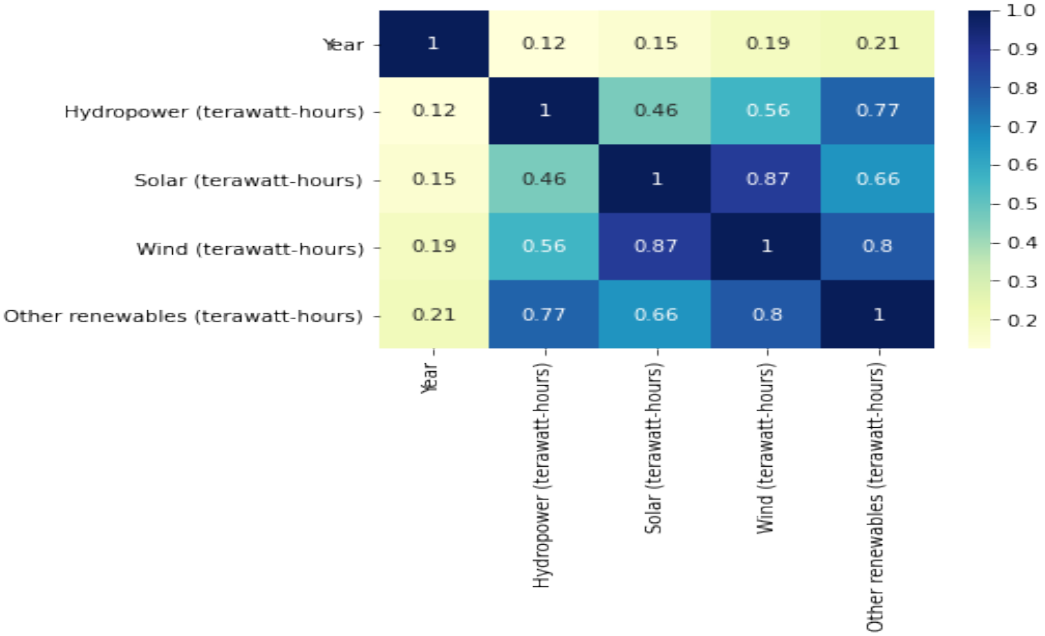
The data visualization

The data visualization - correlation estimation between all variables. It can be seen that the best correlation is between the variables:

- Wind (terawatt-hours) & Solar (terawatt-hours) ~ 87%;
- Wind (terawatt-hours) & Other renewables (terawatt-hours) ~ 79,7%.

Good correlations between:

- Hydropower (terawatt-hours) & Other renewables (terawatt-hours) ~ 76,9 %

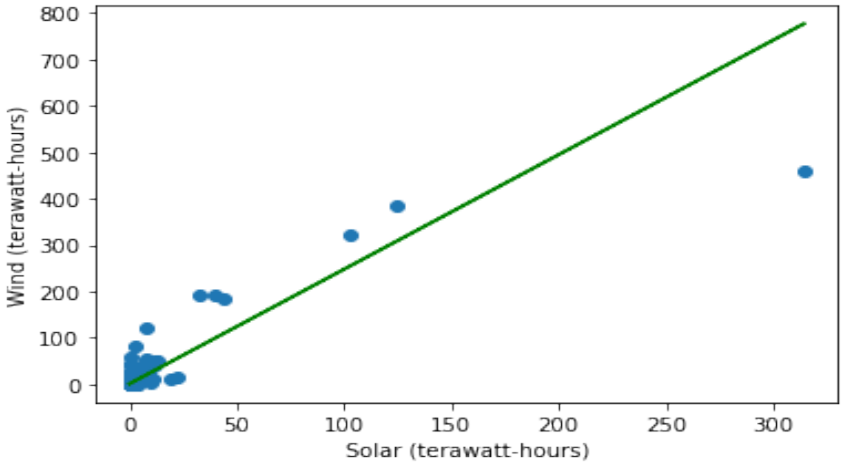


Prognostication

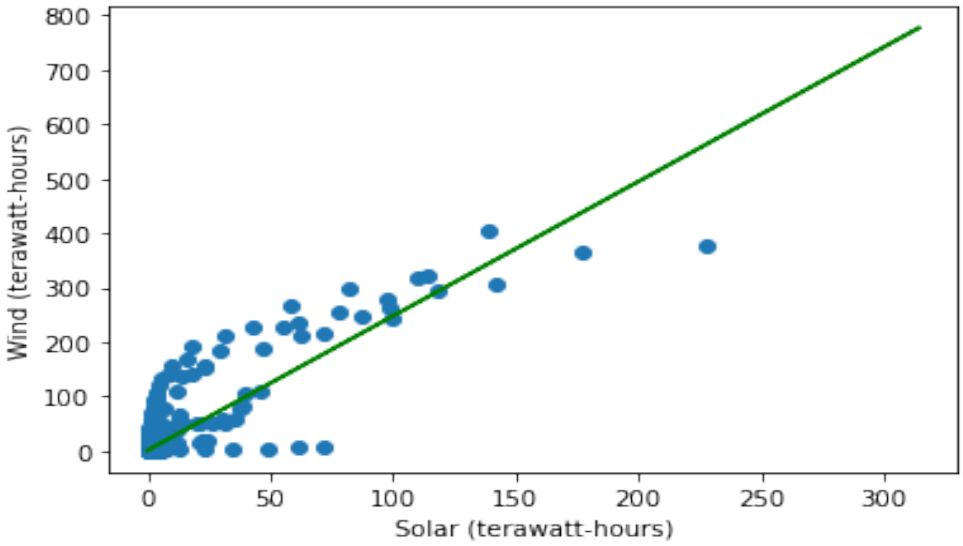
Linear regression is performed, its function is formed and calculated **R-s quared (R2)**, which is the proportion of variation in the outcome that is explained by the predictor variables. The Higher the R-squared, the better the model.

In this analysis, R2 is about 72%, which shows that the created model is good .

Linear regression on data not seen (about 20% of the data)



Linear regression on train data (about 80% of the data)



Conclusions

Based on performed analysis (data of 1965-2018) the following conclusions were made:

1. Hydro energy is taking the lead based on all time averages, then come other sources, then wind and solar energy is last
2. Each year average taken separately shows constant growth and domination of hydro energy

3. In Lithuania hydro power also takes the lead, but analyzing growth, wind and other sources are looking strong
4. Looking and Baltic countries: Latvia is number one in hydro power, Lithuania – in wind, Estonia - in solar.
5. After analyzing data correlation, we see a match in wind growth to solar and wind to other sources, so these sources grow together in very similar rates