Data Analytics

Predicting electricity prices

Marcin Pilarski

Mateusz Saternus

Problem formulation

- The energy market plays a crucial role in the economic development and sustainability of nations worldwide. The energy market in Poland faces multifaceted challenges, including fluctuating energy prices, increasing demand, changing regulatory frameworks, and the need to transition towards a more sustainable energy mix. These complexities pose significant difficulties for market participants, policymakers, and investors in making informed decisions regarding energy production, distribution, and consumption. To tackle these challenges effectively, it is crucial to understand the underlying patterns and interdependencies within the market, which can be achieved through statistical modelling.

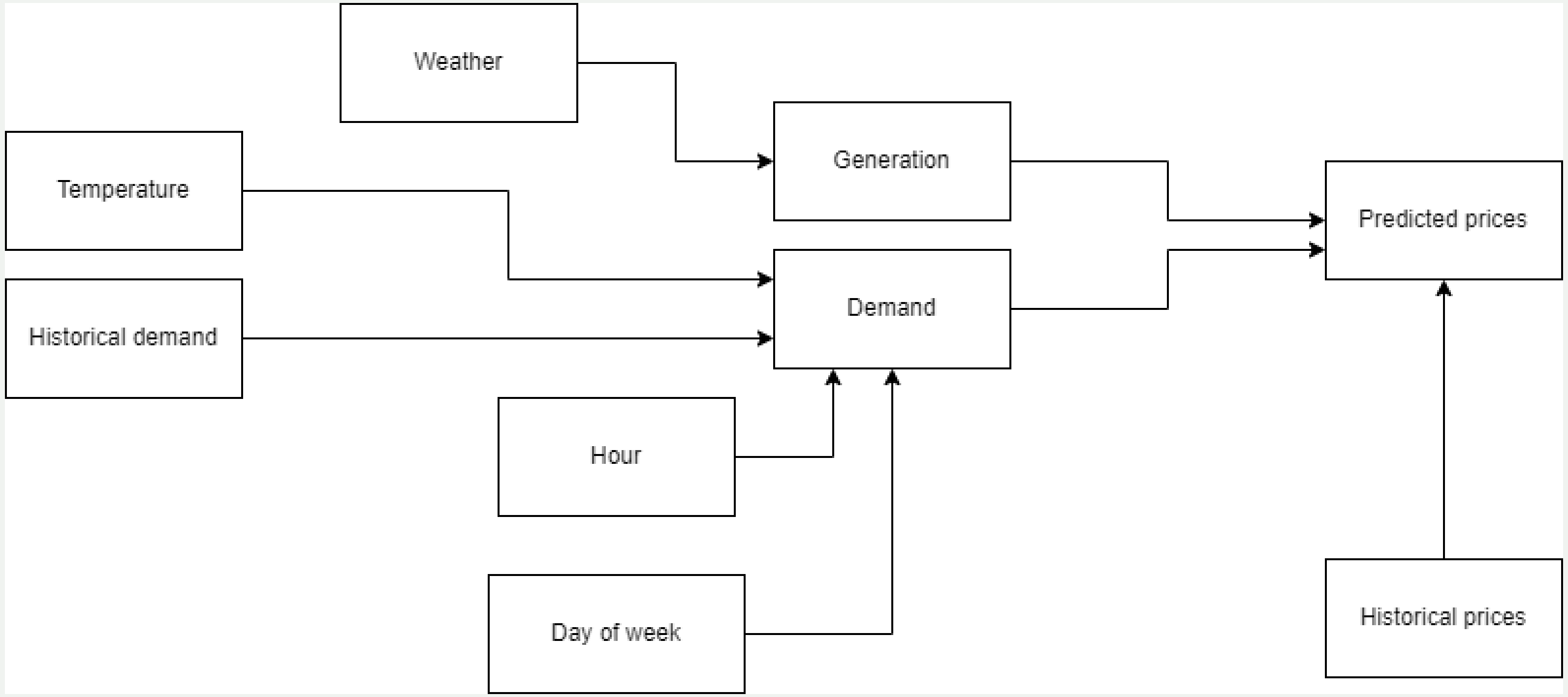
Problem formulation – potencial use cases

- The statistical modelling of the energy market in Poland offers a wide range of potential use cases across different sectors. For energy companies, the model can assist in optimizing production levels, pricing strategies, and investment decisions by providing insights into demand patterns and market conditions. Regulators and policymakers can leverage the model to evaluate the effectiveness of existing policies, design new regulations, and identify areas for promoting renewable energy sources and energy efficiency. Investors can use the model to assess the financial viability and risks associated with energy projects, aiding in informed decision-making.

Problem formulation – data sources

- Data used in model comes from official databases. First database used in project is PVgis, which is tool released by European Commission. (https://re.jrc.ec.europa.eu/pvg_tools/en/) It helps to plan investitions in solar and wind energy sources. The data contains irradiance, wind speed, temperature.
- The information about energy market such as energy prices and system load come from polish energy system operator (PSE). (https://www.pse.pl/obszary-dzialalnosci/rynek-energii/ceny-i-ilosc-energii-na-rynku-bilansujacym; https://www.pse.pl/dane-systemowe/funkcjonowanie-kse/raporty-dobowe-z-pracy-kse/wielkosci-podstawowe)

Problem formulation- Directed Acyclic Grap



Problem formulation - confoundings

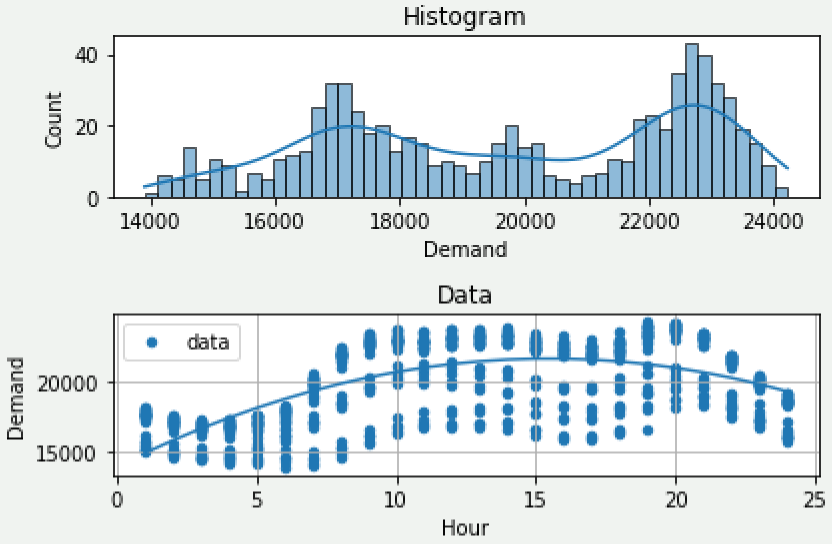
- There are few confoundings detected in the model.
- The pipe type is on relation of weather->generation->price.
- There are also fork type confoundings:
  - Hour -> Load <- Day of week
  - Generation -> Price <- Historical price

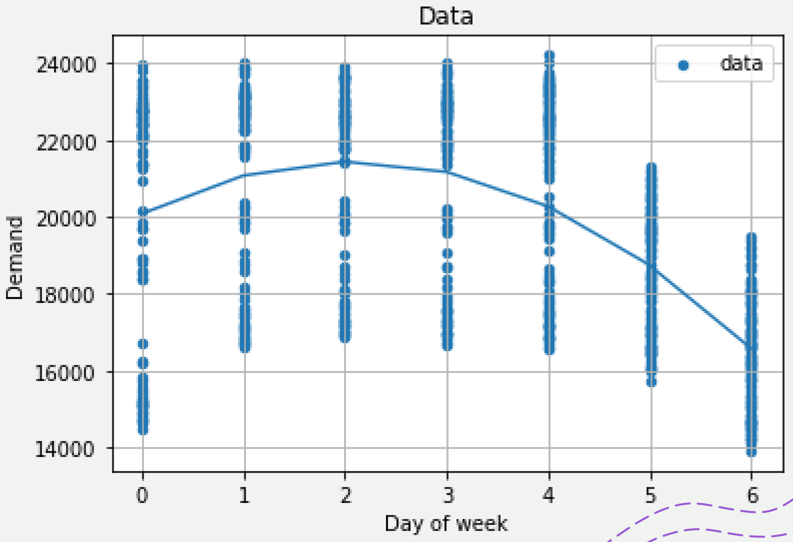
Data preprocessing

+ The data about the weather was downloaded from PVgis database. Although it contains all the weather values that were useful for us, it allows only to collect data locally and we need information for the whole country. Because of that we have chosen 22 points in Poland and calculated mean value at the time. Points are shown on the map. Data is collected every hour and was downloaded separately for year 2019 and 2020.
+ The data about the energy market was downloaded from PSE polish energy system operator. It was downloaded for march 2019 and march 2020 separately. It contained information about the system load and balancing market energy prices. The data was slightly changed to allow easier operations on dates.
+ Data from 2019 was used for prior and from 2020 for posterior

Model 1 – Demand

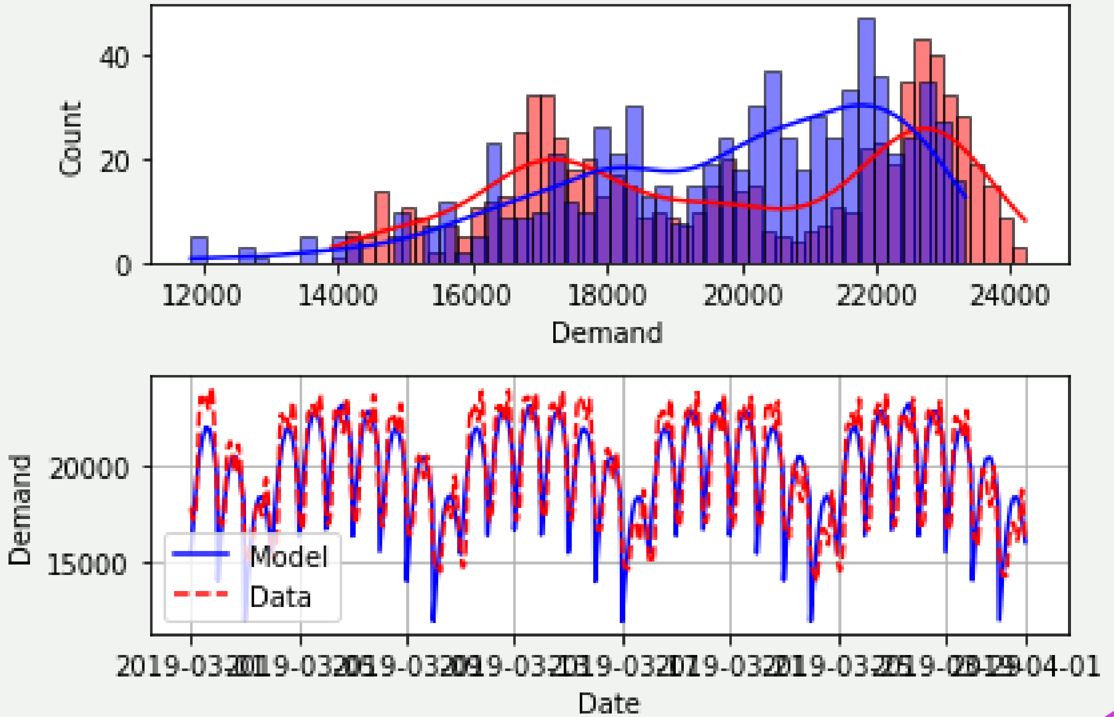
- The load has been fitted with 2nd degree polynomial for each dependence on hour and day of week





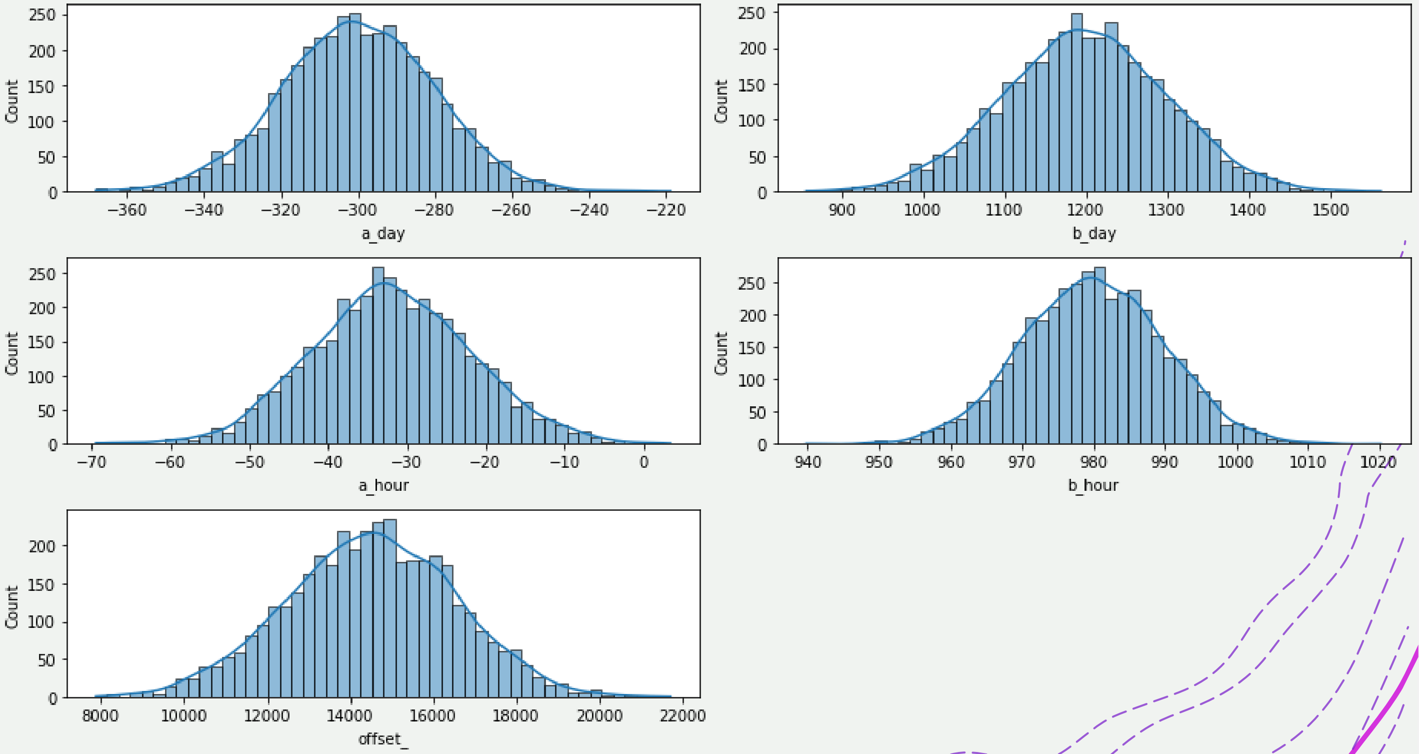
Model 1 – Demand - Prior

+ Fitting time data with 2nd degree polynomial allowed us to get samples with distribution shown below
+ Value of error was:
  RMSE = 1352.998 / (max possible) 19982.844



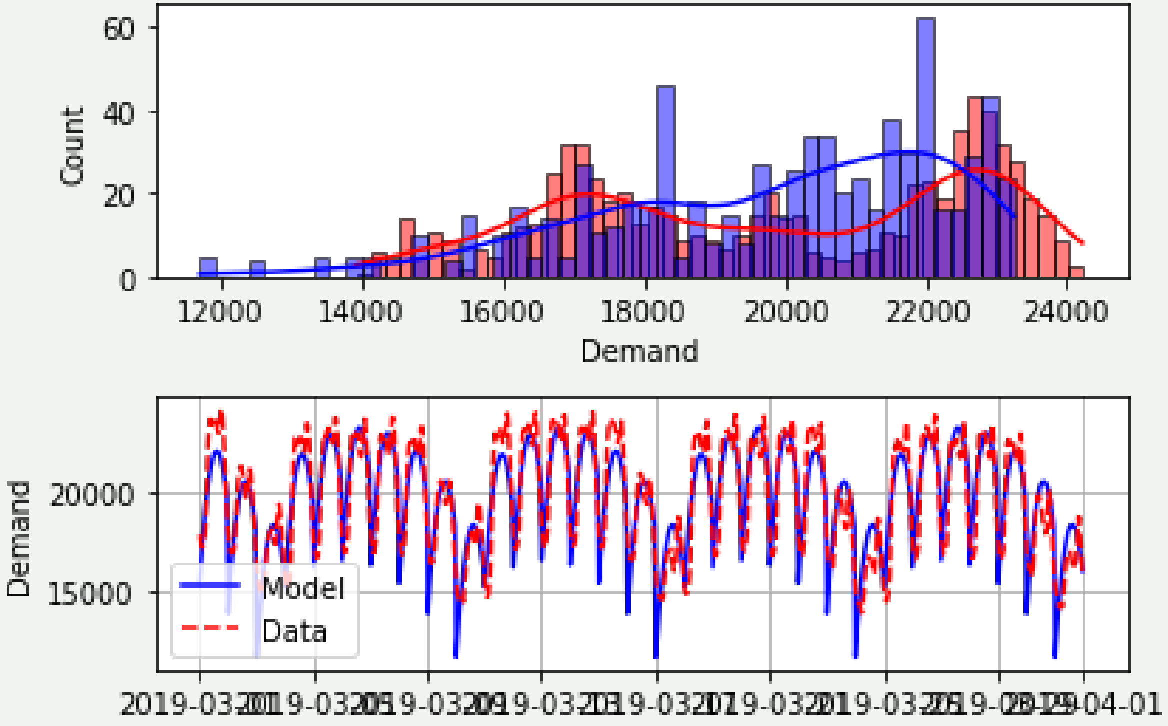
Model 1 – Demand – Prior Summary

name	Mean	MCSE	StdDev	5%	50%	95%	N_Eff	N_Eff/s	R_hat
lp__	0.0	NaN	0.0	0.0	0.0	0.0	NaN	NaN	NaN
a_hour	-32.0	0.16	10.0	-49.0	-32.0	-15.0	4200.0	650.0	1.0
b_hour	980.0	0.15	9.9	960.0	980.0	1000.0	4200.0	650.0	1.0
offset_	14000.0	32.00	2000.0	11000.0	15000.0	18000.0	4000.0	620.0	1.0
a_day	-300.0	0.32	20.0	-330.0	-300.0	-270.0	3700.0	570.0	1.0
...	...	...	...	...	...	...	...	...	...
demand[739]	17680.0	72.00	4590.0	10173.0	17570.0	25313.0	4012.0	621.0	1.0
demand[740]	17344.0	79.00	4886.0	9343.0	17336.0	25360.0	3817.0	590.0	1.0
demand[741]	16908.0	85.00	5336.0	8117.0	16830.0	25644.0	3907.0	604.0	1.0
demand[742]	16492.0	92.00	5773.0	6734.0	16566.0	25916.0	3916.0	606.0	1.0
demand[743]	16013.0	108.00	6317.0	5847.0	15916.0	26478.0	3392.0	525.0	1.0



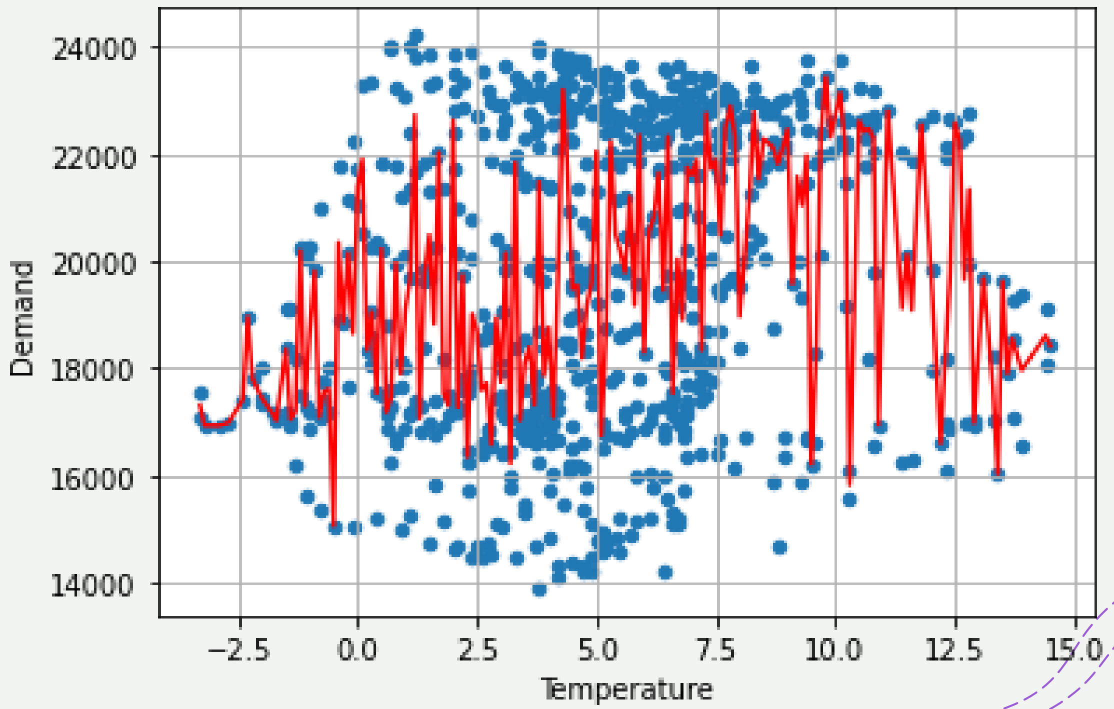
Model 1 – Demand – Posterior

+ Posterior modeling gave the results shown on graph
+ Value of error was: RMSE = 1348.9/ (max possible) 19982.8



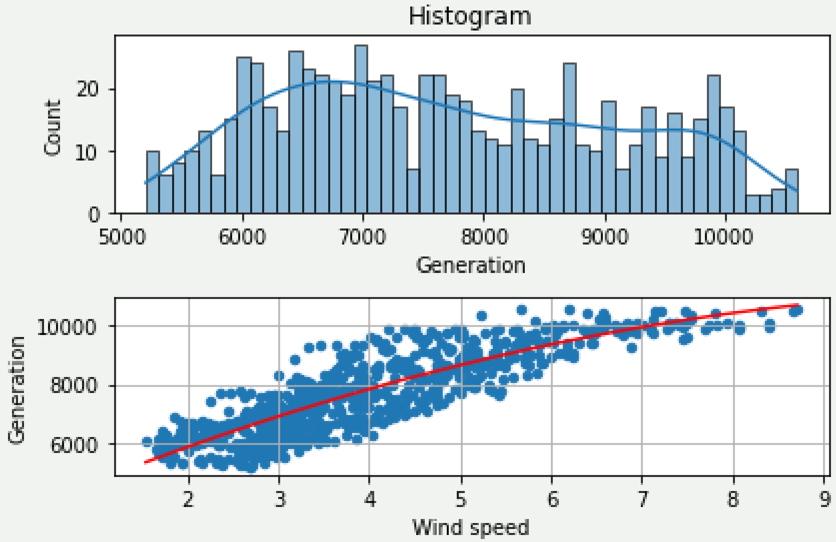
Temperature’s influence on the load

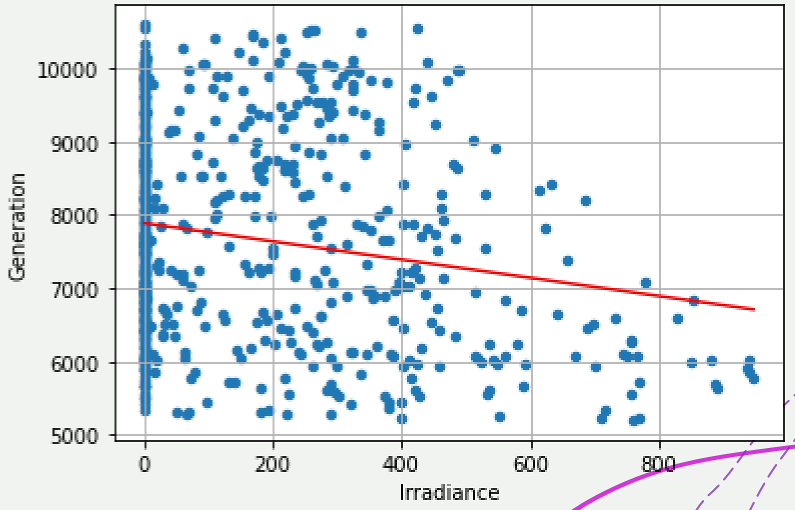
+ It seems that there is no influence of temperature on the demand, contrary to the project's assumptions.
+ Therefore, the influence of temperature was omitted in models



Model 1 – Generation

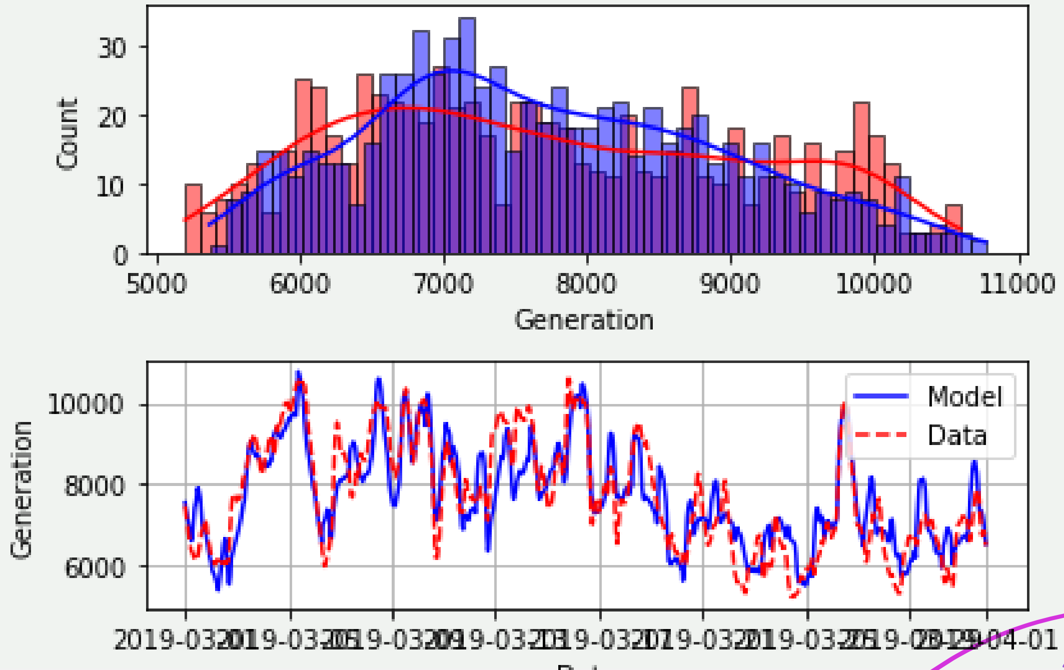
- Generation dependence on wind and solar irradiance has been tested
- As shown below, there is strong dependence on wind speed and almost no dependance on irradiation
- Wind speed has been fitted with quadratic function and irradiance with linear
- Therefore we omitted irradiation influence on generation in 1st model – it is added in the 2nd





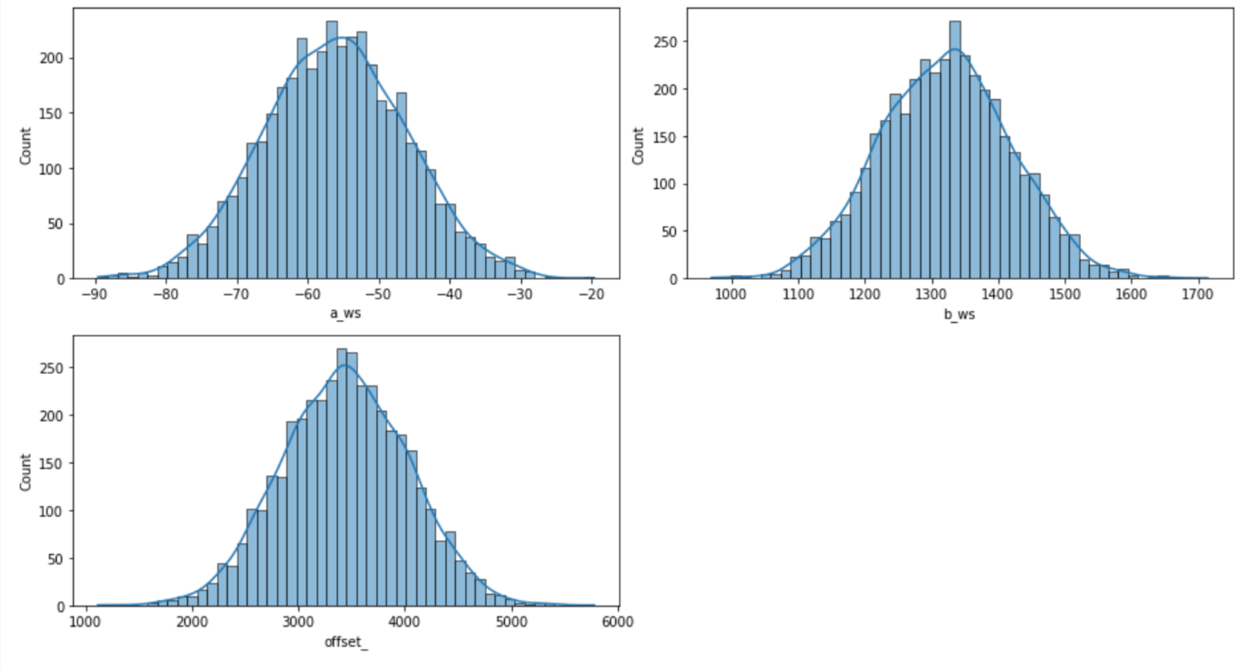
Model 1 – Generation - Prior

+ The prior distribution for Generation in 1st model is shown below
+ The value of error was: RMSE = 694.38 / (max possible) 7836.45



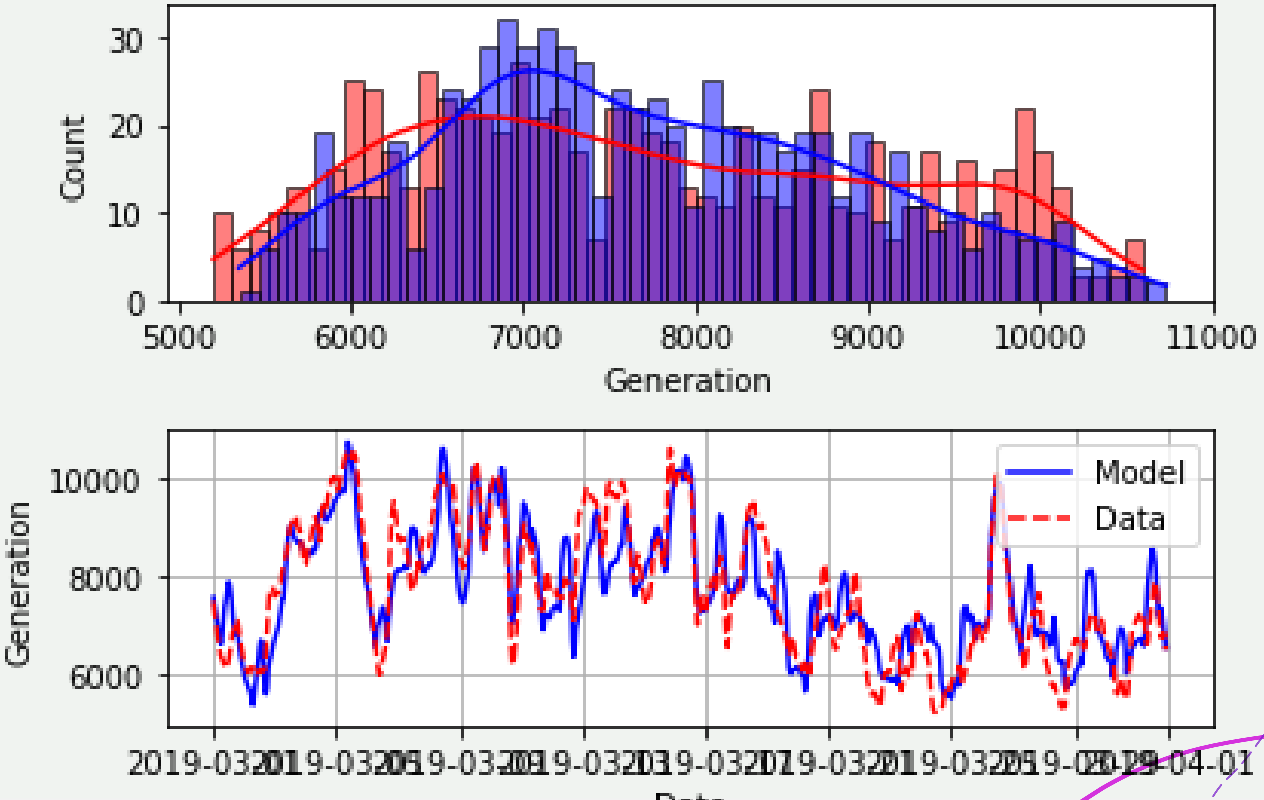
Model 1 – Generation – Prior summary

name	Mean	MCSE	StdDev	5%	50%	95%	N_Eff	N_Eff/s	R_hat
lp__	0.0	NaN	0.0	0.0	0.0	0.0	NaN	NaN	NaN
a_ws	-56.0	0.16	10.0	-72.0	-56.0	-40.0	3700.0	800.0	1.0
b_ws	1300.0	1.60	99.0	1200.0	1300.0	1500.0	3900.0	840.0	1.0
offset_	3400.0	9.20	590.0	2500.0	3400.0	4400.0	4100.0	890.0	1.0
generation[1]	7534.0	19.00	1248.0	5494.0	7503.0	9633.0	4219.0	910.0	1.0
...	...	...	...	...	...	...	...	...	...
generation[739]	7316.0	19.00	1225.0	5360.0	7294.0	9358.0	4107.0	886.0	1.0
generation[740]	7237.0	20.00	1216.0	5239.0	7257.0	9240.0	3854.0	832.0	1.0
generation[741]	7074.0	19.00	1225.0	5034.0	7093.0	9083.0	4003.0	864.0	1.0
generation[742]	6770.0	19.00	1193.0	4828.0	6756.0	8751.0	3888.0	839.0	1.0
generation[743]	6511.0	20.00	1206.0	4505.0	6496.0	8493.0	3736.0	806.0	1.0



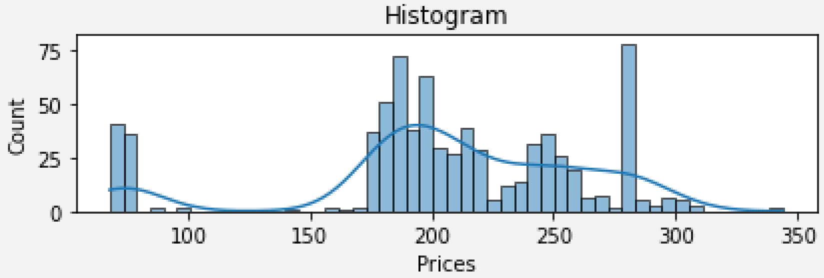
Model 1 – Posterior

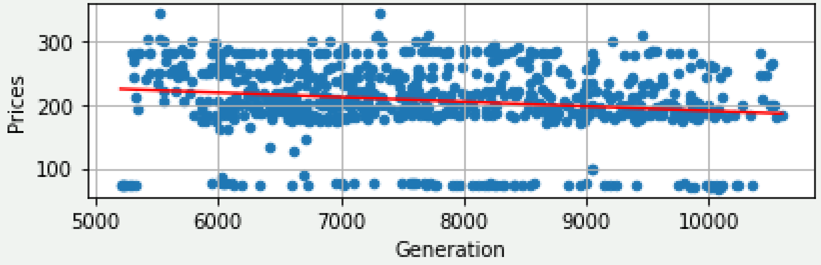
+ Posterior modeling gave the results shown on graph
+ The error value was: RMSE = 694.48 / (max possible) 7836.45

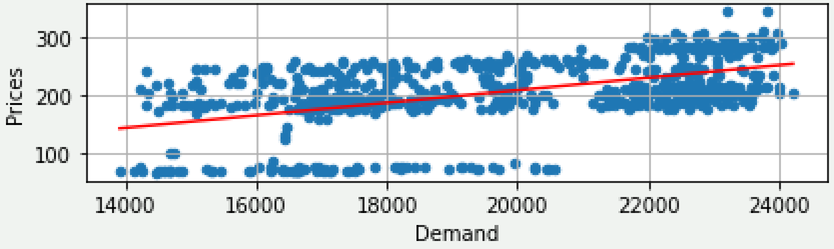


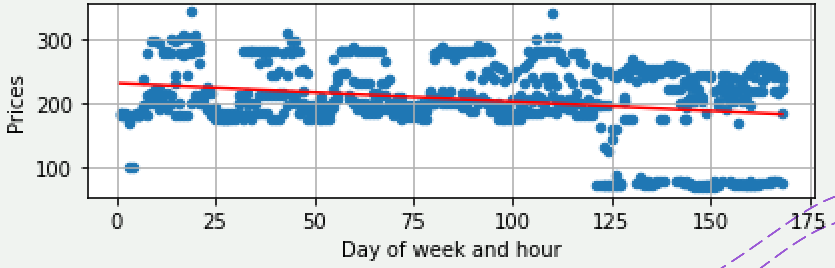
Model 1 - Prices

+ We have tried to determine the relations with load, generation and day of week with linear functions



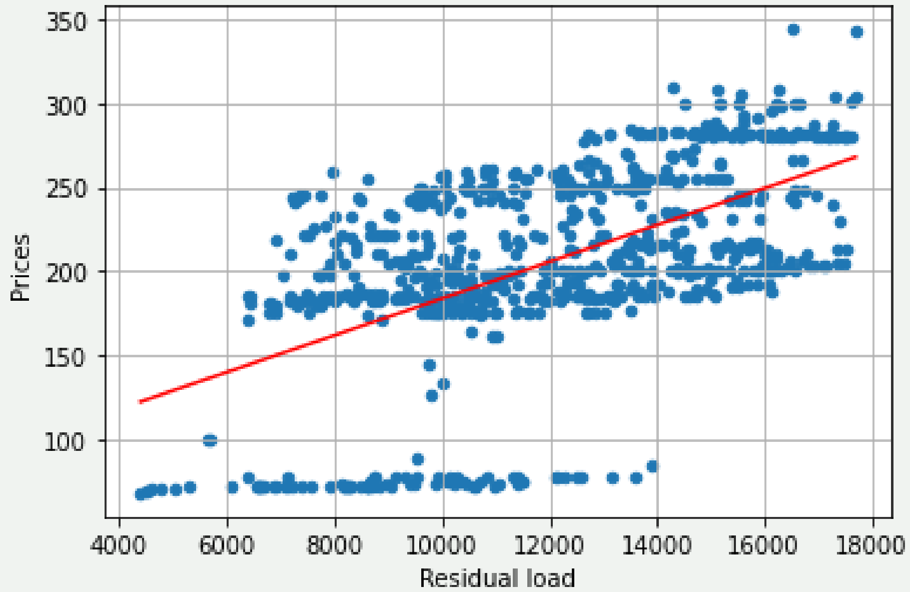






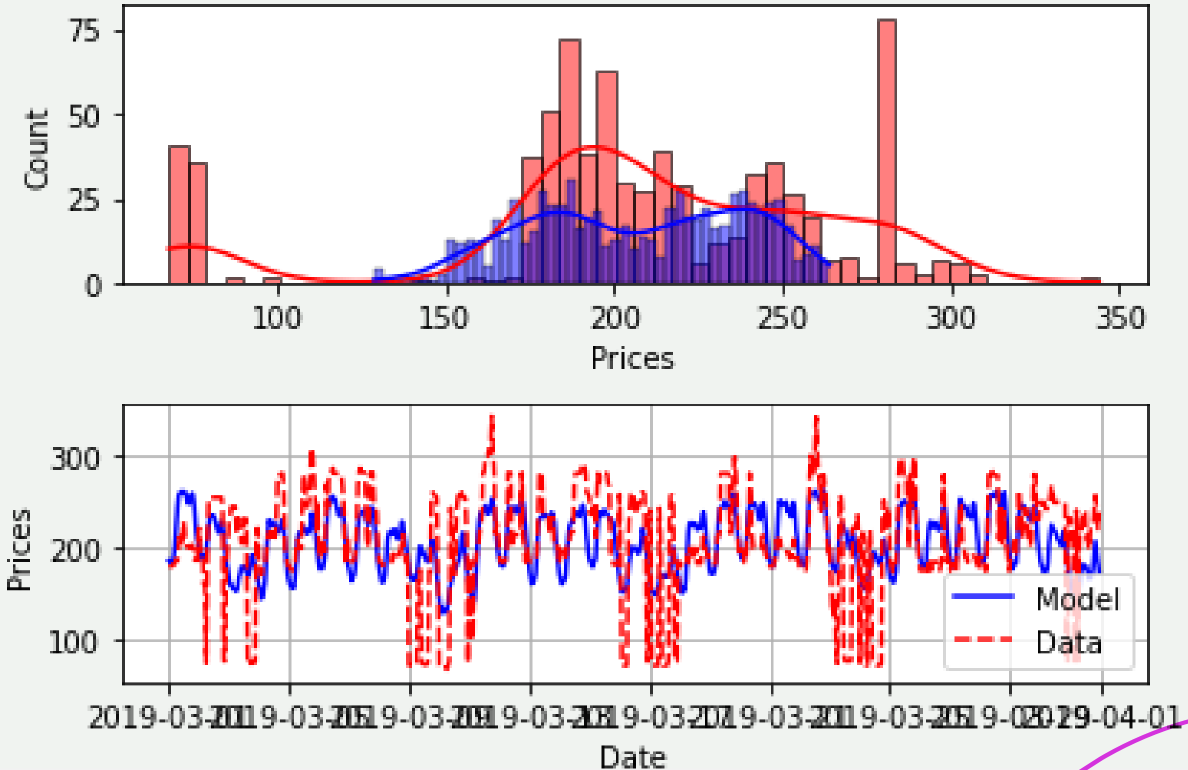
Model 1 - Prices

+ Finally we have decided to use the prices dependence on residulal load
+ Residual load is system load substracted by noncontrollable generation (called simply „Generation" in our project)



Model 1 – Prices - Prior

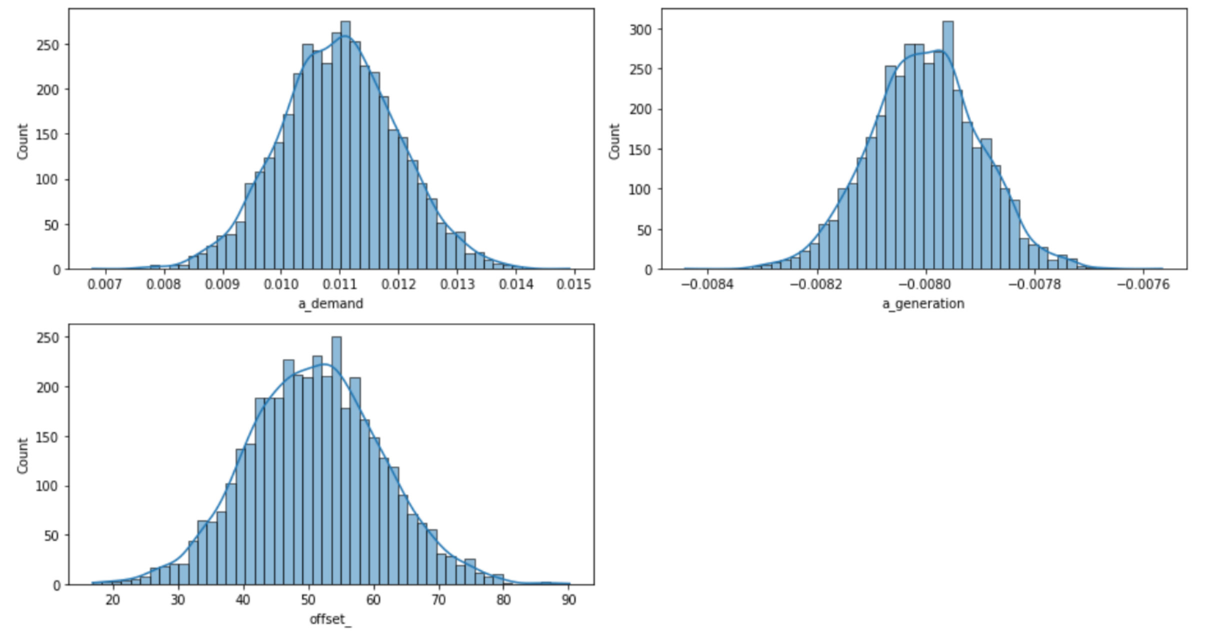
+ Modelling Prior for prices in 1st Model gave results shown below
+ The error value was: RMSE = 48.18 / (max possible) 214.47



Model 1 – Prices – Prior summary

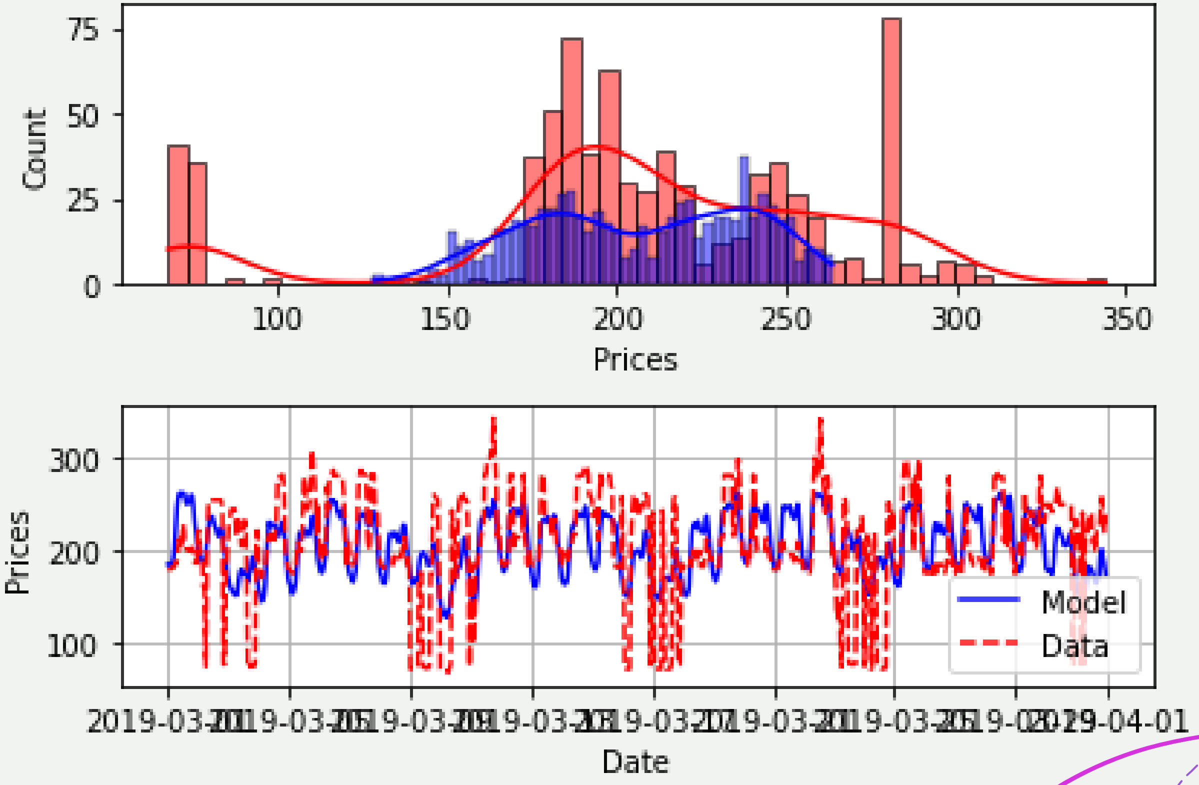
name	Mean	MCSE	StdDev	5%	50%	95%	N_Eff
lp__	0.000	NaN	0.0000	0.0000	0.000	0.0000	NaN
a_demand	0.011	0.000016	0.0010	0.0094	0.011	0.0130	4100.0
a_generation	-0.008	0.000002	0.0001	-0.0082	-0.008	-0.0078	3800.0
offset_	51.000	0.160000	10.0000	34.0000	51.000	68.0000	4000.0
prices[1]	187.000	1.600000	103.0000	19.0000	188.000	357.0000	3970.0
...	...	...	...	...	...	...	...
prices[739]	194.000	1.600000	101.0000	31.0000	193.000	361.0000	3826.0
prices[740]	208.000	1.600000	103.0000	41.0000	207.000	381.0000	4241.0
prices[741]	197.000	1.600000	102.0000	29.0000	197.000	362.0000	4091.0
prices[742]	185.000	1.600000	101.0000	16.0000	185.000	349.0000	4038.0
prices[743]	174.000	1.700000	102.0000	7.0000	174.000	346.0000	3798.0

name	N_Eff/s	R_hat
lp__	NaN	NaN
a_demand	810.0	1.0
a_generation	750.0	1.0
offset_	780.0	1.0
prices[1]	786.0	1.0
...	...	...
prices[739]	757.0	1.0
prices[740]	839.0	1.0
prices[741]	810.0	1.0
prices[742]	799.0	1.0
prices[743]	752.0	1.0



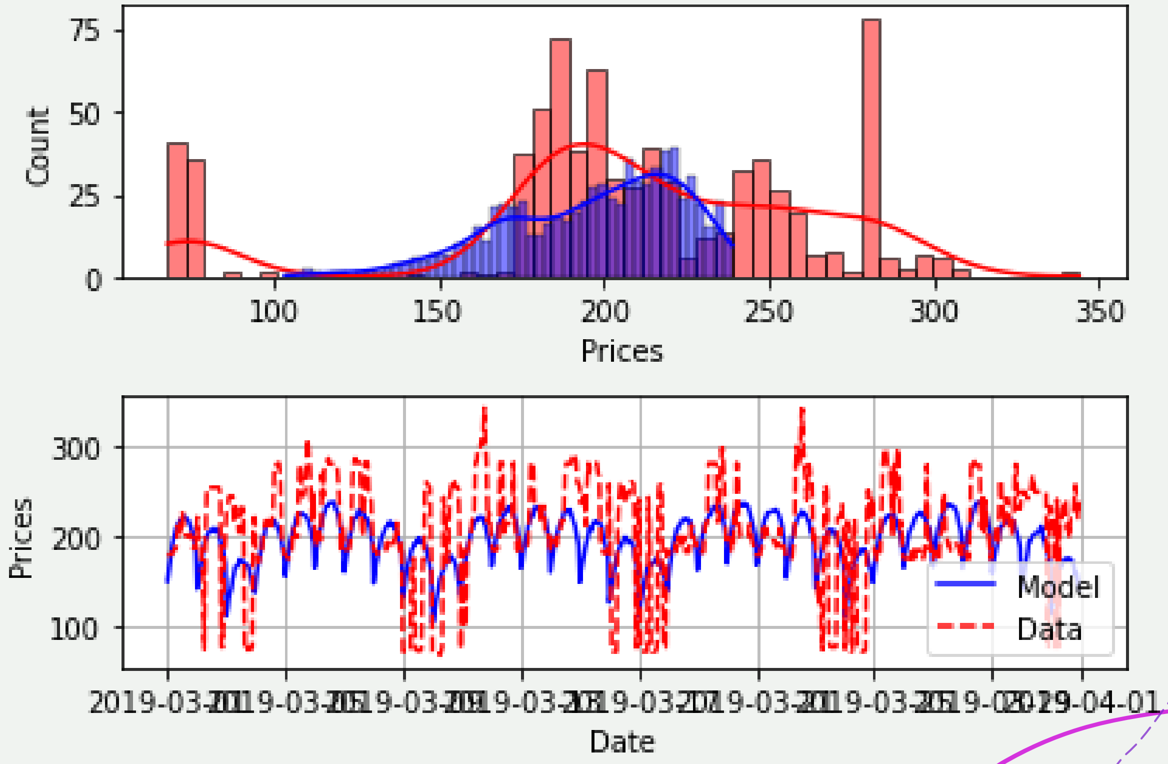
Model 1 – Prices - Posterior

+ Posterior model gave us results shown below
+ The error was: RMSE = 48.14 / (max possible) 214.47



Model 1 – Overall posterior

- Overall posterior results are shown below
- The error value was: RMSE = 53.27 / (max possible) 214.47

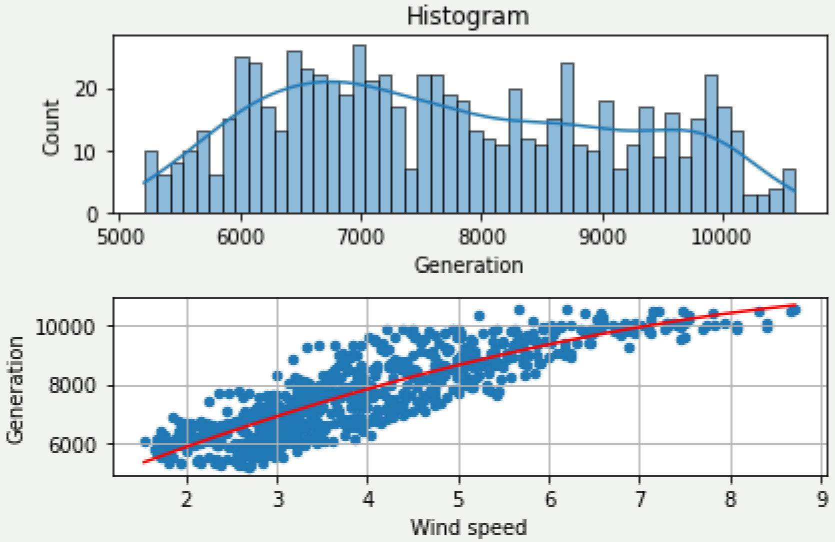


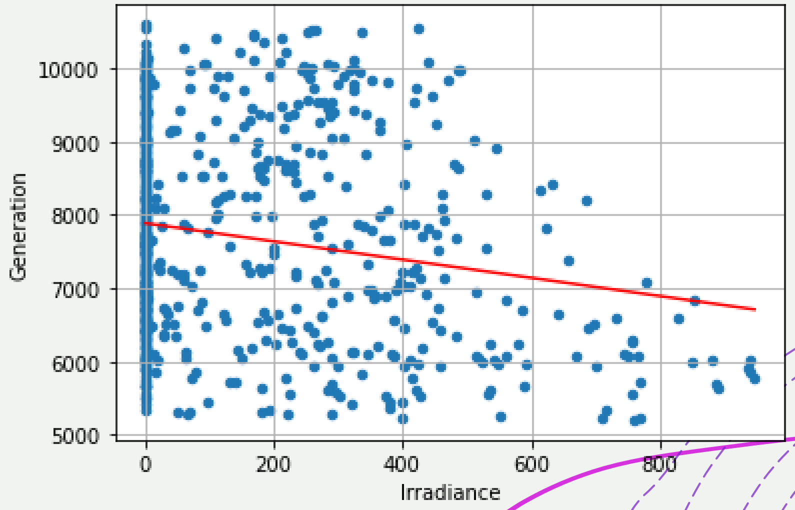
Model 2

- Demand prior and posterior for both models are the same
- 2nd model differs from the first one in generation – in the 2nd one the irradiance influence is calculated, while in the first it is omitted

Model 2 – Generation

+ Generation dependence on wind is the same as in the first model
+ Wind speed has been fitted with quadratic function and irradiance with linear
+ The irreadiance influence on generation is negative, which is a bit couterintuitive
+ The reason for that can be explained with the fact that not only renewable Energy sources are included in non-controllable energy generation data, but also small conventional power plants





Demand dependance on day of week and hour – additional model

- Hourly demand distriution has been done separately for each weekday