# Data for Ecological Transition – Using open data to assess EU renewable energy context

Data Hackathon – a.y. 2021/2022

Proff.s:

A. Longo;
A. Martella,
M. Zappatore

Students:

Marco De Carlo,
Simone Lavria,
Simone Olimpico,
Andrea Panico

Team:

Pennic Hellas

# 1 - INTRODUCTION

## 1.1 - THE MISSION

UN Agenda 2030, Italian PNRR and the increase of our energy bills demand for a general awareness towards ecological transition to create a more sustainable world.
Open data is one of the resources available to understand our reality, analyse the context and support decision making.
The aim of the contest is to provide well-grounded evidence and support decision makers on how to improve the exploitation of renewable energy in the EU . Therefore, it is requested to
1. Analyze a big dataset describing renewable power plants in different EU countries by reverse engineering and querying it.
2. Transform the dataset in order to support situation assessment and decision making by developing adequate analytics and data dashboards.

## 1.2 - THE DATASET

The dataset is available on the website https://data.open-power-system-data.org/renewable_power_plants/2020-08-25.
It contains a list of renewable energy power plants in lists of renewable energy-based power plants of:

- Czechia,
- Denmark,
- France,
- Germany,
- Poland,
- Sweden,
- Switzerland,
- United Kingdom

# 2 - REVERSE ENGINEERING

The dataset contains information regarding electricity plants that exploit some renewable sources.
The different datasets (divided by country) provide information regarding the electrical capacity of the individual plants, details relating to the renewable source used, the location, the manufacturing company and some contractual information.
There are obvious differences related to the definition of the place where the plant resides. This is highly dependent on the country being considered. However, in each dataset the punctual information (municipality) and the relative region are reported. So it was decided to model these two pieces of information in the reconciled database.
As for the technical information, some nations better detail the wind and photovoltaic plants used. Although this information is not reported in all datasets, it was decided to include it anyway as it could be useful for the analysis.
Some datasets provide information regarding the identifiers of the plants and the personnel who manage them. This information will not be processed as it is not related to the objectives of the project.

## 2.1 - EER DIAGRAM

The reverse engineering performed led to the EER diagram shown in fig. 2.1.

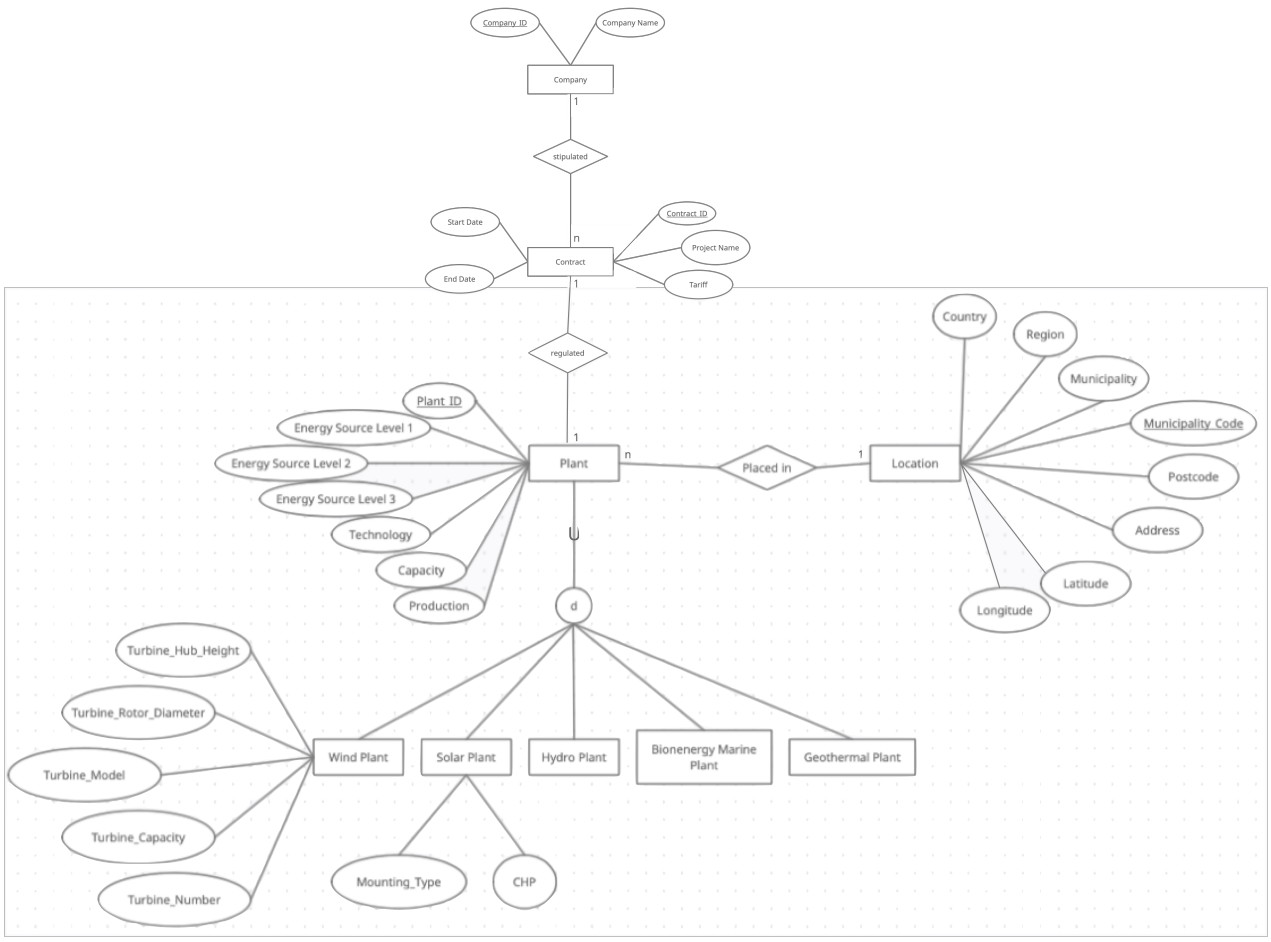



Fig. 2.1 EER Diagram for the renewable power plants of UE

In the model it is possible to observe how a generic renewable electricity plant can specialize in five types of specific plant.
Among these we find wind and solar plants which in the dataset are detailed in the technical aspects as opposed to geothermal, water and bioenergy plants.

### 2.1.1 - Company Entity

The "Company" entity models the company that manages the electricity plant or that, in general, enters into a contract relating to a plant with a country. It is modeled by its name and a surrogate key attribute.
A company can enter into several contracts.

#### 2.1.1.1 - Dominio degli attributi

- Company_ID {int}
- Company Name {string}

### 2.1.2 - Contract Entity

The "Contract" entity models and defines the aspects relating to the duration of the contract (through the start date and end date attributes), the rates applied and the name associated with the project.
Each contract refers to a single plant.

#### 2.1.2.1 - Dominio degli attributi

- Contract_ID {int}
- Project Name {string}
- Tariff {float}
- Start Date {Date, YYYY-MM-DD}
- End Date {Date, YYYY-MM-DD}

### 2.1.3 - Plant Entity

The "Plant" entity is the core entity of the model. It describes the properties common to all renewable energy production plants.
Energy source level 1 describes the belonging itself to the class of plants that exploit renewable sources to produce electricity.
Energy source level 2 describes the energy source used (solar, wind, etc.).
Energy source level 3 specifies the details of the energy source used (for example biomass and biogas).
Technology describes the implementation technology used.
Capacity describes the plant's ability to produce energy.
Production tracks actual energy production.
This entity specializes in the following five subclasses through the "Energy source level 2" attribute:
- Wind power plant
- Solar system
- Hydroelectric
- Geothermal plant

- Marine Bioenergetic Plant

Of these subclasses, the most detailed are the first two.
In a wind power plant, the capacity of the individual turbines, the height of the hubs, the diameter of the turbines, the turbine model used and the number of turbines present in a plant are tracked.
The solar plant is described through the attributes "Mounting Type" and "CHP" which respectively describe the positioning of the installation (on the ground or on all) and if the plant is CHP or not and therefore if it is only responsible for the production of electricity or if it also produces heat.
A plant can be located in a single place but a place can host several plants.

#### 2.1.3.1 - Attributes domain

- Plant_ID {int}
- Energy Source Level 1 {string}
- Energy Source Level 2 {string}
- Energy Source Level 3 {string}
- Technology {string}
- Capacity {float}
- Production {float}
- Turbine_Hub_Height {float}
- Turbine_Rotor_Diameter {float}
- Turbine_Model {string}
- Turbine_Capacity {float}
- Turbine_Number {int}
- Mounting_Type {string}
- CHP {string, YES or NOT}

### 2.1.4 - Location Entity

"Location" entity describes the place where a plant is settled and in particular the country that manages it.
A place is uniquely identified through the "Municipality Code" attribute which is unique.
"Region" attribute describes the geopolitical subset immediately below the nation. This is named differently from country to country but in this model it is standardized under the same attribute (Region).
For example, the Swiss cantons, the German federal states and the counties of Sweden fall into this attribute.

The municipality and the related municipality code accurately describe the location of an installation.
Address describes the address where the plant can be found.
Latitude and longitude identify the location of the plant on a planetary level.

#### 2.1.4.1 - Dominio degli attributi

- Country {string}
- Regione {string}

- Municipality {string}
- Municipality Code {int}
- Postcode {int}
- Address {string}
- Latitude {real}
- Longitude {real}

# 3 - DIMENSIONAL FACT MODEL

The dataset lends itself to various interpretations and different possibilities of analysis. Surely the main objective of the report is the production of electricity from renewable sources.
This fact is modeled by the DFM reported in fig. 3.1.

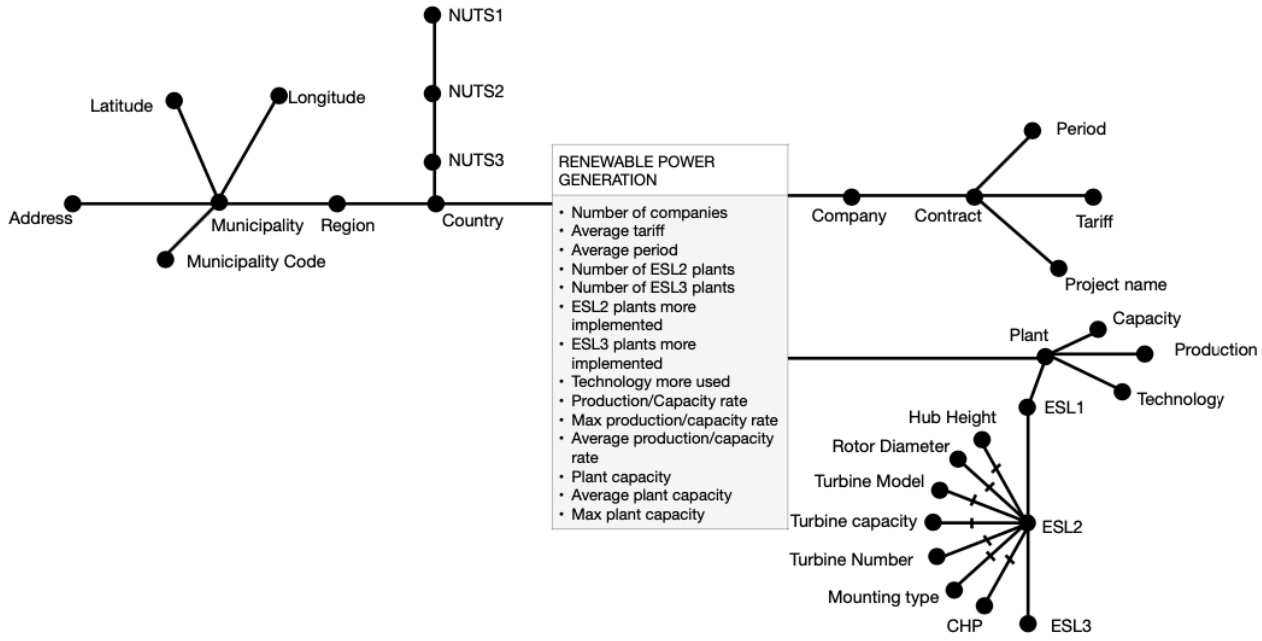



Fig. 3.1 DFM relating to the production of electricity

In this model we find three dimensions: Country, Company and Plant.
Country specifies the place where electricity is produced and is detailed up to the precise information (address and town hall) where the plant is located.
Company models the company that operates the plant. This hierarchy is detailed with the contract that regulates the supply period and the tariff applied.
Finally, Plant provides all the details relating to the plant. This hierarchy details the renewable source used, the capacity and production of the plant.
This hierarchy is characterized by seven optional arcs validated in the case of a wind or solar plant.
For what has been said before, it was considered appropriate to detail the "production of electricity" by identifying four facts that describe the production of wind energy (fig. 3.2), solar (fig. 3.3), bioenergetics (fig. 3.4) and geothermal (fig. 3.5).
Actually these four models are a subset of the model illustrated in figure 3.1 which allow a greater focus on the characteristics of each different type of system.

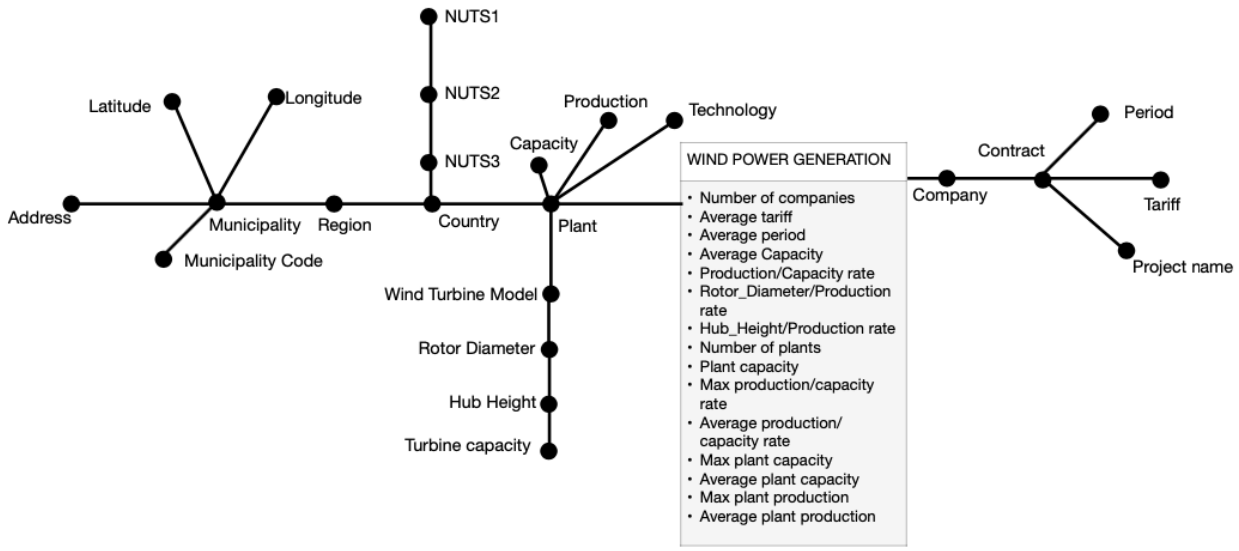



Fig. 3.2 DFM: Wind Power Generation

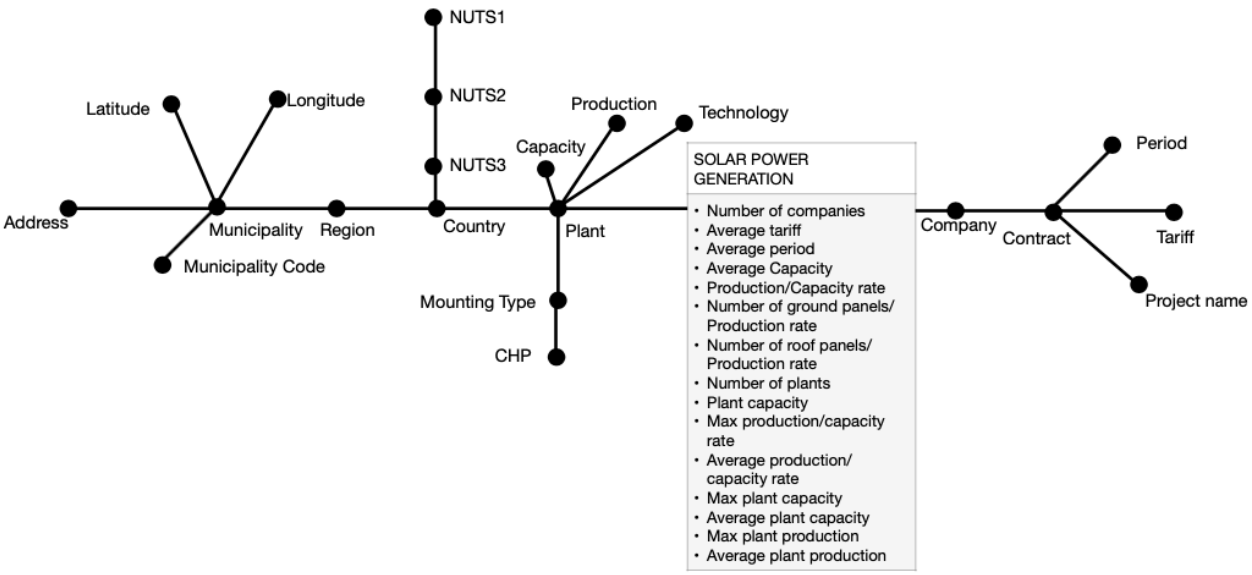



Fig. 3.3 DFM: Solar Power Generation

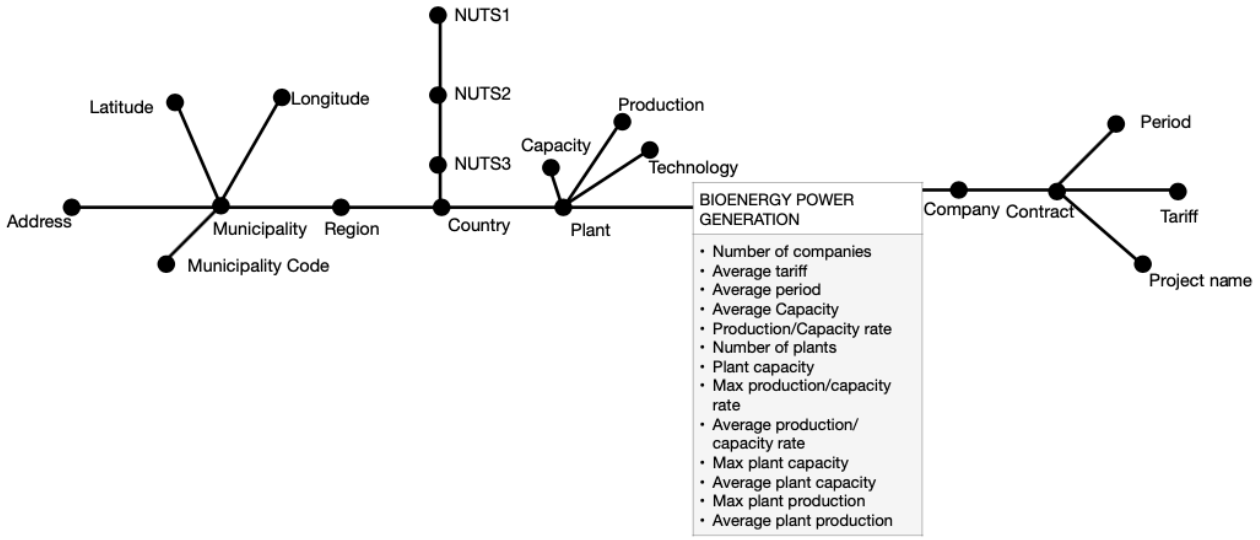



Fig. 3.4 DFM: Bioenergy Power Generation

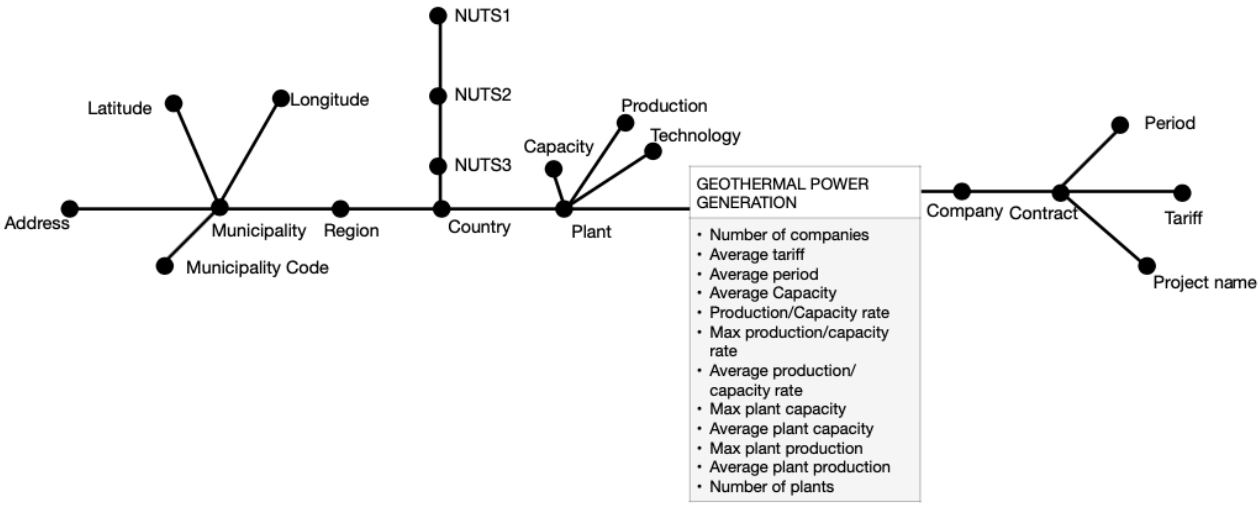



Fig. 3.5 DFM: Geothermal Power Generation

By analyzing the dataset further, this can be interpreted as a valuable source of data for political and commercial analysis.
The same could be used by a head of state to assess how many electricity generation plants are installed in their own country or in other countries of the European community. This aspect is modeled by the DFM shown in figure 3.6 which describes the installation of a plant in one place.

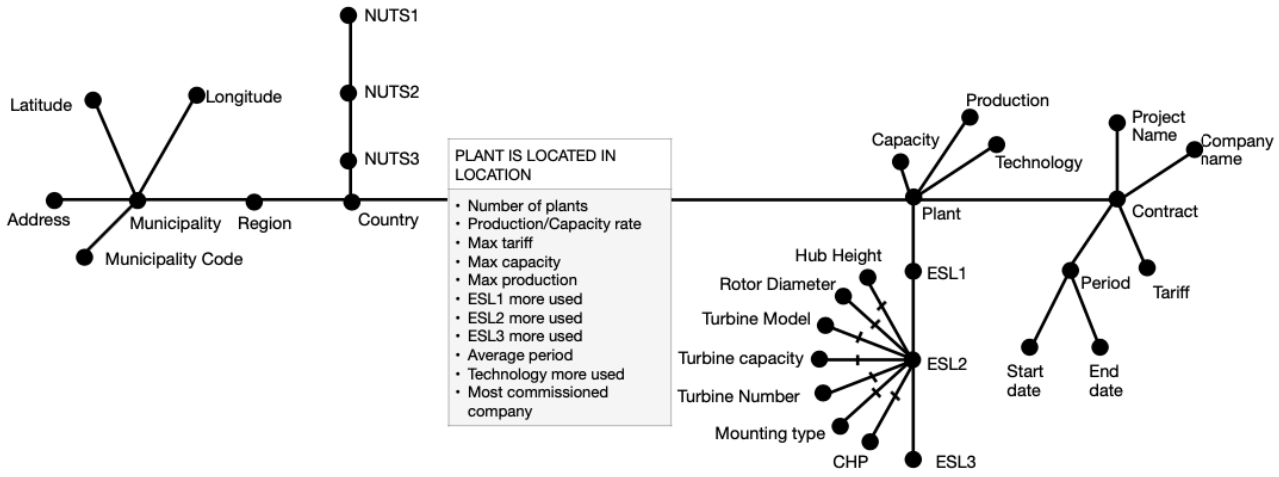



Fig. 3.6 DFM: Plant is located in a place

Similarly, the dataset could be valuable for the CEO of an electricity production company in evaluating the tariffs (and periods) applied by its competitors in the sector. This aspect suggested the modeling of the fact "Contract regulates a plant" (fig. 3.7) in order to be able to trace the company that stipulates the contract and the rates it applies (in relation to the plant treated).

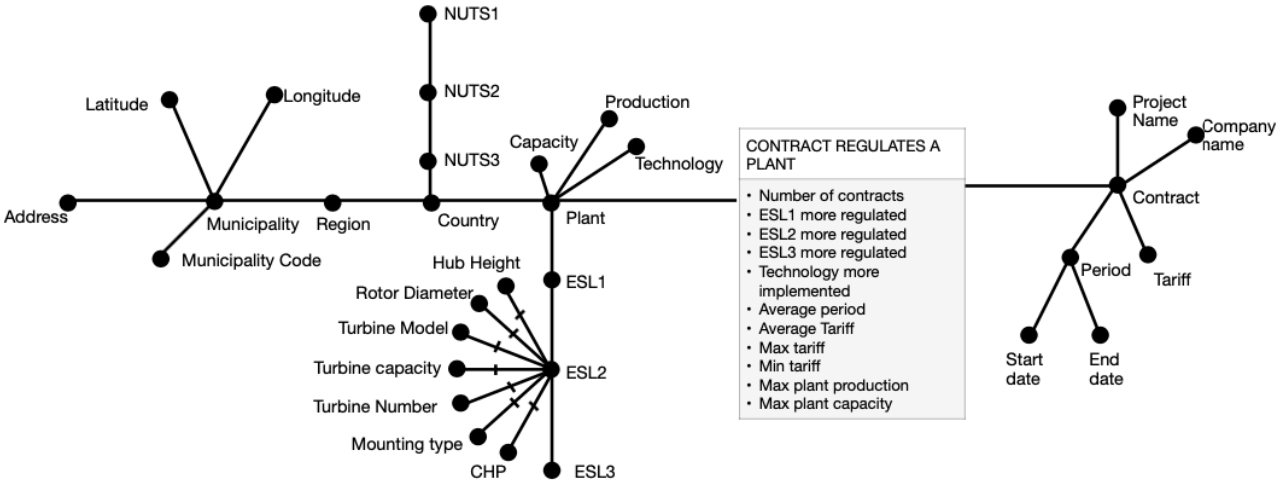



Fig. 3.7 DFM: Contract regulates a plant

## 3.1 - STAR SCHEMAS

The related star / snowflake patterns are shown below.
In fig. 3.1.1 the model relating to the DFM “Renewable power generation”.
In fig. 3.1.2 the model relating to the DFM “Wind power generation”.
In fig. 3.1.3 the model relating to the DFM “Solar power generation”.
In fig. 3.1.4 the model relating to the DFM “Bioenergy power generation”.
In fig. 3.1.5 the model relating to the DFM “Geothermal power generation”.
In fig. 3.1.6 the model relating to the DFM “Contract regulates a plant”.
In fig. 3.1.7 the model relating to the DFM “Plant is located in location”.

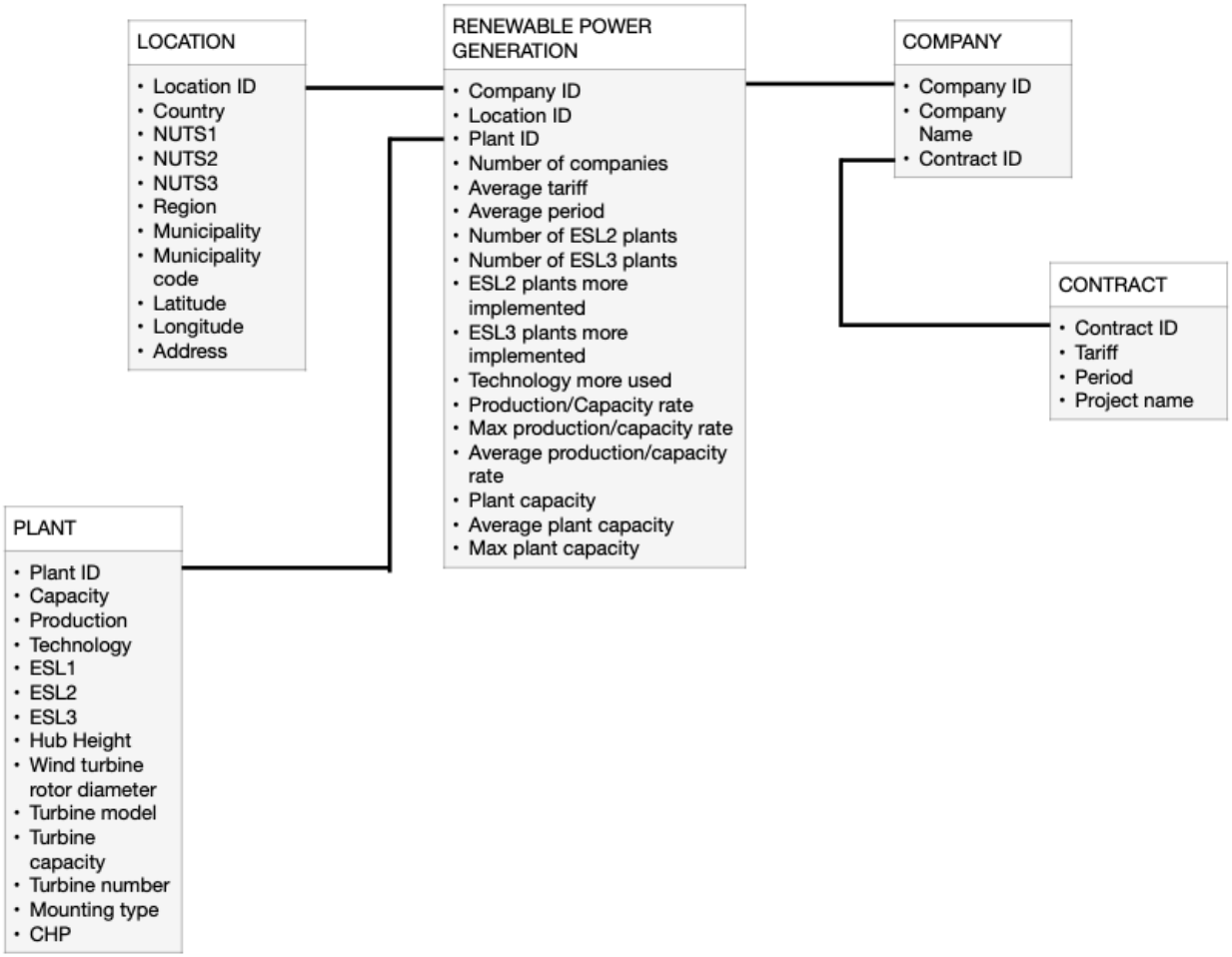



Fig. 3.1.1 Star schema: “Renewable power generation”

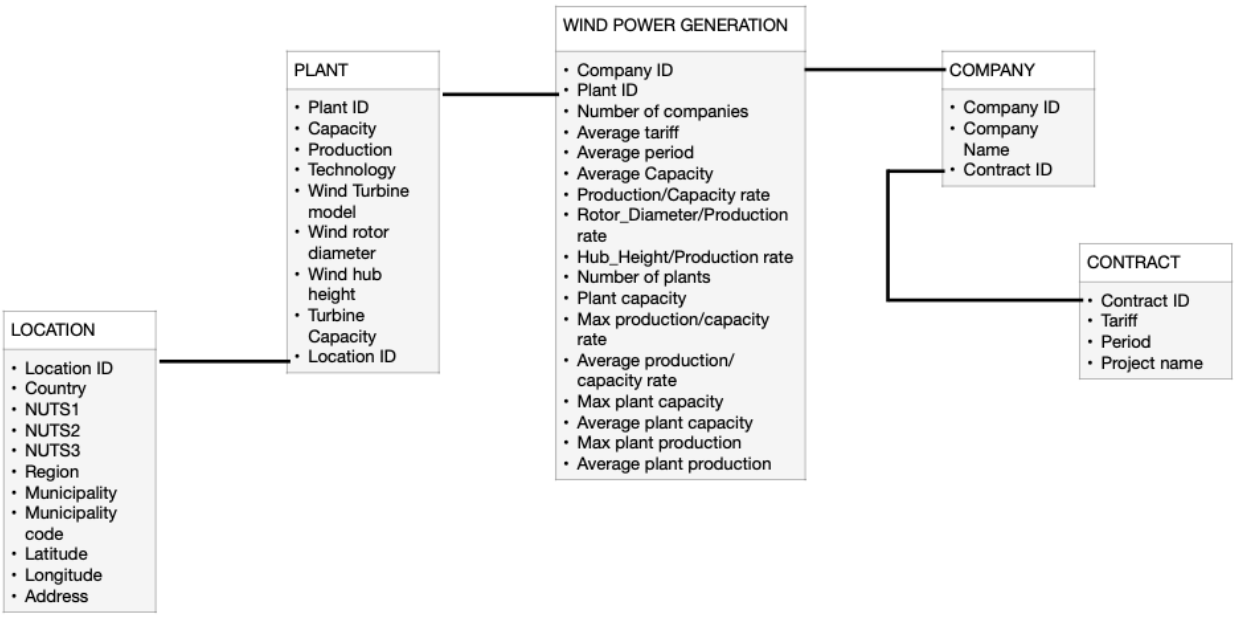



Fig. 3.1.2 Star schema: “Wind power generation”

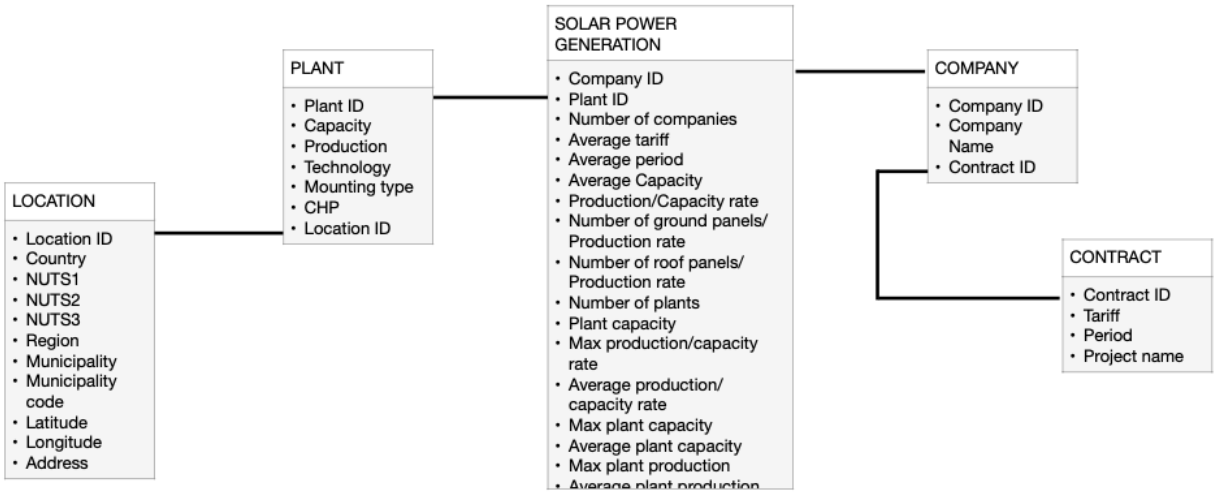



FIg. 3.1.3 Star schema: “Solar power generation”

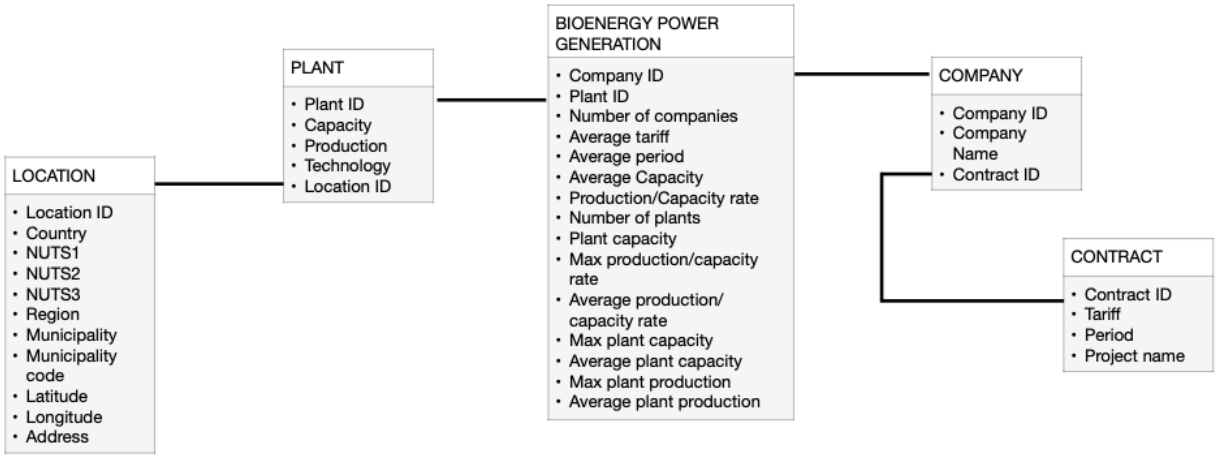



Fig. 3.1.4 Star schema: “Bioenergy power generation”

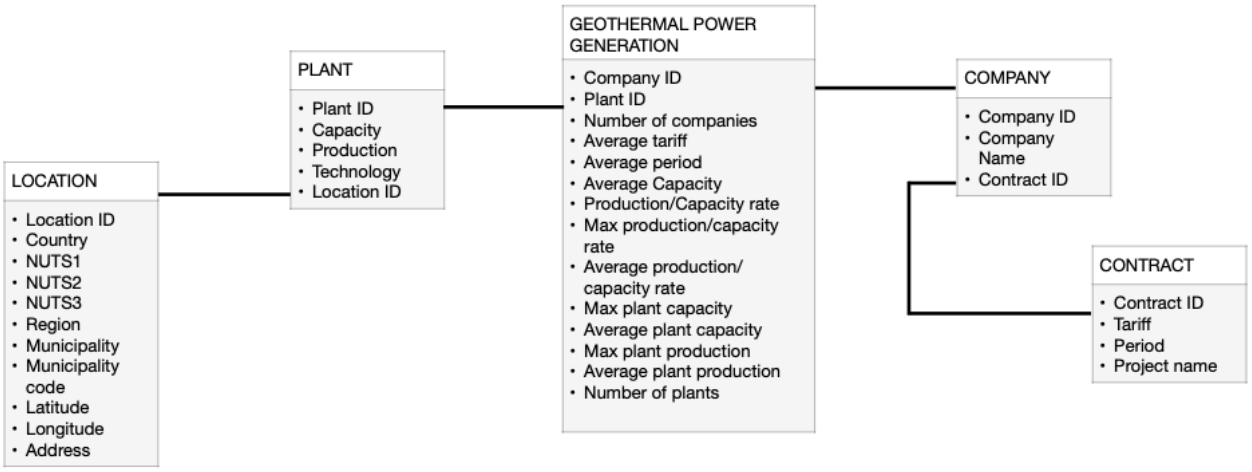



Fig. 3.1.5 Star schema: “Geothermal power generation”

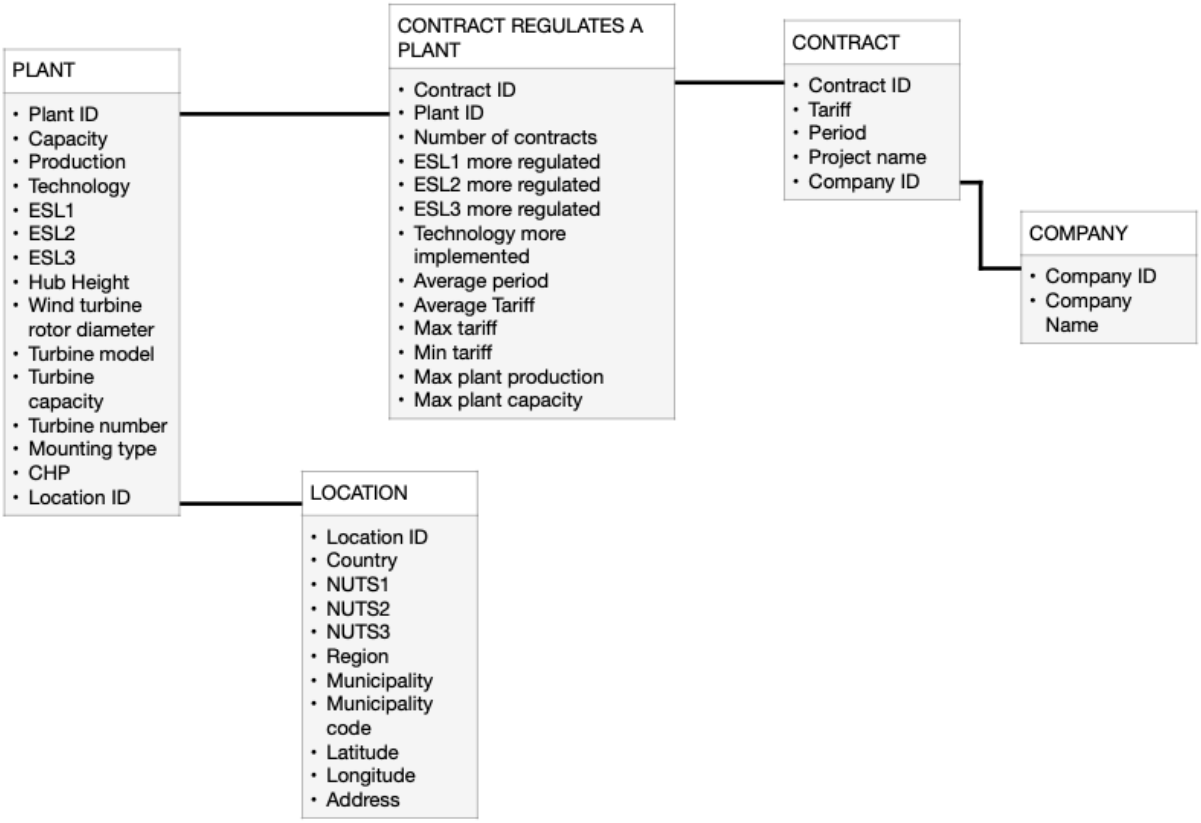



Fig. 3.1.6 Star schema: “Contract regulates a plant”

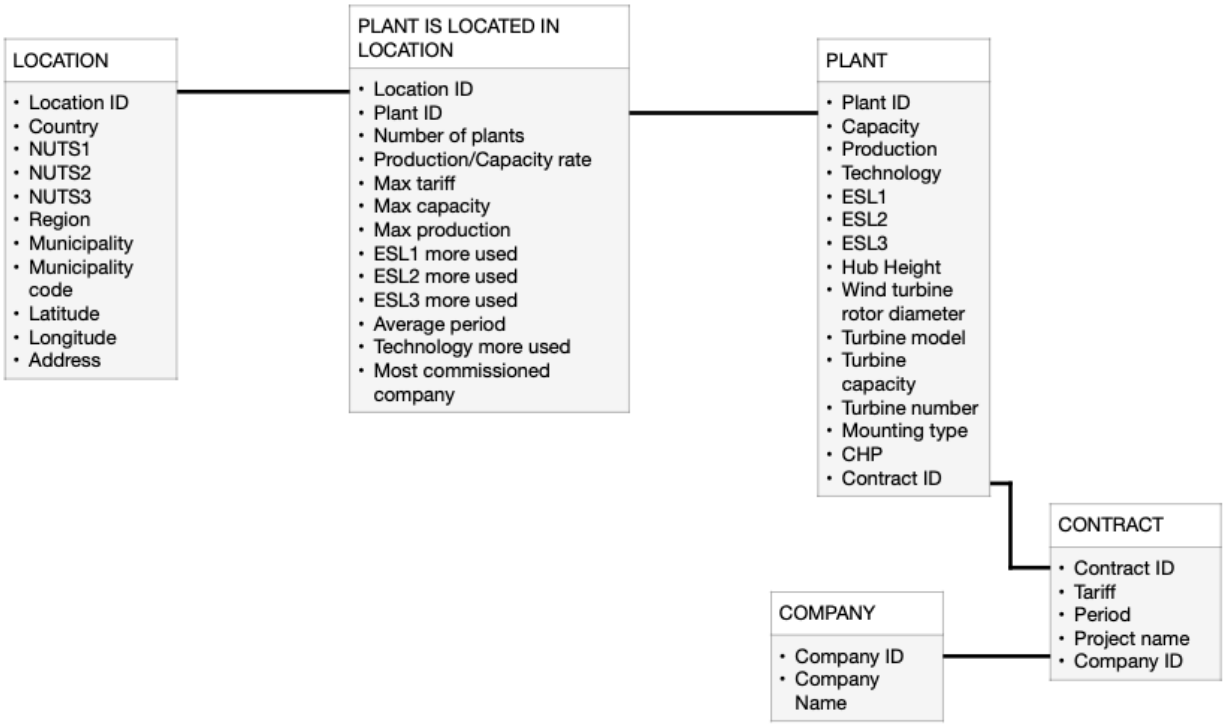



Fig. 3.1.7 Star schema: “Plant is located in location”

# 4 - USE CASE

In this chapter the most relevant use cases (fig. 4.1) of the three identified stackholders will be analyzed.
The dataset can be used by a European Chief Energy Officer to assess the extent to which renewable energy sources are used within the community.
Similarly, a president of state can use the same dataset to analyze the number of sustainable plants used within his own nation, possibly comparing the results with those of neighboring nations.
A CEO of a company can use this data to analyze the offerings of the companies in the sector and understand what strategic choices can be made to be competitive in the market.

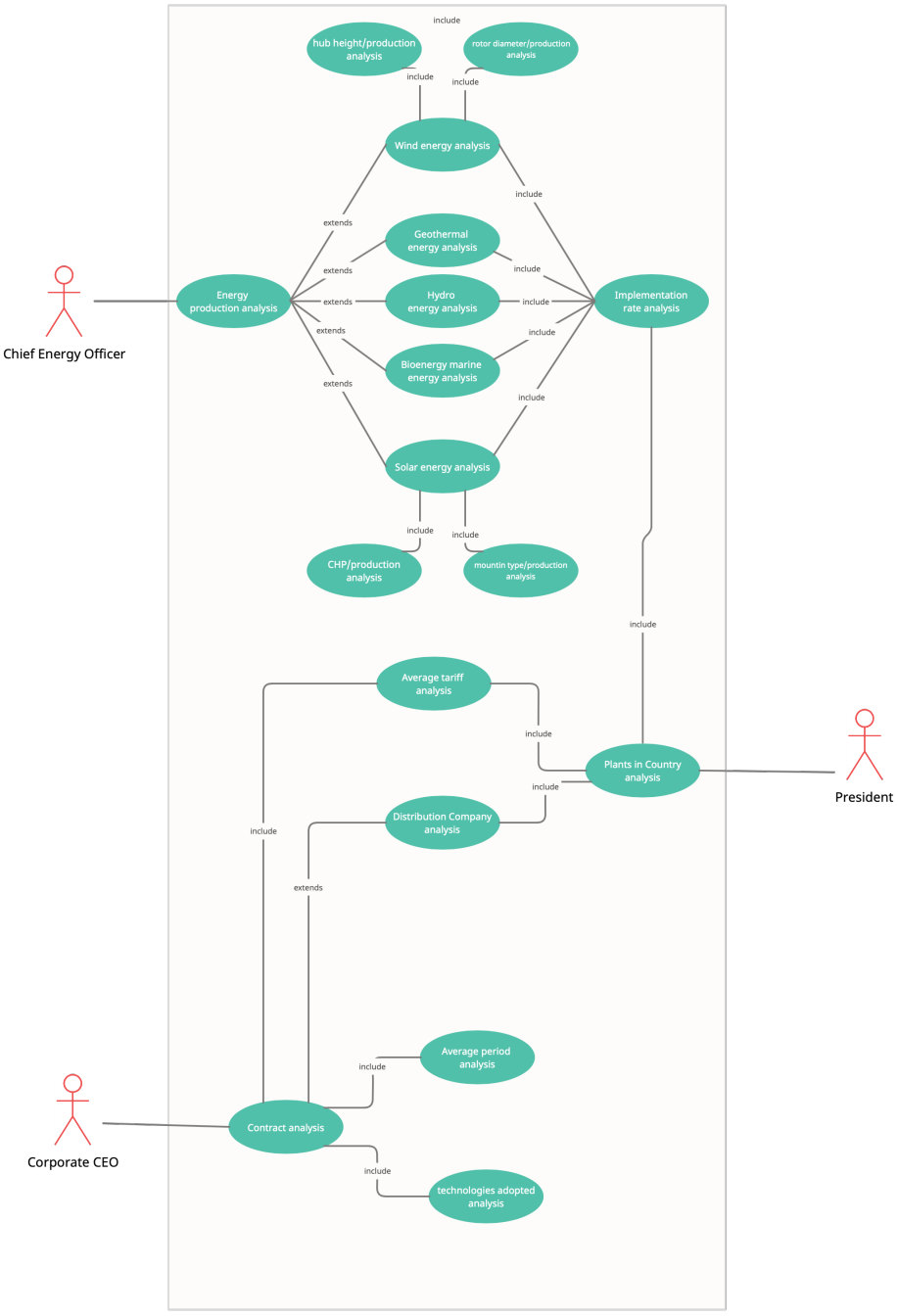



Fig. 4.1 Use case diagram

## 4.1 - CHIEF ENERGY OFFICER

The main objective of the Chief Energy Officer is to analyze the different renewable sources used in Europe. The use case "Energy production analysis" is detailed, therefore, with the analysis of the individual types of production. These in turn include different types of analyzes depending on the type of renewable source considered. For example, the production of wind energy can be further analyzed by evaluating how some technological choices (height of the hubs, diameter of the rotors) can influence the production of electricity. Similarly, the analysis on photovoltaic systems can be deepened by assessing the production levels in relation to the type of panel used and its positioning.

## 4.2 - PRESIDENT OF COUNTRY

A country president may be interested in analyzing the utilization rate of facilities within their country and in other countries.
This type of analysis can be further detailed by evaluating the companies that manage several plants in the area and the average rate requested.
The analysis of the plants in the area includes the analysis relating to the implementation rate by type of renewable sources.

## 4.3 - CORPORATE CEO

Finally, a company's CEO can leverage the data to perform useful analytics at the business strategic level.
By analyzing the contracts proposed by the various companies, he can evaluate his positioning on the market and undertake any business choices to improve the position of his company.
The analysis of the contracts provided includes and takes into account the analysis relating to the technologies used, the average periods of the contracts and the tariffs applied by other companies.
Surely this type of analysis can be extended to the analysis of companies in the sector.

## 4.4 GOALS/STACKHOLDERS MODEL

At the end of this chapter, consistently with what has been said above, figure 4.4.1 shows the associated goals / stackholders model.

| | Wind Energy Analysis | Solar Energy Analysis | Hidro Energy Analysis | Geoth. Energy Analysis | Bioen.Marine Analysis | AVG tariff analysis | Distribution Company Analysis | AVG Period Analysis | Technology Analysis |
|---|---|---|---|---|---|---|---|---|---|
| Chief Energy Officer | ✅ | ✅ | ✅ | ✅ | ✅ | ❌ | ❌ | ❌ | ❌ |
| President of a Country | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | ❌ | ❌ |
| CEO of a Company | ❌ | ❌ | ❌ | ❌ | ❌ | ✅ | ✅ | ✅ | ✅ |

Fig. 4.4.1 Goals/Stackholders model

# 5 - DATA ANALYSIS

## 5.1 - NULL VALUES ANALYSIS

### 5.1.1 Czechia

As you can see in figures 5.1.1.1 and 5.1.1.2, the values with the highest number of nulls are Energy Source Level 3 and owner. This creates serious gaps in information that do not allow the correct association between the plant and the company that manages it.

| CZ Repubblica Ceca | rows | 31604 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 30308 | 95,90 |
| technology | 1300 | 4,11 |
| data_source | 0 | 0,00 |
| nuts_1_region | 253 | 0,80 |
| nuts_2_region | 253 | 0,80 |
| nuts_3_region | 253 | 0,80 |
| lon | 253 | 0,80 |
| lat | 253 | 0,80 |
| municipality | 1303 | 4,12 |
| postcode | 0 | 0,00 |
| region | 11 | 0,03 |
| locality | 10 | 0,03 |
| owner | 22561 | 71,39 |
| site_name | 0 | 0,00 |

Fig. 5.1.1.1 Table of null values in CZ Dataset

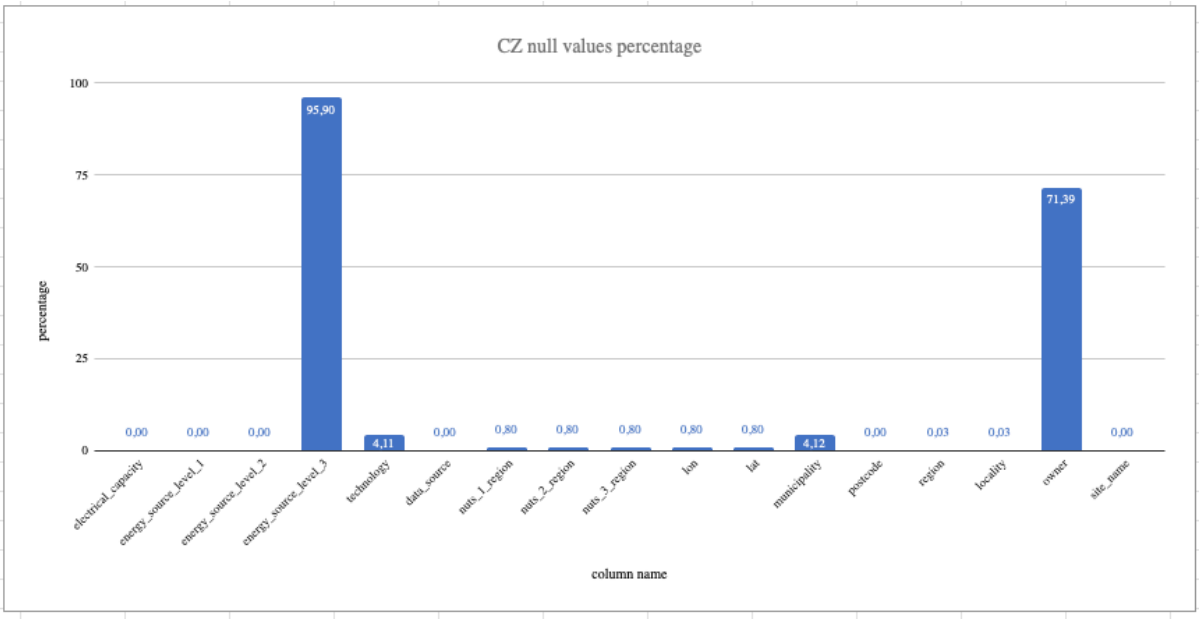



Fig. 5.1.1.2 Bar chart of percentages of null values in CZ

### 5.1.2 Switzerland

As shown below (fig. 5.1.2.1 and fig. 5.1.2.2) the fields with multiple null values are Energy Source Level 3 and Address.

The lack of the address does not allow the precise identification of the location of the system, in this case we can stop at the single identification of the town hall.

| CH Svizzera | rows | 12694 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 12417 | 97,82 |
| technology | 131 | 1,03 |
| data_source | 0 | 0,00 |
| nuts_1_region | 0 | 0,00 |
| nuts_2_region | 0 | 0,00 |
| nuts_3_region | 0 | 0,00 |
| lon | 0 | 0,00 |
| lat | 0 | 0,00 |
| municipality | 0 | 0,00 |
| municipality_code | 0 | 0,00 |
| postcode | 0 | 0,00 |
| address | 8777 | 69,14 |
| canton | 0 | 0,00 |
| commissioning_date | 0 | 0,00 |
| contract_period_end | 0 | 0,00 |
| company | 0 | 0,00 |
| tariff | 0 | 0,00 |
| project_name | 0 | 0,00 |
| production | 0 | 0,00 |

Fig. 5.1.2.1 Table of null values in CH Dataset

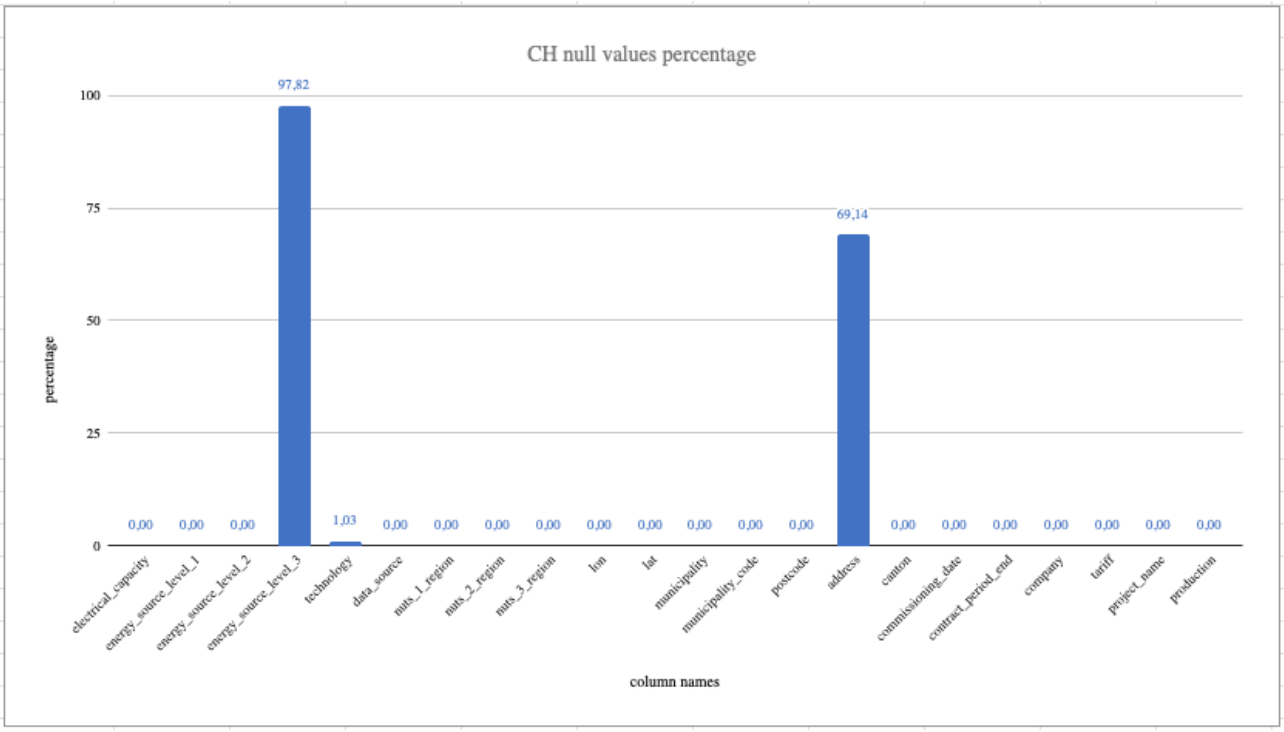



Fig. 5.1.2.2 Bar chart of percentages of null values in CH

### 5.1.3 Denmark

Denmark is the nation that provides the dataset with the highest number of null values (fig. 5.1.3.1 and fig. 5.1.3.2).
This dataset suffers from various information gaps.
Not having information about the municipality, it will not be possible to understand in detail in which place the plant is located. A possible analysis can only use the NUTS values to identify the location but these are less easy to interpret.
The plants are also poorly detailed due to the null values present in the hub height, rotor diameter, and model.
Furthermore, there is a lack of association between the system and the manufacturer of the system.

| DK Danimarca | rows | 84353 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 84353 | 100,00 |
| technology | 0 | 0,00 |
| data_source | 0 | 0,00 |
| nuts_1_region | 4 | 0,00 |
| nuts_2_region | 4 | 0,00 |
| nuts_3_region | 4 | 0,00 |
| lon | 596 | 0,71 |
| lat | 596 | 0,71 |
| municipality | 78152 | 92,65 |
| municipality_code | 78148 | 92,64 |
| postcode | 6205 | 7,36 |
| address | 79618 | 94,39 |
| commissioning_date | 41 | 0,05 |
| hub_height | 78148 | 92,64 |
| rotor_diameter | 78148 | 92,64 |
| model | 78153 | 92,65 |
| grsn_id | 78148 | 92,64 |
| dso | 6205 | 7,36 |
| manufacturer | 78277 | 92,80 |

Fig. 5.1.3.1 Table of null values in DK Dataset

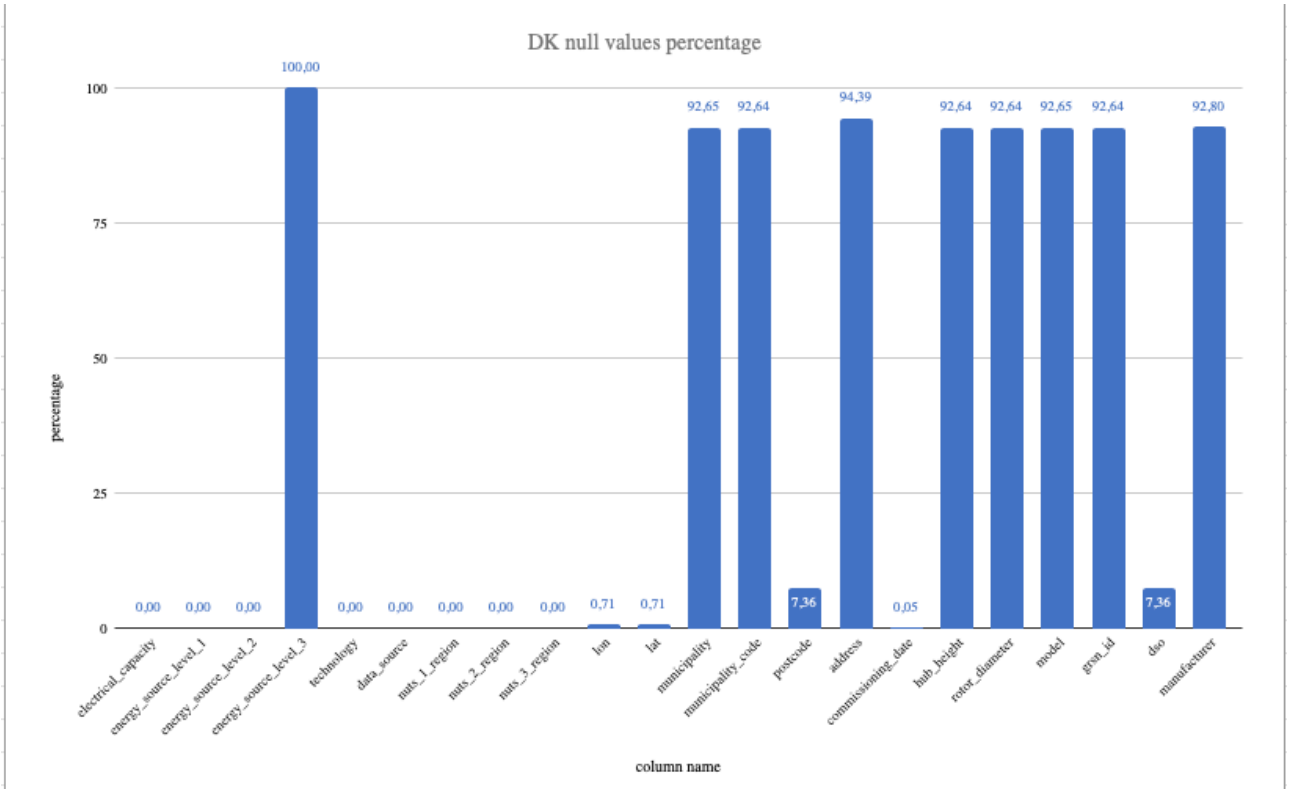



Fig. 5.1.3.2 Bar chart of percentages of null values in DK

### 5.1.4 France

In the French dataset, in addition to the null values present in the Energy source level 3 attribute (as happens in all datasets), there are about a third of null values for geographic and temporal information (fig. 5.1.4.1 and fig. 5.1. 4.2).
The lack of the start date and the end date do not allow to fully describe the contract that regulates it.

| FR Francia | rows | 56097 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 55585 | 99,09 |
| technology | 515 | 0,92 |
| data_source | 0 | 0,00 |
| nuts_1_region | 236 | 0,42 |
| nuts_2_region | 236 | 0,42 |
| nuts_3_region | 236 | 0,42 |
| lon | 241 | 0,43 |
| lat | 241 | 0,43 |
| municipality | 200 | 0,36 |
| municipality_code | 405 | 0,72 |
| region | 15310 | 27,29 |
| region_code | 15310 | 27,29 |
| municipality_group | 15537 | 27,70 |
| municipality_group_code | 15536 | 27,69 |
| departement | 15536 | 27,69 |
| departement_code | 15536 | 27,69 |
| commissioning_date | 15429 | 27,50 |
| connection_date | 15373 | 27,40 |
| disconnection_date | 56087 | 99,98 |
| number_of_installations | 0 | 0,00 |
| site_name | 21367 | 38,09 |
| IRIS_code | 26300 | 46,88 |
| EIC_code | 29291 | 52,21 |
| as_of_year | 0 | 0,00 |

Fig. 5.1.4.1 Table of null values in FR Dataset

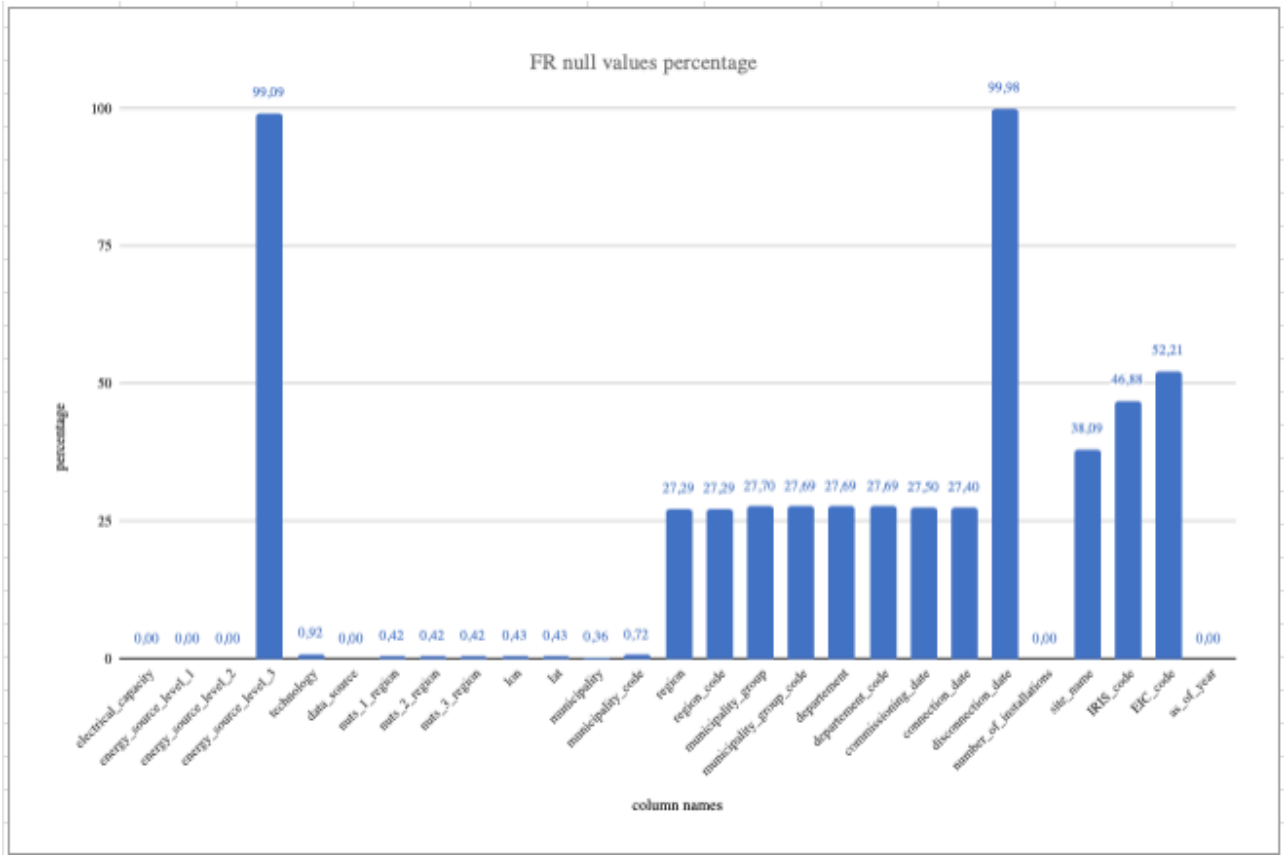



Fig. 5.1.4.2 Bar chart of percentages of null values in FR

### 5.1.5 - Poland

In the dataset provided by Poland there are no serious gaps in information other than those relating to Energy Source Level 3.
The results of the analysis can be seen in figures 5.1.5.1 and 5.1.5.2.

| PL Polonia | rows | 3451 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 3082 | 89,31 |
| technology | 369 | 10,69 |
| data_source | 0 | 0,00 |
| nuts_1_region | 0 | 0,00 |
| nuts_2_region | 0 | 0,00 |
| nuts_3_region | 0 | 0,00 |
| region | 0 | 0,00 |
| district | 0 | 0,00 |
| URE_id | 0 | 0,00 |
| as_of_year | 0 | 0,00 |

Fig. 5.1.5.1 Table of null values in PL Dataset

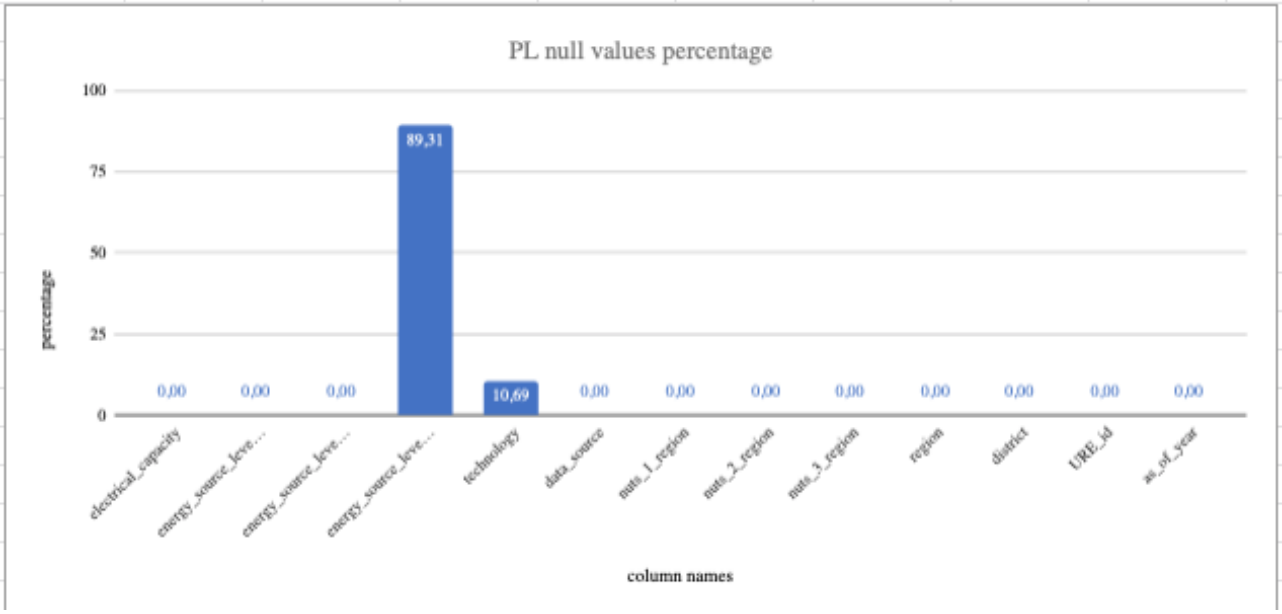



Fig. 5.1.5.2 Bar chart of percentages of null values in PL

### 5.1.6 - United Kingdom

As reported in the analysis of France, although the dataset provides for the detail of some technical aspects of the system such as the type of photovoltaic panel and the point in which it is positioned, these fields are not adequately corroborated.
The result of the analysis can be seen in figures 5.1.6.1 and 5.1.6.2.

| UK Regno unito | rows | 2620 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 2049 | 78,21 |
| technology | 339 | 12,94 |
| data_source | 0 | 0,00 |
| nuts_1_region | 2 | 0,08 |
| nuts_2_region | 2 | 0,08 |
| nuts_3_region | 2 | 0,08 |
| lon | 2 | 0,08 |
| lat | 2 | 0,08 |
| municipality | 2 | 0,08 |
| postcode | 1420 | 54,20 |
| address | 7 | 0,27 |
| region | 1 | 0,04 |
| country | 1 | 0,04 |
| commissioning_date | 0 | 0,00 |
| solar_mounting_type | 1491 | 56,91 |
| chp | 2055 | 78,44 |
| capacity_individual_turbine | 1840 | 70,23 |
| number_of_turbines | 1840 | 70,23 |
| site_name | 1 | 0,04 |
| uk_beis_id | 2 | 0,08 |
| operator | 40 | 1,53 |
| comment | 2617 | 99,89 |

Fig. 5.1.6.1 Table of null values in UK Dataset

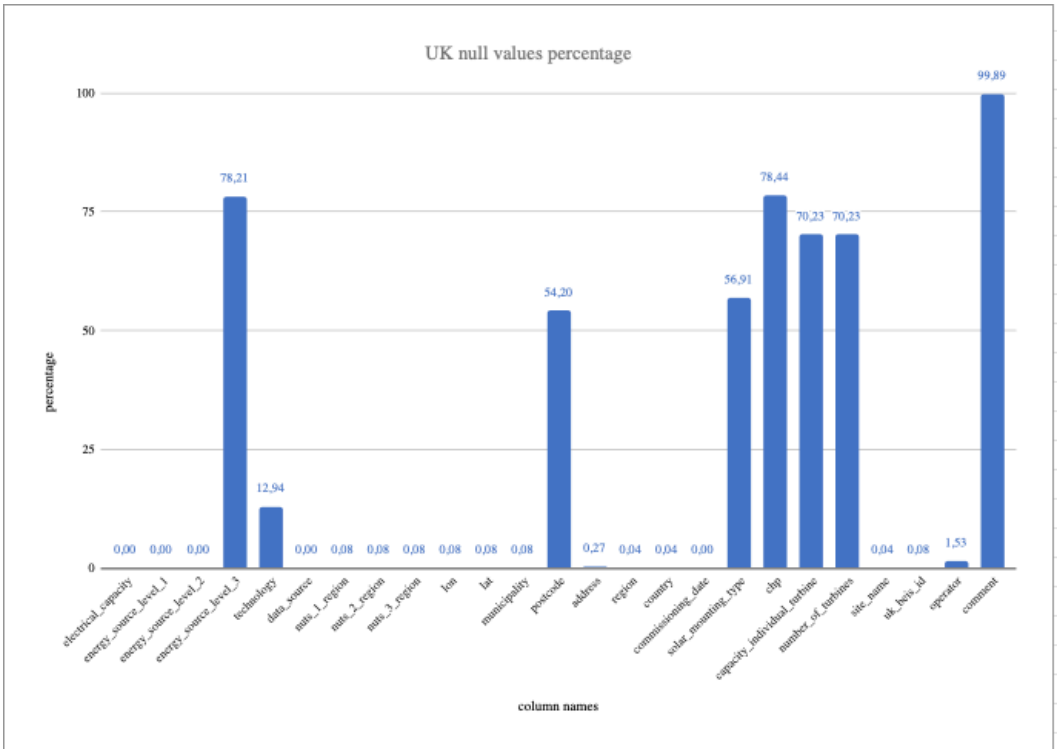



Fig. 5.1.6.2 Bar chart of percentages of null values in UK

### 5.1.7 - Sweden

Sweden provides a relatively complete dataset (fig. 5.1.7.1 and fig. 5.1.7.2). The only information with a high rate of null values is the Commissioning date. This does not allow to identify the times that characterize the contract that regulates the plant.

| SE Svezia | rows | 5529 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 5529 | 100,00 |
| technology | 0 | 0,00 |
| data_source | 0 | 0,00 |
| nuts_1_region | 0 | 0,00 |
| nuts_2_region | 0 | 0,00 |
| nuts_3_region | 0 | 0,00 |
| lon | 0 | 0,00 |
| lat | 0 | 0,00 |
| municipality | 0 | 0,00 |
| county | 0 | 0,00 |
| commissioning_date | 1308 | 23,66 |
| site_name | 1 | 0,02 |
| se_vindbrukskollen_id_null | 0 | 0,00 |
| manufacturer | 1440 | 26,04 |

Fig. 5.1.7.1 Table of null values in SE Dataset

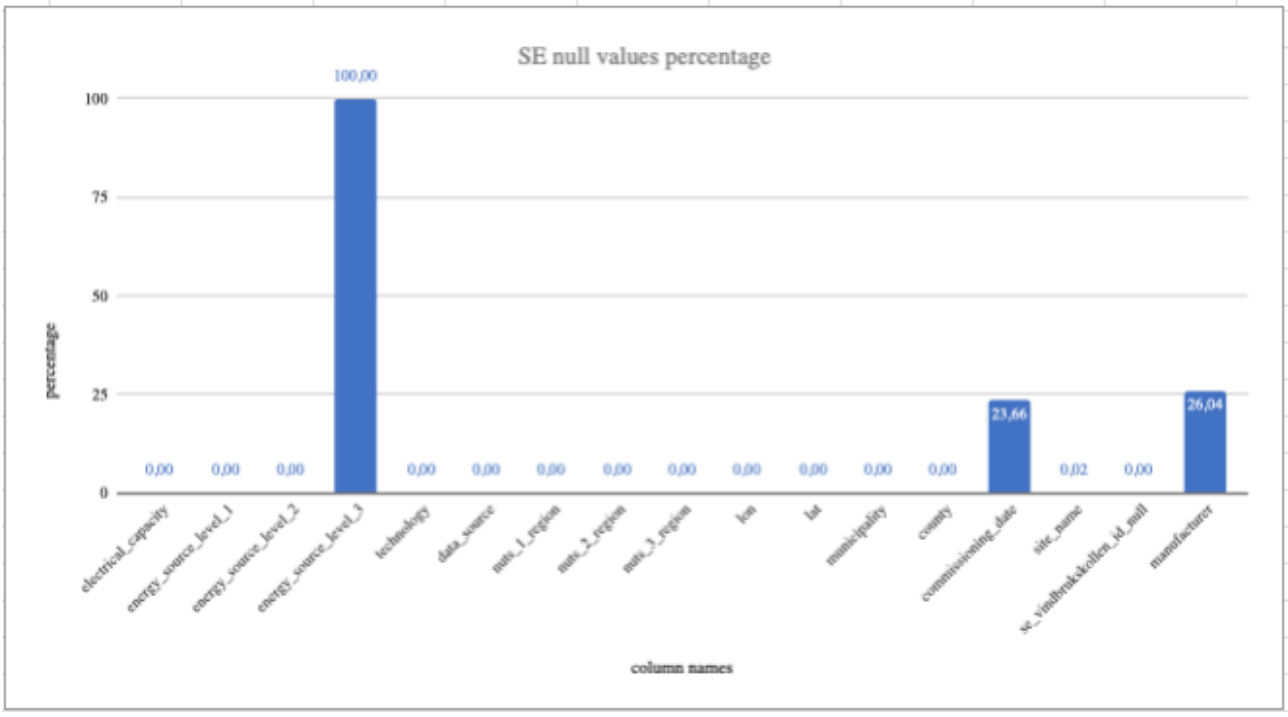



Fig. 5.1.7.2 Bar chart of percentages of null values in SE

### 5.1.8 - Germany

The German dataset, as can be seen in figure 5.1.8.1 and in figure 5.1.8.2, does not allow a correct association between plant and site for 60% of the plants.
Furthermore, it does not allow in any way to identify the time window in which the plant is regulated by a contract (due to 100% of the null values of "decomissioning date" field).
Again, energy source level 3 has a high null rate.
It should be noted that Germany is the contributing country with the highest number of data in the entire dataset. Any shortcomings in this dataset will have repercussions on the overall dataset.

| DE Germania | rows | 1768744 |
|---|---|---|
| column name | null values | percentage |
| electrical_capacity | 0 | 0,00 |
| energy_source_level_1 | 0 | 0,00 |
| energy_source_level_2 | 0 | 0,00 |
| energy_source_level_3 | 1754345 | 99,19 |
| technology | 0 | 0,00 |
| data_source | 0 | 0,00 |
| nuts_1_region | 675 | 0,04 |
| nuts_2_region | 675 | 0,04 |
| nuts_3_region | 675 | 0,04 |
| lon | 1286 | 0,07 |
| lat | 1286 | 0,07 |
| municipality | 1073177 | 60,67 |
| municipality code | 33859 | 1,91 |
| postcode | 15 | 0,00 |
| address | 1764641 | 99,77 |
| federal_state | 5 | 0,00 |
| commissioning_date | 0 | 0,00 |
| decommissioning_date | 1768744 | 100,00 |
| voltage_level | 33141 | 1,87 |
| eeg_id | 1768226 | 99,97 |
| dso | 33140 | 1,87 |
| dso_id | 33899 | 1,92 |
| tso | 33886 | 1,92 |

Fig. 5.1.8.1 Table of null values in DE Dataset

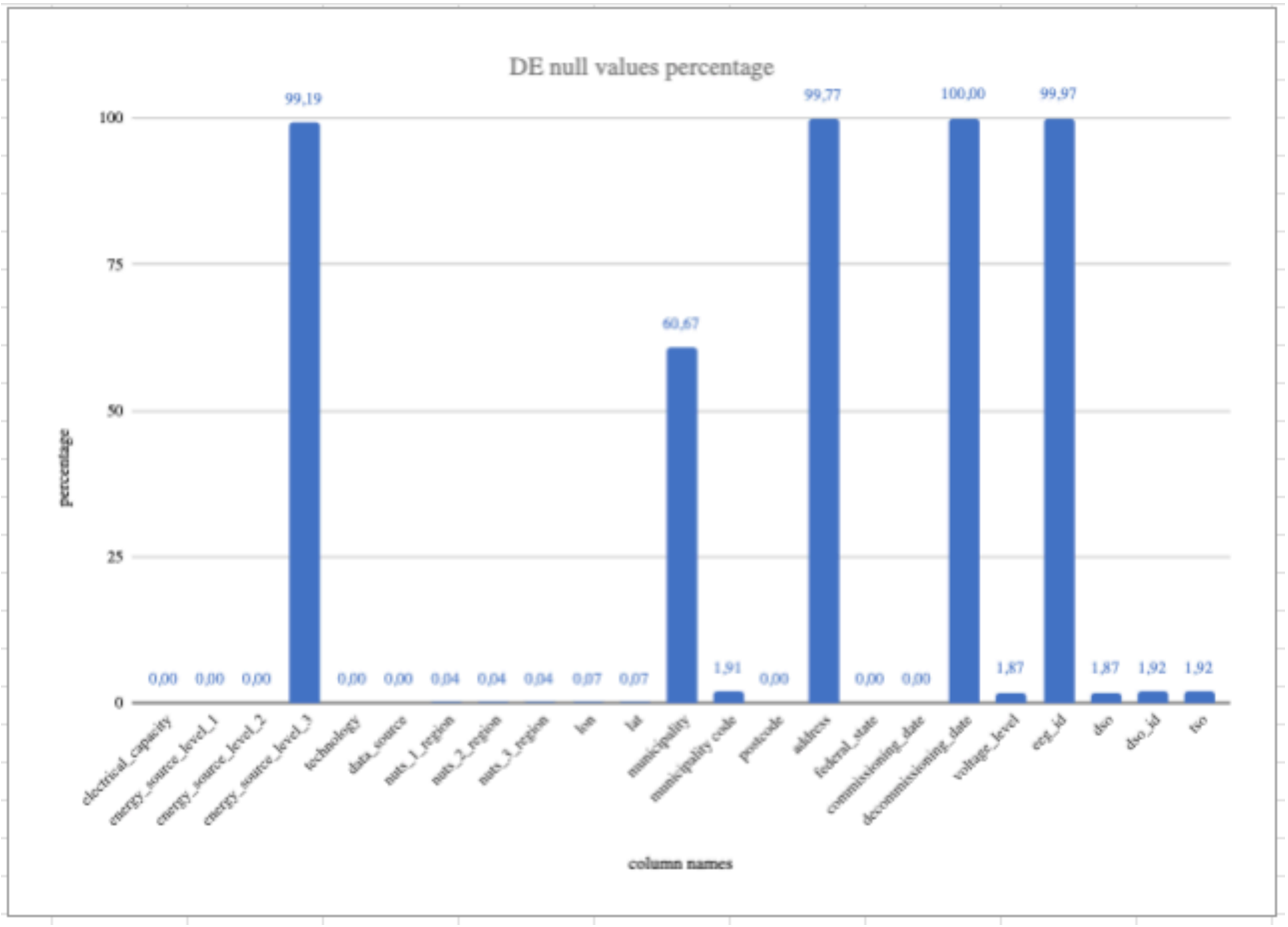
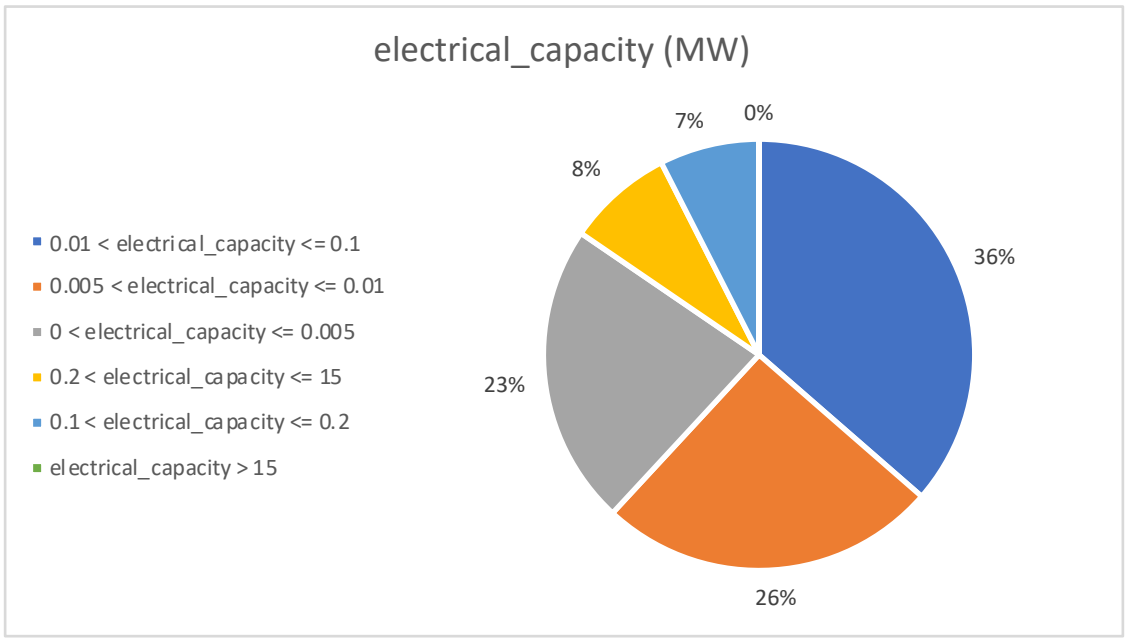


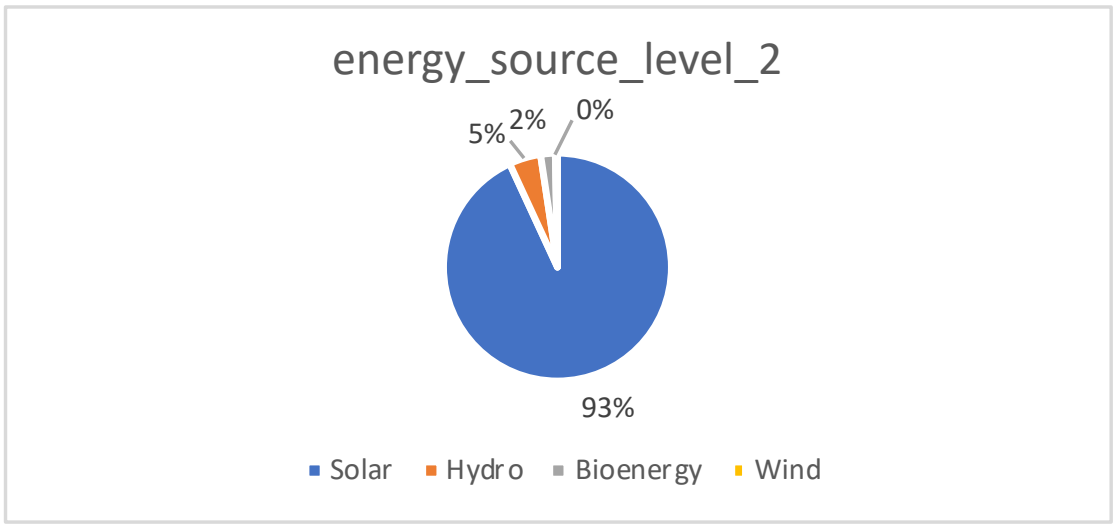

Fig. 5.1.8.2 Bar chart of percentages of null values in DE

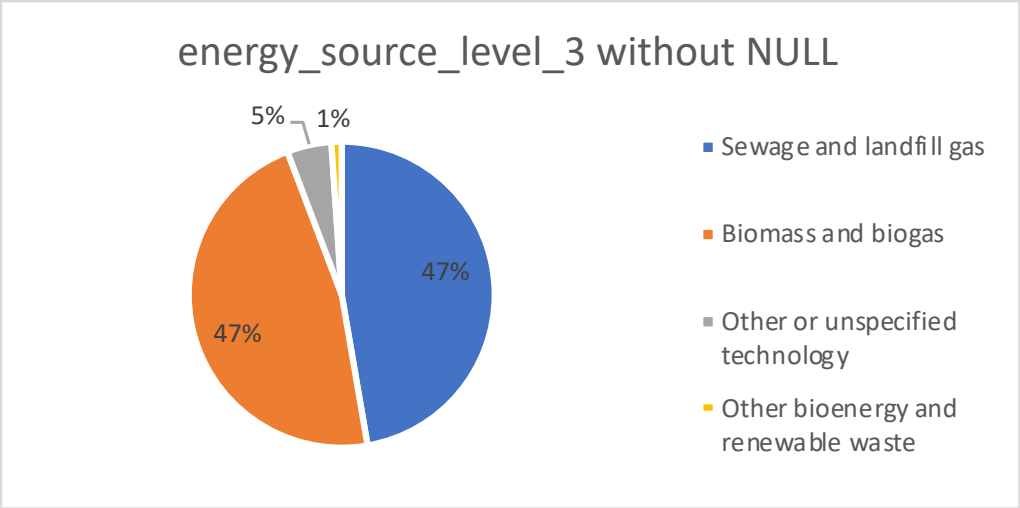
## 5.2 - DOMAIN ANALYSIS

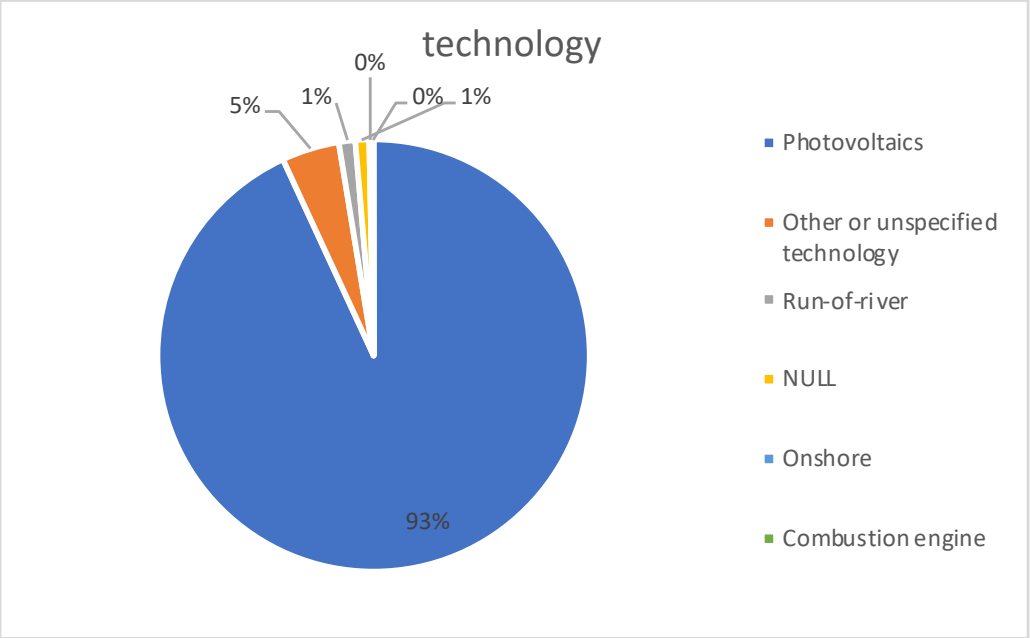
The details of the domain analysis are available at the following link: https://docs.google.com/spreadsheets/d/1oLaxNsv6oqkJ_lNd87jVQdUxKkFHQgoQ/edit?usp=sharing&ouid=107988850620628418604&rtpof=true&sd=true

### 5.2.1 - Switzerland

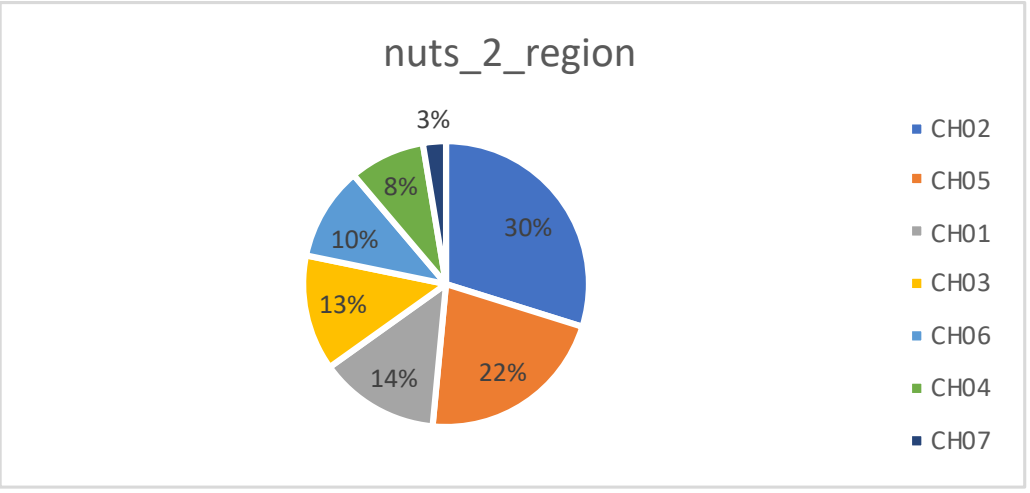


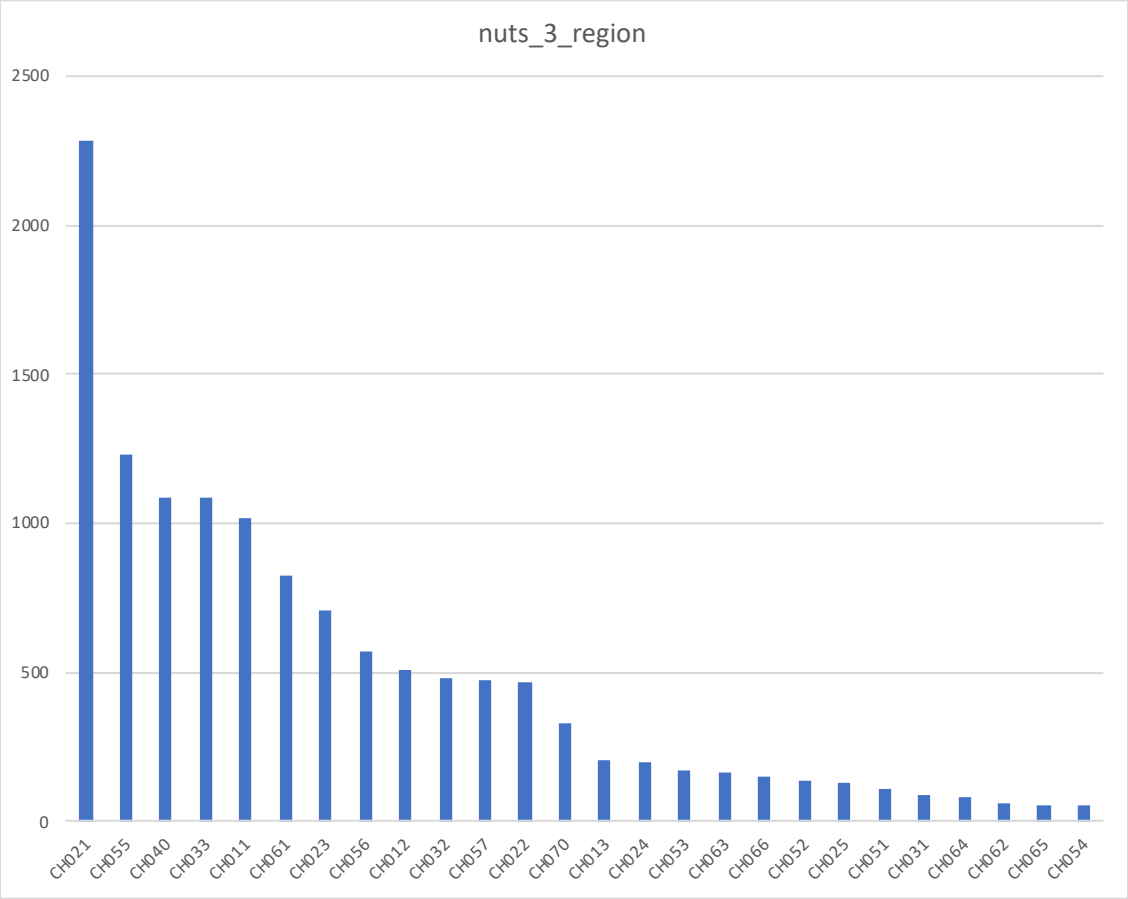



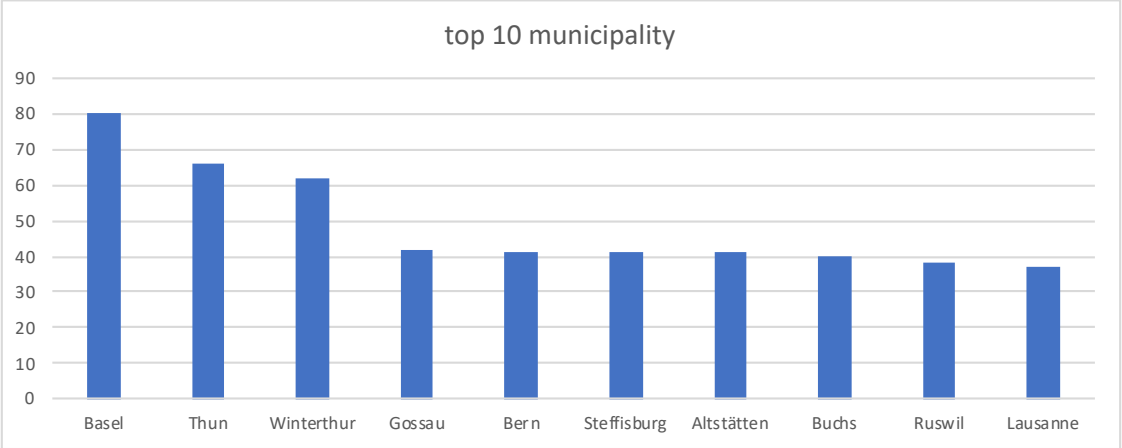


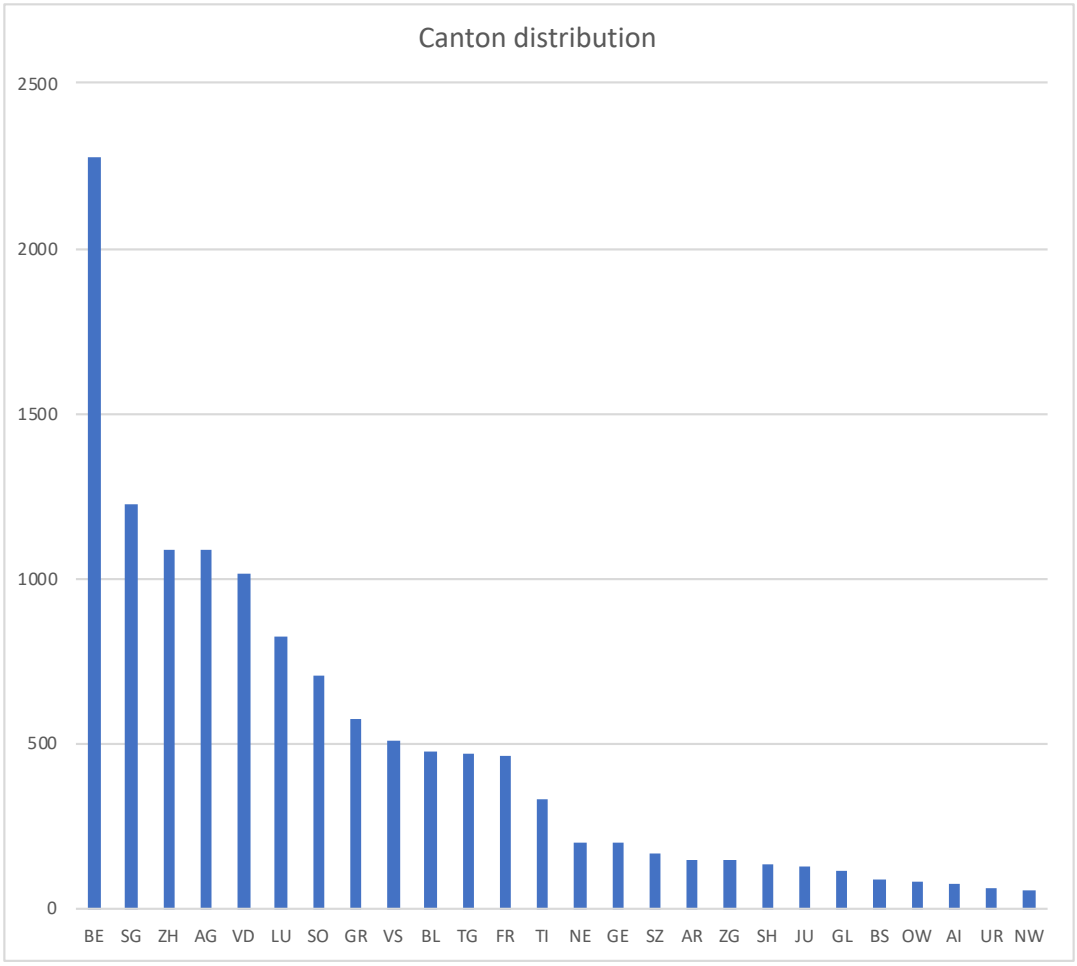



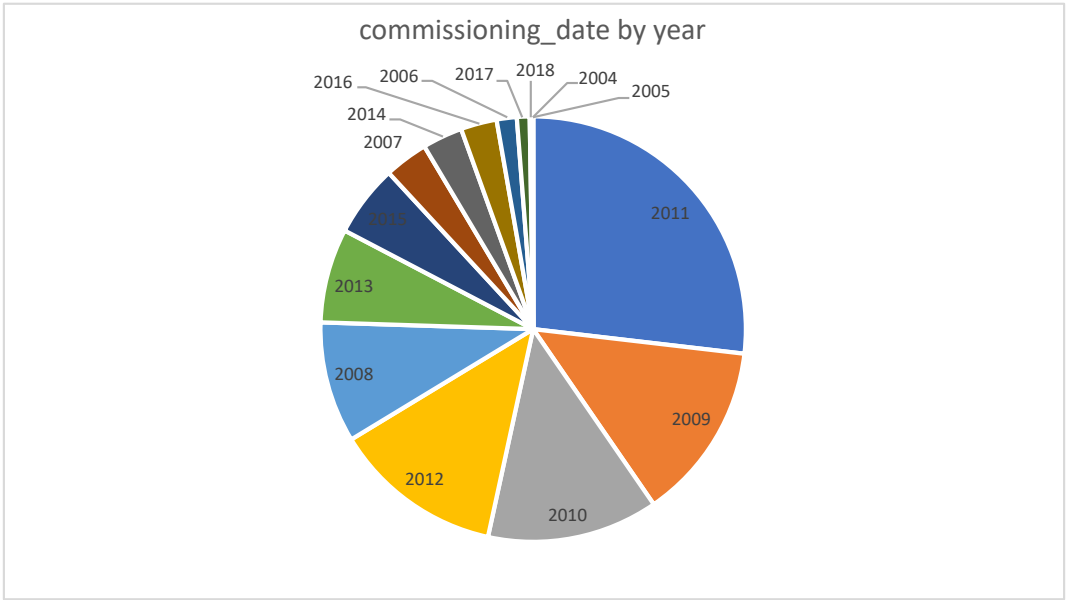

| nuts_1_region | count |
|---|---|
| CH0 | 12694 |

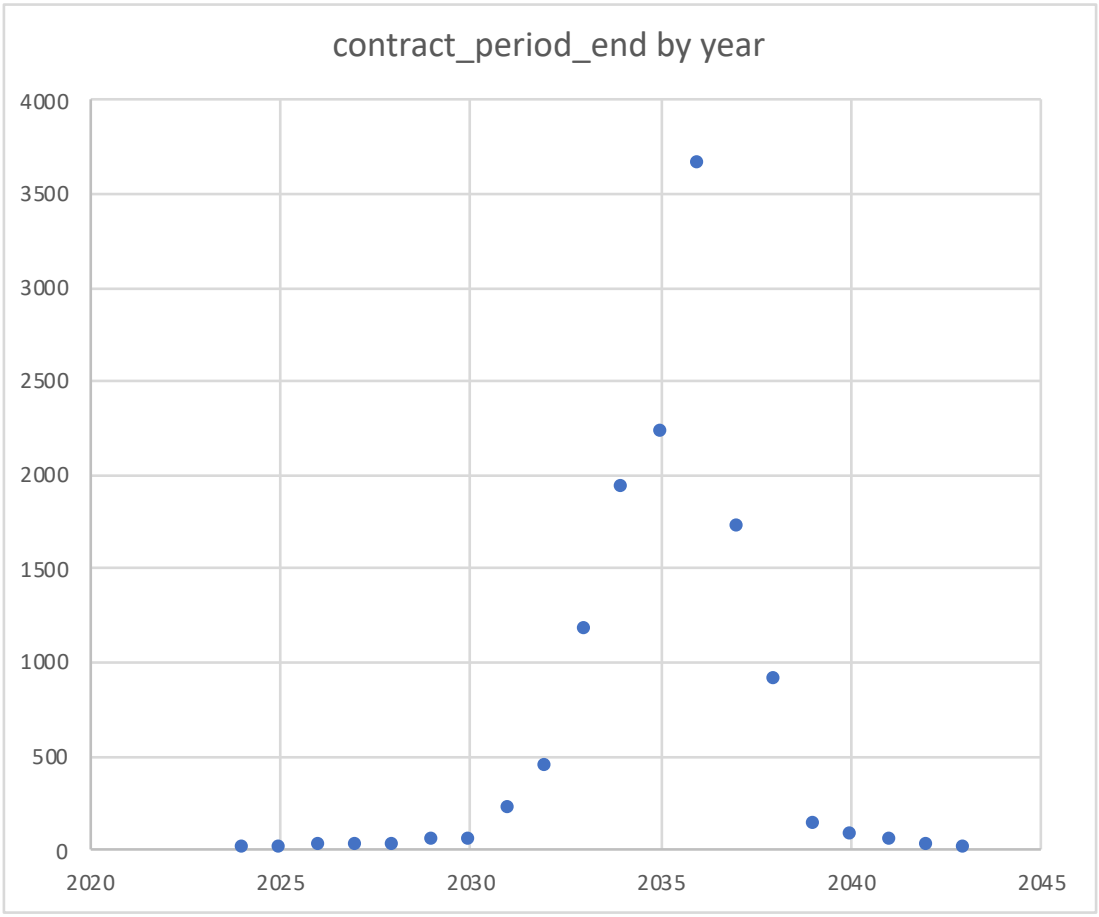


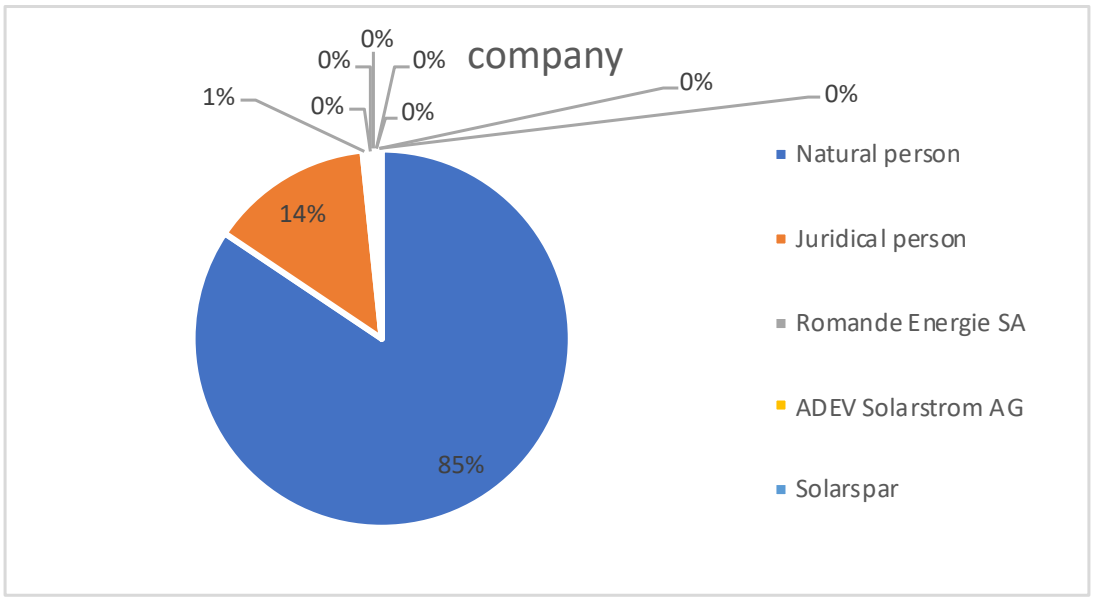

nuts_3_region
2500
2000
1500
1000
500
0
CH021
CH055
CH040
CH033
CH011
CH061
CH023
CH056
CH012
CH032
CH057
CH022
CH070
CH013
CH024
CH053
CH063
CH066
CH052
CH025
CH051
CH031
CH064
CH062
CH065
CH054

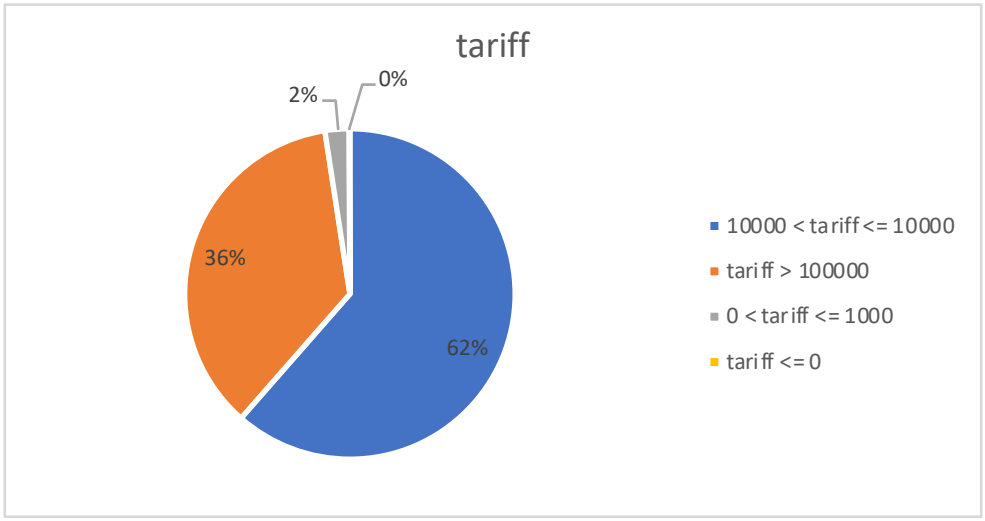

top 10 municipality
90
80
70
60
50
40
30
20
10
0
Basel
Thun
Winterthur
Gossau
Bern
Steffisburg
Altstätten
Buchs
Ruswil
Lausanne

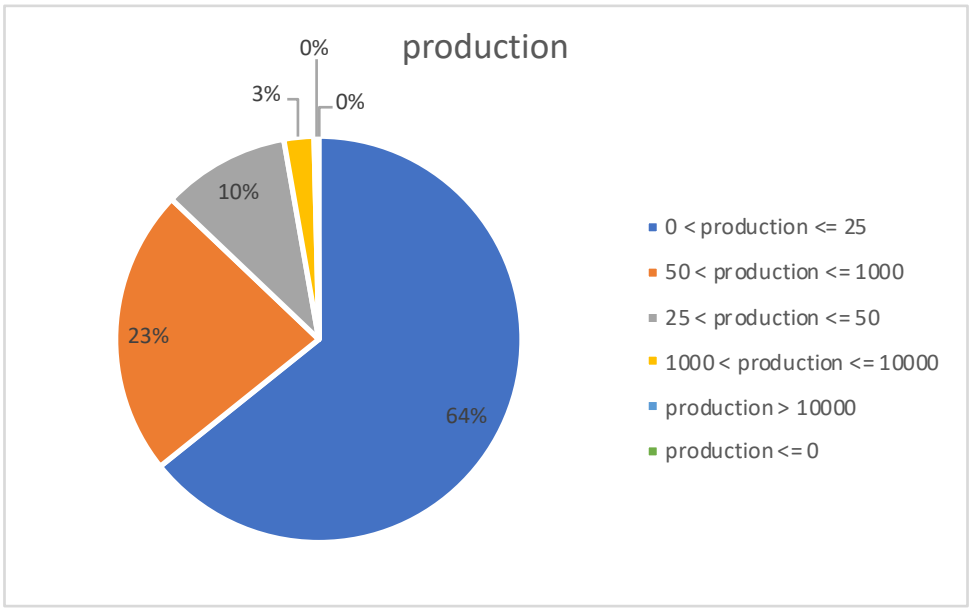

Canton distribution
2500
2000
1500
1000
500
0
BE SG ZH AG VD LU SO GR VS BL TG FR TI NE GE SZ AR ZG SH JU GL BS OW AI UR NW


commissioning_date by year
2016
2006
2017
2018
2004
2005
2014
2007
2015
2011
2013
2008
2009
2012
2010


contract_period_end by year
4000
3500
3000
2500
2000
1500
1000
500
0
2020
2025
2030
2035
2040
2045


company
0%
0%
0%
1%
0%
0%
0%
0%
14%
85%
Natural person
Juridical person
Romande Energie SA
ADEV Solarstrom AG
Solarspar



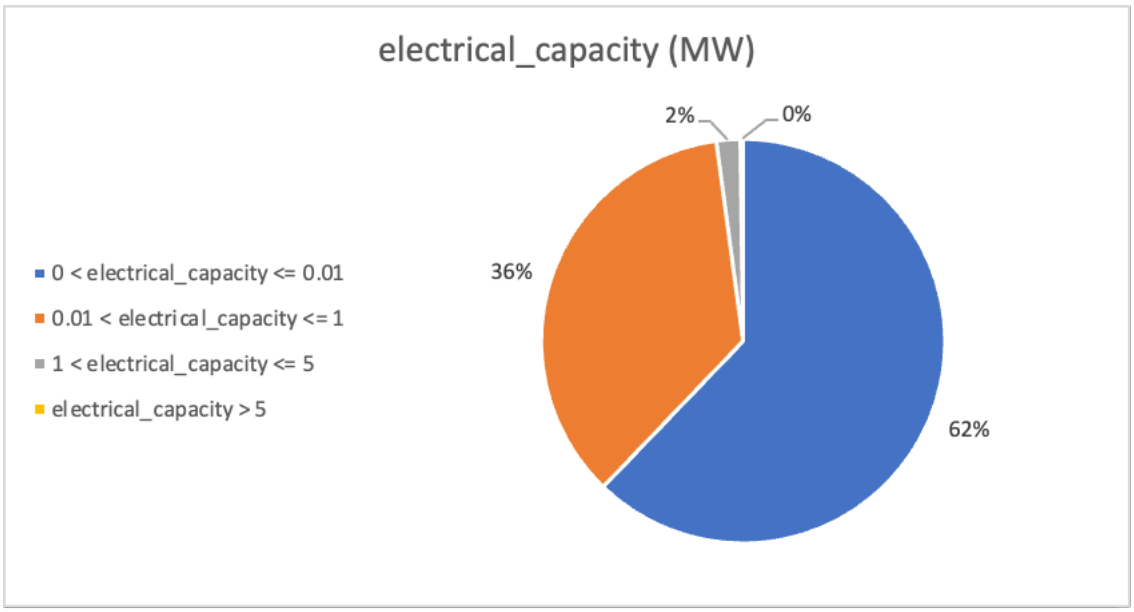

#### 5.2.1.1 - Domain Correctness

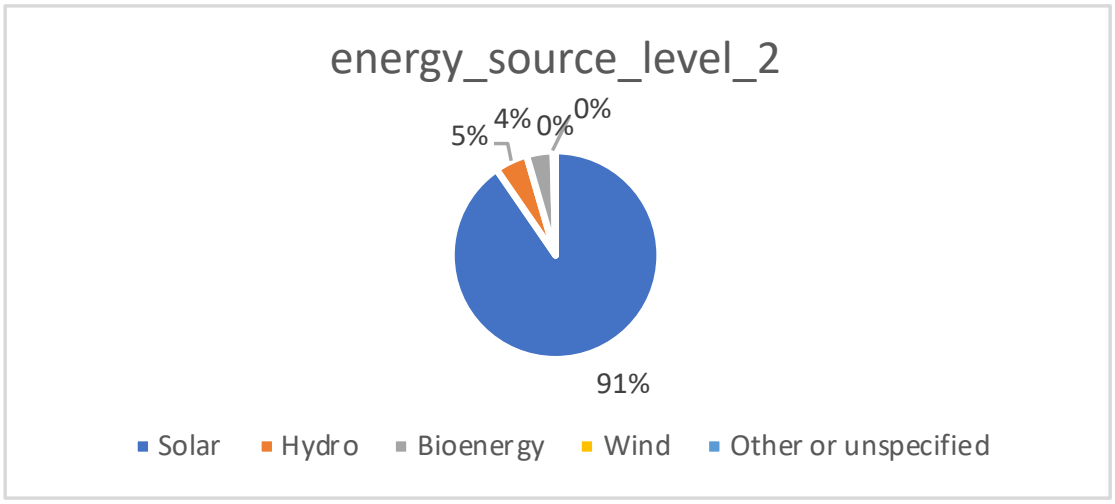
- Energy Source Level 2: no value out of domain
- Municipality Code: no value equal to zero

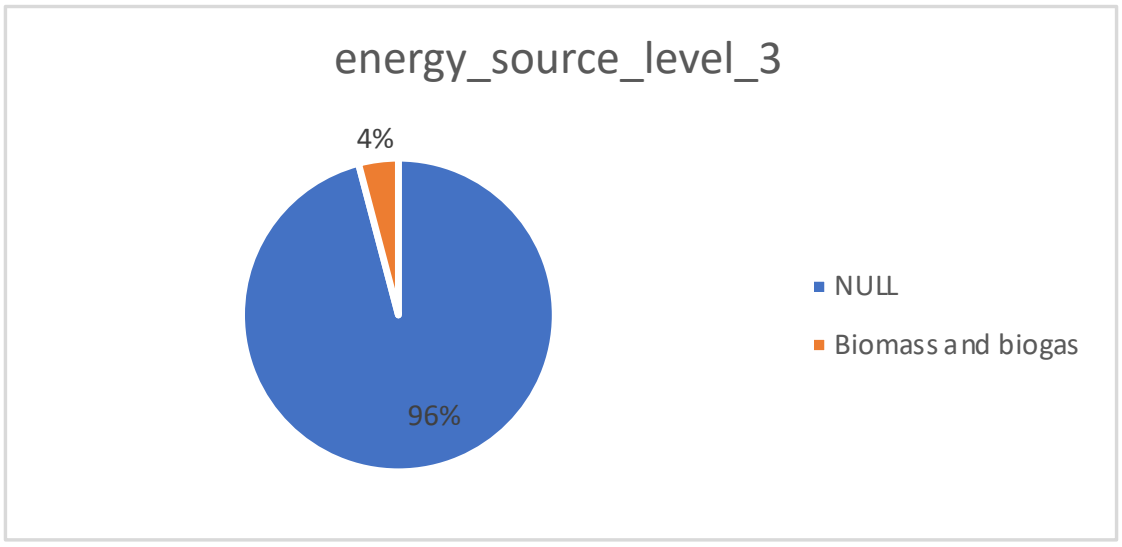
#### 5.2.1.2 - Duplication

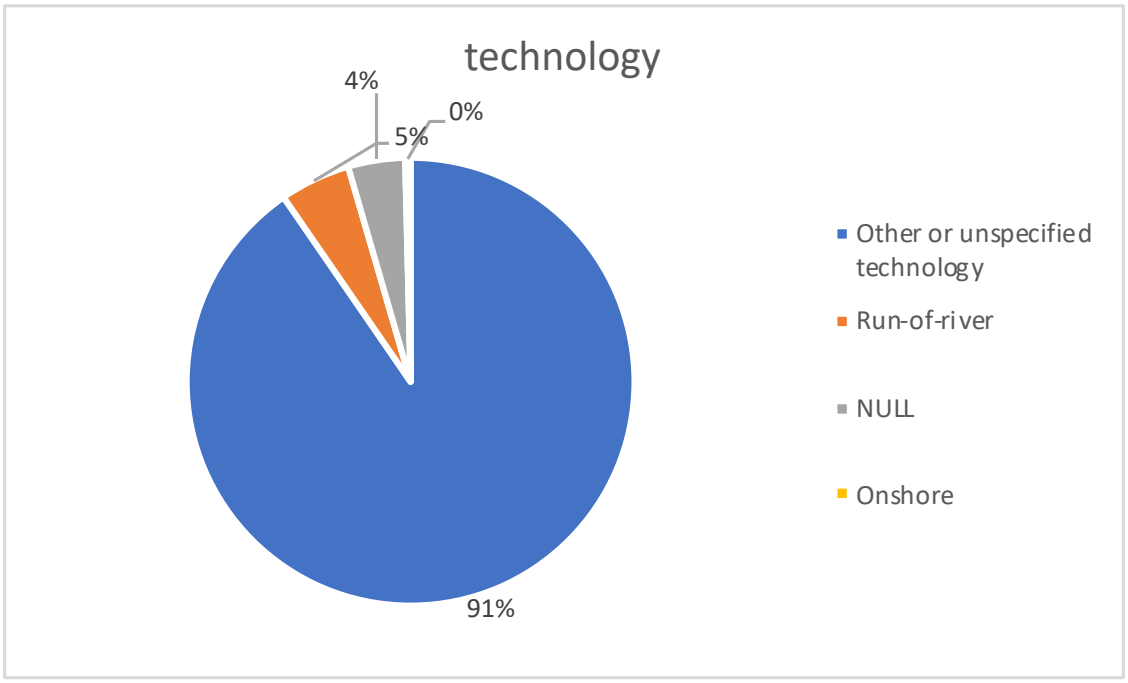
No duplicated rows

### 5.2.2 - Czechia

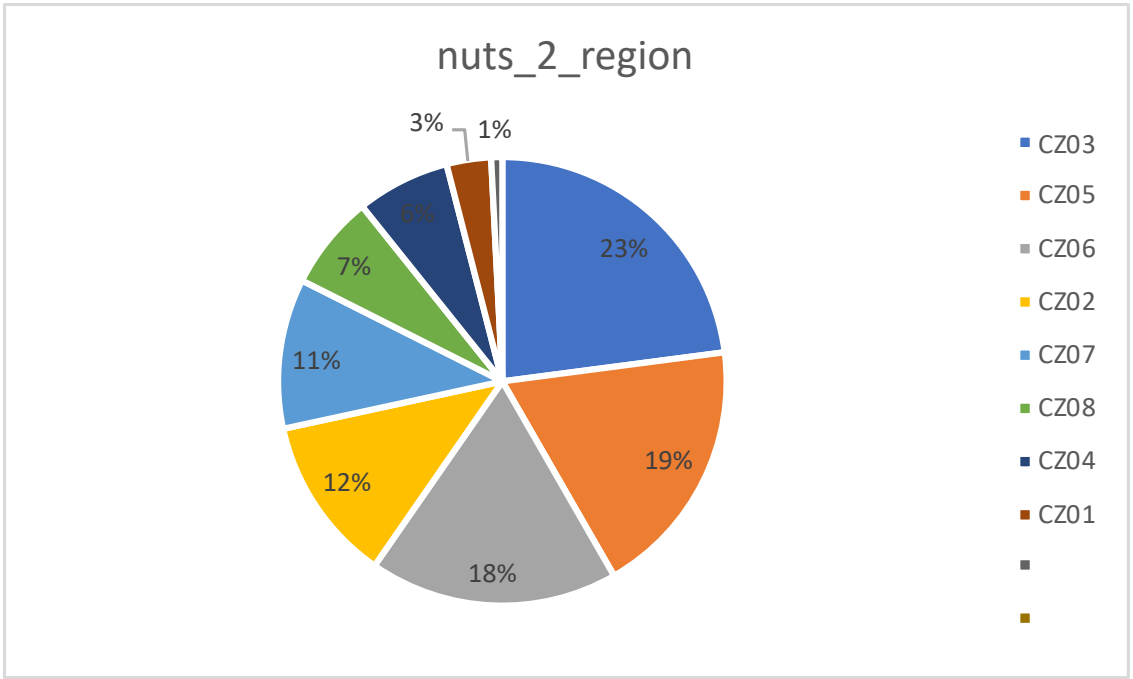

electrical_capacity (MW)
0 < electrical_capacity <= 0.01
0.01 < electrical_capacity <= 1
1 < electrical_capacity <= 5
electrical_capacity > 5
2%
0%
36%
62%

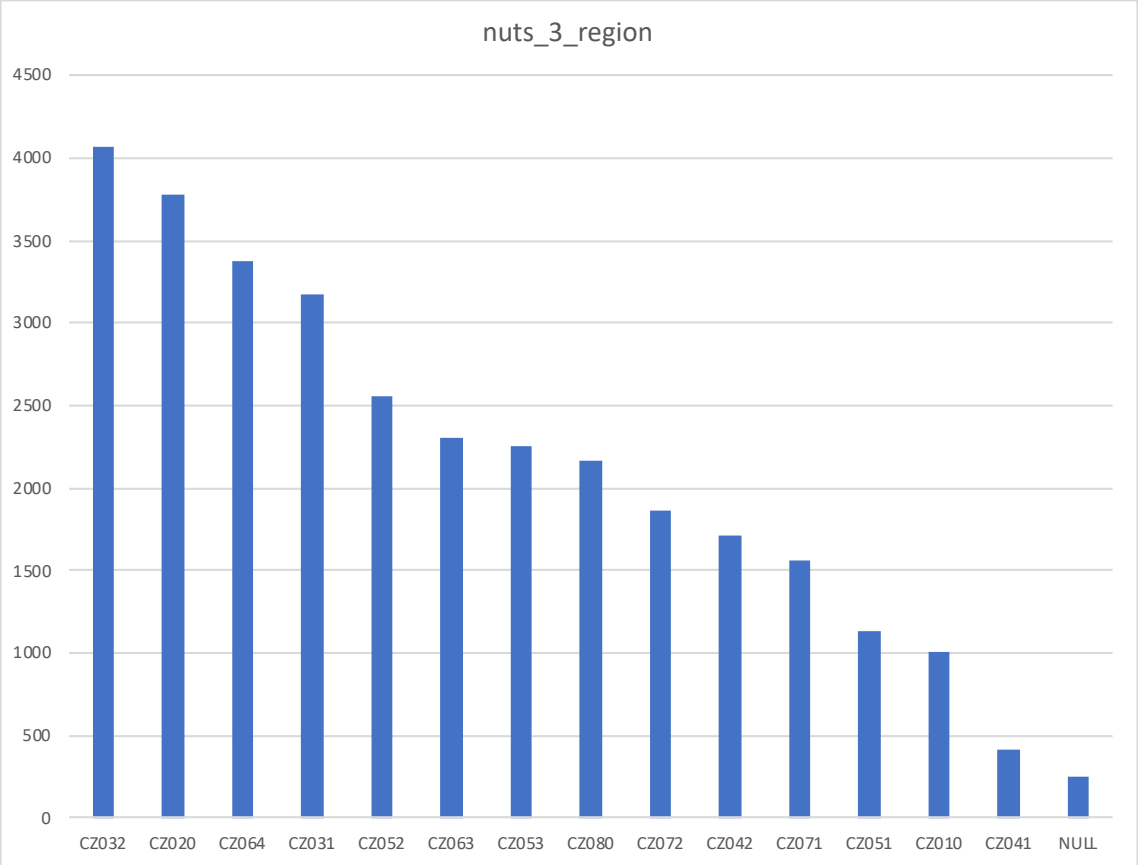

energy_source_level_2
5%
4%
0%
0%
91%
Solar
Hydro
Bioenergy
Wind
Other or unspecified

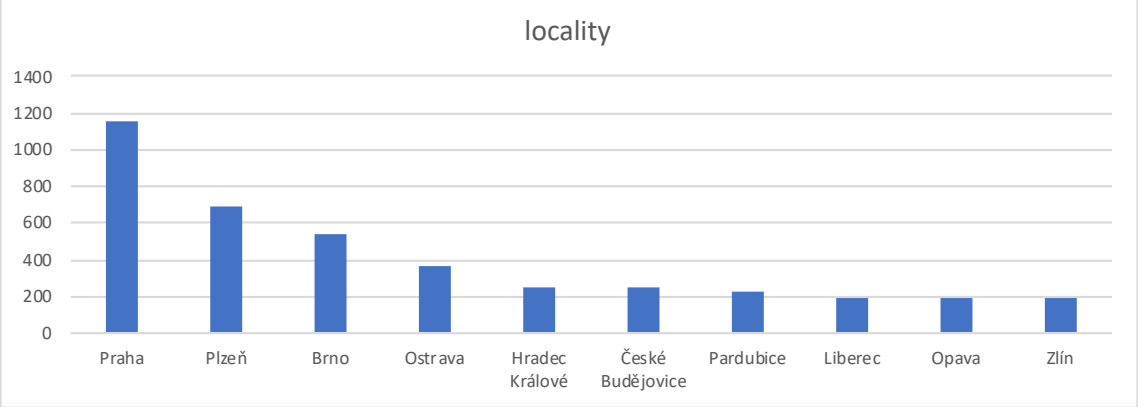

energy_source_level_3
4%
96%
NULL
Biomass and biogas

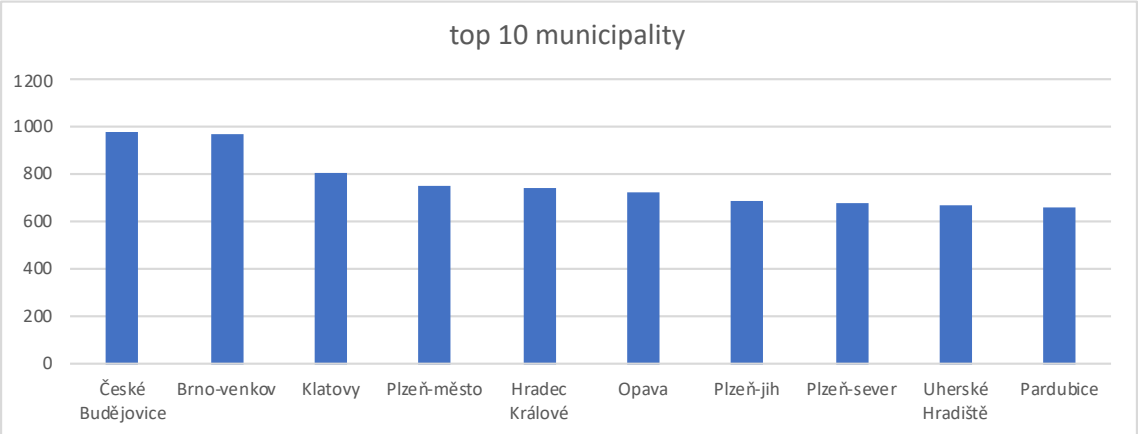


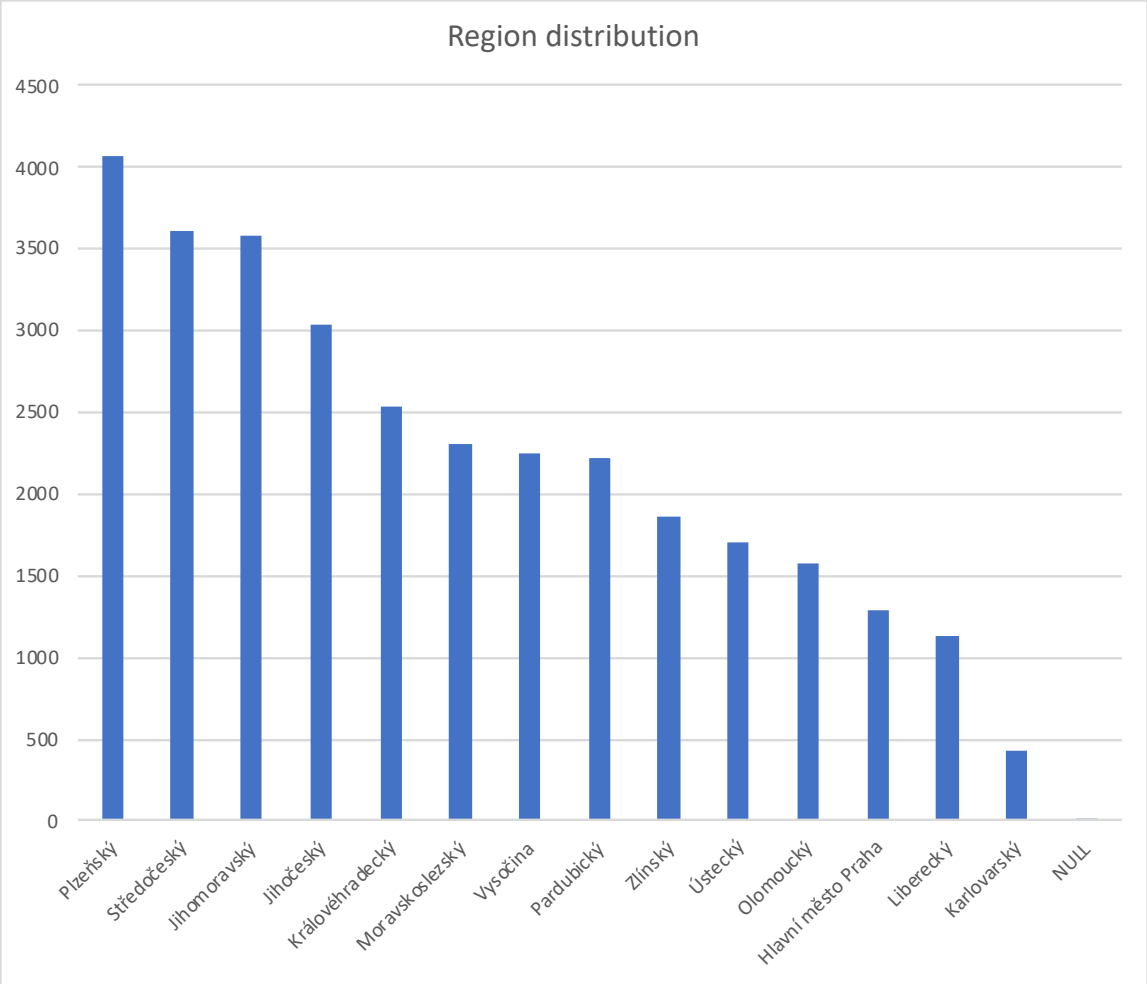

| nuts_1_region | count |
|---|---|
| CZ0 | 31351 |
| NULL | 253 |

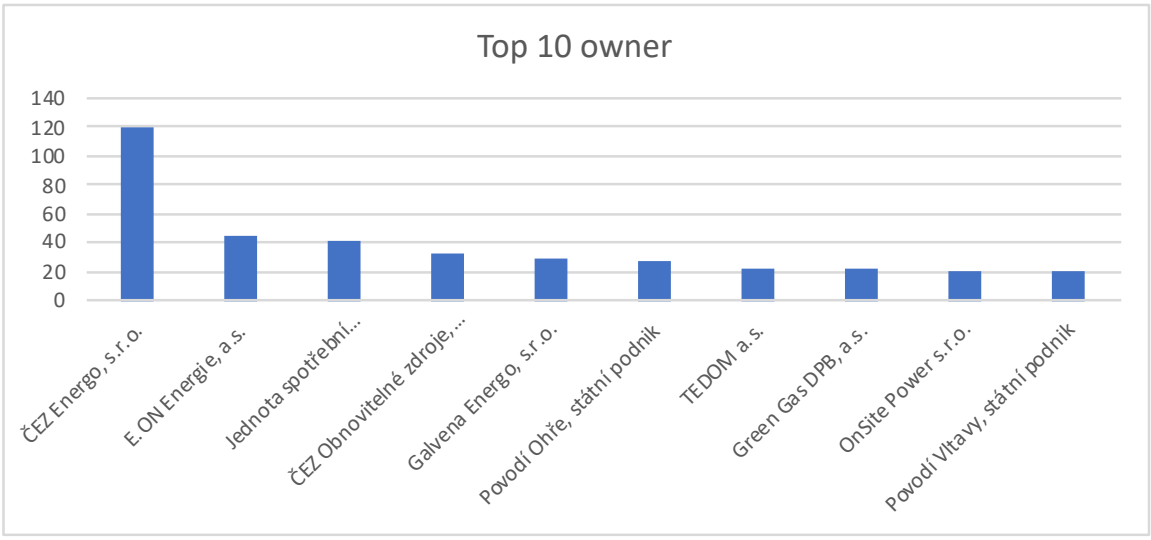



nuts_3_region
4500
4000
3500
3000
2500
2000
1500
1000
500
0
CZ032
CZ020
CZ064
CZ031
CZ052
CZ063
CZ053
CZ080
CZ072
CZ042
CZ071
CZ051
CZ010
CZ041
NULL


locality
1400
1200
1000
800
600
400
200
0
Praha
Plzeň
Brno
Ostrava
Hradec Králové
České Budějovice
Pardubice
Liberec
Opava
Zlín


top 10 municipality
1200
1000
800
600
400
200
0
České Budějovice
Brno-venkov
Klatovy
Plzeň-město
Hradec Králové
Opava
Plzeň-jih
Plzeň-sever
Uherské Hradiště
Pardubice



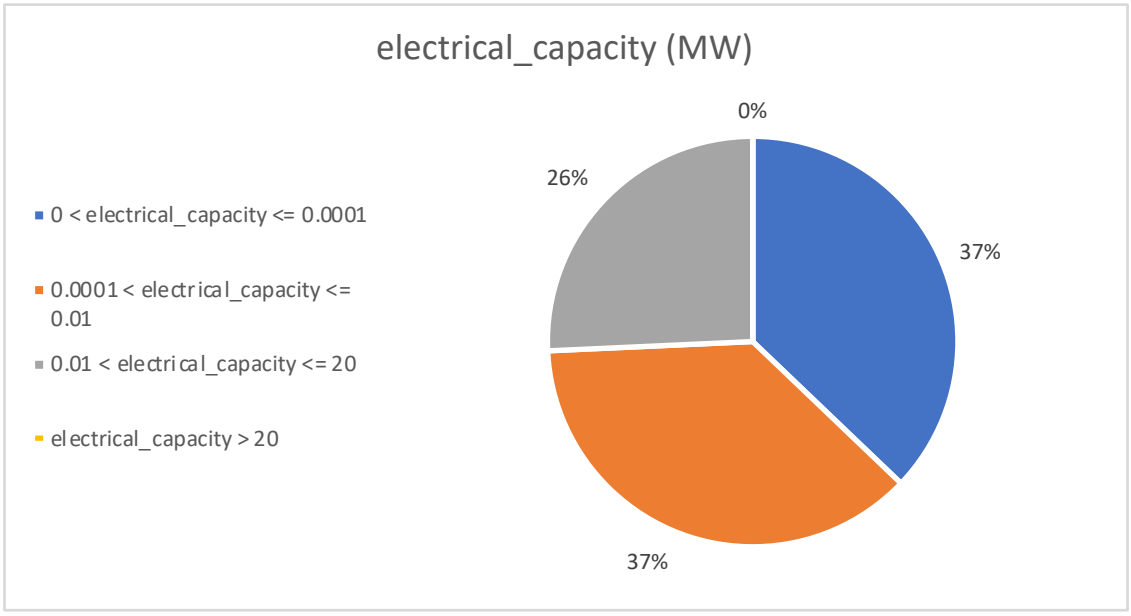

#### 5.2.2.1 - Domain Correctness

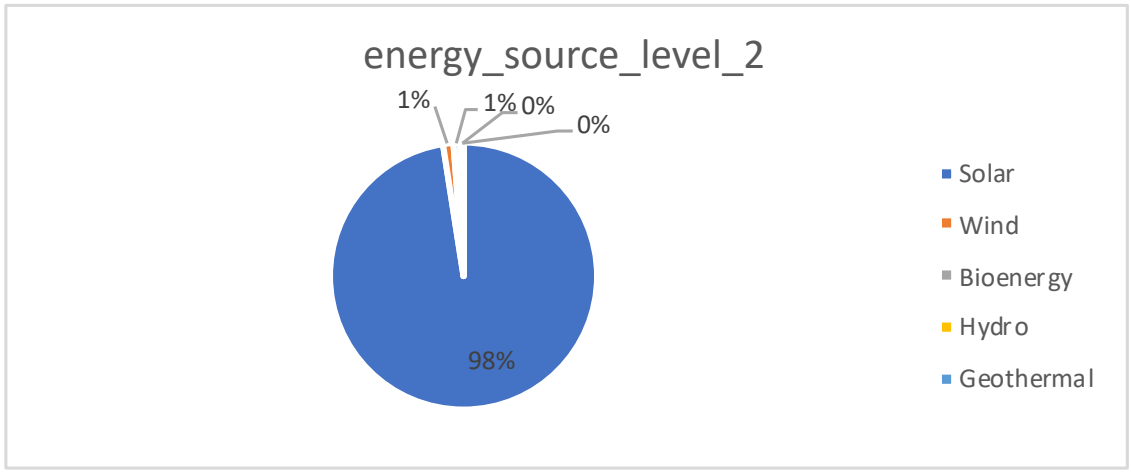
- Energy Source Level 2: four value out of domain (value = “Other or unspecified”)
- Municipality Code: no value equal to zero

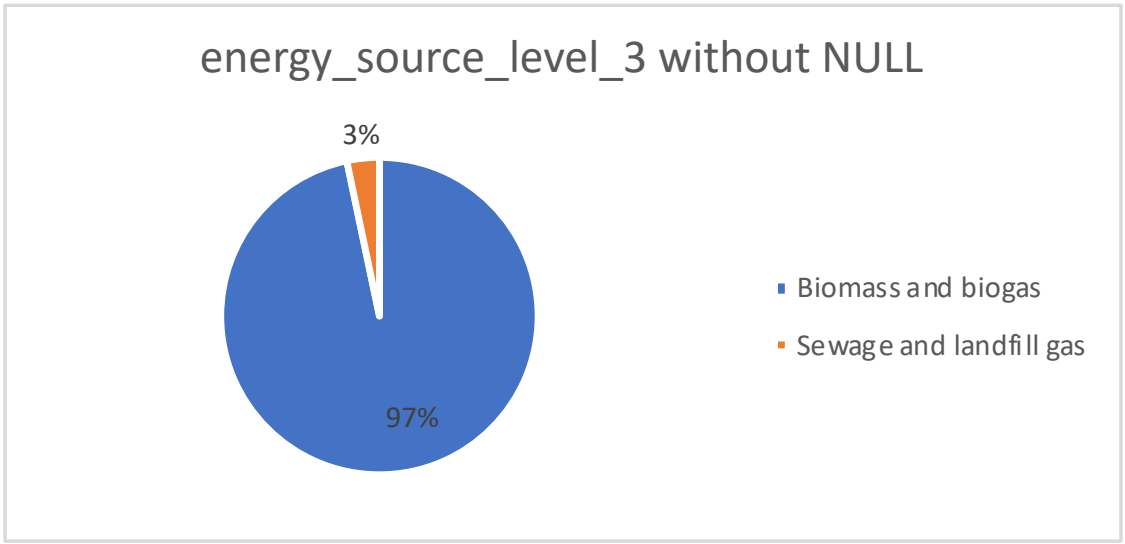
#### 5.2.2.2 - Duplication

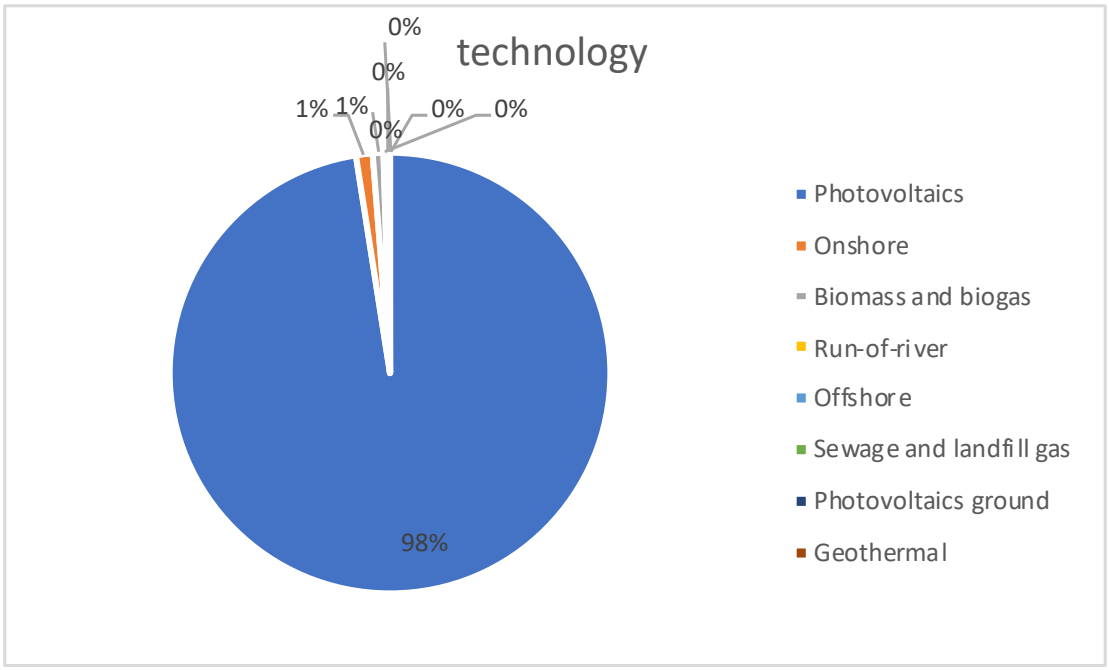
16 duplicated rows

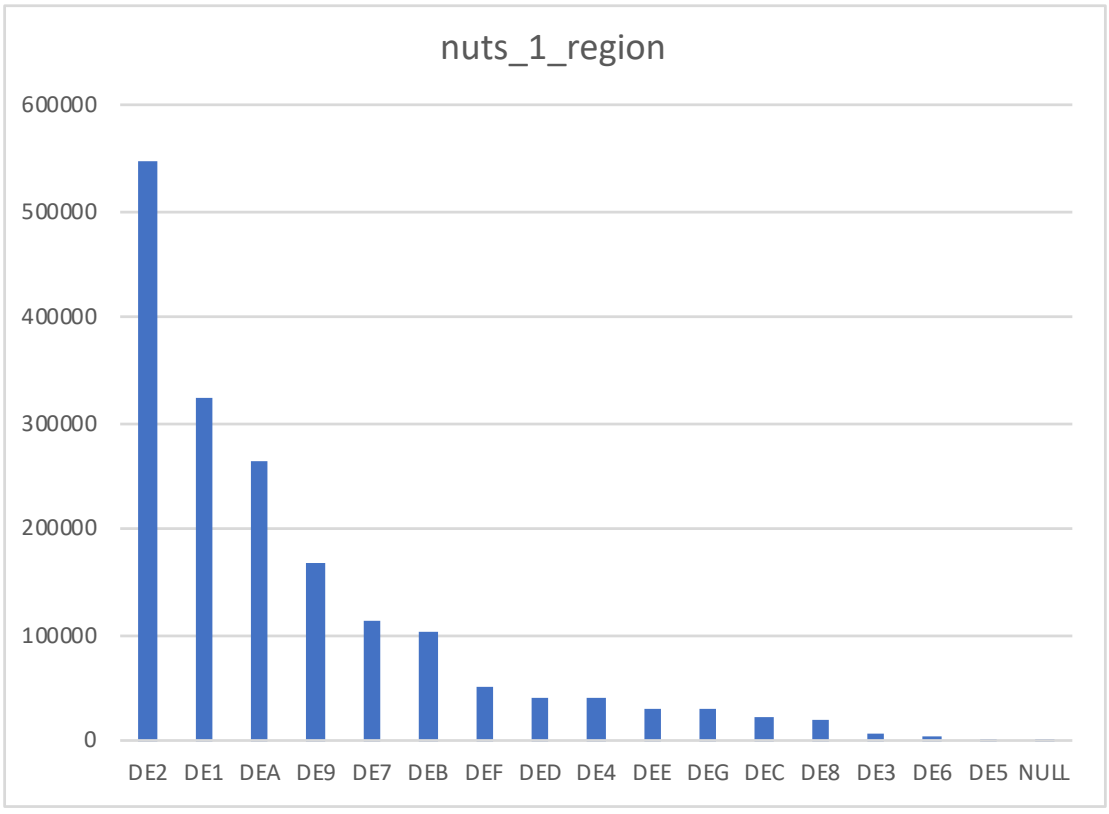
### 5.2.3 - Germany

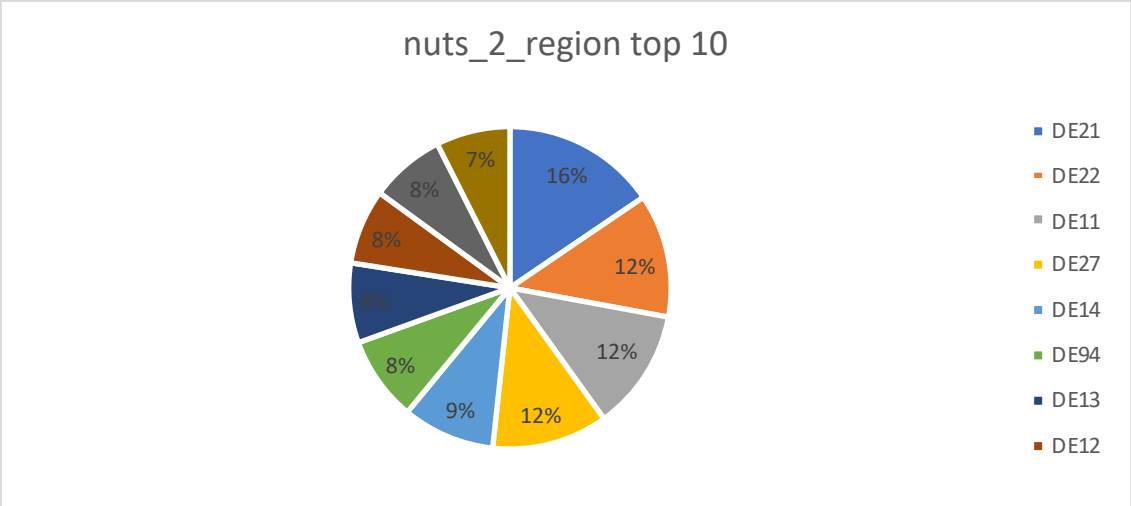

electrical_capacity (MW)
0 < electrical_capacity <= 0.0001
0.0001 < electrical_capacity <= 0.01
0.01 < electrical_capacity <= 20
electrical_capacity > 20
0%
26%
37%
37%

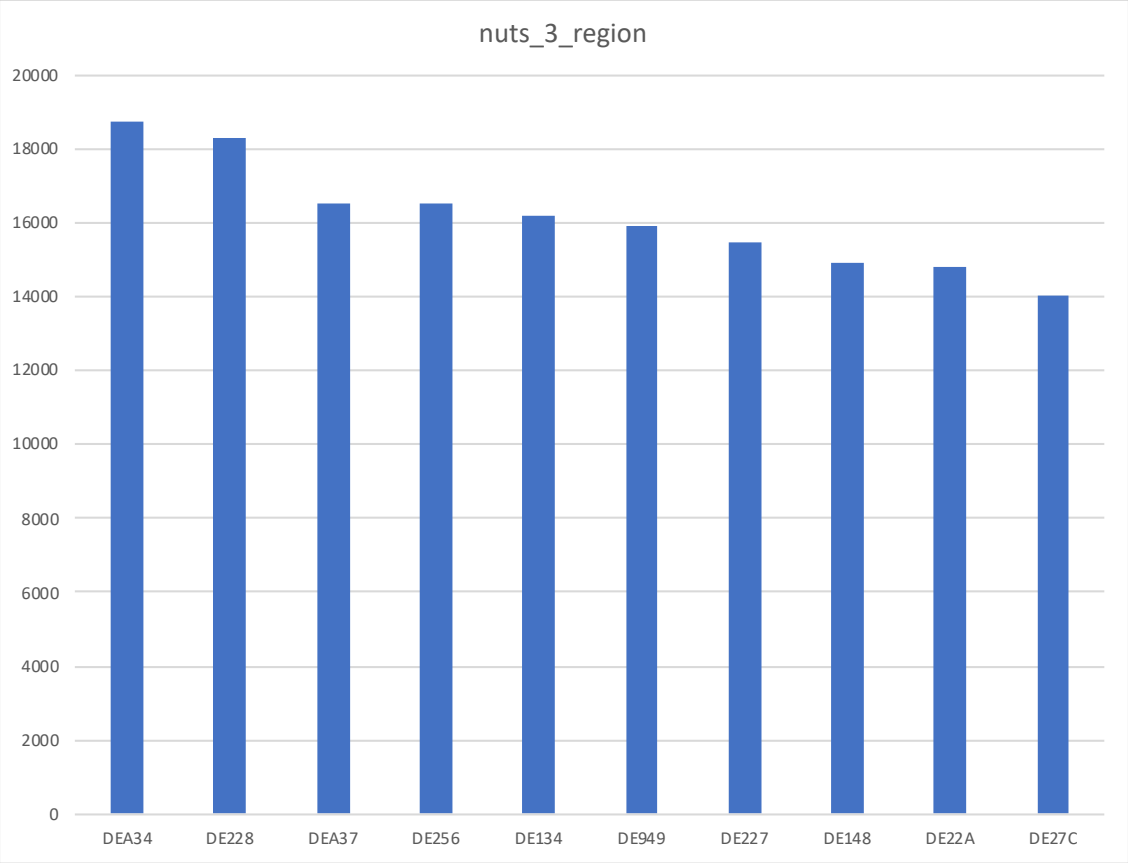

energy_source_level_2
1%
1%
0%
0%
98%
Solar
Wind
Bioenergy
Hydro
Geothermal

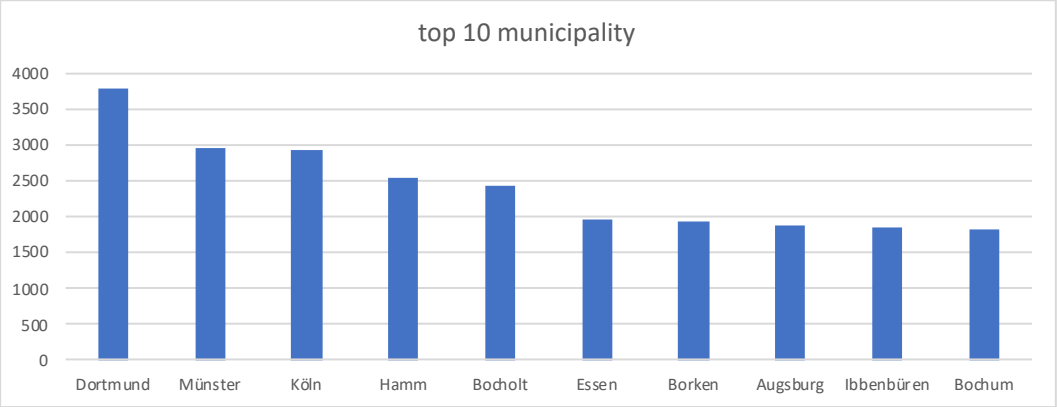

energy_source_level_3 without NULL
3%
97%
Biomass and biogas
Sewage and landfill gas

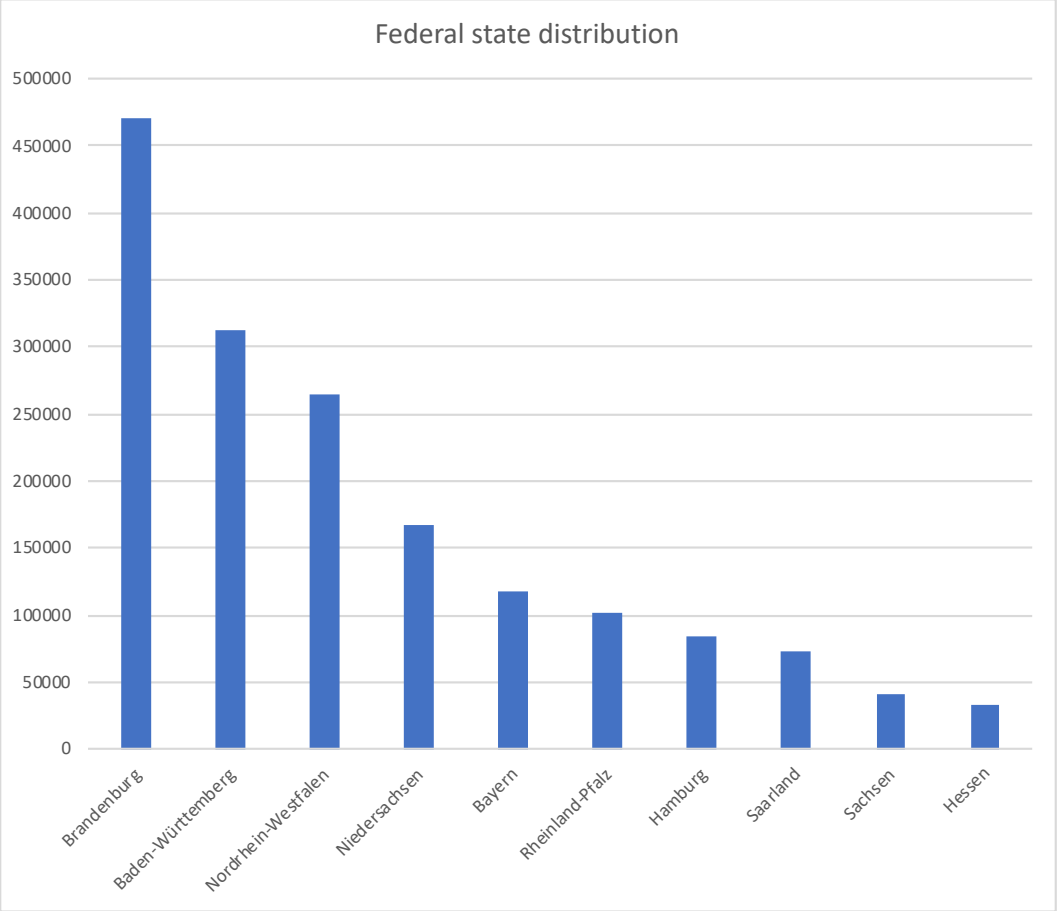

technology
0%
0%
1%
1%
0%
0%
0%
98%
Photovoltaics
Onshore
Biomass and biogas
Run-of-river
Offshore
Sewage and landfill gas
Photovoltaics ground
Geothermal

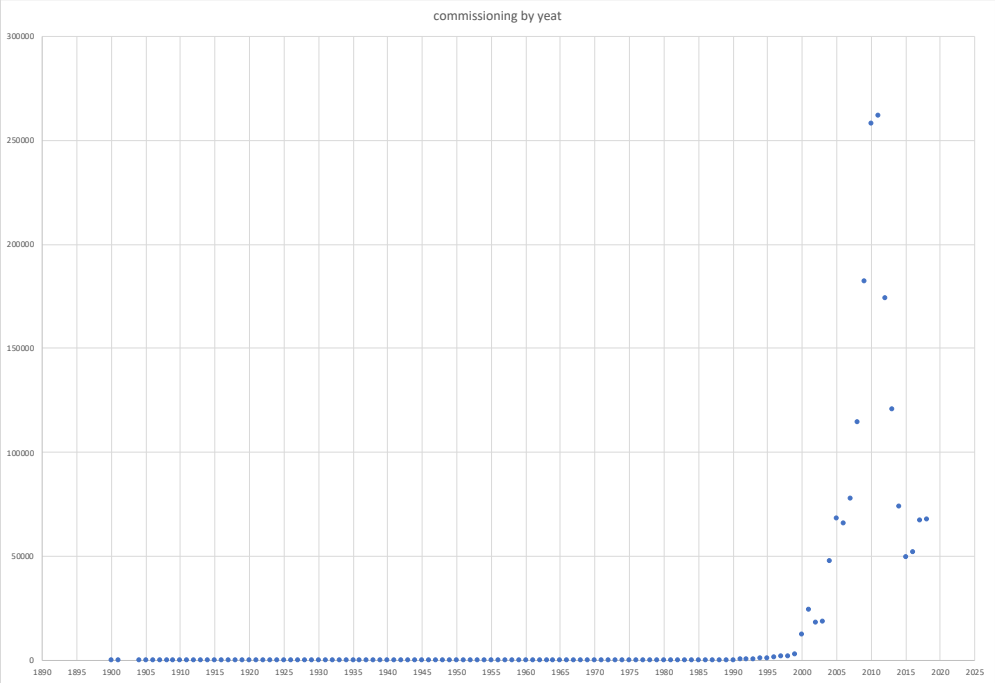

nuts_1_region
600000
500000
400000
300000
200000
100000
0
DE2
DE1
DEA
DE9
DE7
DEB
DEF
DED
DE4
DEE
DEG
DEC
DE8
DE3
DE6
DE5
NULL

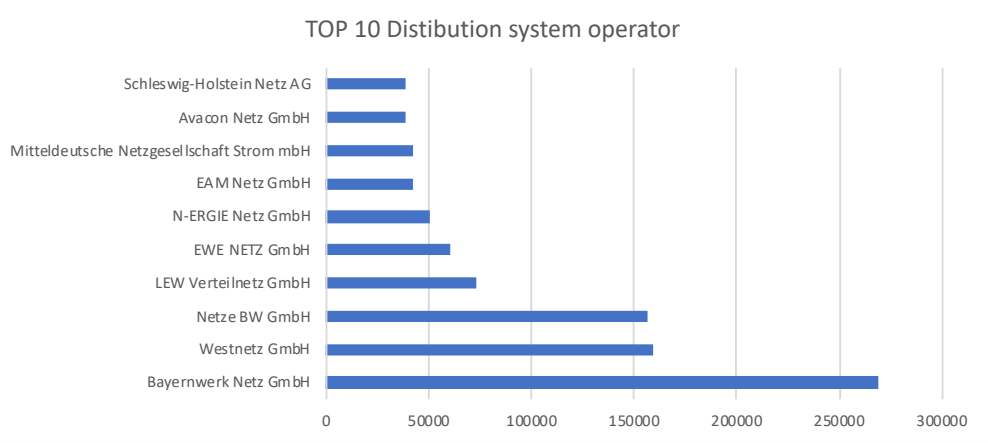

nuts_2_region top 10
16%
12%
12%
12%
9%
8%
8%
8%
7%
DE21
DE22
DE11
DE27
DE14
DE94
DE13
DE12


nuts_3_region
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
DEA34
DE228
DEA37
DE256
DE134
DE949
DE227
DE148
DE22A
DE27C


top 10 municipality
4000
3500
3000
2500
2000
1500
1000
500
0
Dortmund
Münster
Köln
Hamm
Bocholt
Essen
Borken
Augsburg
Ibbenbüren
Bochum


Federal state distribution
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
Brandenburg
Baden-Württemberg
Nordrhein-Westfalen
Niedersachsen
Bayern
Rheinland-Pfalz
Hamburg
Saarland
Sachsen
Hessen




| production | values |
|---|---|
| min | -68,943 |
| max | 66.058,391 |

| production | count |
|---|---|
| 0 < production <= 25 | 8156 |
| 50 < production <= 1000 | 2901 |
| 25 < production <= 50 | 1286 |
| 1000 < production <= 10000 | 303 |
| production > 10000 | 39 |
| production <= 0 | 9 |

### 5.2.3.1 - Domain Correctness

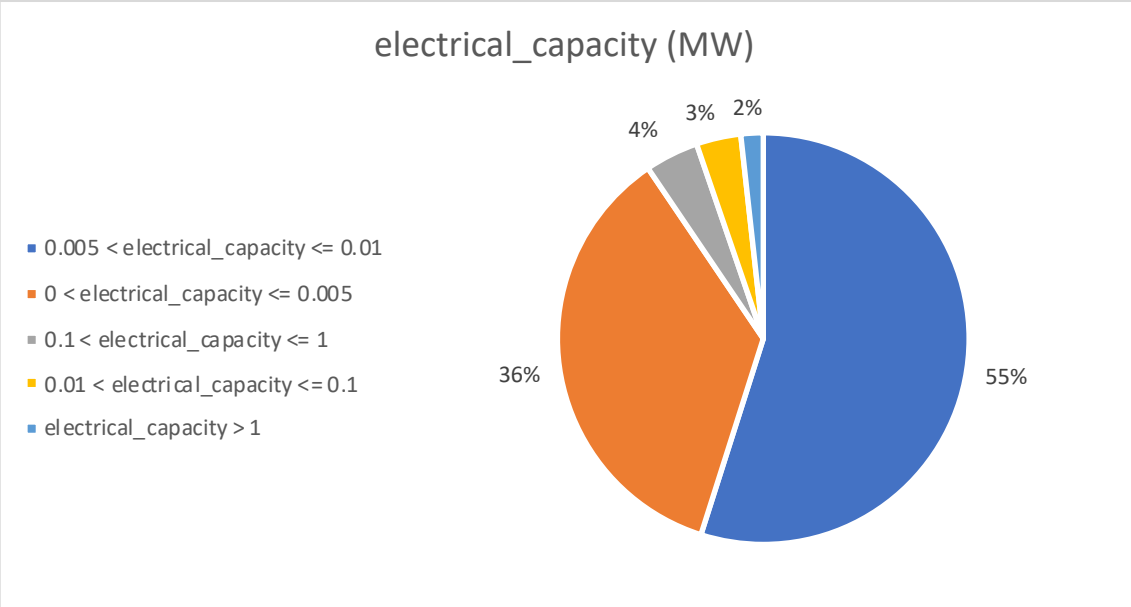
- Energy Source Level 2: no value out of domain
- Municipality Code: no value equal to zero
- tso: no value out of domain

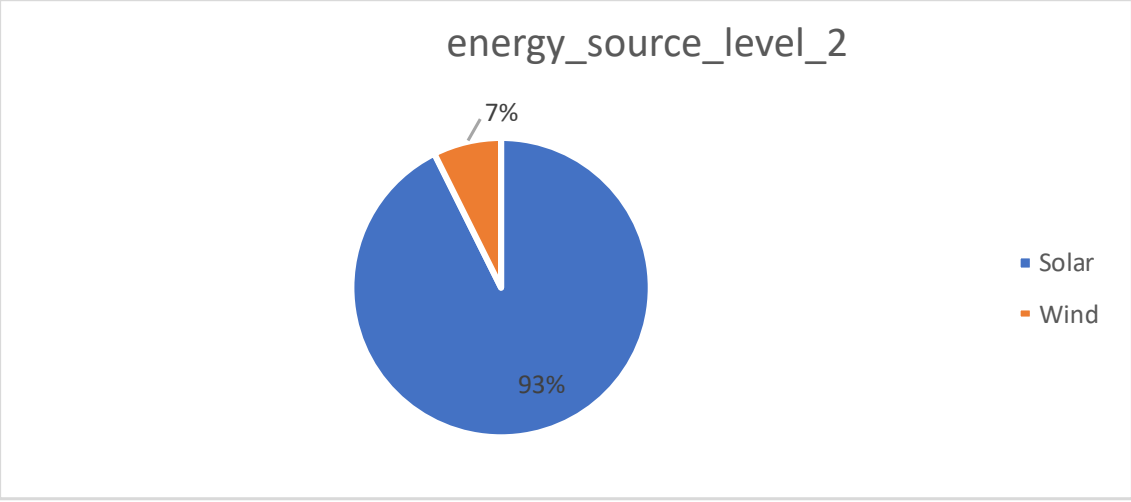
### 5.2.3.2 - Duplication

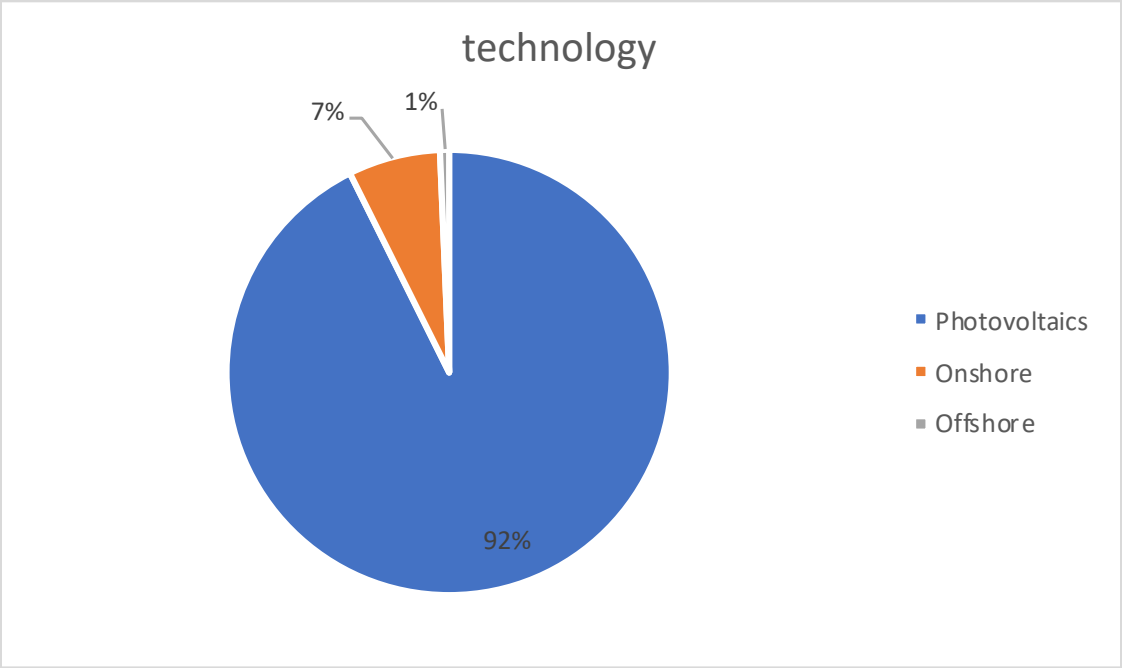
463 duplicated rows

## 5.2.4 - Denmark

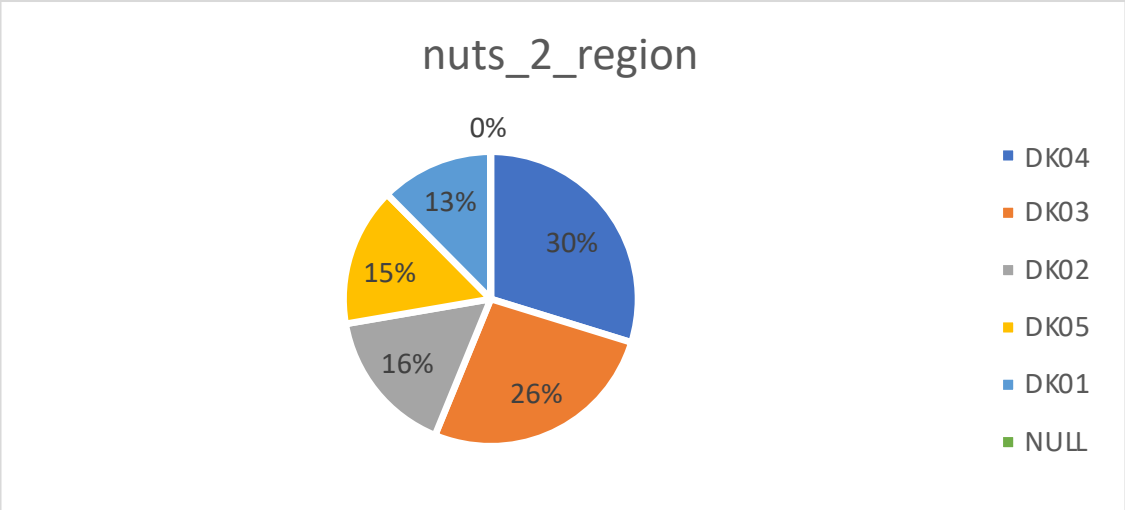
| electrical_capacity | value |
|---|---|
| min | 0 |
| max | 10 |
| | |

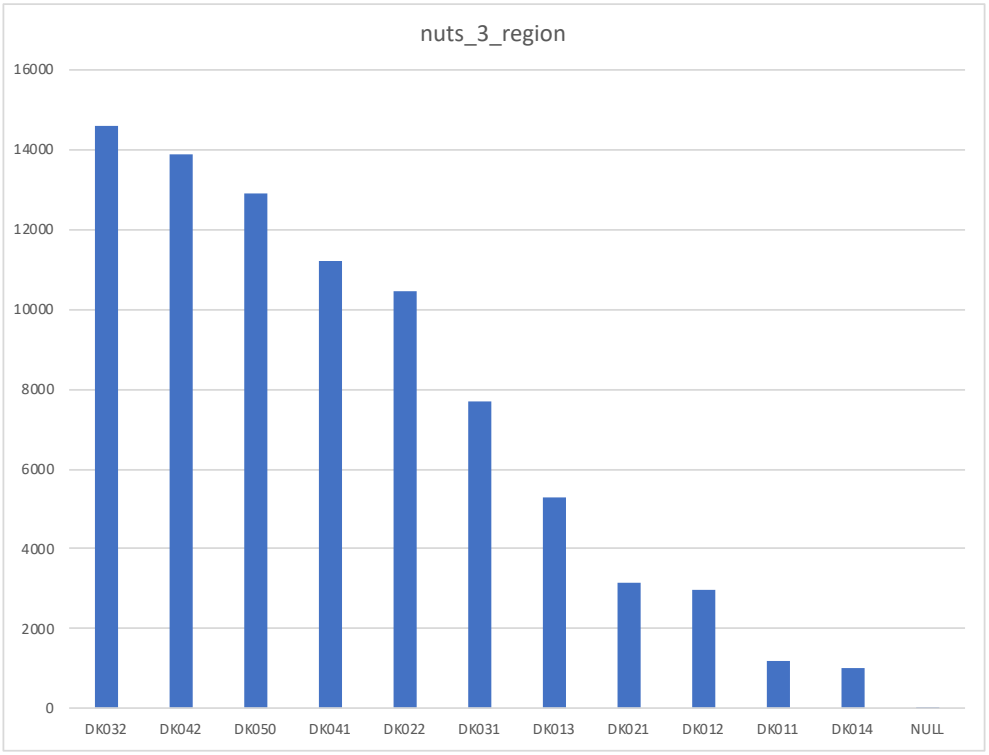


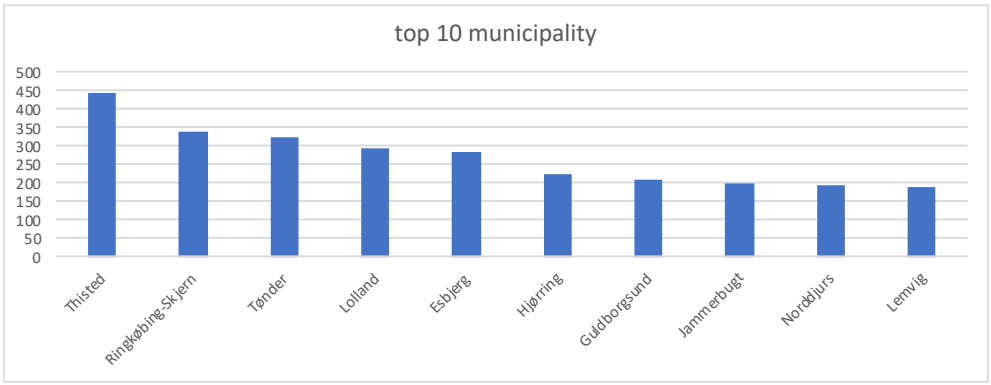



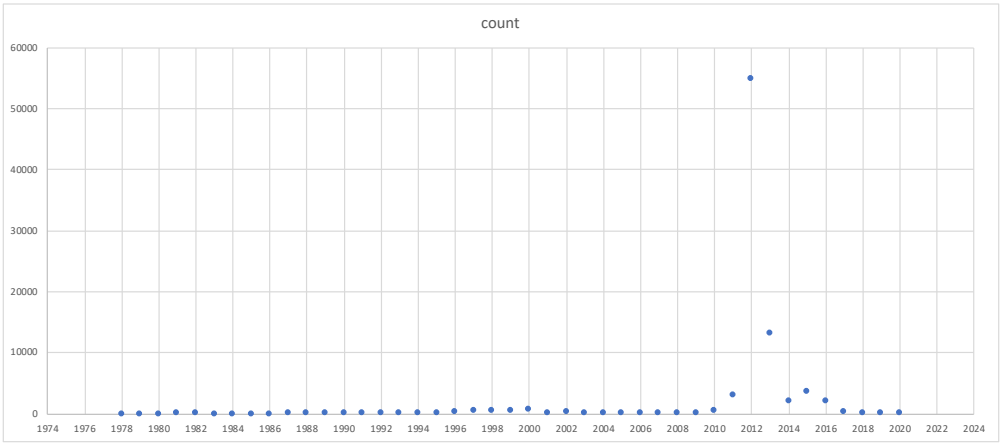


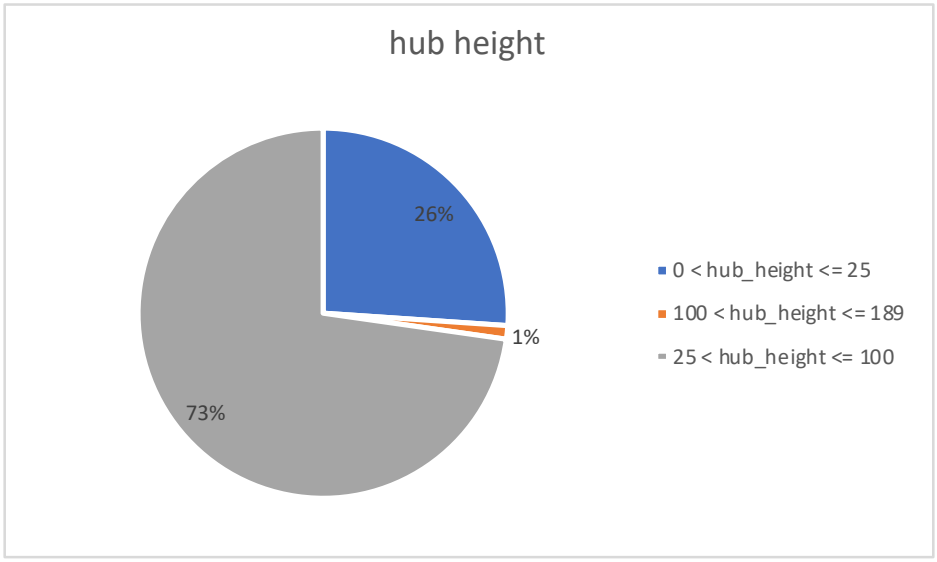

| nuts_1_region | count |
|---|---|
| DK0 | 84349 |
| NULL | 4 |

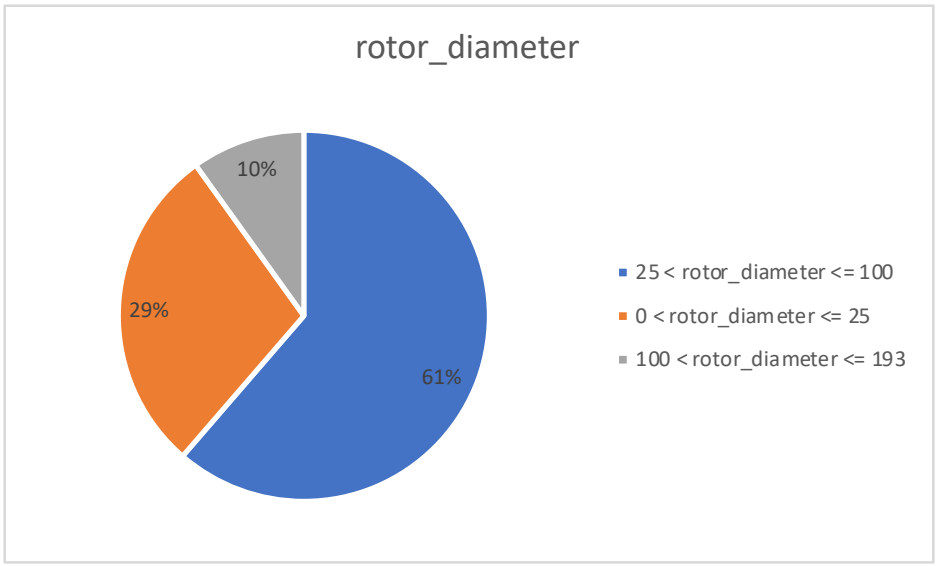


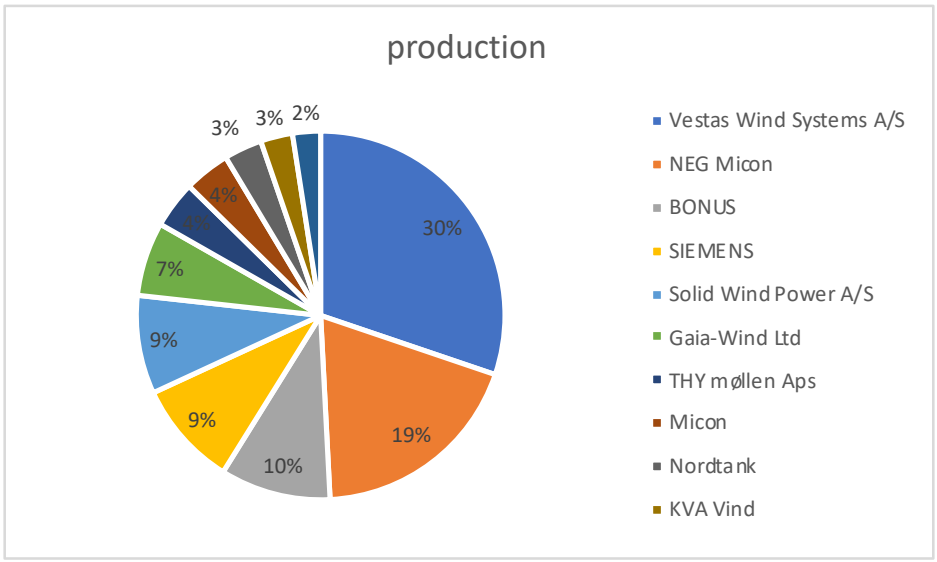

nuts_3_region
16000
14000
12000
10000
8000
6000
4000
2000
0
DK032
DK042
DK050
DK041
DK022
DK031
DK013
DK021
DK012
DK011
DK014
NULL


top 10 municipality
500
450
400
350
300
250
200
150
100
50
0
Thisted
Ringkøbing-Skjern
Tønder
Lolland
Esbjerg
Hjørring
Guldborgsund
Jammerbugt
Norddjurs
Lemvig


count
60000
50000
40000
30000
20000
10000
0
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
2024




**5.2.4.1 - Domain Correctness**

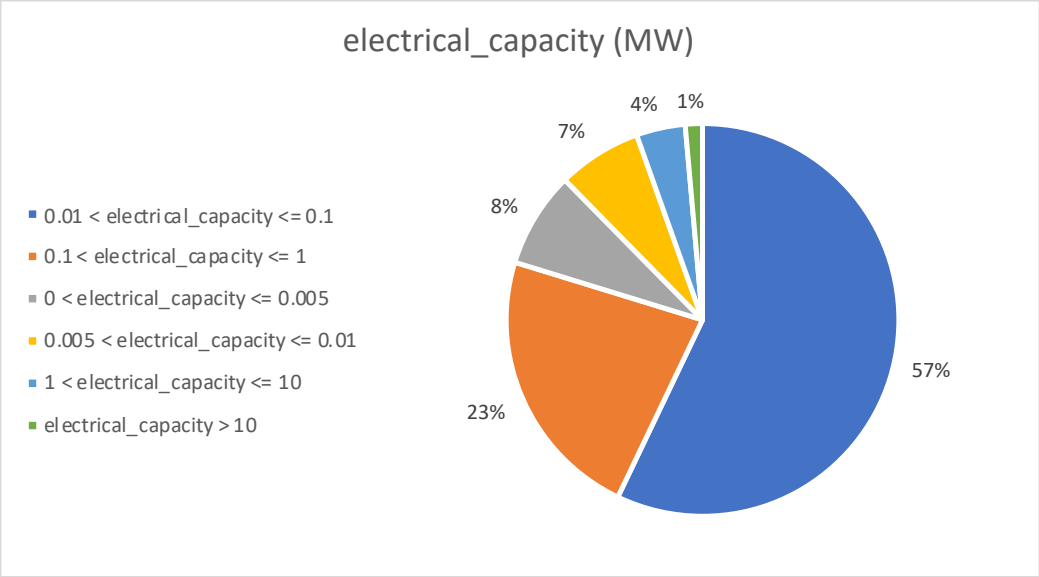
- Energy Source Level 2: no value out of domain
- Municipality Code: no value equal to zero
- hub_height: 78.216 value < 1m
- rotor_diameter: 78.217 value < 0.5m

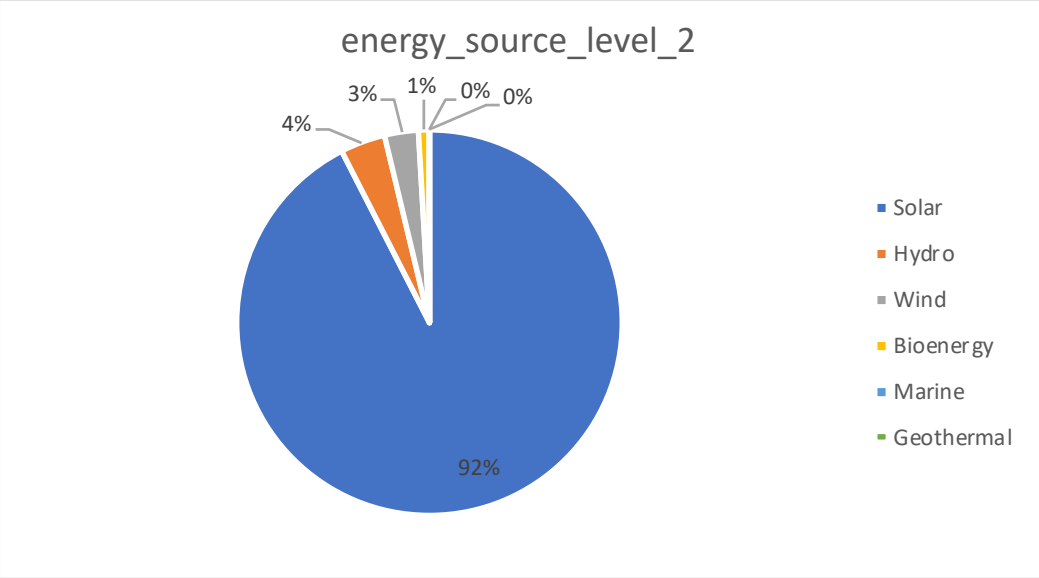
**5.2.4.2 - Duplication**

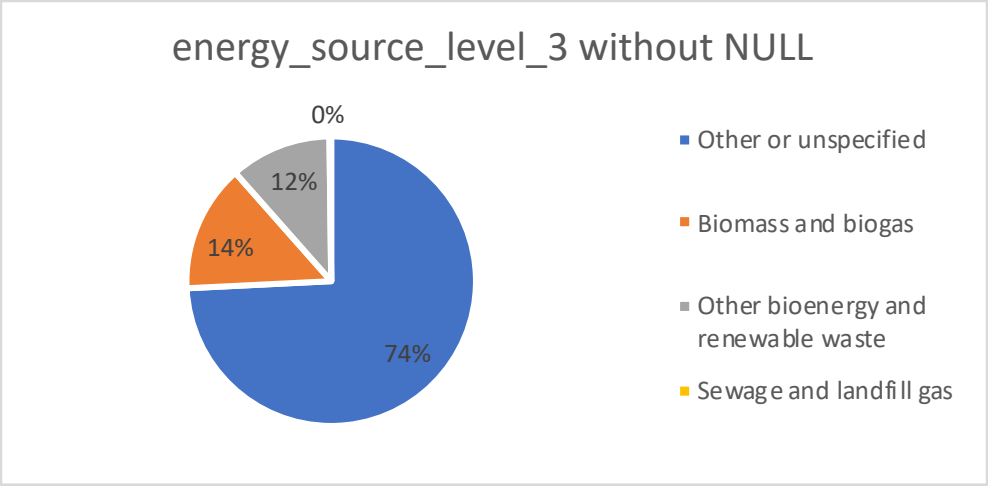
37 duplicated rows

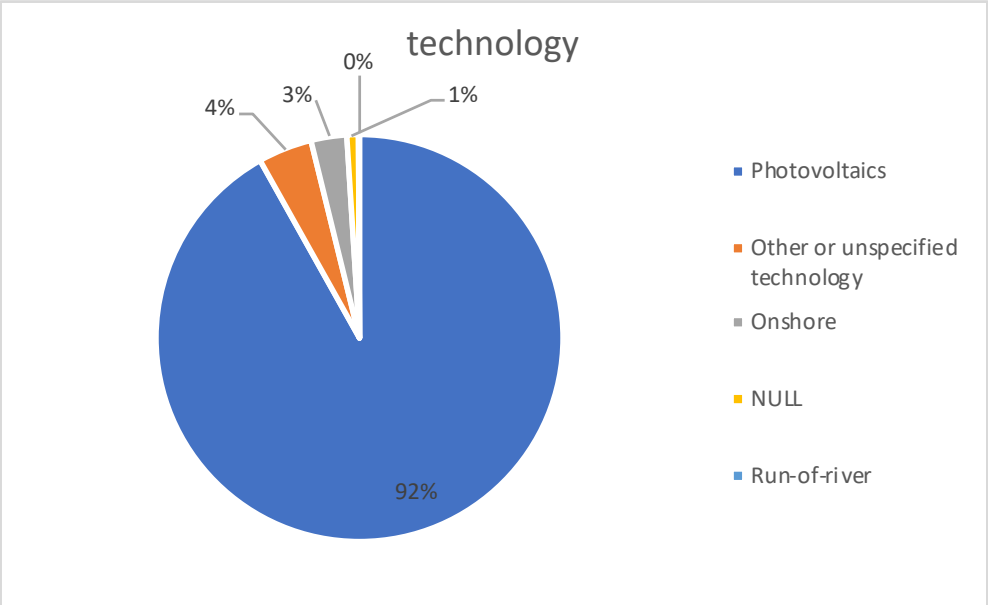
### 5.2.5 - France

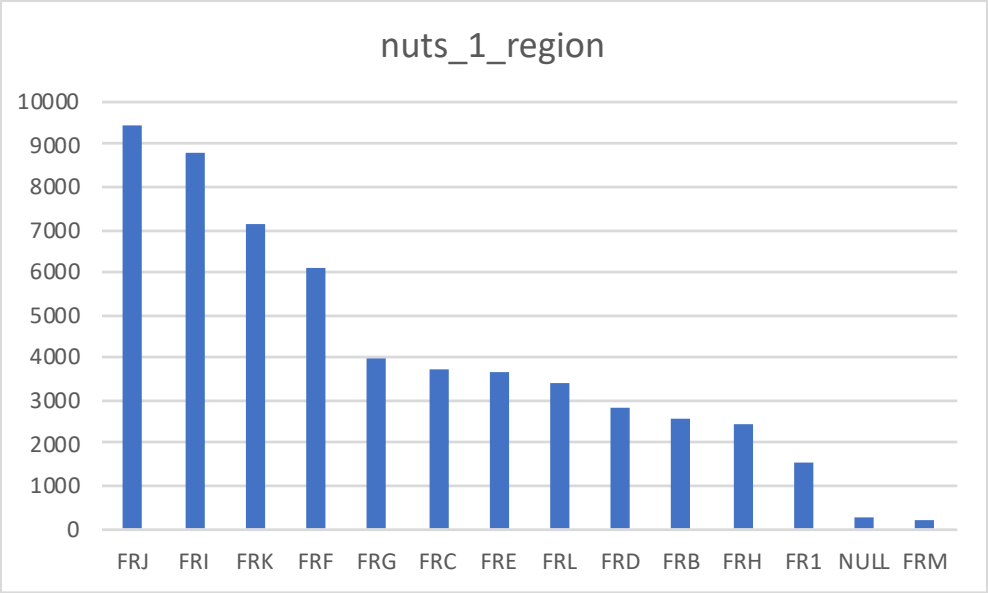
| electrical_capacity | value |
|---|---|
| min | 1 |
| max | 80 |

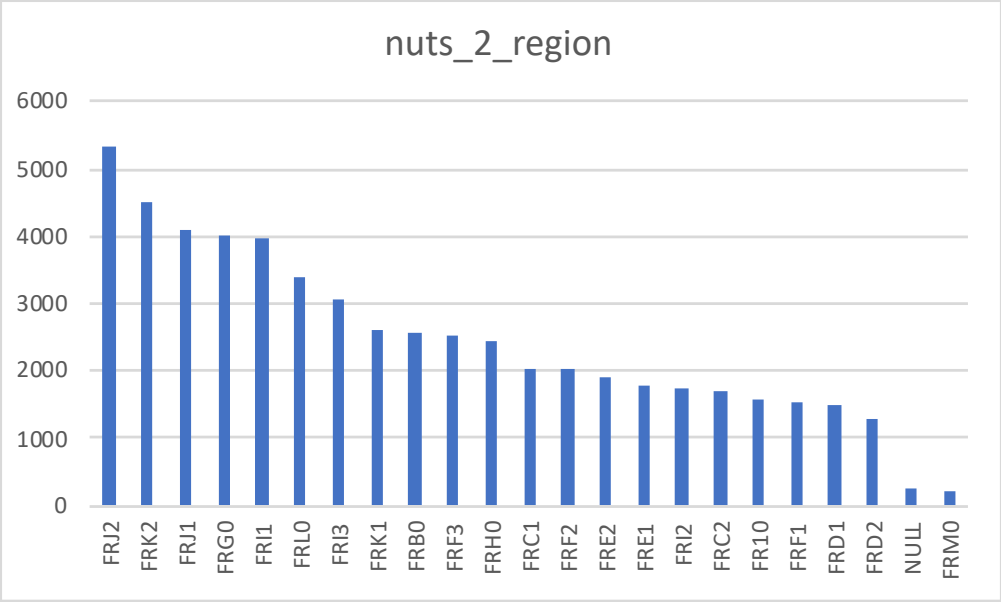


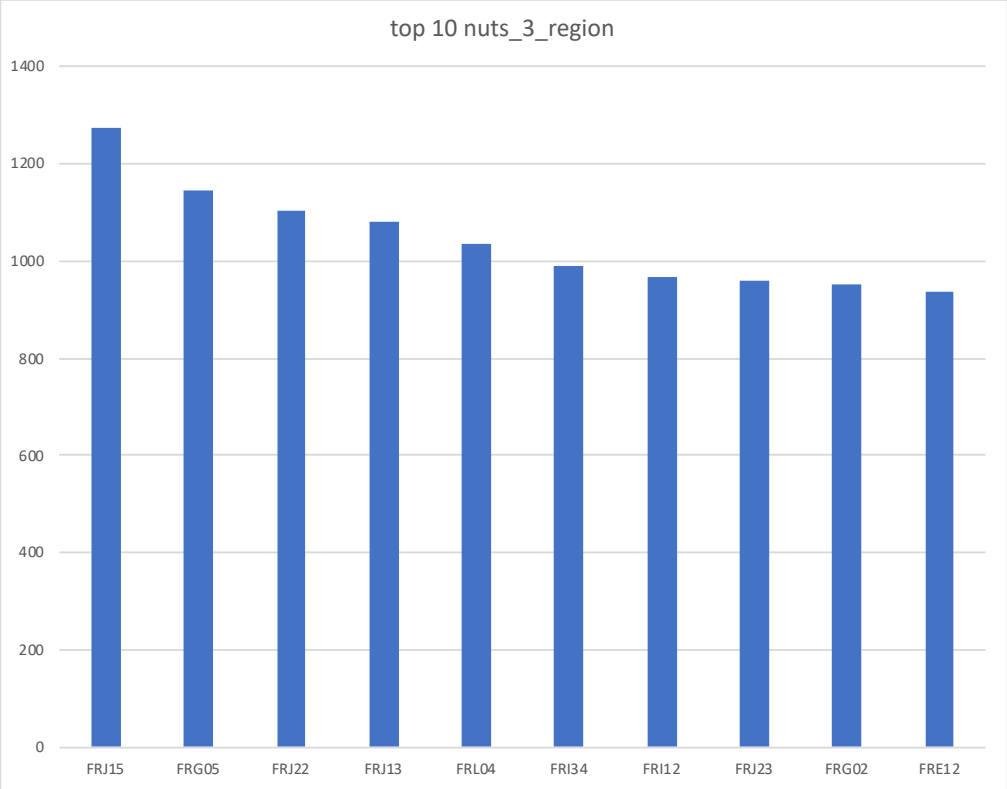



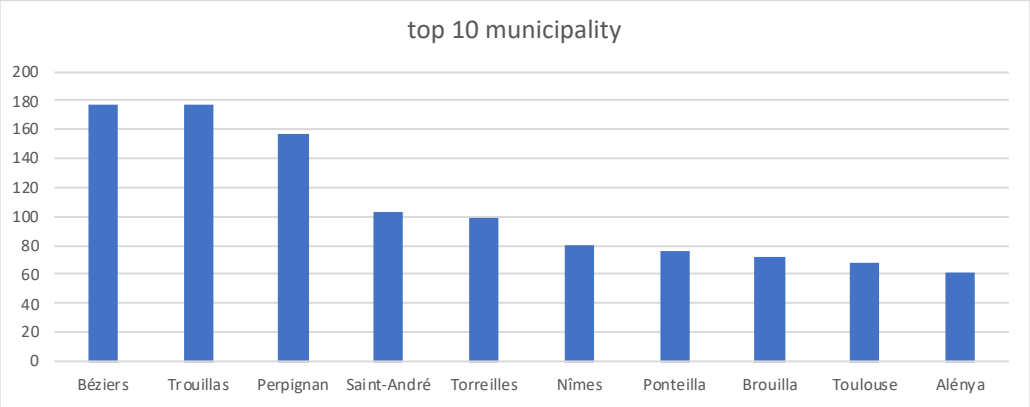

energy_source_level_3 without NULL
0%
12%
14%
74%
Other or unspecified
Biomass and biogas
Other bioenergy and renewable waste
Sewage and landfill gas

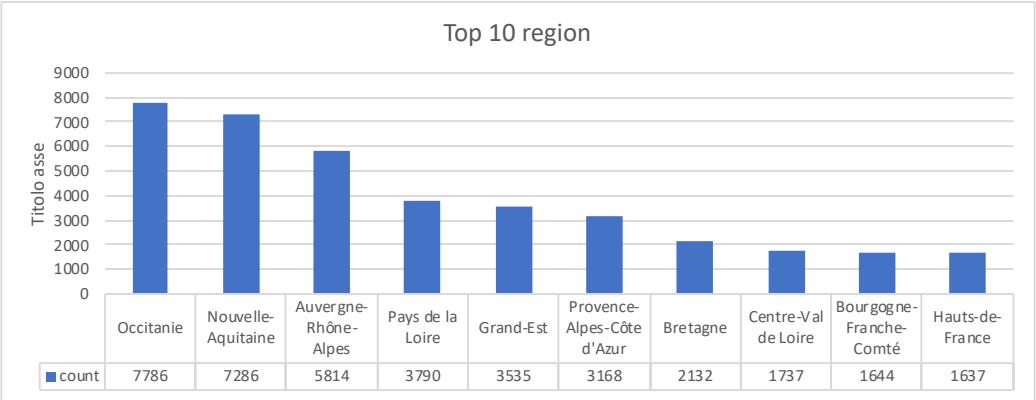

technology
0%
4%
3%
1%
92%
Photovoltaics
Other or unspecified technology
Onshore
NULL
Run-of-river

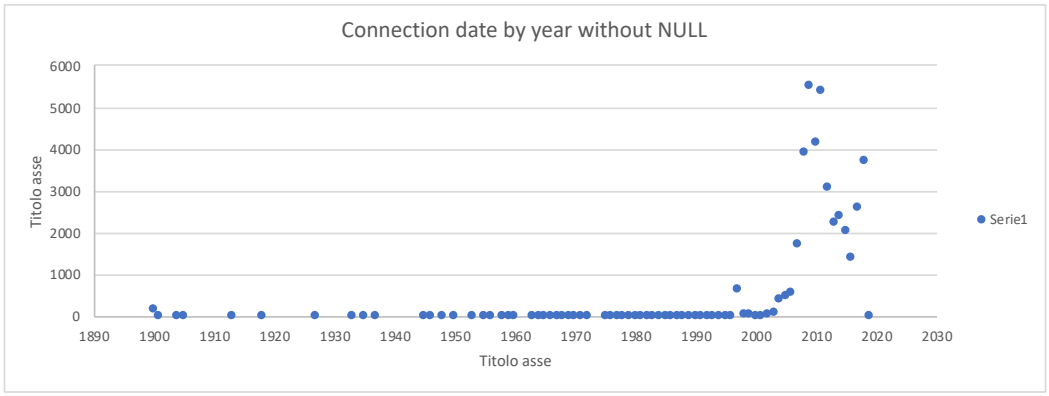

nuts_1_region
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
FRJ
FRI
FRK
FRF
FRG
FRC
FRE
FRL
FRD
FRB
FRH
FR1
NULL
FRM


nuts_2_region
6000
5000
4000
3000
2000
1000
0
FRJ2
FRK2
FRJ1
FRG0
FRI1
FRL0
FRI3
FRK1
FRB0
FRF3
FRH0
FRC1
FRF2
FRE2
FRE1
FRI2
FRC2
FR10
FRF1
FRD1
FRD2
NULL
FRM0

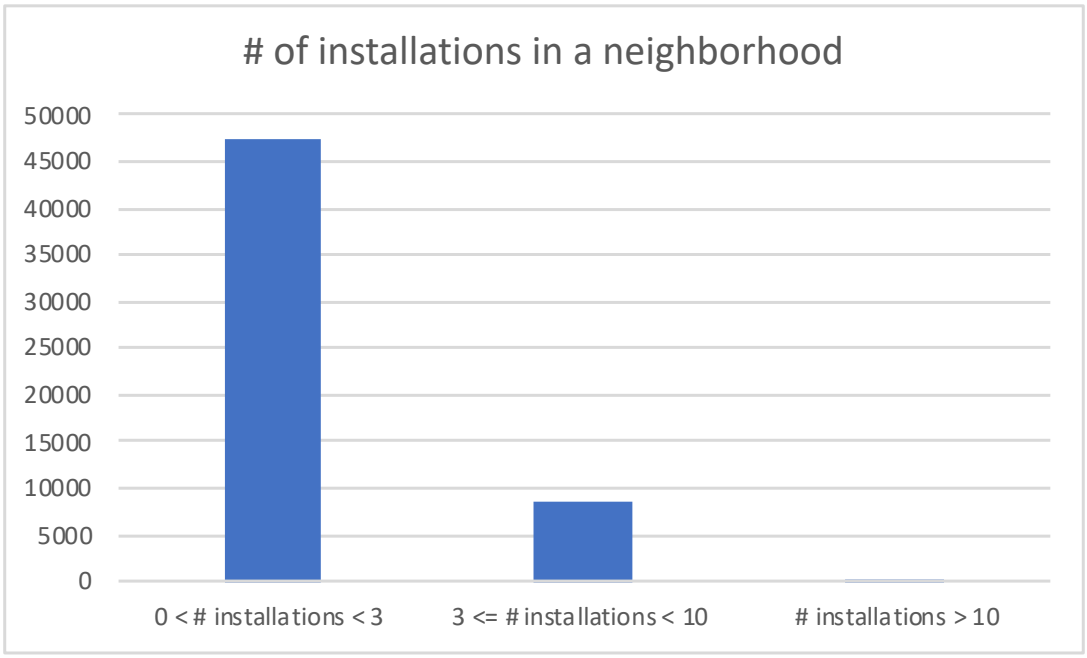

top 10 nuts_3_region
1400
1200
1000
800
600
400
200
0
FRJ15
FRG05
FRJ22
FRJ13
FRL04
FRI34
FRI12
FRJ23
FRG02
FRE12


top 10 municipality
200
180
160
140
120
100
80
60
40
20
0
Béziers
Trouillas
Perpignan
Saint-André
Torreilles
Nîmes
Ponteilla
Brouilla
Toulouse
Alénya


Top 10 region
Titolo asse
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Occitanie
Nouvelle-Aquitaine
Auvergne-Rhône-Alpes
Pays de la Loire
Grand-Est
Provence-Alpes-Côte d'Azur
Bretagne
Centre-Val de Loire
Bourgogne-Franche-Comté
Hauts-de-France
count
7786
7286
5814
3790
3535
3168
2132
1737
1644
1637


Connection date by year without NULL
Titolo asse
6000
5000
4000
3000
2000
1000
0
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
Titolo asse
Serie1

| disconnection_date | count |
|---|---|
| NULL | 56087 |
| 2018 | 3 |
| 2031 | 2 |
| 2029 | 1 |
| 2025 | 1 |
| 2030 | 1 |
| 2034 | 1 |
| 2037 | 1 |




#### 5.2.5.1 - Domain Correctness

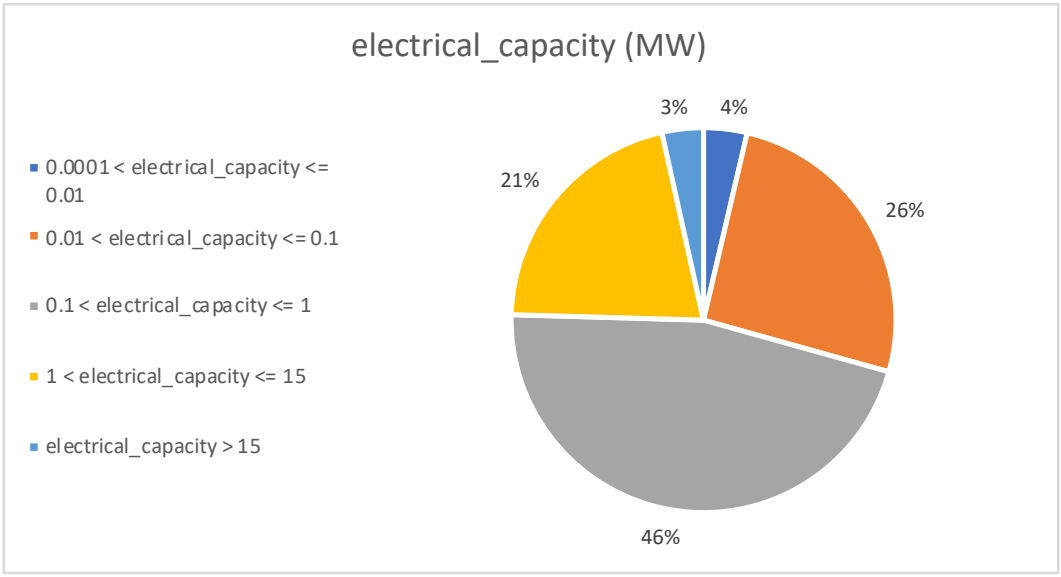
- Energy Source Level 2: no value out of domain
- Municipality Code: no value equal to zero

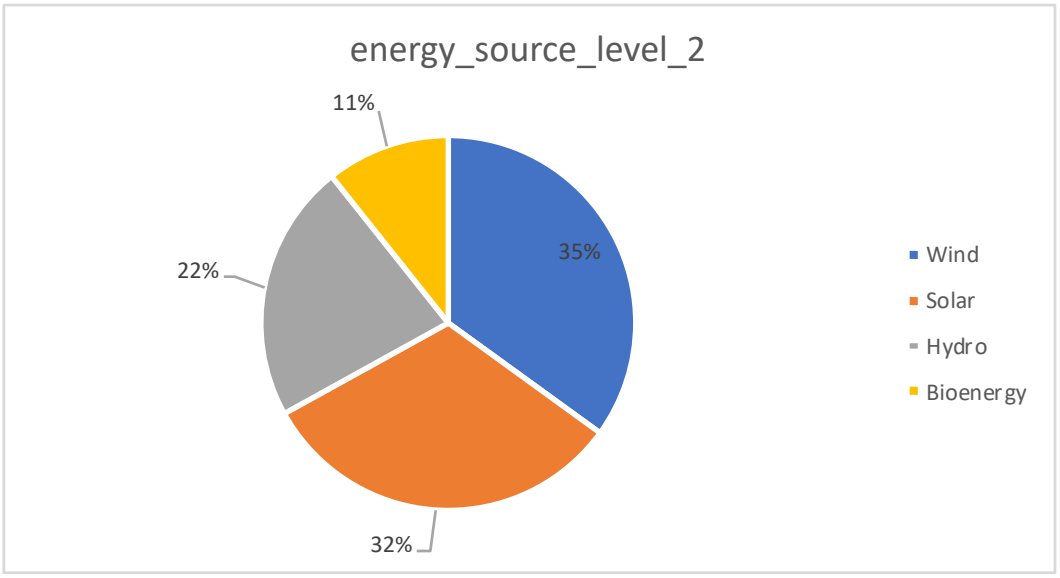
#### 5.2.5.2 - Duplication

260 duplicated rows

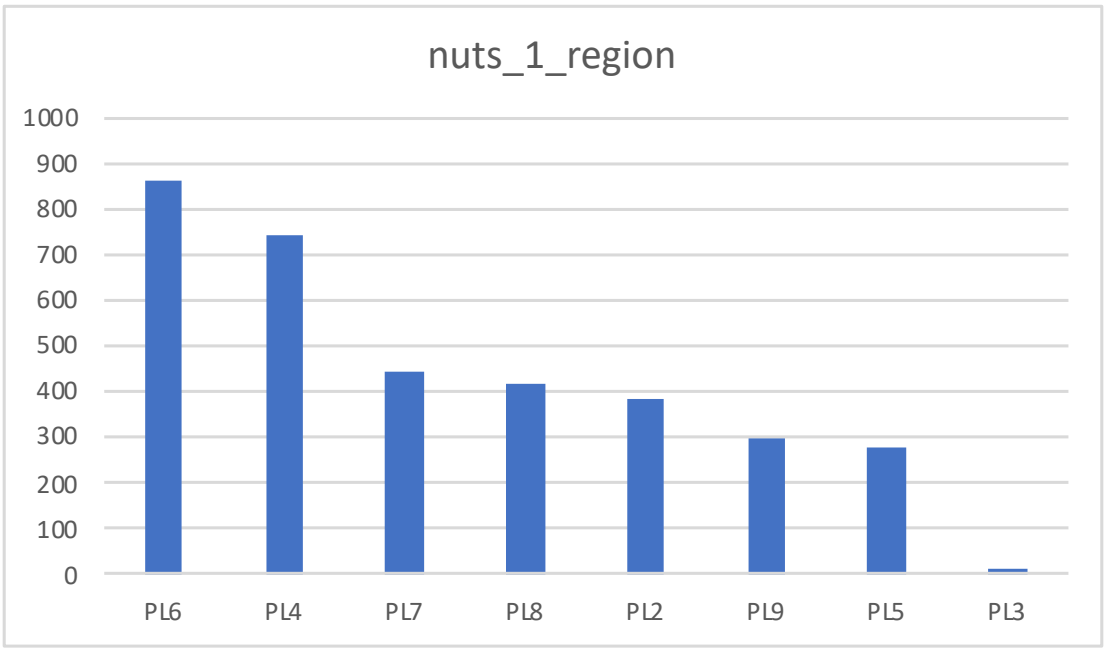
## 5.2.6 - Poland

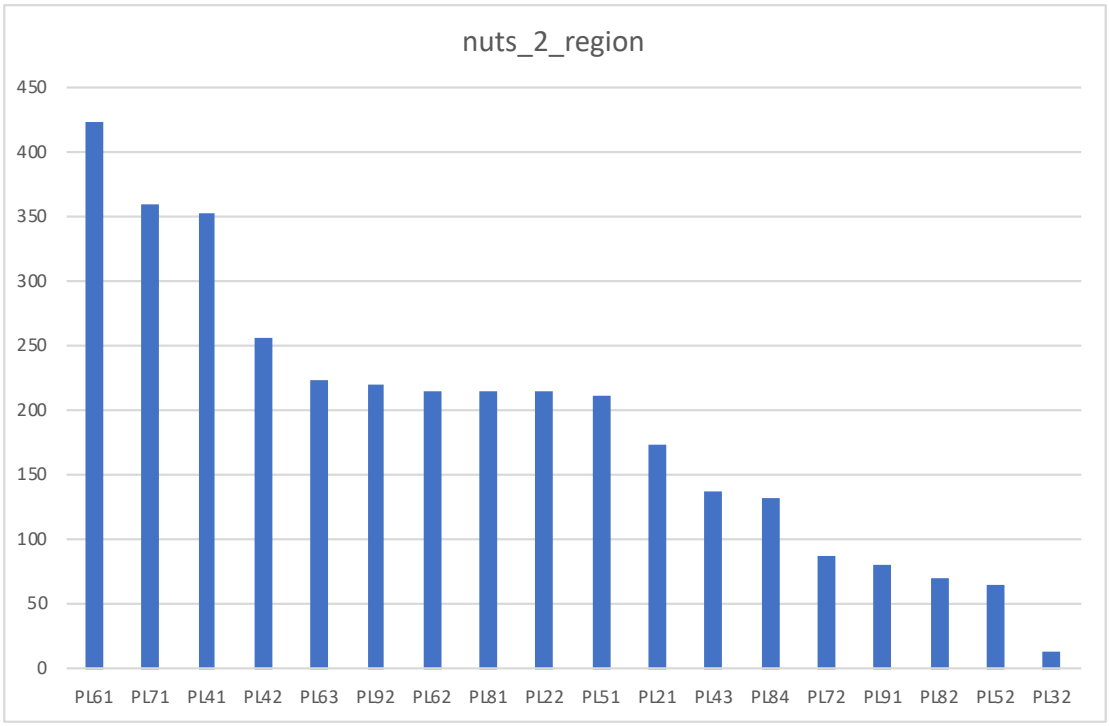
| electrical_capacity | value |
|---|---|
| min | 1 |
| max | 24.3 |

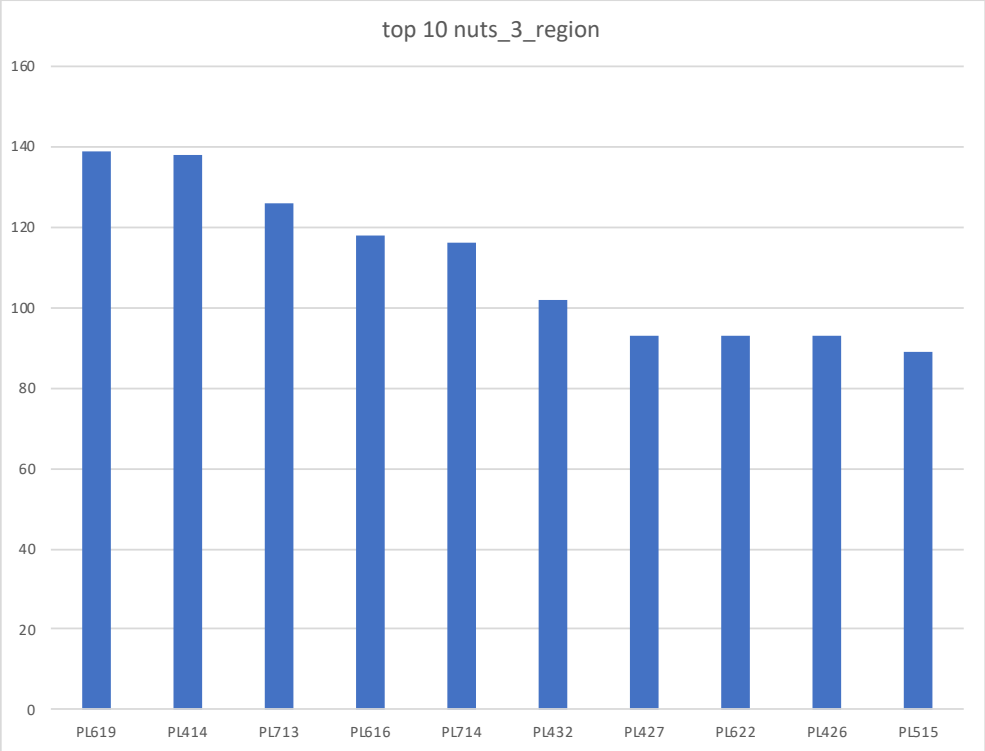


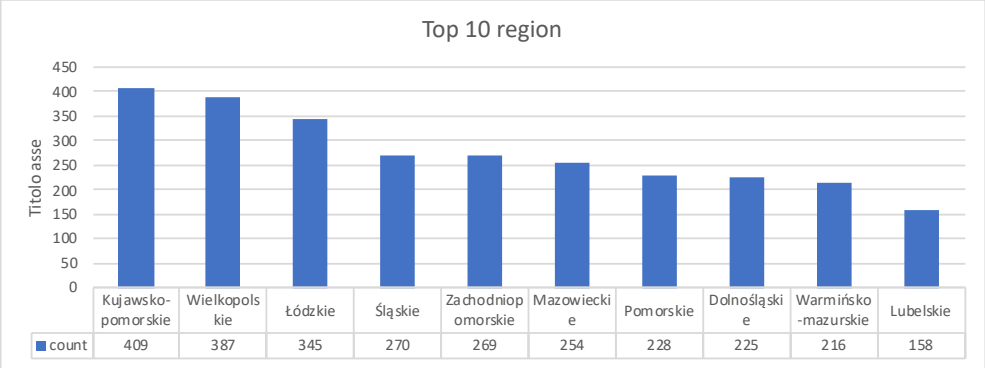



| energy_source_level_3 | count |
|---|---|
| NULL | 3082 |
| Biomass and biogas | 369 |






### 5.2.6.1 - Domain Correctness

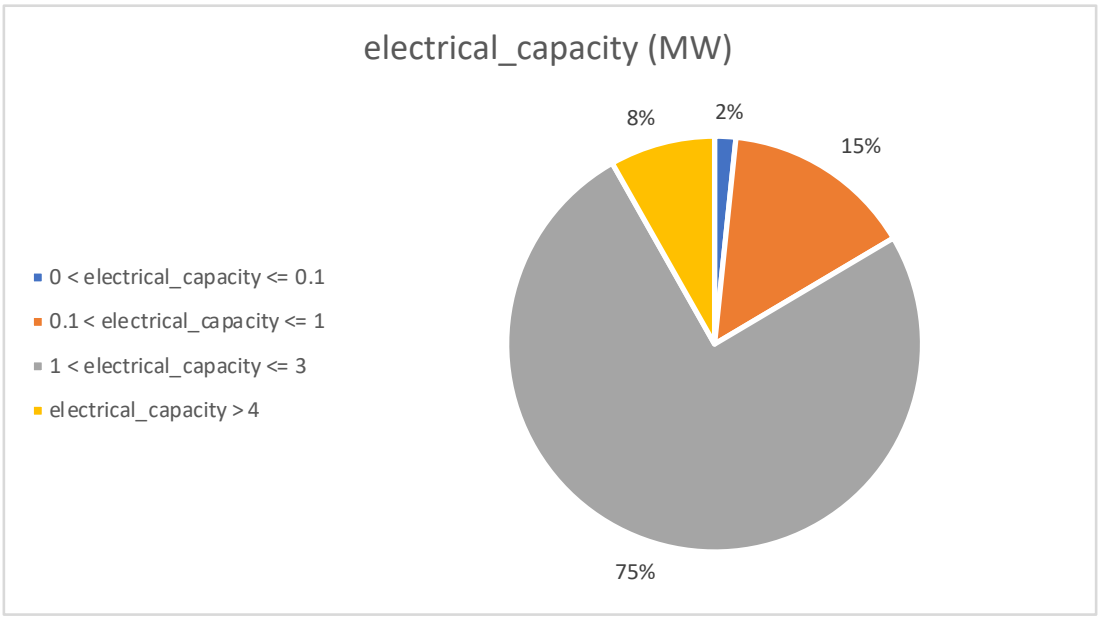
- Energy Source Level 2: no value out of domain

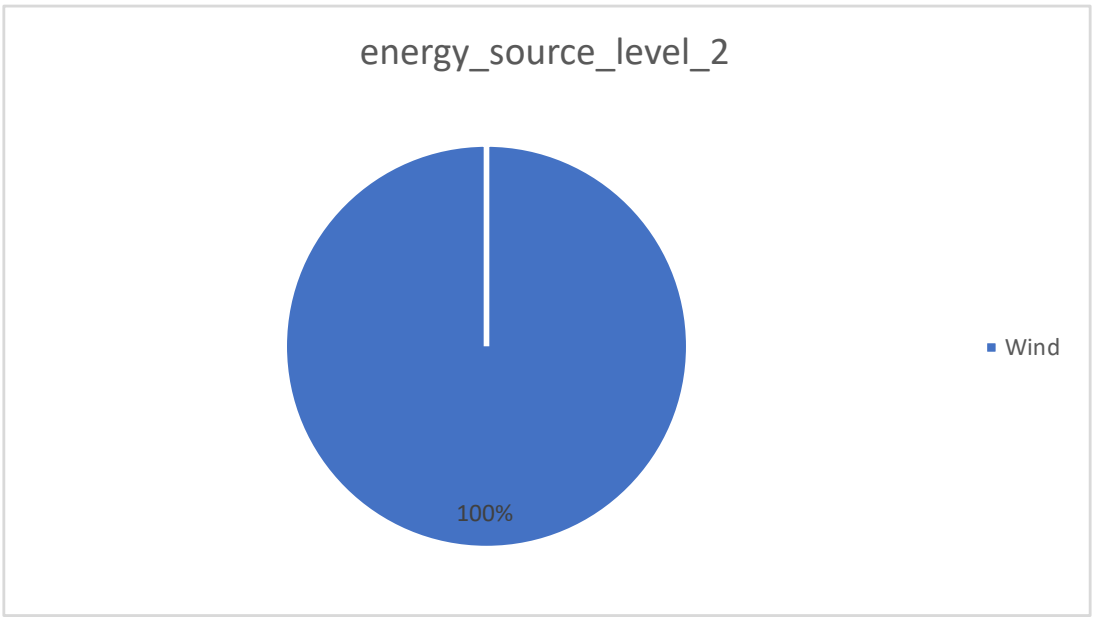
### 5.2.6.2 - Duplication

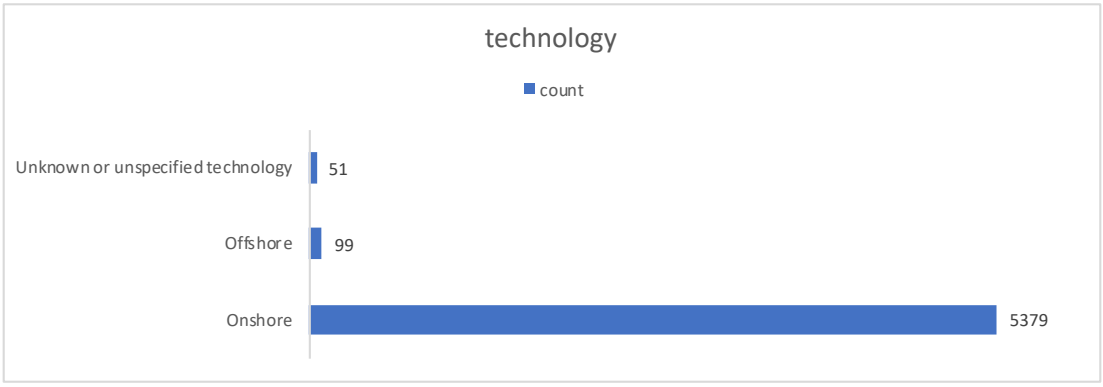
437 duplicated rows

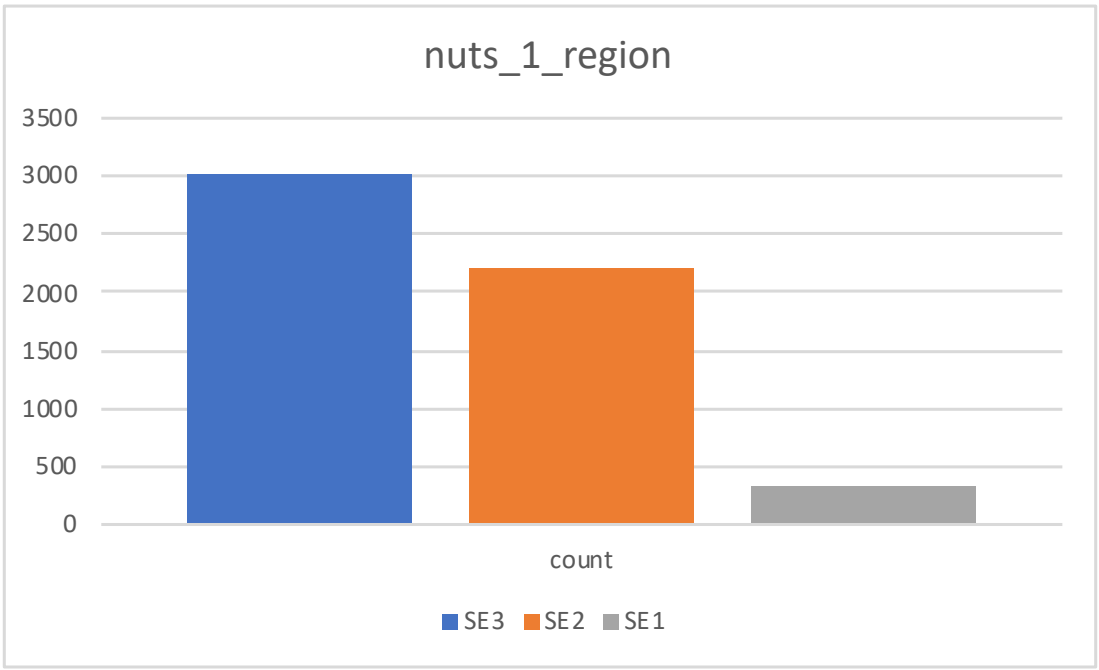
### 5.2.7 - Sweden

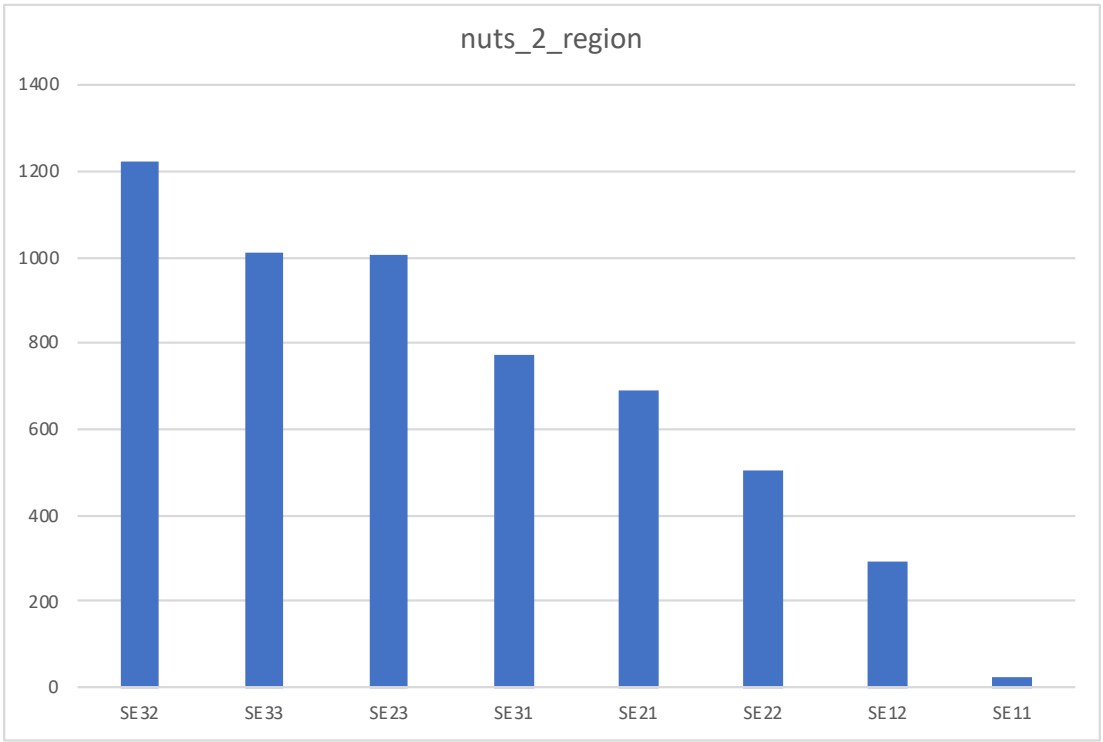
| electrical_capacity | value |
|---|---|
| min | 0 |
| max | 13 |

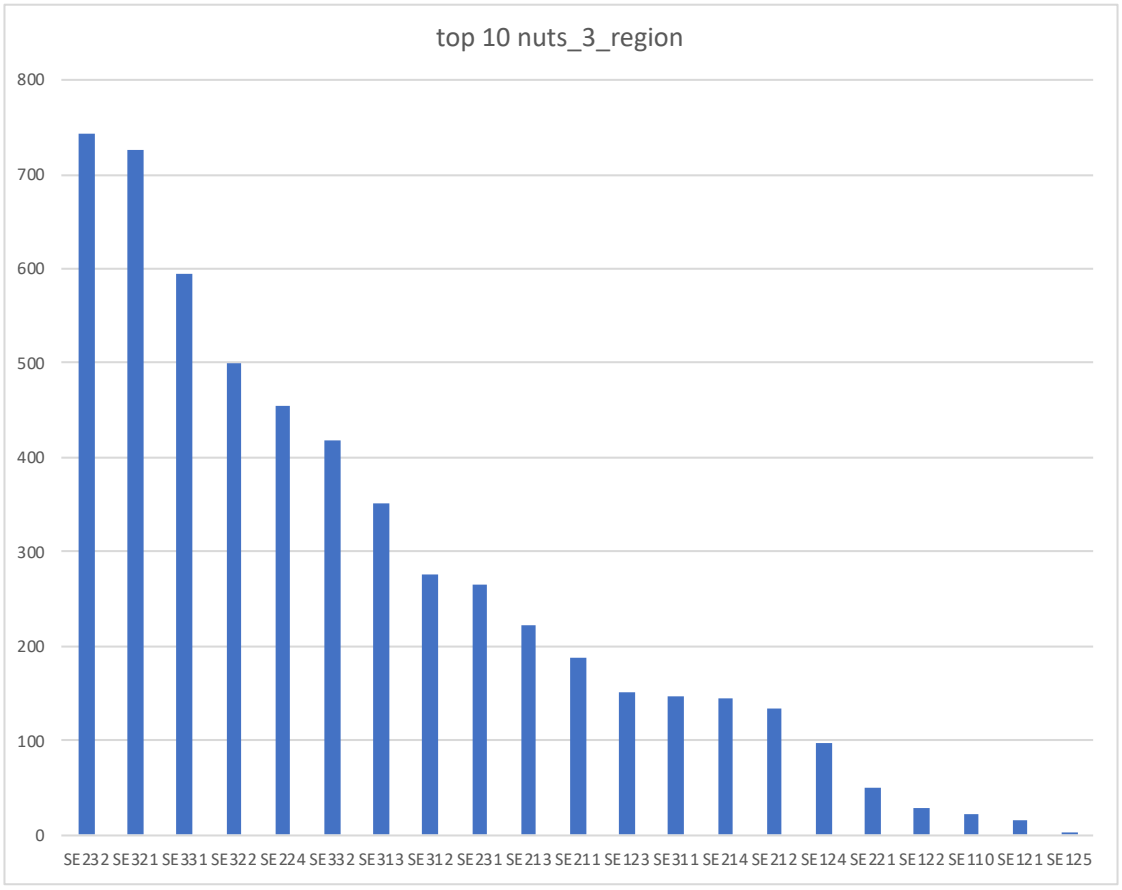


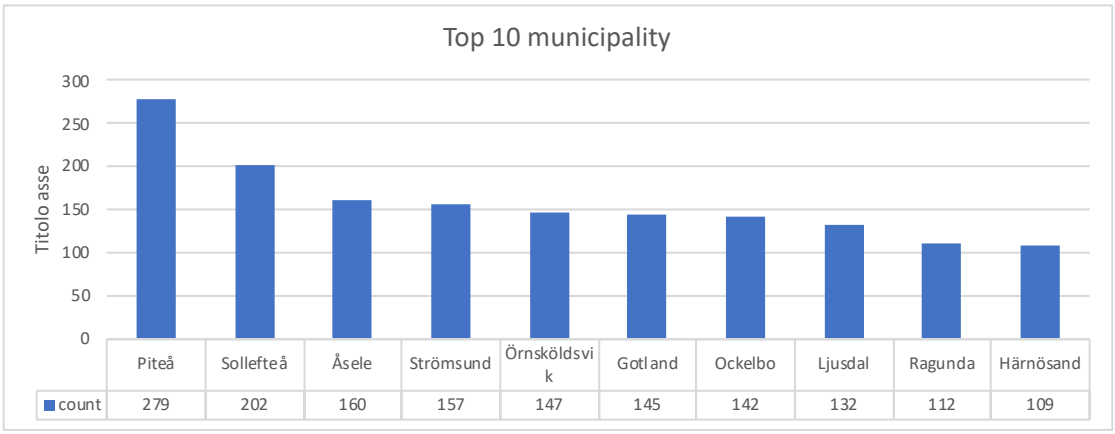



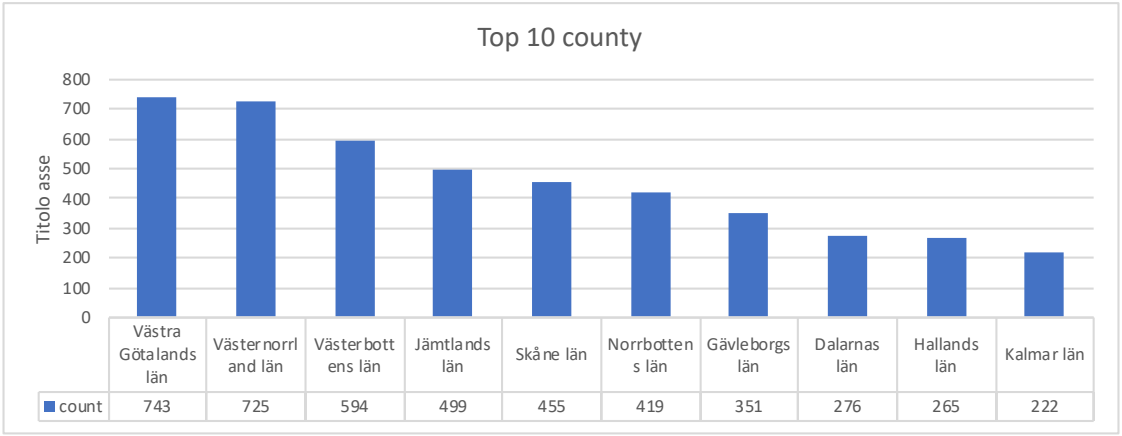

technology
count
Unknown or unspecified technology
51
Offshore
99
Onshore
5379

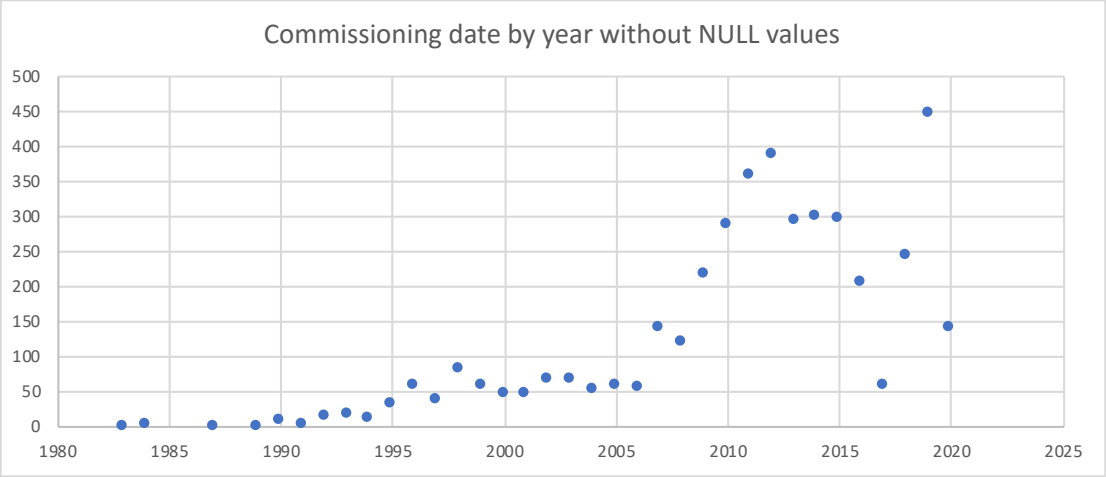

nuts_1_region
3500
3000
2500
2000
1500
1000
500
0
count
SE3
SE2
SE1

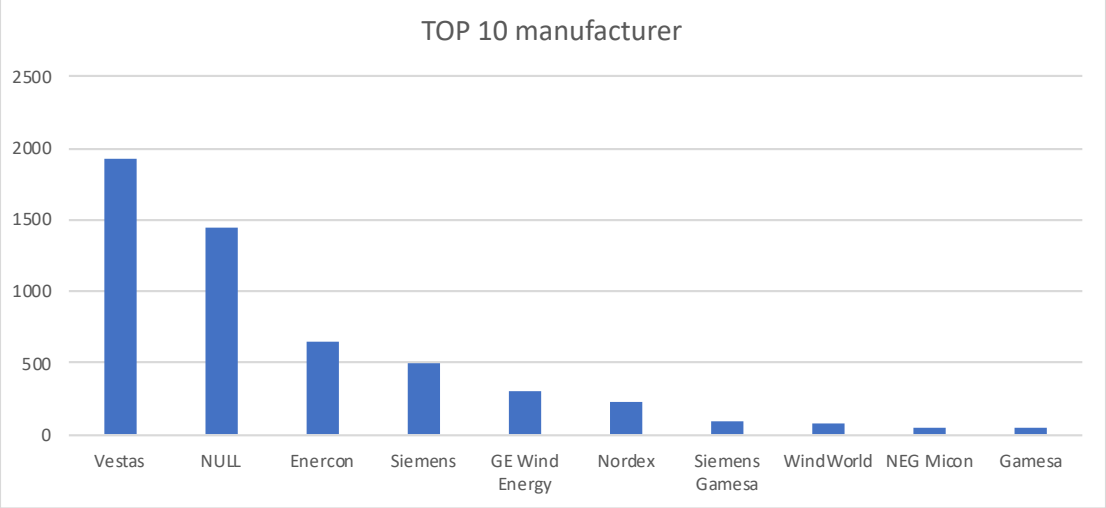

nuts_2_region
1400
1200
1000
800
600
400
200
0
SE32
SE33
SE23
SE31
SE21
SE22
SE12
SE11


top 10 nuts_3_region
800
700
600
500
400
300
200
100
0
SE232 SE321 SE331 SE322 SE224 SE332 SE313 SE312 SE231 SE213 SE211 SE123 SE311 SE214 SE212 SE124 SE221 SE122 SE110 SE121 SE125


Top 10 municipality
Titolo asse
300
250
200
150
100
50
0
Piteå
Sollefteå
Åsele
Strömsund
Örnsköldsvik
Gotland
Ockelbo
Ljusdal
Ragunda
Härnösand
count
279
202
160
157
147
145
142
132
112
109


Top 10 county
Titolo asse
800
700
600
500
400
300
200
100
0
Västra Götalands län
Västernorrland län
Västerbottens län
Jämtlands län
Skåne län
Norrbottens län
Gävleborgs län
Dalarnas län
Hallands län
Kalmar län
count
743
725
594
499
455
419
351
276
265
222




### 5.2.7.1 - Domain Correctness

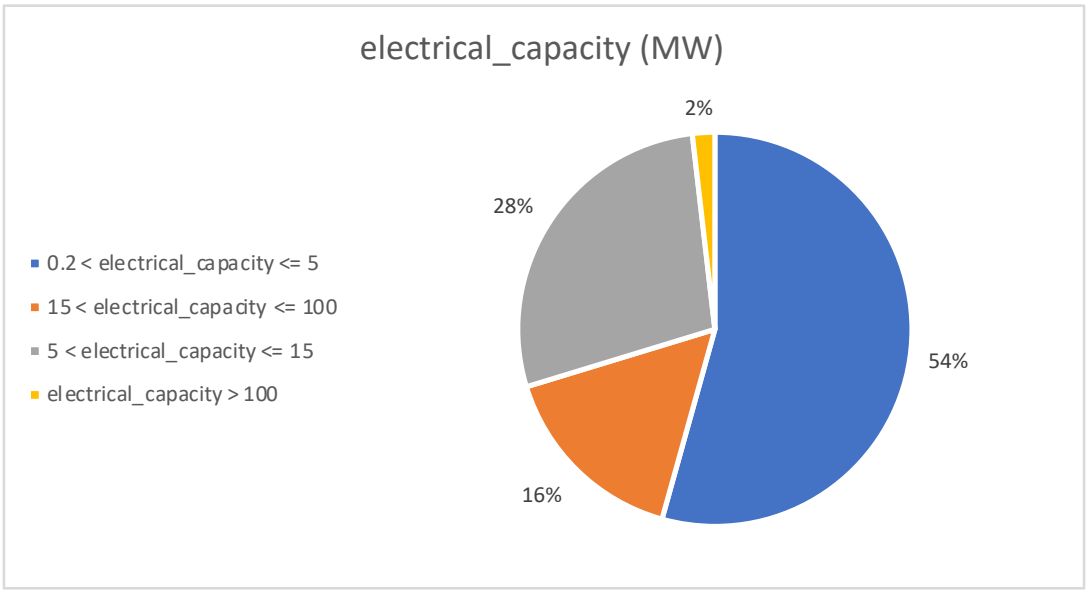
- Energy Source Level 2: no value out of domain

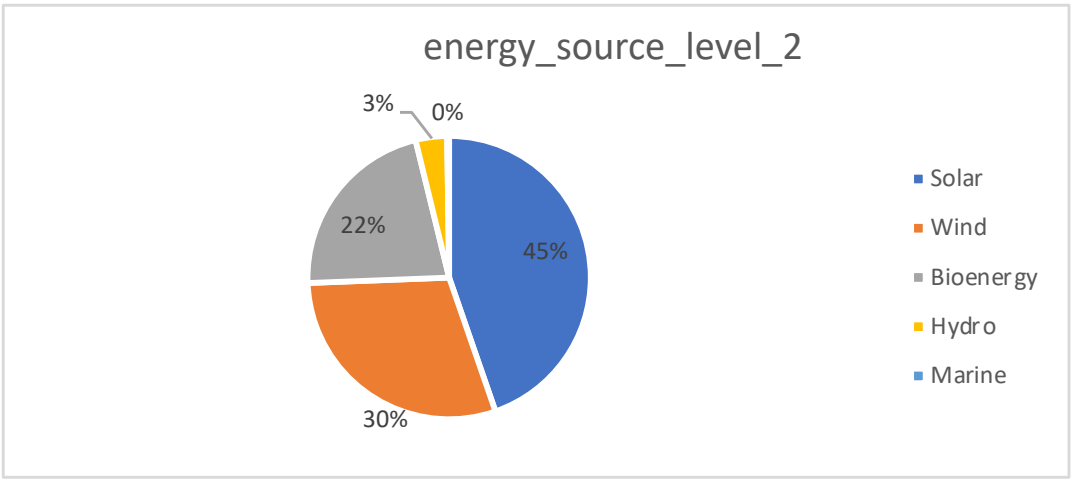
### 5.2.7.2 - Duplication

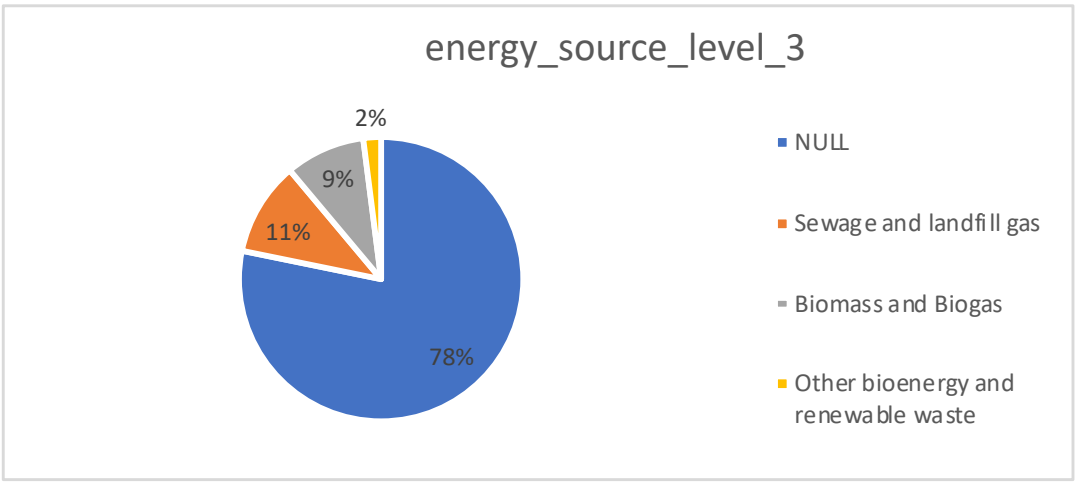
0 duplicated rows

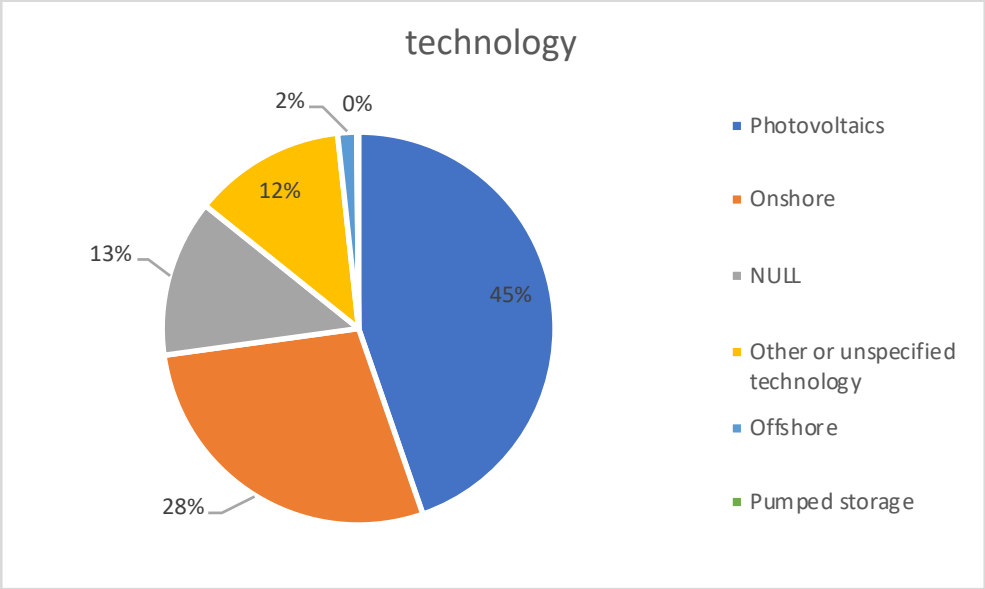
### 5.2.8 - United Kingdom

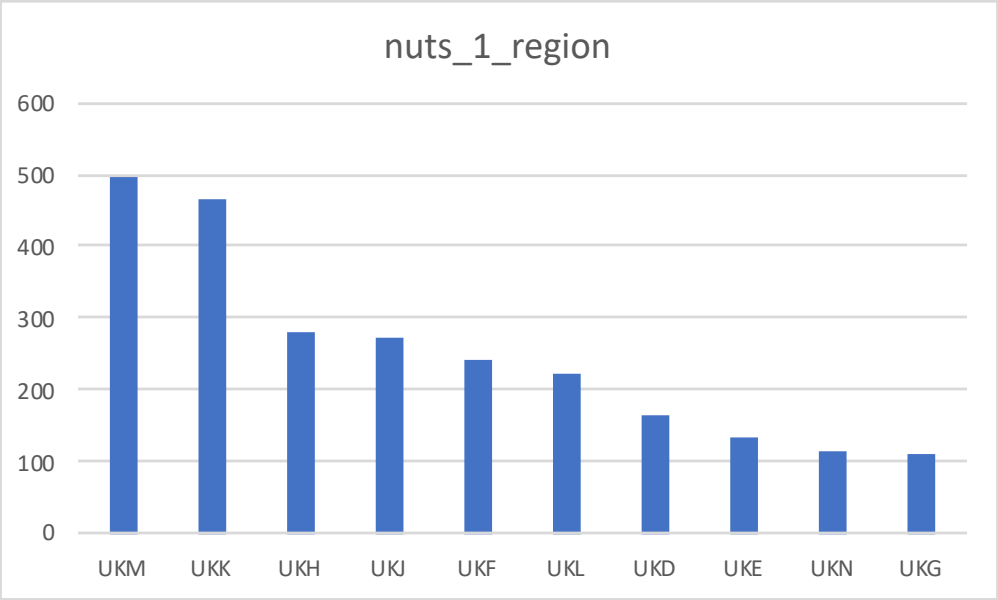
| electrical_capacity | value |
|---|---|
| min | 0.2 |
| max | 1728 |

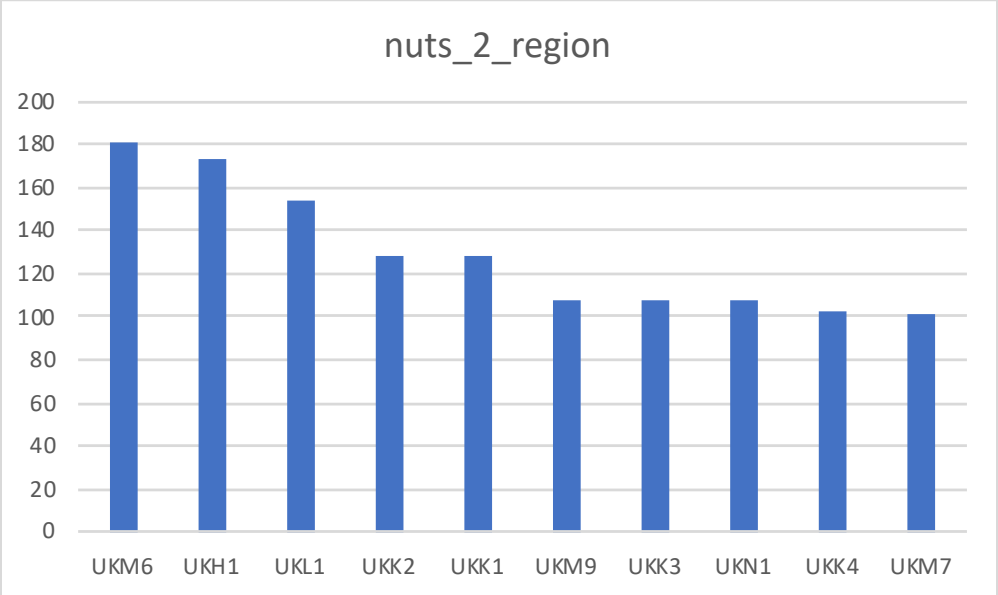


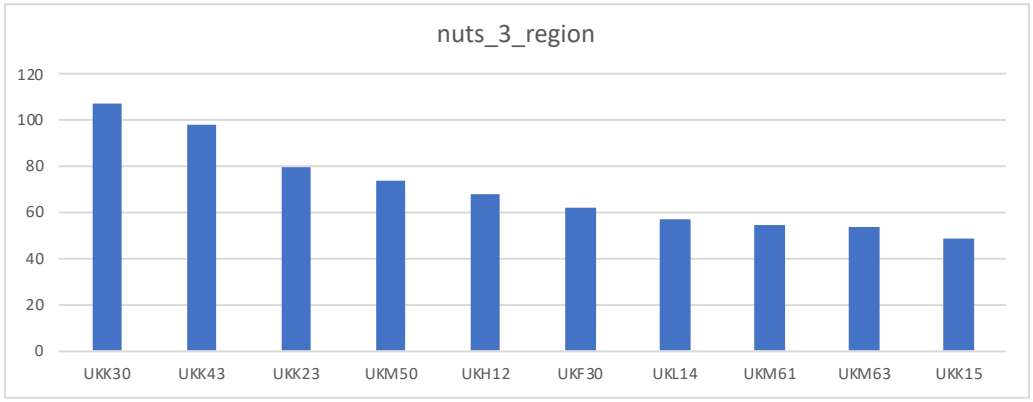



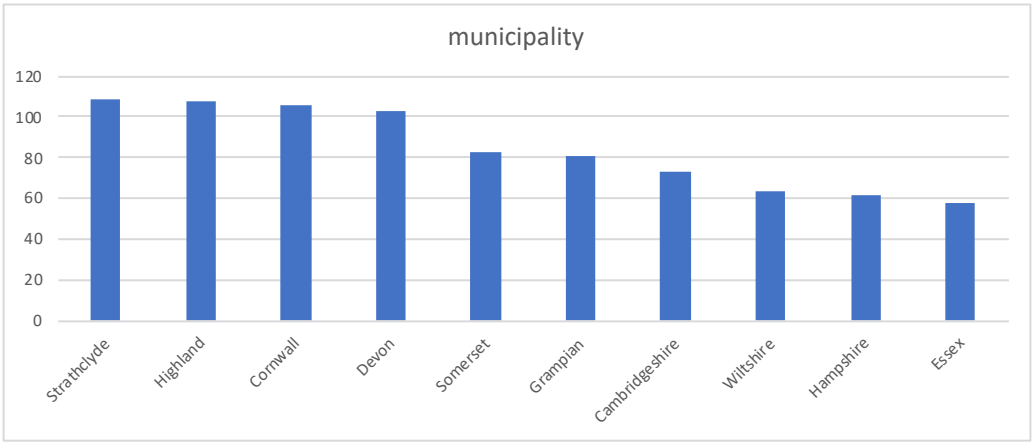



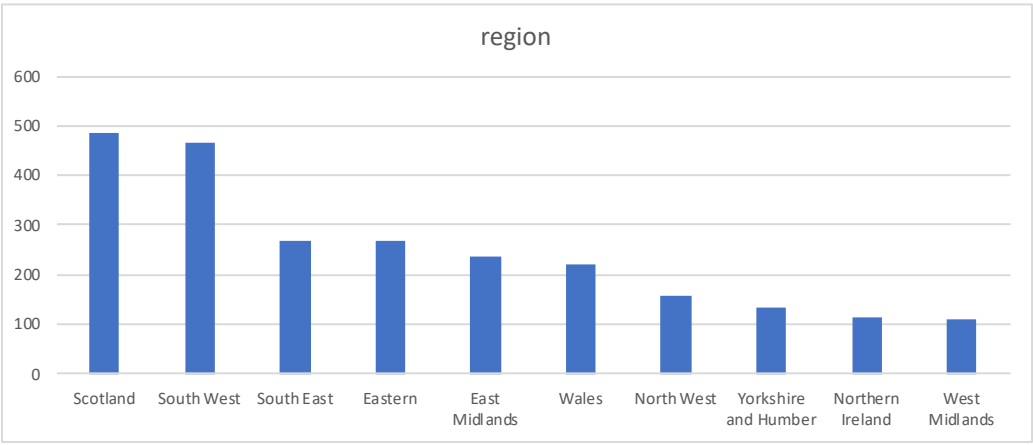

technology
2%
0%
12%
13%
45%
28%
Photovoltaics
Onshore
NULL
Other or unspecified technology
Offshore
Pumped storage

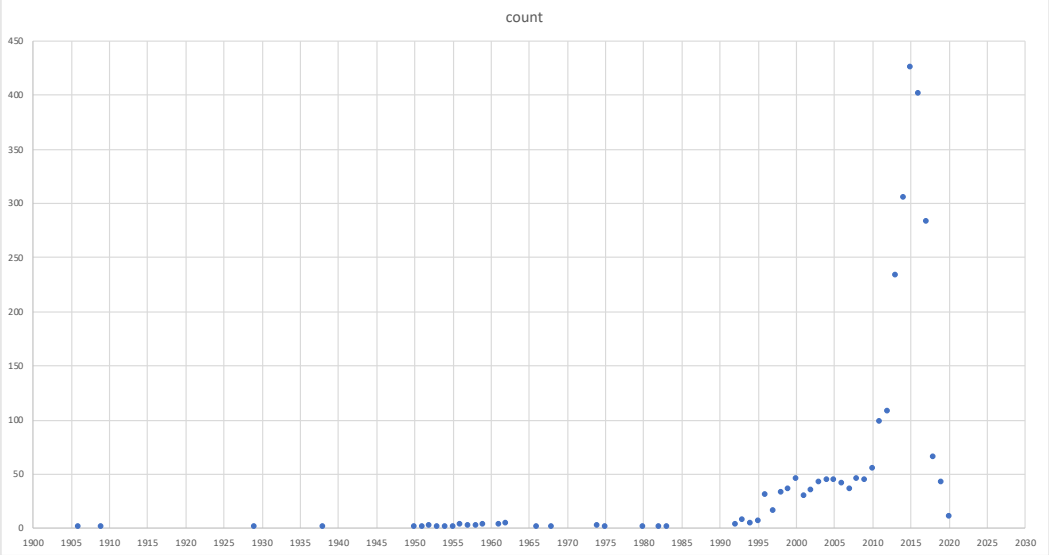

nuts_1_region
600
500
400
300
200
100
0
UKM
UKK
UKH
UKJ
UKF
UKL
UKD
UKE
UKN
UKG

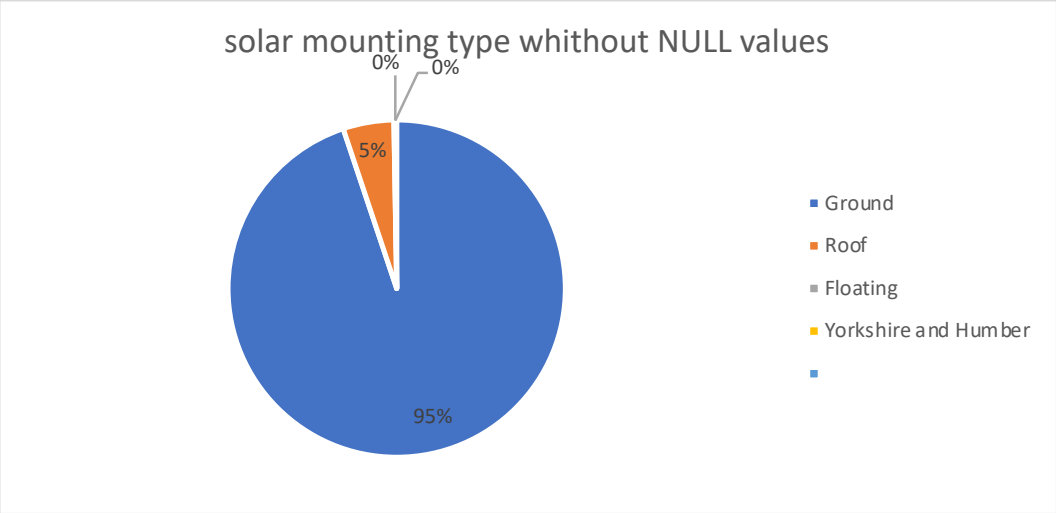

nuts_2_region
200
180
160
140
120
100
80
60
40
20
0
UKM6
UKH1
UKL1
UKK2
UKK1
UKM9
UKK3
UKN1
UKK4
UKM7

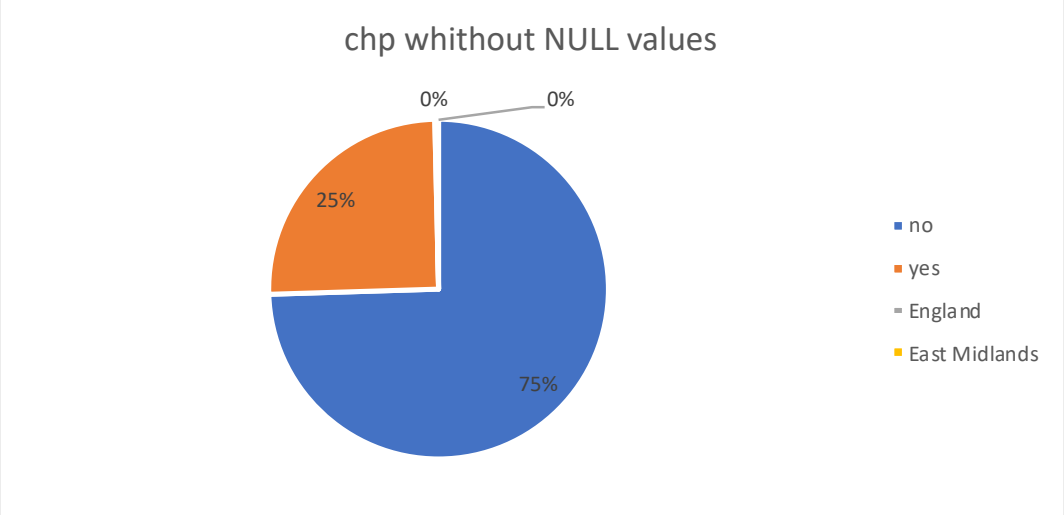

nuts_3_region
120
100
80
60
40
20
0
UKK30
UKK43
UKK23
UKM50
UKH12
UKF30
UKL14
UKM61
UKM63
UKK15


municipality
120
100
80
60
40
20
0
Strathclyde
Highland
Cornwall
Devon
Somerset
Grampian
Cambridgeshire
Wiltshire
Hampshire
Essex

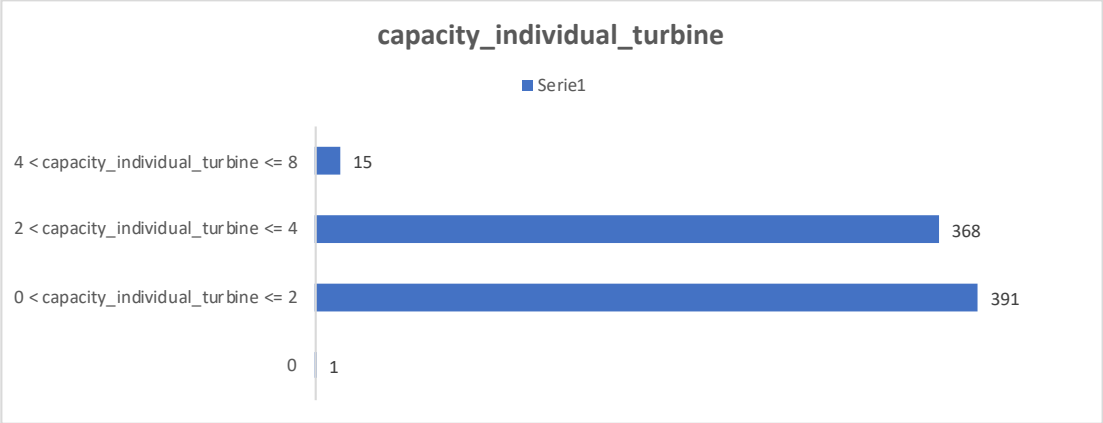

region
600
500
400
300
200
100
0
Scotland
South West
South East
Eastern
East Midlands
Wales
North West
Yorkshire and Humber
Northern Ireland
West Midlands



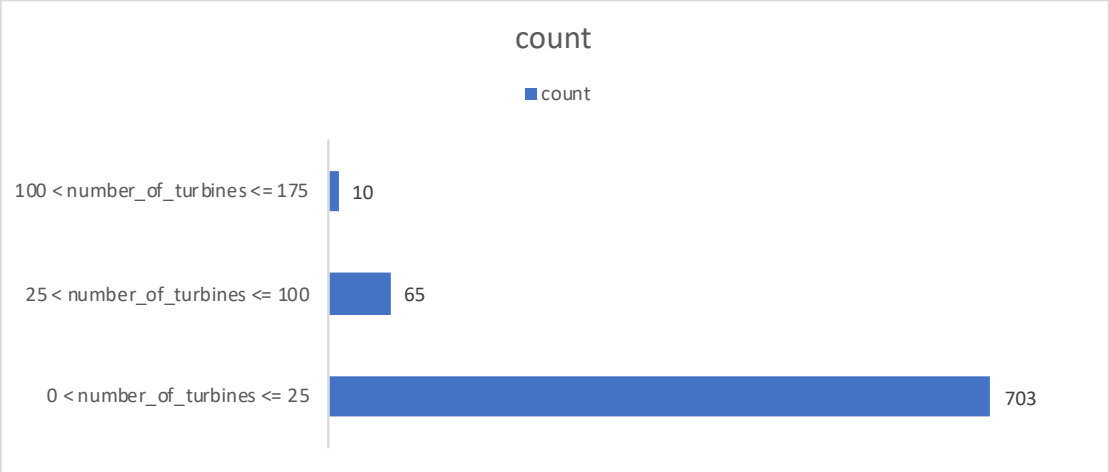



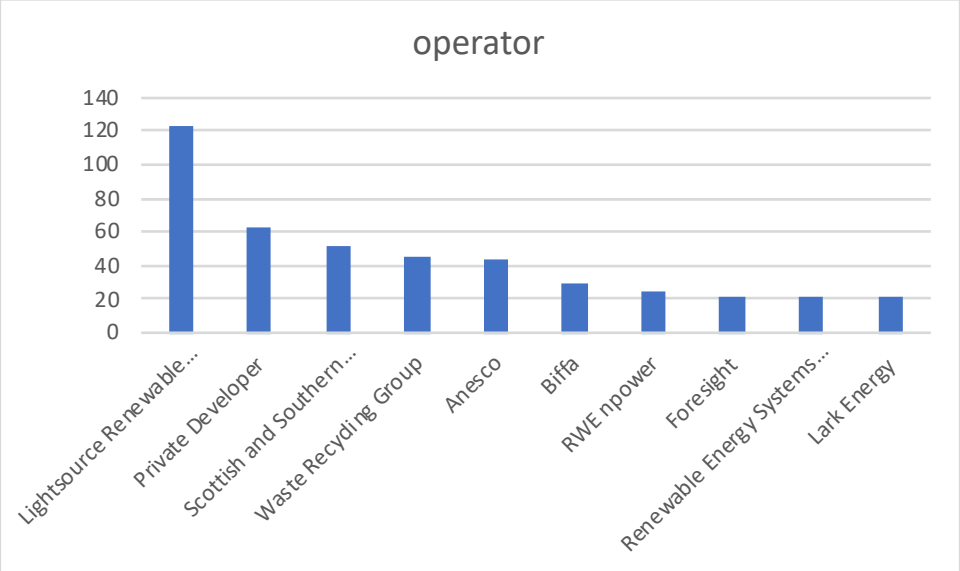




| capacity_individual_turbine | values |
|---|---|
| min | 0.0 |
| max | 8.5 |




| number_of_turbines | values |
|---|---|
| min | 1 |
| max | 175 |




### 5.2.8.1 - Domain Correctness

- Energy Source Level 2: no value out of domain
- Municipality Code: no value equal to 0
- CHP: 2.057 value out of domain
- Solar Mounting Type: 1.491 value out of domain (empty)

### 5.2.8.2 - Duplication

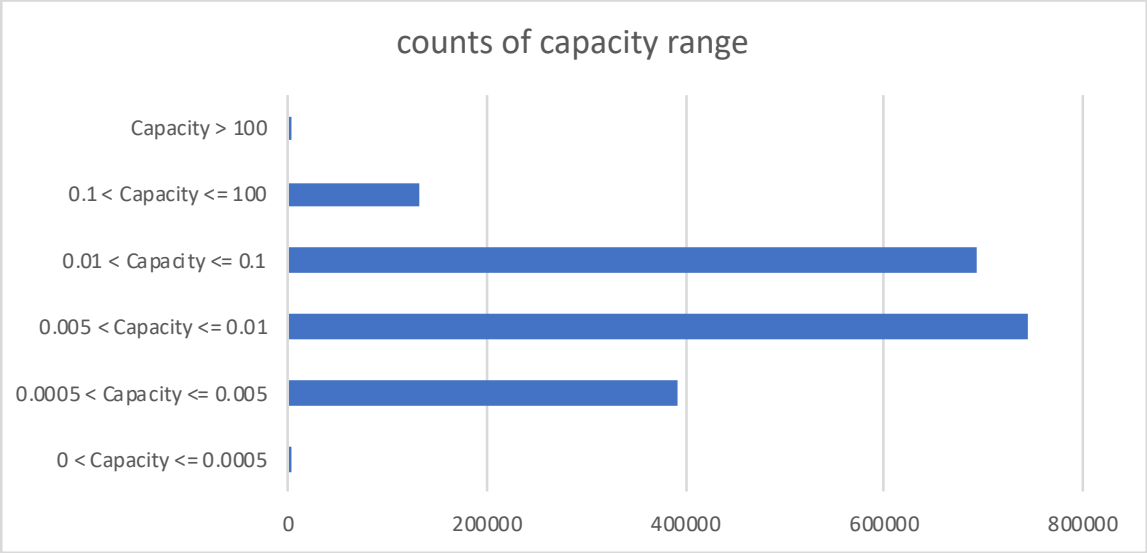
0 duplicated rows

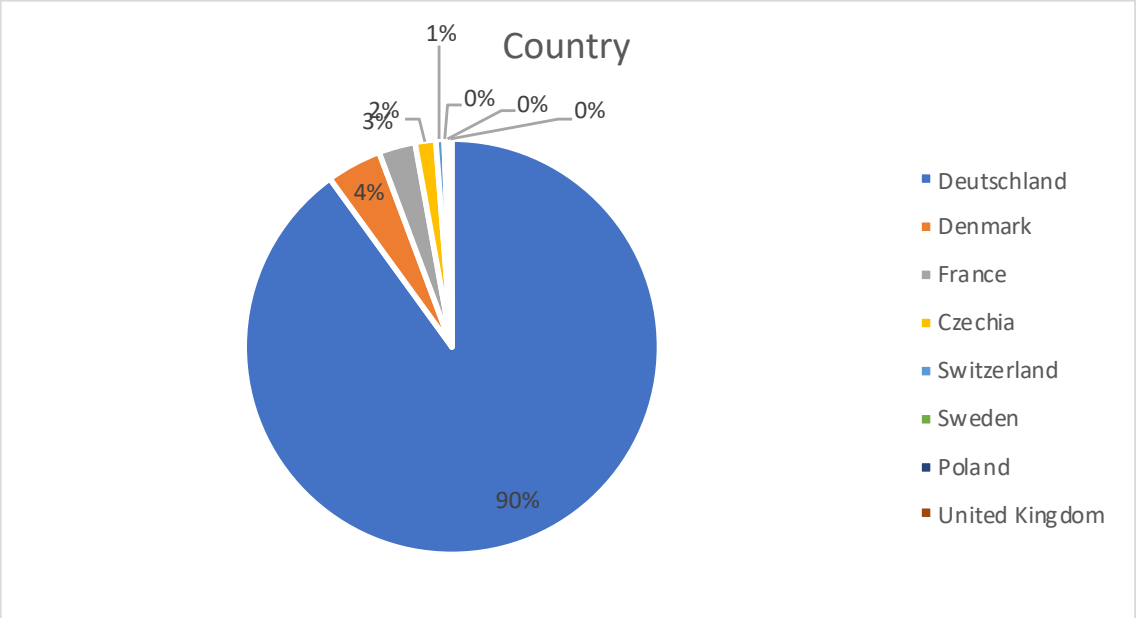
## 5.3 - RICONCILED DATASET

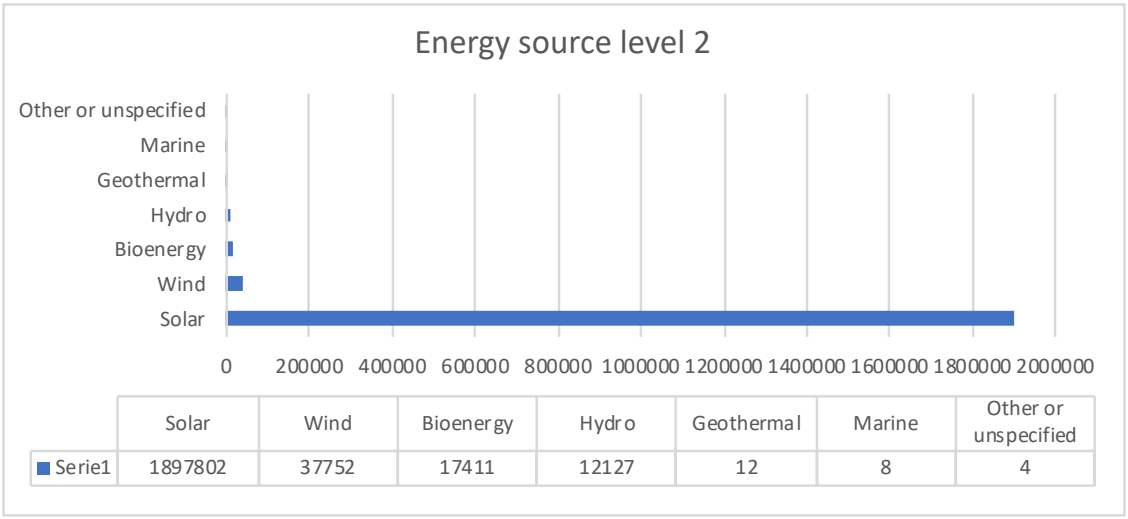
In this paragraph we will perform the null and domain analysis on the reconciled dataset. The reconciled dataset was created by taking the information common to all national datasets and selecting some attributes that allow you to detail the wind and solar plants. All the details relating to the associations between the reconciled attributes and the attributes of the individual datasets are available at the following link:https://docs.google.com/document/d/10B2QhqPwg5bR8Yn_4yd2jtjCb4wUojfD/edit?usp=sharing&ouid=107988850620628418604&rtpof=true&sd=true

### 5.3.1 - Domain Analysis

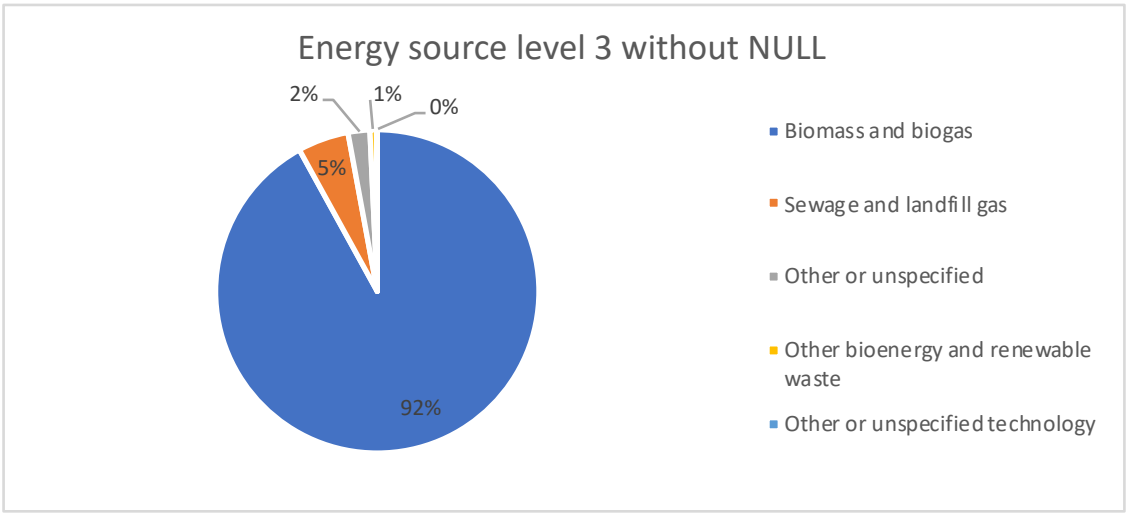
| electrical capacity in MW | value |
|---|---|
| min | 0 |
| max | 1728 |

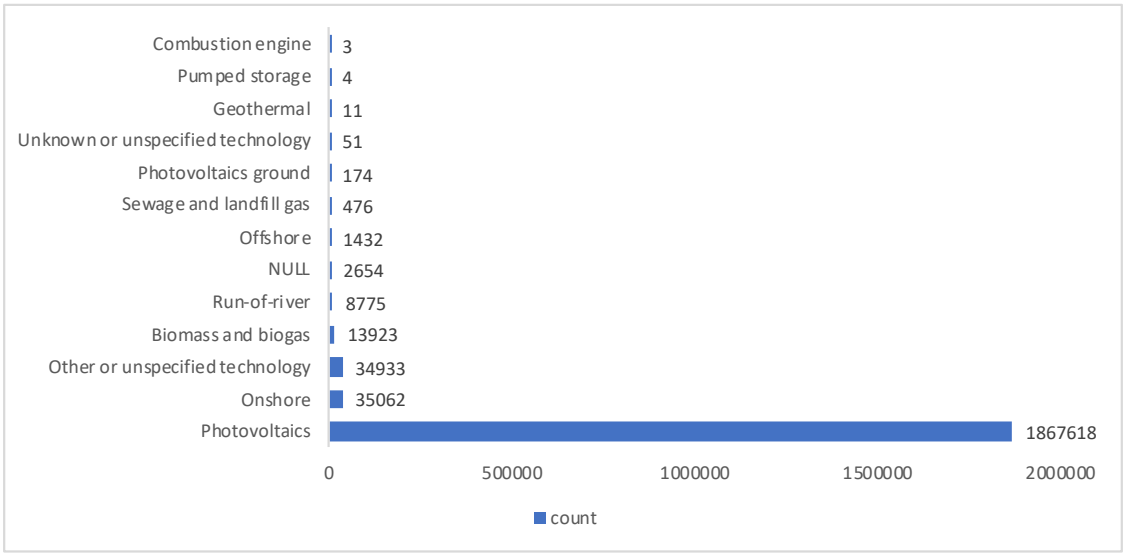


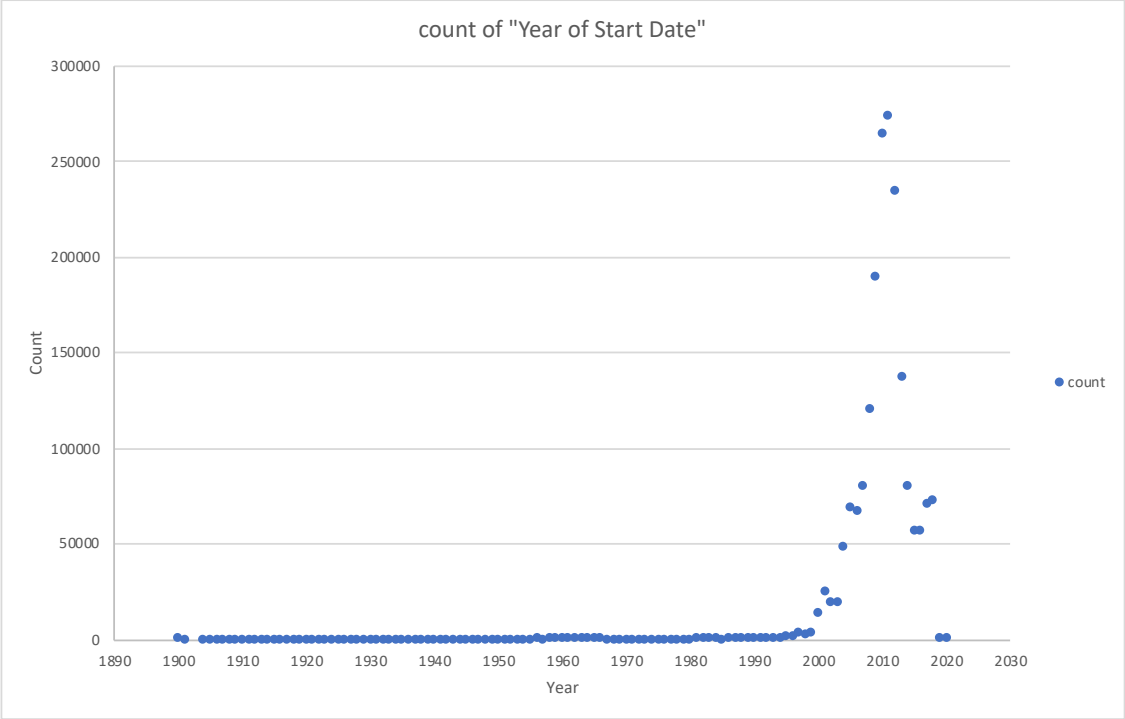



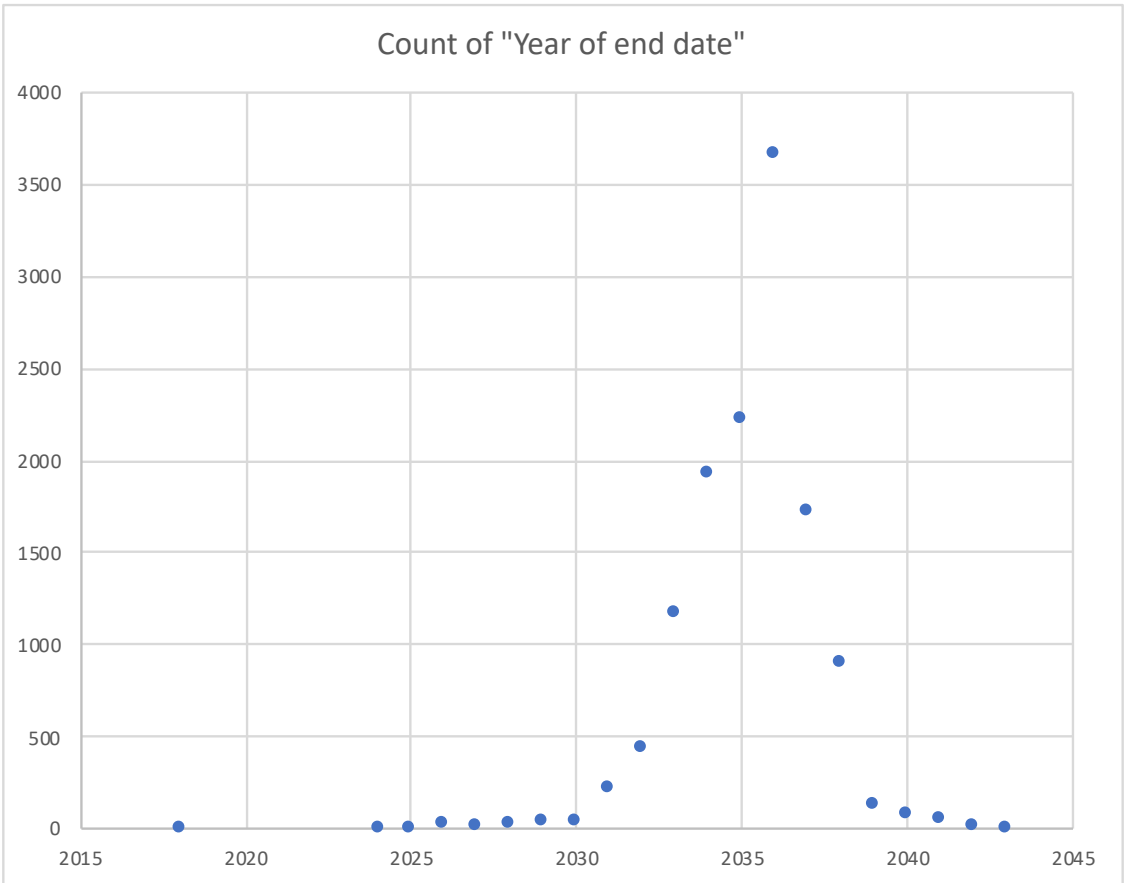



| | Solar | Wind | Bioenergy | Hydro | Geothermal | Marine | Other or unspecified |
|---|---|---|---|---|---|---|---|
| Serie1 | 1897802 | 37752 | 17411 | 12127 | 12 | 8 | 4 |




count of "Year of Start Date"
300000
250000
200000
150000
100000
50000
0
Count
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
Year
count


Count of "Year of end date"
4000
3500
3000
2500
2000
1500
1000
500
0
2015
2020
2025
2030
2035
2040
2045

| Top 10 Company | plant count |
|---|---|
| Bayernwerk Netz GmbH | 269228 |
| Westnetz GmbH | 158856 |
| Netze BW GmbH | 156355 |
| NULL | 127113 |
| LEW Verteilnetz GmbH | 72932 |
| EWE NETZ GmbH | 60741 |
| N-ERGIE Netz GmbH | 50168 |
| EAM Netz GmbH | 42609 |
| Mitteldeutsche Netzgesellschaft Strom mbH | 41899 |
| Avacon Netz GmbH | 38898 |

As anticipated in the previous chapters, the dominant nation within the dataset is Germany therefore the results of the analyzes performed on the respective dataset influence the reconciliation dataset.
Details are available at the following link: https://docs.google.com/spreadsheets/d/1_MTevOk2u8r7Rijif8QvcXkVZbyacc6-/edit?usp=sharing&ouid=107988850620628418604&rtpof=true&sd=true

### 5.3.2 - Null Value Analysis

As can be seen from figures 5.3.2.1 and 5.3.2.2, the dataset has a very high concentration of null values in all the attributes that differ from the location of the plant.
The high presence of null values in the tariff attribute severely limits the analysis relating to the contracts stipulated by the various distribution companies.
Similar speeches are valid for the technical attributes of the plant (rotor diameter, hub height, etc.) whose lack of completeness does not allow to identify any correlations between production and details.

| column name | null values | percentage |
|---|---|---|
| Capacity | 0 | 0,00 |
| Energy_source_level_2 | 0 | 0,00 |
| Energy_source_level_3 | 1947692 | 99,11 |
| Technology | 2654 | 0,14 |
| NUTS1 | 1180 | 0,06 |
| NUTS2 | 1180 | 0,06 |
| NUTS3 | 1180 | 0,06 |
| Lon | 5853 | 0,30 |
| Lat | 5853 | 0,30 |
| Municipality | 1154992 | 58,77 |
| Municipality_code | 155616 | 7,92 |
| Postcode | 72741 | 3,70 |
| Address | 1949736 | 99,22 |
| Region | 99680 | 5,07 |
| Start_date | 51777 | 2,63 |
| End_date | 1952388 | 99,35 |
| Company | 127113 | 6,47 |
| Tariff | 1952398 | 99,35 |
| Project_name | 1952398 | 99,35 |
| Production | 1952398 | 99,35 |
| Wind_turbine_Hub_Height | 1958911 | 99,68 |
| Wind_turbine_Rotor_Diameter | 1958911 | 99,68 |
| Wind_turbine_Model | 1958916 | 99,68 |
| Wind_Turbine_Capacity | 1964336 | 99,96 |
| Wind_Turbine_Number | 1964336 | 99,96 |
| Solar_mounting_type | 1963987 | 99,94 |
| CHP | 1964551 | 99,97 |
| Country | 0 | 0,00 |

Fig. 5.3.2.1 Null value for Riconciled Dataset

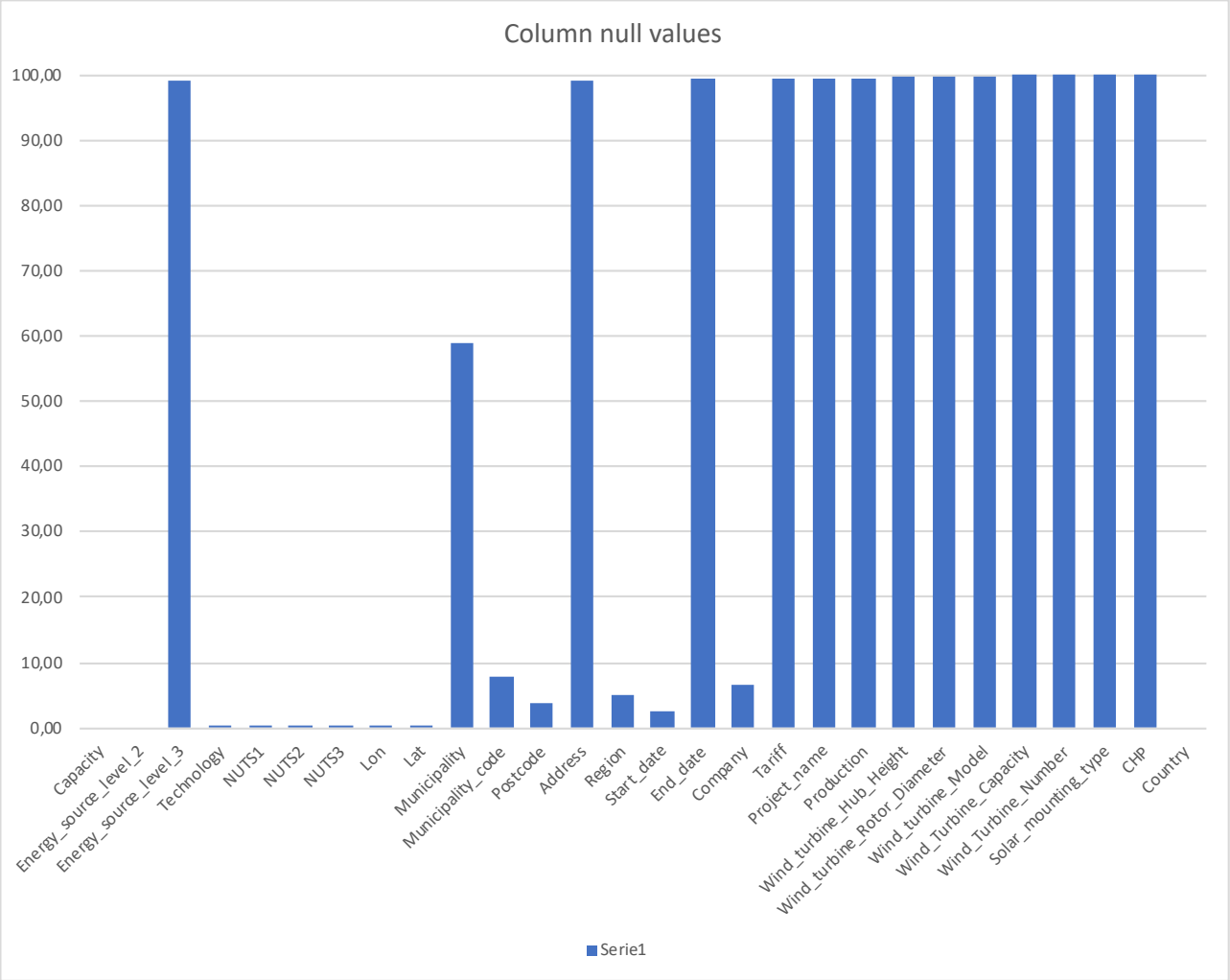



Fig. 5.3.2.2 Bar chart of percentages of null values in Riconciled

Consistently with what has been said previously, the dataset allows a discrete description regarding the distribution of the plants on the European territory with details relating to the type of plant.
Evidently, however, it does not allow to detail secondary aspects relating, for example, to the use case "Analysis of contracts" for which a stackholder such as the CEO of a company could be interested.
Furthermore, in figure 5.3.2.3 it is possible to observe the null values relating to the capacity of the various plants.

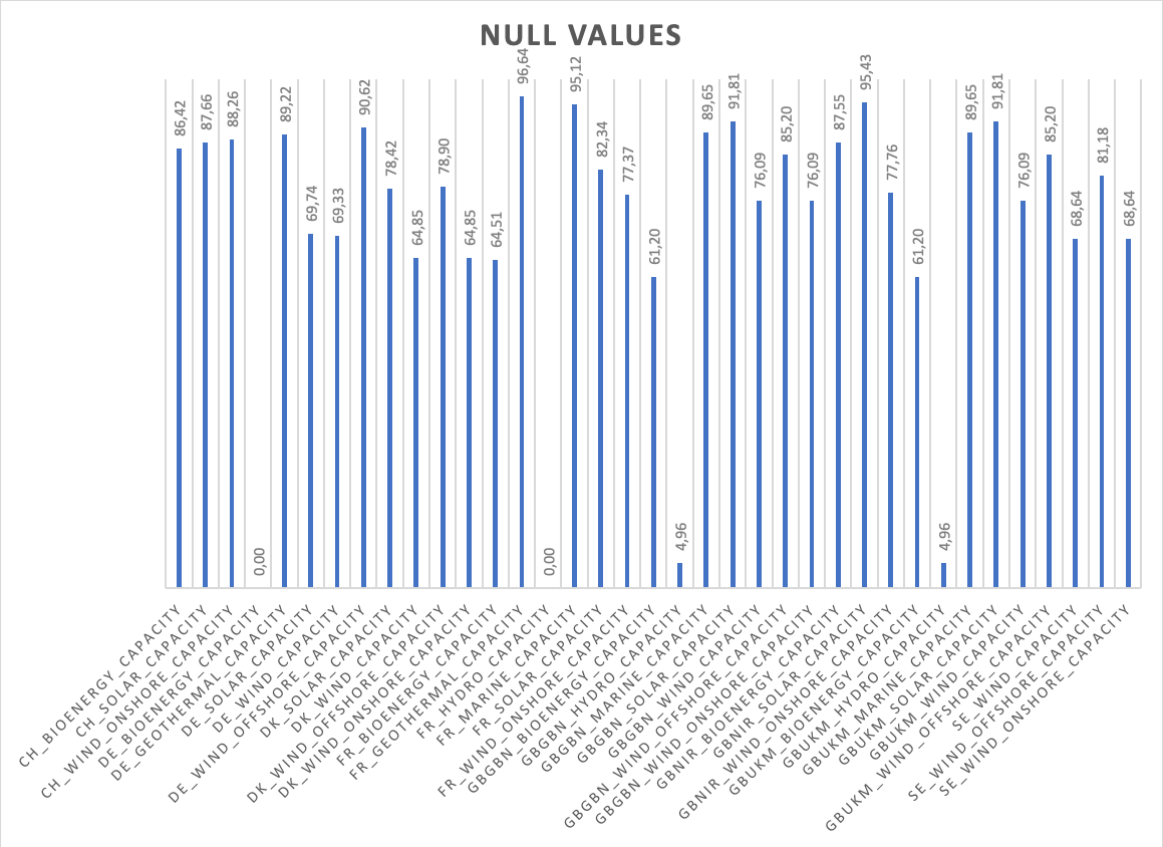



Fig. 5.3.2.3 Null values for capacities

All the SQL queries performed to analyze the datasets are available at the following link: https://drive.google.com/drive/folders/1FKf4y8HNPc9oRt0abkgtCibPm3DssEhm?usp=sharing

# 6 - DATA TRASFORMATION

Jupyter was used for cleaning and transforming the data.
The jupyter file is available at the following link: https://drive.google.com/file/d/1kaV1dnqtDLw3w4xchaDEz2gKG8E0Jv1g/view?usp=sharing

The following libraries were used for data transformation:
- Pandas
- Numpy

The main information of the dataset is the existence of an electricity plant, therefore we proceeded with checking (fig. 6.1) for the possible presence of lines with the fields "Energy source level 2", "Production" and "Capacity" to null at the same time. As evidenced by the data analysis, this case is not present.

```
#Verifico se in una riga ci sono production, ESL2 e capacity posti a null contemporaneamnte
df_null_comuni = df[(df['Production'].isnull()&df['Energy_Source_Level_2'].isnull()&df['Capacity'].isnull())]
df_null_comuni.info(verbose=True)
#Non ce ne sono :)
```

Fig. 6.1 Check if production, capacity and production values are null at the same time

Subsequently, the NAN values were replaced with 0 to facilitate data visualization.
When importing the csv file, the <string> fields were imported as <object> and therefore they were converted to the original format (fig. 6.2).

```
#REPLACE NAN CON 0 in tutto il csv
df = df.replace(to_replace = np.nan, value = 0)

#Conversione delle colonne <object> in colonne <string>
for y in df.columns:
    if df[y].dtype==object:
        df[y] = df[y].astype('string')
```

Fig. 6.2 Replace NAN values with 0 and type conversion

The previous fields set to 0 in the <string> columns have been set to 'NA' to identify any missing information (fig. 6.3)

```
#nelle colonne di stringhe, sostituzione di 0 con ND (Non Definito)
for y in df.columns:
    if df[y].dtype=='string[python]':
        df[y].replace({"0": "ND"}, inplace=True)
```

Fig. 6.3 Replace 0 values with 'ND' in string fields

As can be seen in figure 6.4, duplicate rows were eliminated using the pandas function: drop_duplicates ().

```
#Elimino righe duplicate
df.drop_duplicates(subset=None,
                           inplace=True,
                           ignore_index=False)
df.shape
```

Fig. 6.4 Drop duplicates rows

According to what emerged from the data analysis, the "Energy Source Level 1" column always reports the same value (Renewable Energy) and therefore it was decided to eliminate it using the drop () function of pandas (fig. 6.5)

```
#Se la colonna ESL1 ha sempre lo stesso valore la elimino
boolean_ESL1 = df['Energy_Source_Level_1'].str.contains('Renewable energy')
total_occurence = boolean_ESL1.sum()
if total_occurence==df.shape[0]:
    print("Energy Source Level 1 ha sempre lo stesso valore, cancello la colonna")
    df.drop('Energy_Source_Level_1', inplace=True, axis=1)
df.shape
```

Fig. 6.5 Drop Energy Source Level 1 column

Checks have been applied on the correctness of some numerical values such as capacity, height. diameter that cannot be negative. If they are identified they are set to 0 (fig. 6.6).

```
#Le tariffe negative sono un dato errato, lo imposto a 0
boolean_negativeTariff = df['Tariff'] < 0
total_occurenceTariff = boolean_negativeTariff.sum()
if(total_occurenceTariff>0):
    for value in df['Tariff']:
        if value < 0:
            df = df.replace(to_replace = value, value = 0)
    boolean_negativeTariff = df['Tariff'] < 0
    total_occurenceTariff = boolean_negativeTariff.sum()
    print(total_occurenceTariff)

#Le capacità negative sono un dato errato, lo imposto a 0
boolean_negativeCapacity = df['Capacity'] < 0
total_occurenceCapacity = boolean_negativeCapacity.sum()
if(total_occurenceCapacity>0):
    for value in df['Capacity']:
        if value < 0:
            df = df.replace(to_replace = value, value = 0)
    boolean_negativeCapacity = df['Capacity'] < 0
    total_occurenceCapacity = boolean_negativeCapacity.sum()
    print(total_occurenceCapacity)


boolean_negativeTurbineCapacity = df['Turbine Capacity'] < 0
total_occurenceTurbineCapacity = boolean_negativeTurbineCapacity.sum()
if(total_occurenceTurbineCapacity>0):
    for value in df['Turbine Capacity']:
        if value < 0:
            df = df.replace(to_replace = value, value = 0)
    boolean_negativeTurbineCapacity = df['Turbine Capacity'] < 0
    total_occurenceTurbineCapacity = boolean_negativeCapacity.sum()
    print(total_occurenceTurbineCapacity)

#Valori negativi di diametro sono errati, impostiamo a 0
boolean_negativeDiameter = df['Turbine Rotor Diameter'] < 0
total_occurenceDiameter = boolean_negativeDiameter.sum()
if(total_occurenceDiameter>0):
    for value in df['Turbine Rotor Diameter']:
        if value < 0:
            df = df.replace(to_replace = value, value = 0)
    boolean_negativeDiameter = df['Turbine Rotor Diameter'] < 0
    total_occurenceDiameter = boolean_negativeDiameter.sum()
```

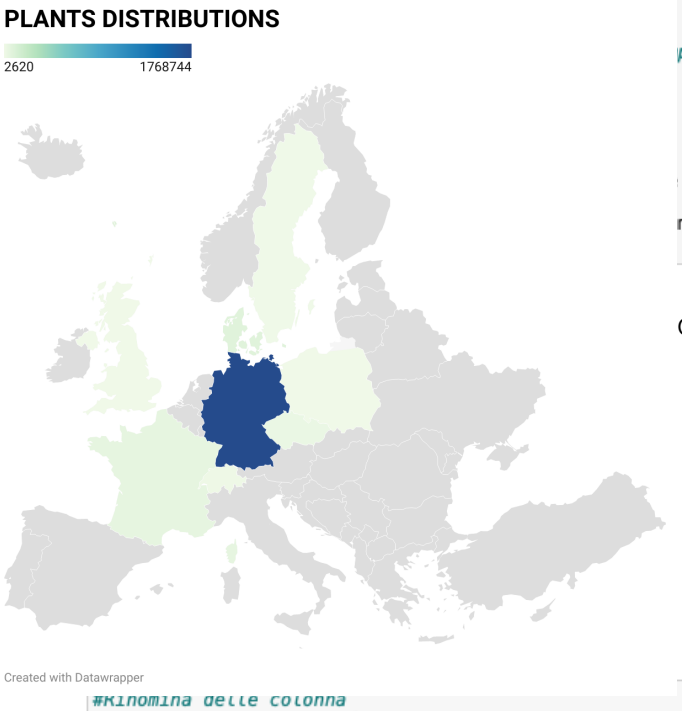



Fig. 7.2.1 Plants distribution nation map

ctness for numerical values

The data transformation ended with the modification of the column names with values more suitable for the data visualization (fig. 6.7).

```
#Rinomina delle colonna
df = df.rename(columns={'Company_Name':'Company', 'Project_Name':'Project','Start_Date':'Start', 'End_Date':'End',
```

Fig. 6.7 Renaming of columns

# 7 - DATA VISUALIZATION

In this chapter some displays will be reported in accordance with what is described in chapter 3.
In figures 7.1, 7.2 and 7.3 it is possible to observe a possible representation of the analyzed data with respect to the location.

```
[12718, 1768292, 84314, 54195, 2508, 5528, 2614, 22349]
['Switzerland', 'Germany', 'Denmark', 'France', 'Poland', 'Sweden', 'United Kingdom', 'Czechia']
```

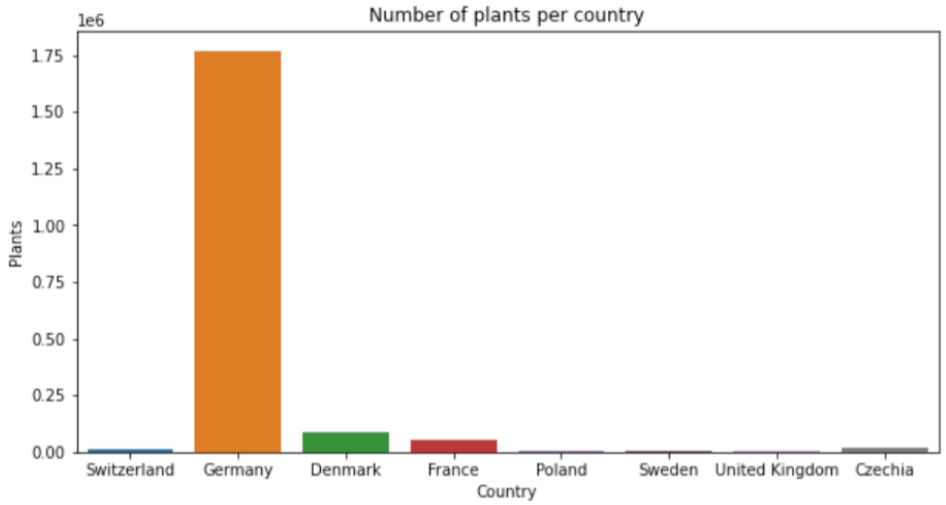



Fig. 7.1 Number of plants for Country

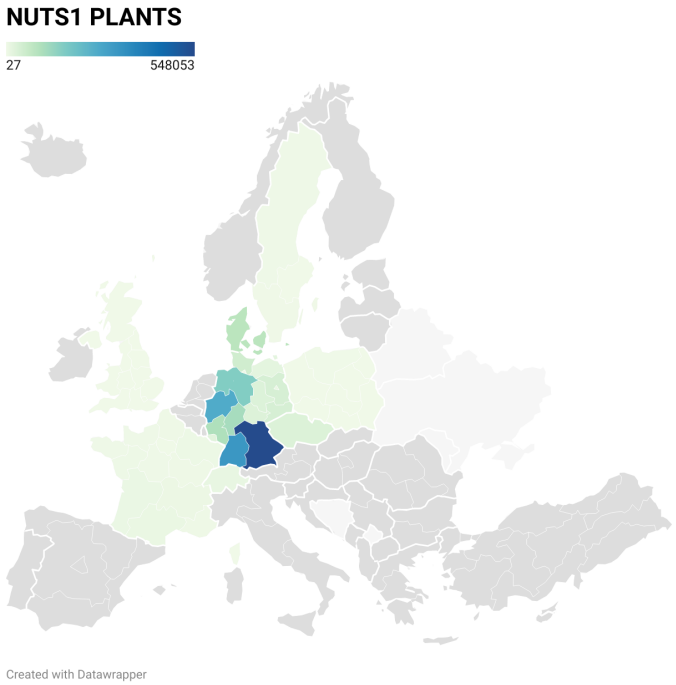



Fig. 7.2.2 Plants distribution NUTS1 map

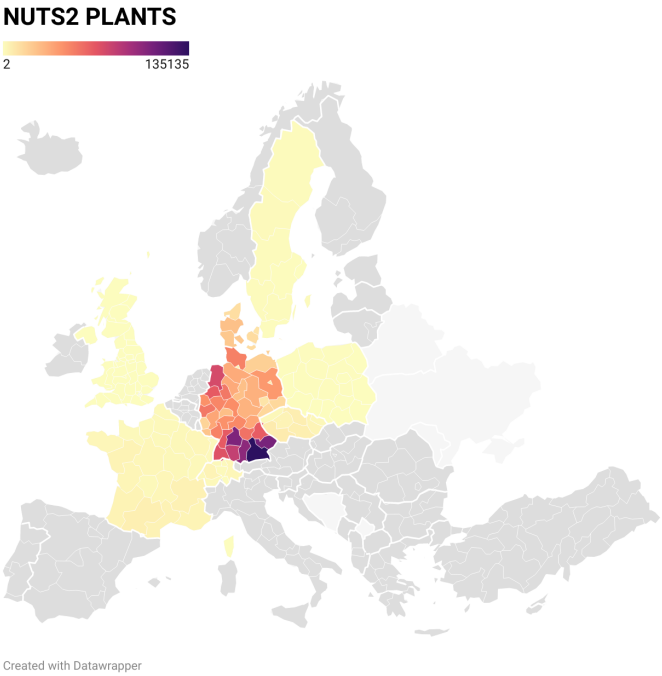



Fig. 7.2.3 Plants distribution NUTS2 map

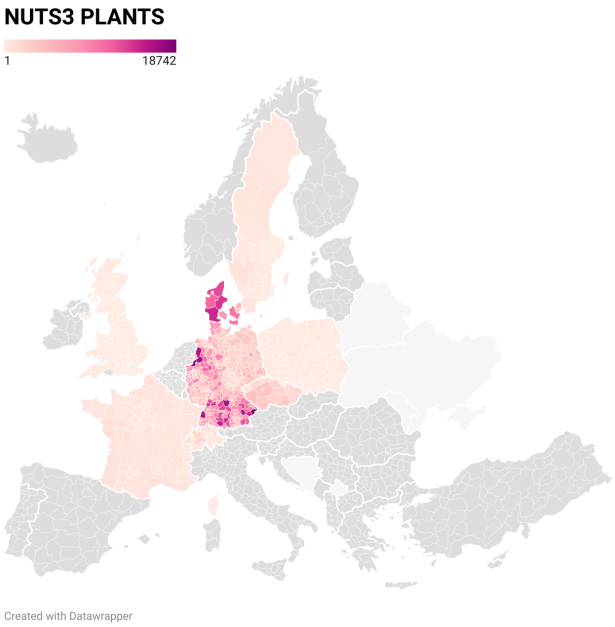



Fig. 7.2.4 Plants distribution NUTS3 map

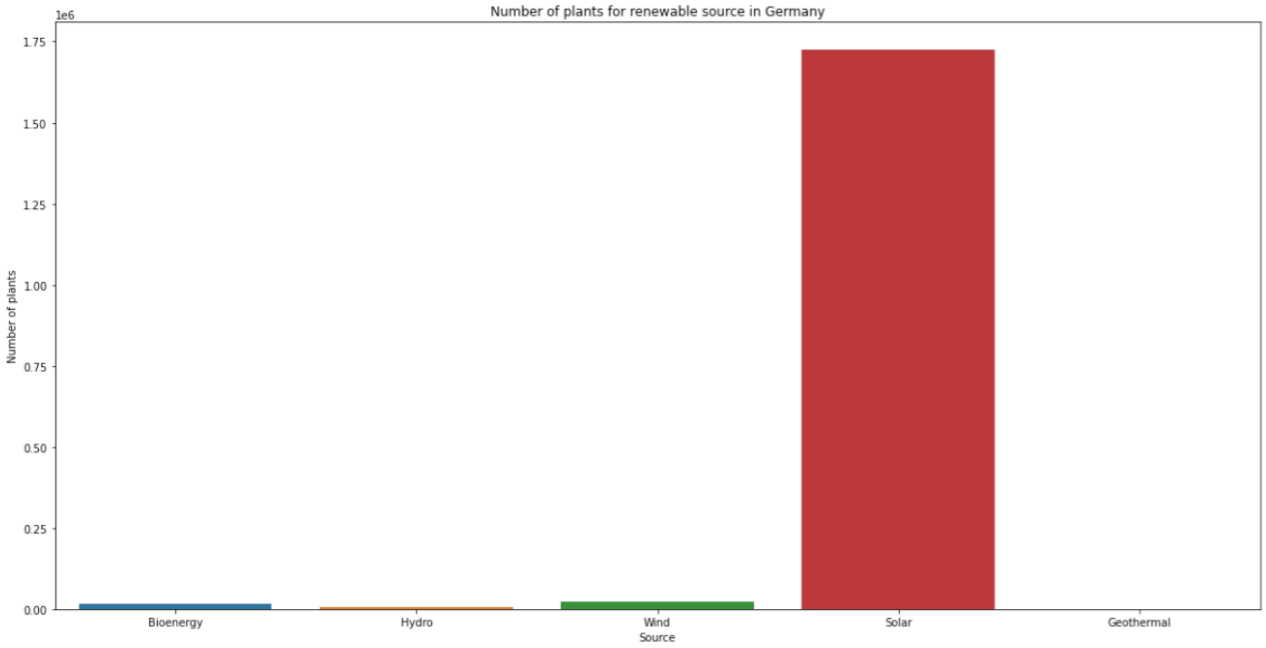



Fig. 7.3 Number of plants for renewable source in Germany

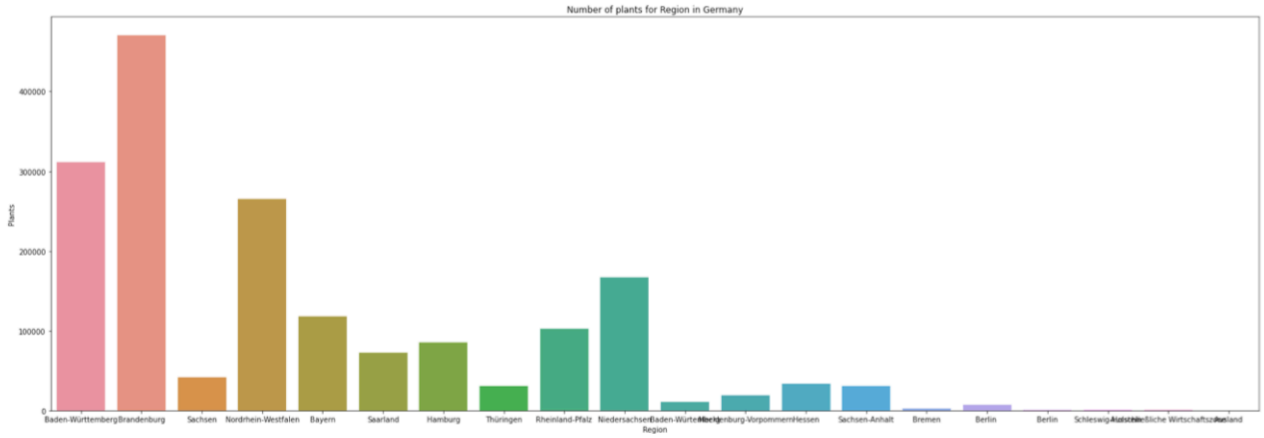



Fig. 7.4 Numbers of plants for regione in Germany

Subsequently it is possible to view the data in relation to the technical aspects of the systems.
For example, in figure 7.4 it is possible to observe the technologies most used in the case of wind energy.
In figure 7.5 it is possible to observe the turbine models most used in wind power plants while in figure 7.6 it is possible to observe the trend relative to the height of the hubs
While figures 7.7 and 7.8 show some technical details relating to photovoltaic systems. In particular, the cases in which panels capable of producing heat (as well as energy) or not are used and the place where they are installed are reported.

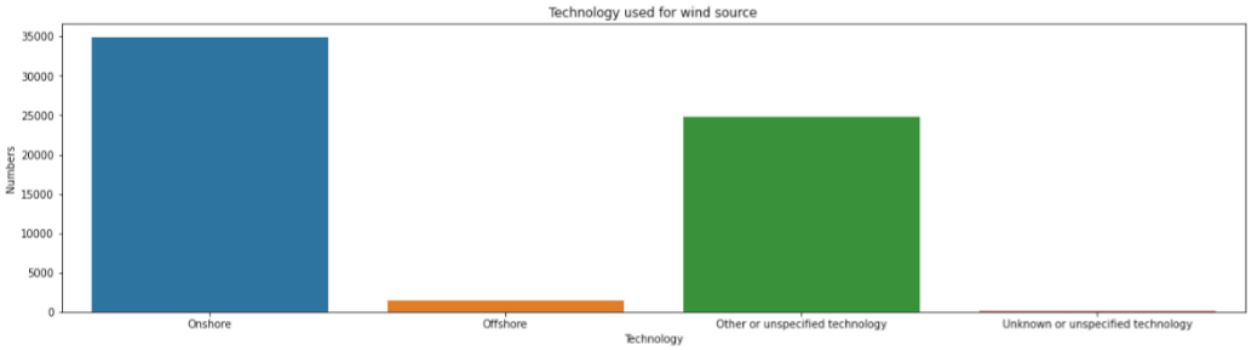



Fig. 7.5 Technology used for wind plants

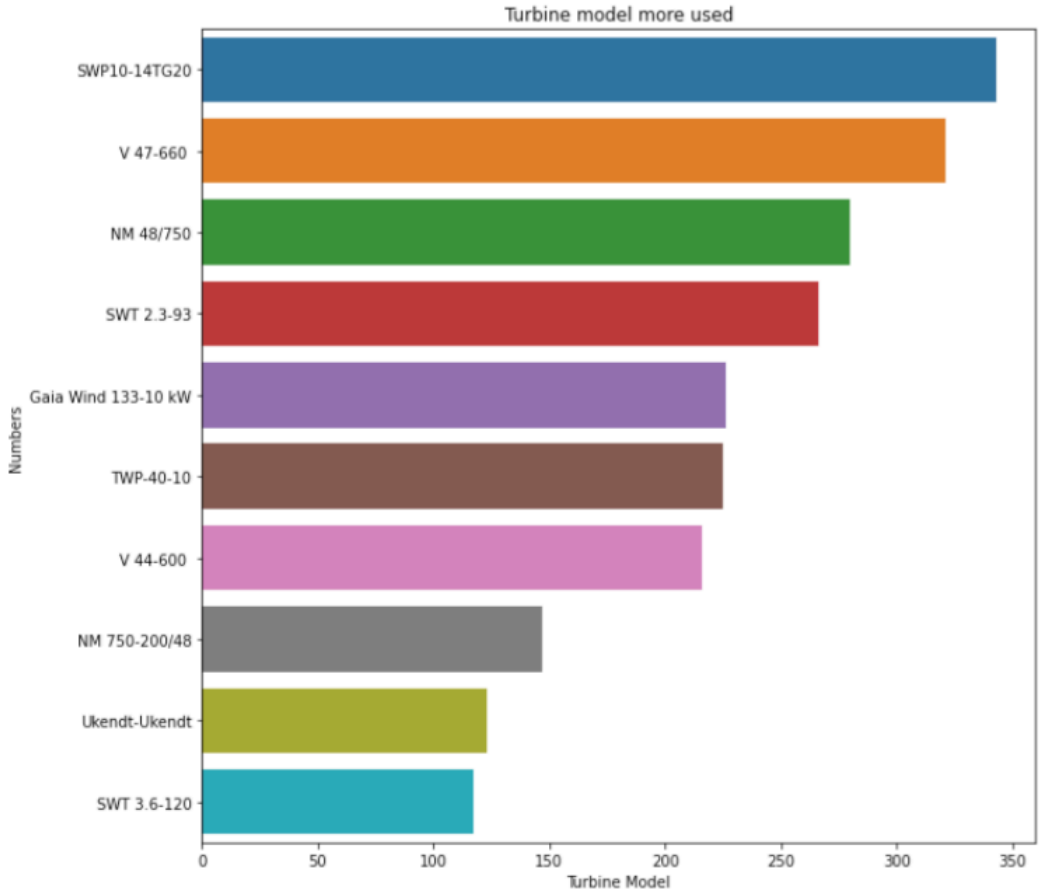



Fig. 7.6 Turbine model most used

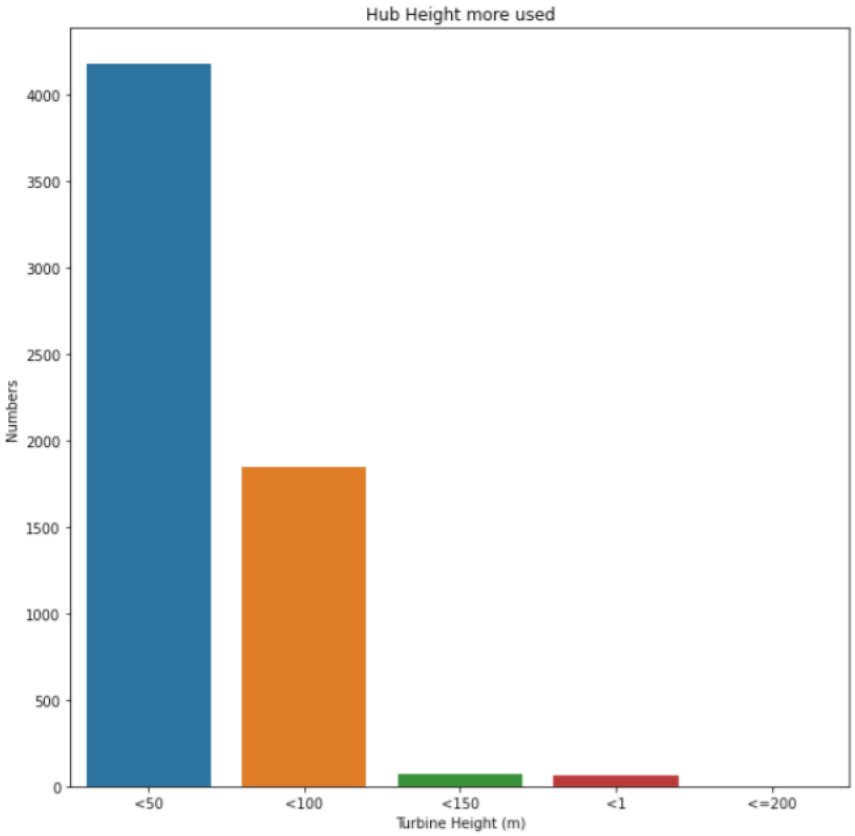



Fig. 7.7 Number of hubs with a certain height range

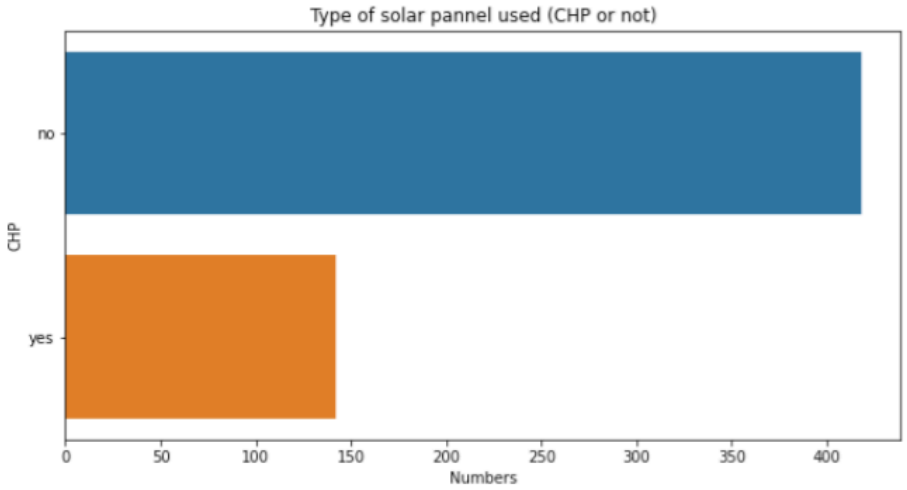



Fig. 7.8 Number of installation of photovoltaic panels with respect to the type

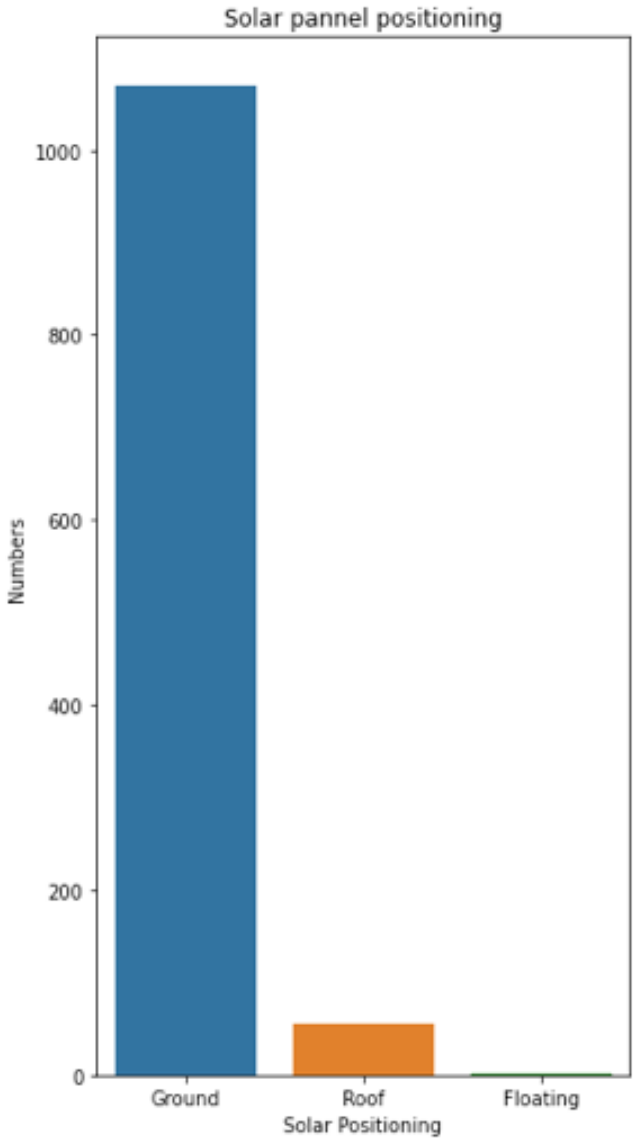
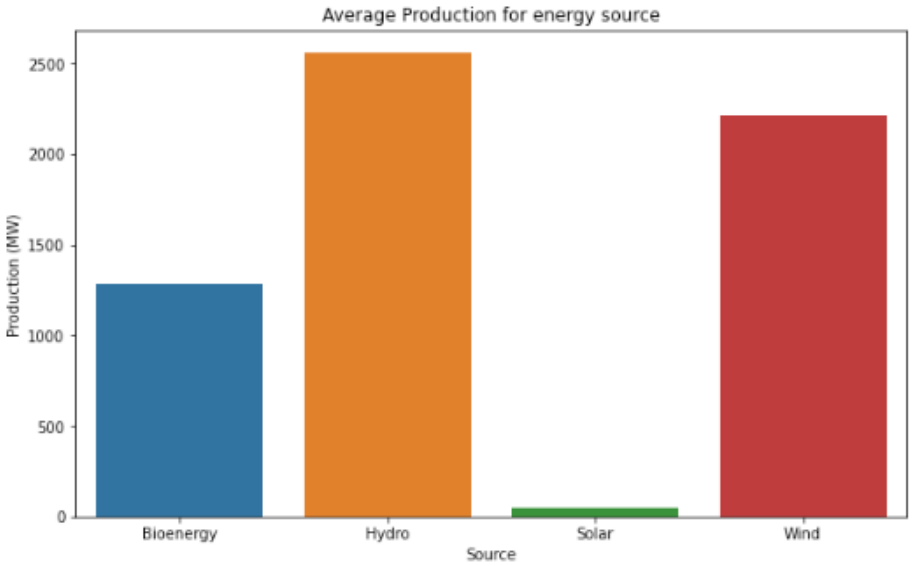



Fig. 7.9 Number of installation of photovoltaic panels with respect to the positioning

For an analysis from a commercial point of view (but also of interest to those interested in the aspects relating to efficiency) it is possible to use the graphs below.
Then in figure 7.10 we can observe the average energy production, expressed in MW, for each different renewable source.
In this case, particular attention must be paid to the data analysis carried out previously. From this, in fact, it emerged that production is a data that is reported only from Switzerland and therefore, really, the proposed graph is calculated only on the basis of Swiss data.




Flg. 7.10 Average production for renewable source

On the contrary, in figure 7.11 it is possible to observe the average production capacity per renewable source. The capacity is a datum managed by all the datasets and therefore the proposed graph allows a much more accurate analysis than the one related to the production.

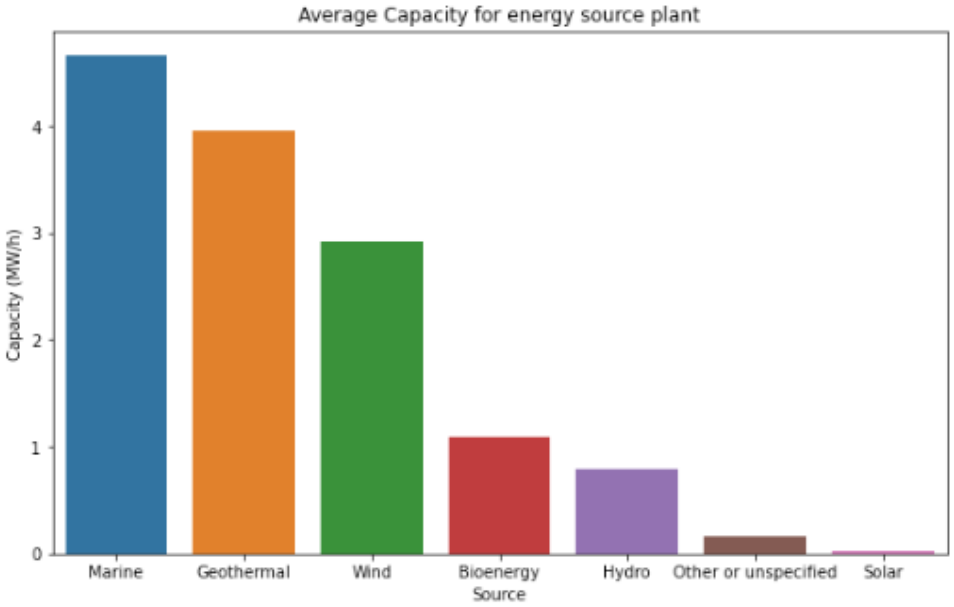



Fig. 7.11 Average plant capacity for different sources

In figure 7.12 it is possible to see a graph of the correlations between production, tariff and capacity. The results should not surprise us as obviously a plant that produces more will be subject to a higher rate.

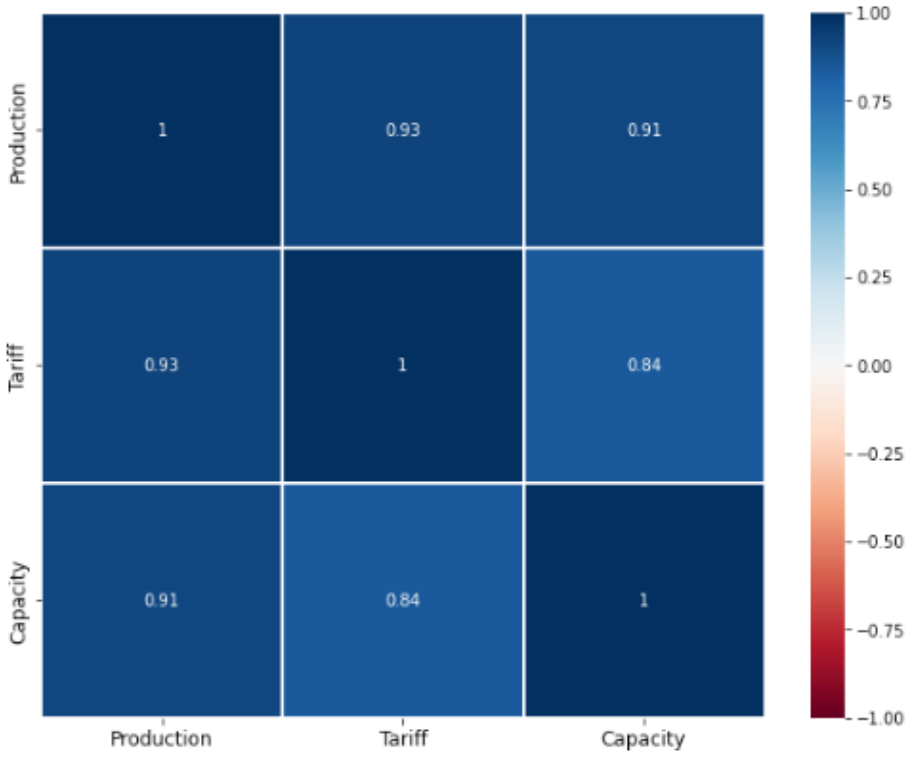



Fig. 7.12 Correlation map between production, tariff and capacity

Figure 7.14 shows the graph relating to the ten dominant companies in the energy production sector using bioenergy.

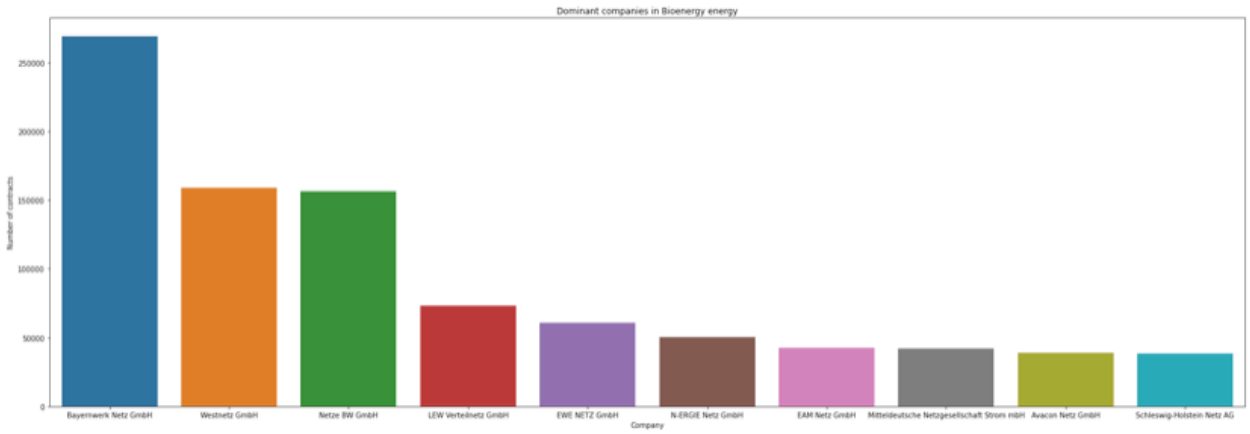



Fig. 7.14 Top ten companies in bioenergy production

Data analysis can also be developed from a temporal point of view. In this case, graphics such as those shown in figure 7.15 could be of great help.
In figure 7.15 it is possible to observe the number of installations carried out over the years, categorized by renewable source. Evidently the production of solar energy is the fashion.

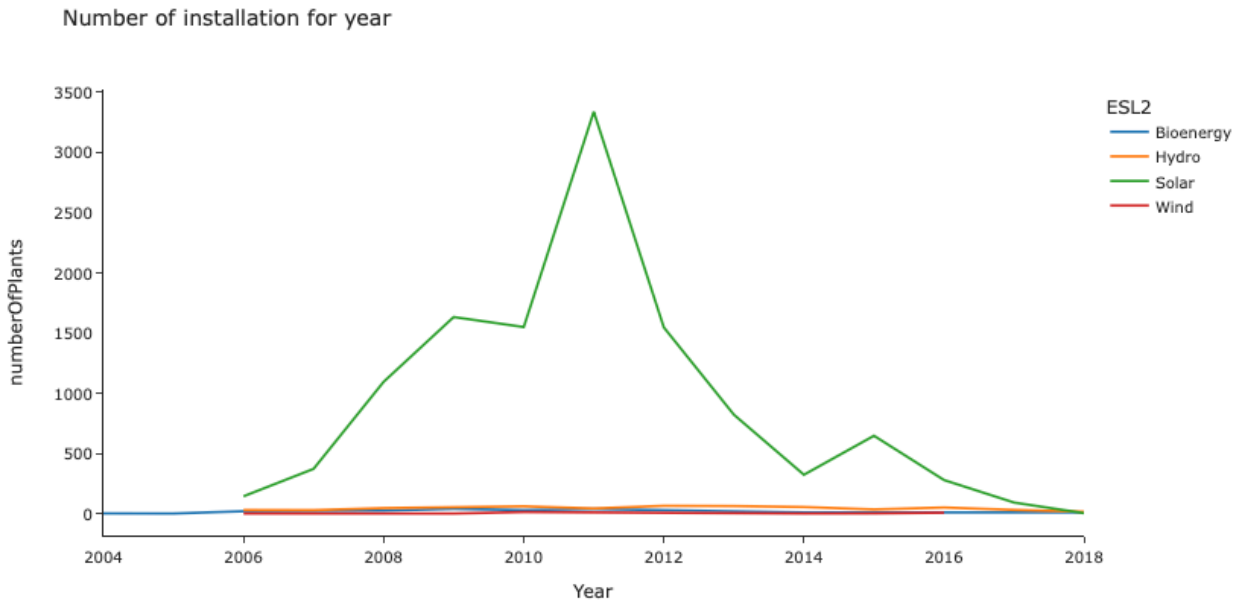



Fig. 7.15 Number of installations of plants for source in different years

By defining the convenience ratio (ratio betweenduration of a contract in years and tariff for year), it is also possible to evaluate how this ratio has changed over the years for the different energy sources (fig. 7.15) and which companies have this rate lowest by sector (Figure 7.16 shows the case of the bioenergy source).

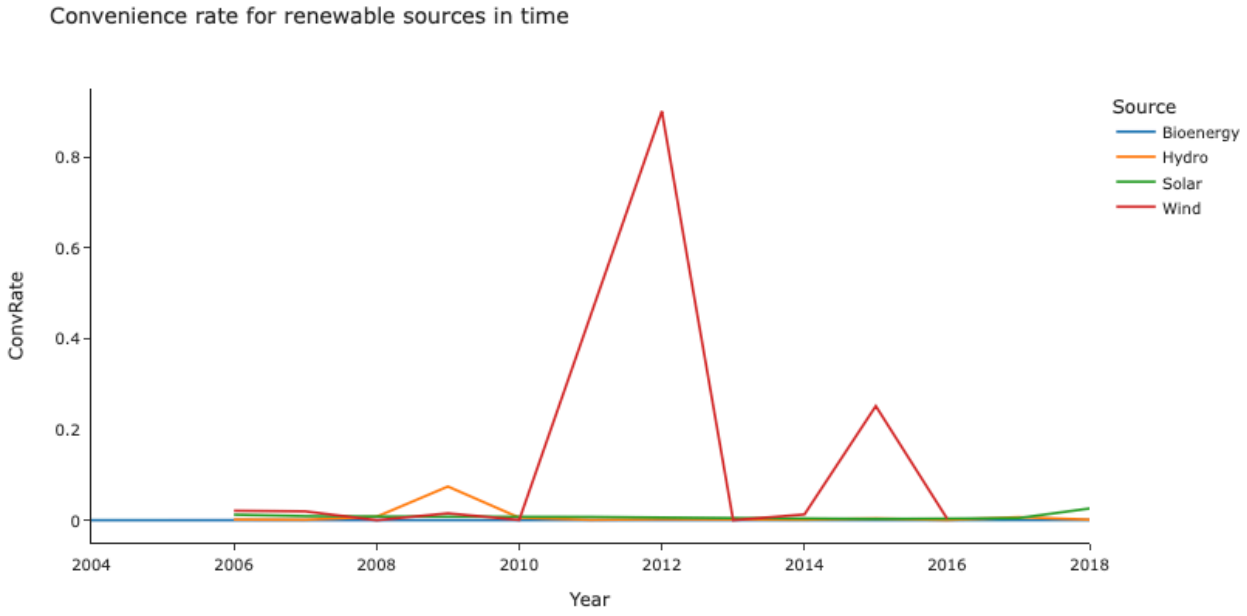



Fig. 7.16 Average convenience ratio for source in different years

Finally, we can evaluate the renewable source which, on average, leads to a higher convenience rate.
As can be seen from the graph in figure 7.17, the greater adoption of photovoltaic systems, despite being among those that produce the least, is essentially due to a purely economic factor.

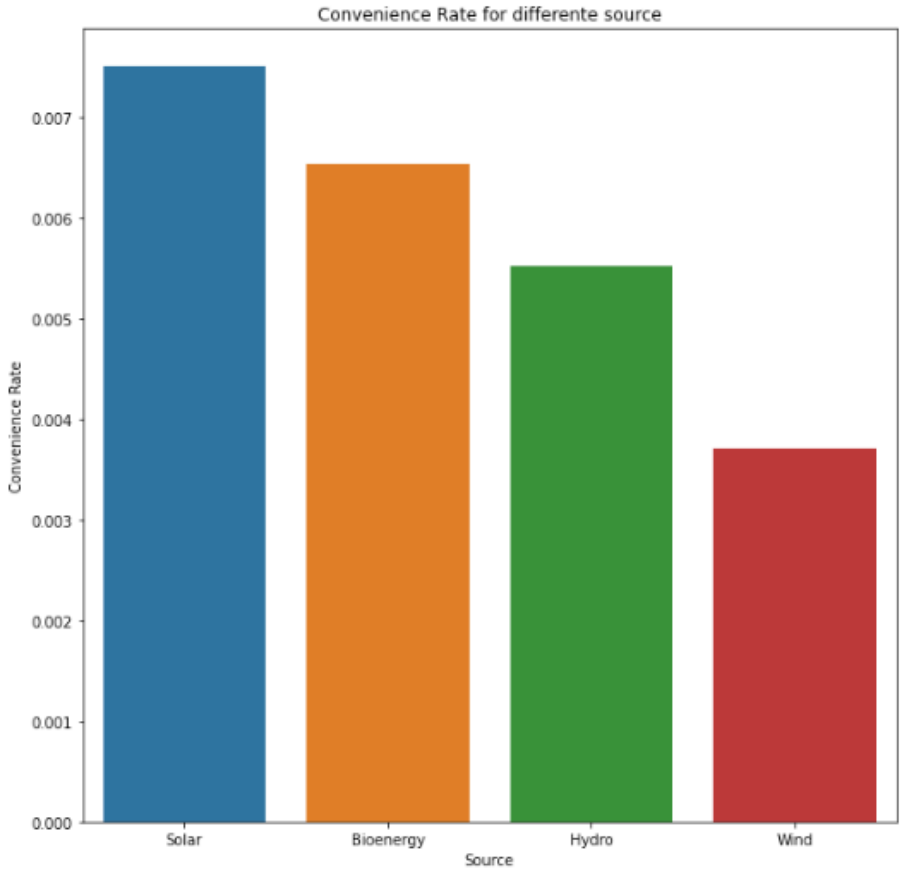



Fig. 7.17 Convenience ratio for renewable sources

The csv which evaluated all the capacities was analyzed and it emerged that the highest capacity is recorded in Germany for solar. Furthermore, it is noted that in general solar and wind are the energies with more capacity in relation to their geographical location ( obviously sunny areas rather than windy).

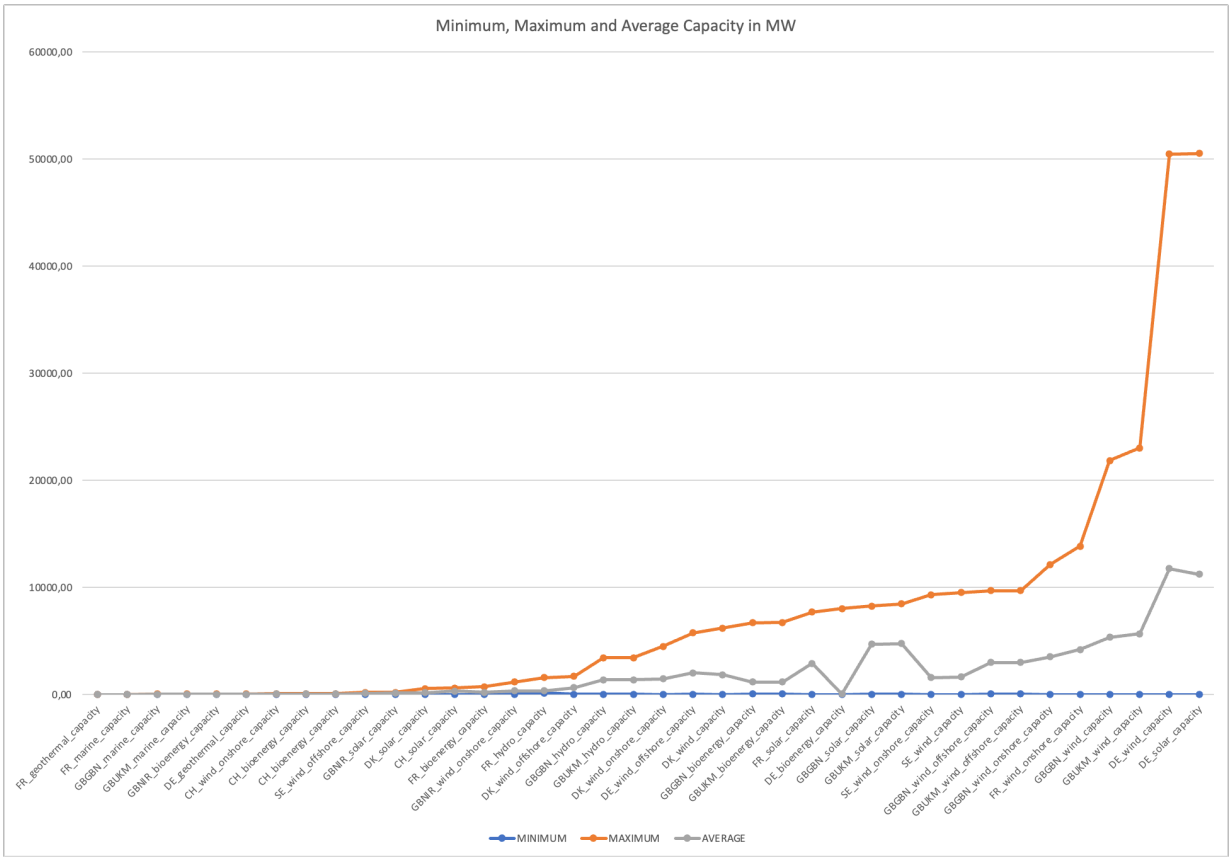



Fig. 7.18 capacity minimum, maximum and average

The comparison and relationship between declared capacity and actual production was analyzed where production data were provided and therefore only in Switzerland.

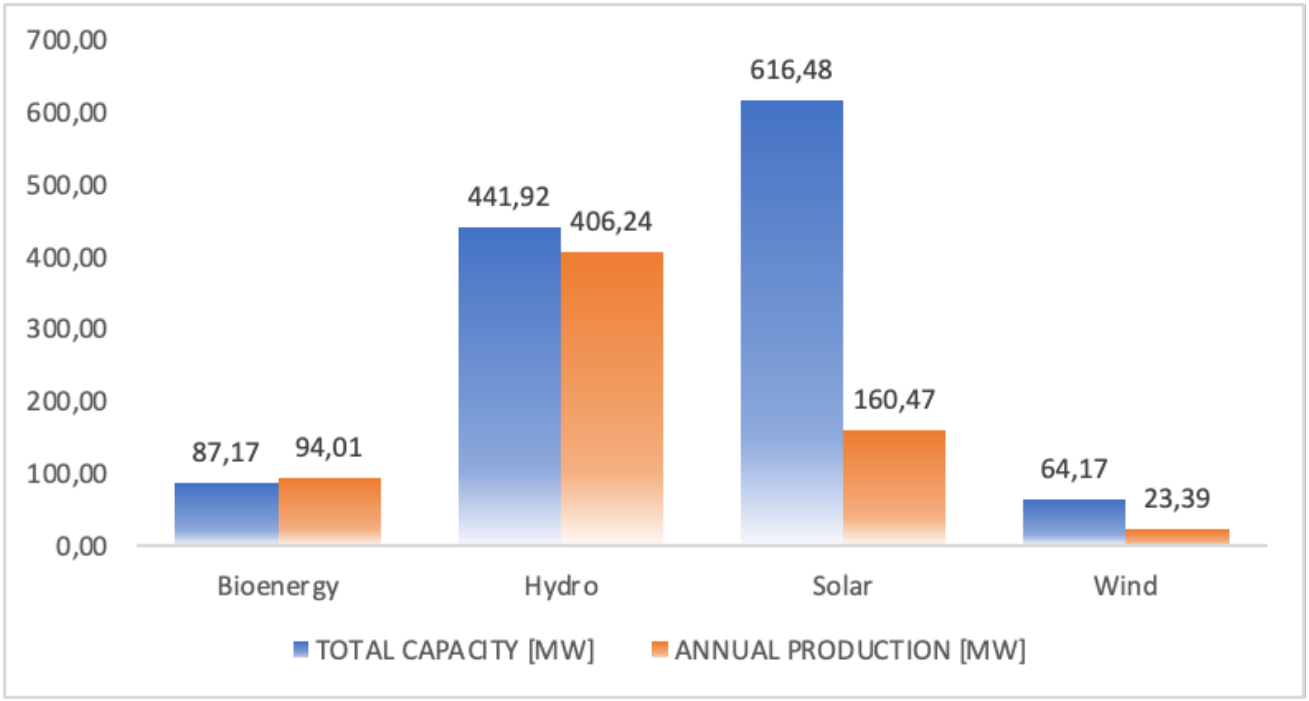



Fig. 7.19 Capacity - Production relationship

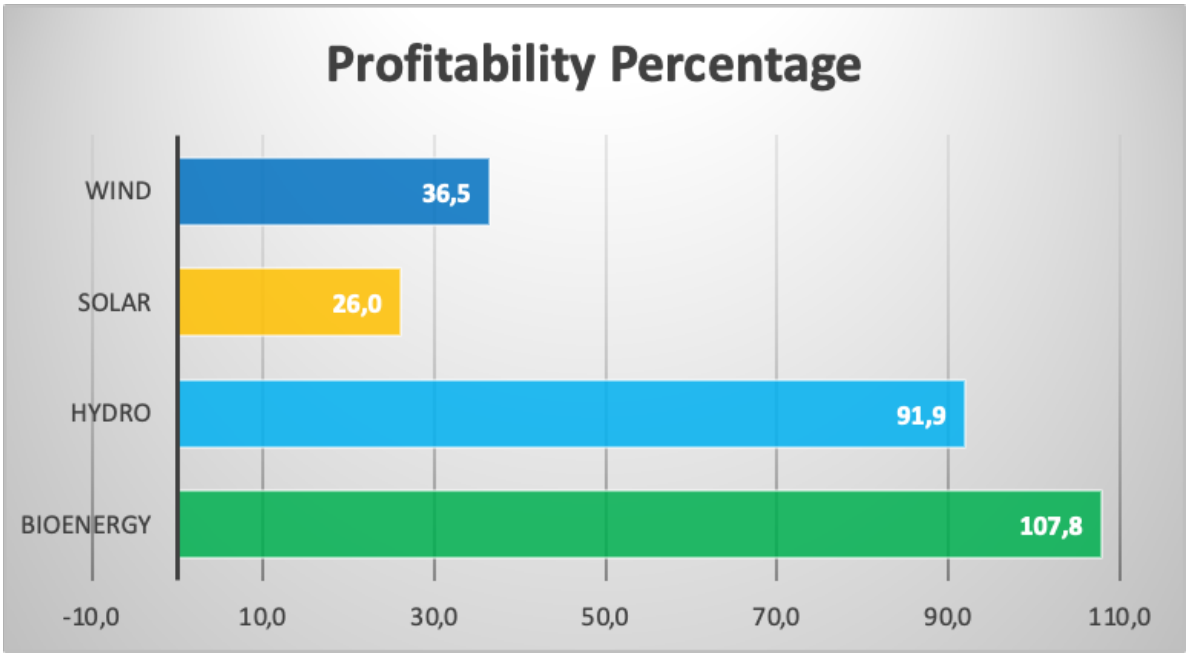



Fig. 7.20 Ratio Production over capacity * 100

In fig. 7.21 it can be seen the analysis on the cost per MW actually produced where the production data was available, and therefore only in Switzerland.
Technically, the production had MWh as a metric, so dividing by 3600 s it is obtained MW.
Finally, the ratio was made between the sum of the tariffs grouped by type of energy and the sum of production.

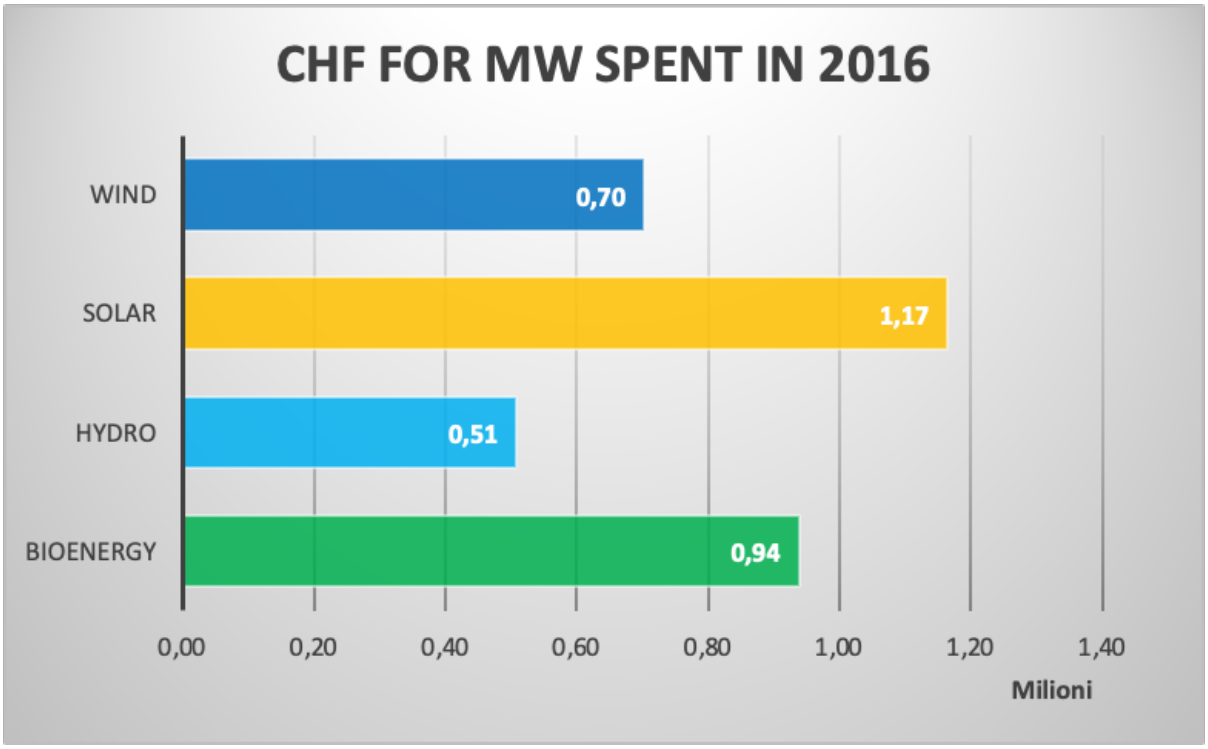



Fig. 7.21 CHF/ MW

The energy that produces energy most efficiently is hydro, while the worst is solar.
This could be worse than the generalized reality, because:

- solar energy is strongly influenced by the particular geographic location;
- incomplete or incorrect data.

Conversely, however, solar energy is that energy that requires less initial expenditure that does not depend on energy production.

According to the data provided, it appears that although the least productive and the most expensive, solar energy is the most used.
However, this is conditioned by the particular nation, Switzerland, which is not ideal for exploiting solar energy.

The data visualization section concluded with analyzes carried out with a datawrapper on the geographical distribution of the types of energy and technologies.

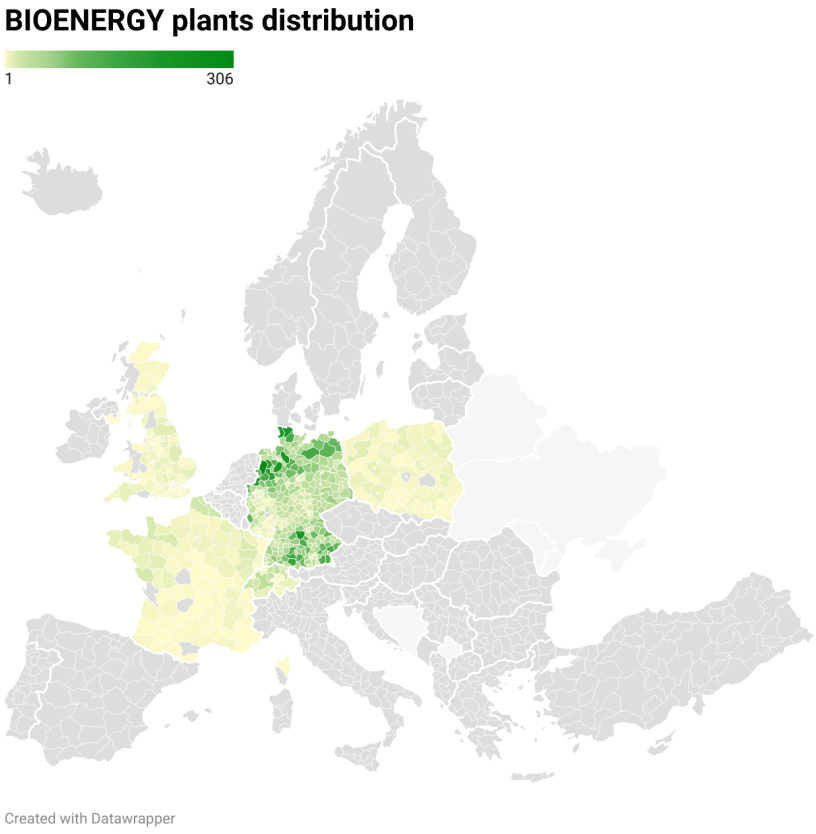



Fig. 7.22 Bioenergy plants distribution

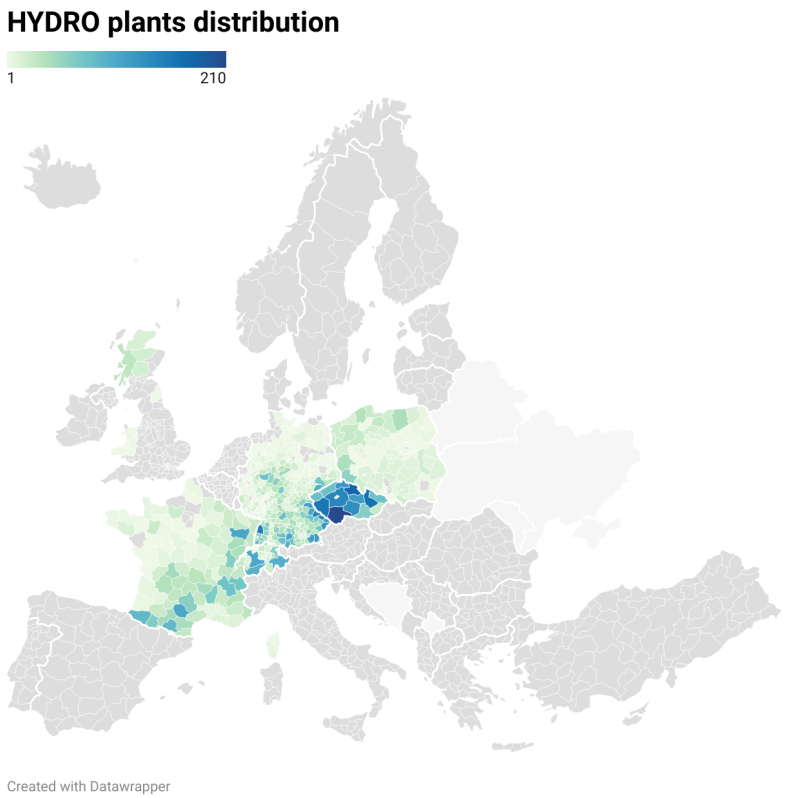



Fig. 7.23 Hydro plants distribution

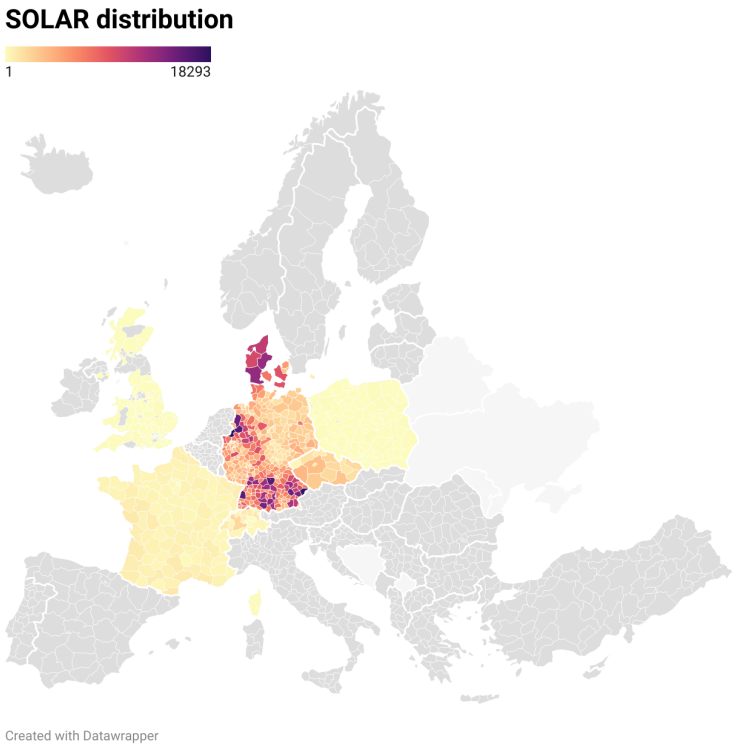



Fig. 7.24 Solar plants distribution

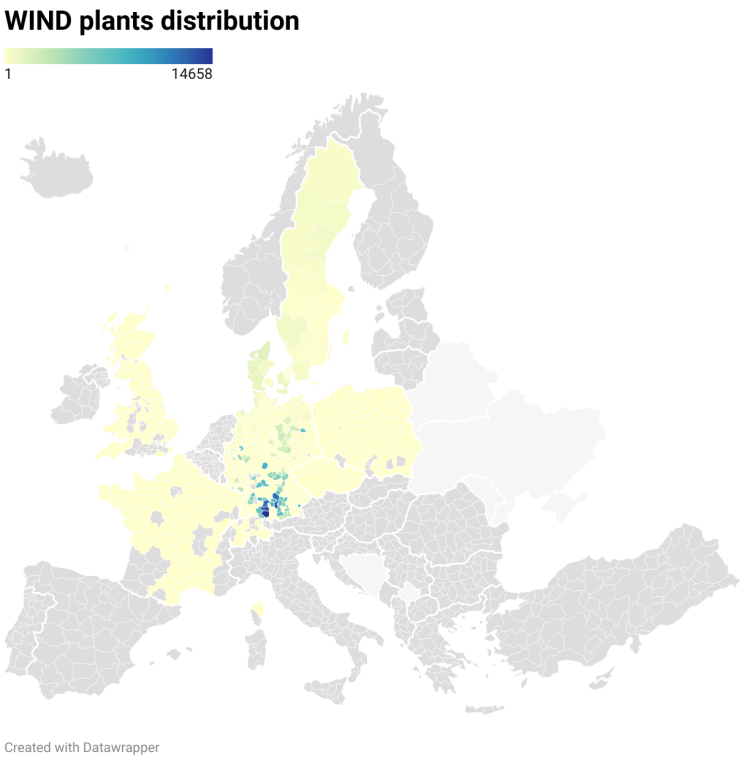



Fig. 7.25 Wind plants distribution

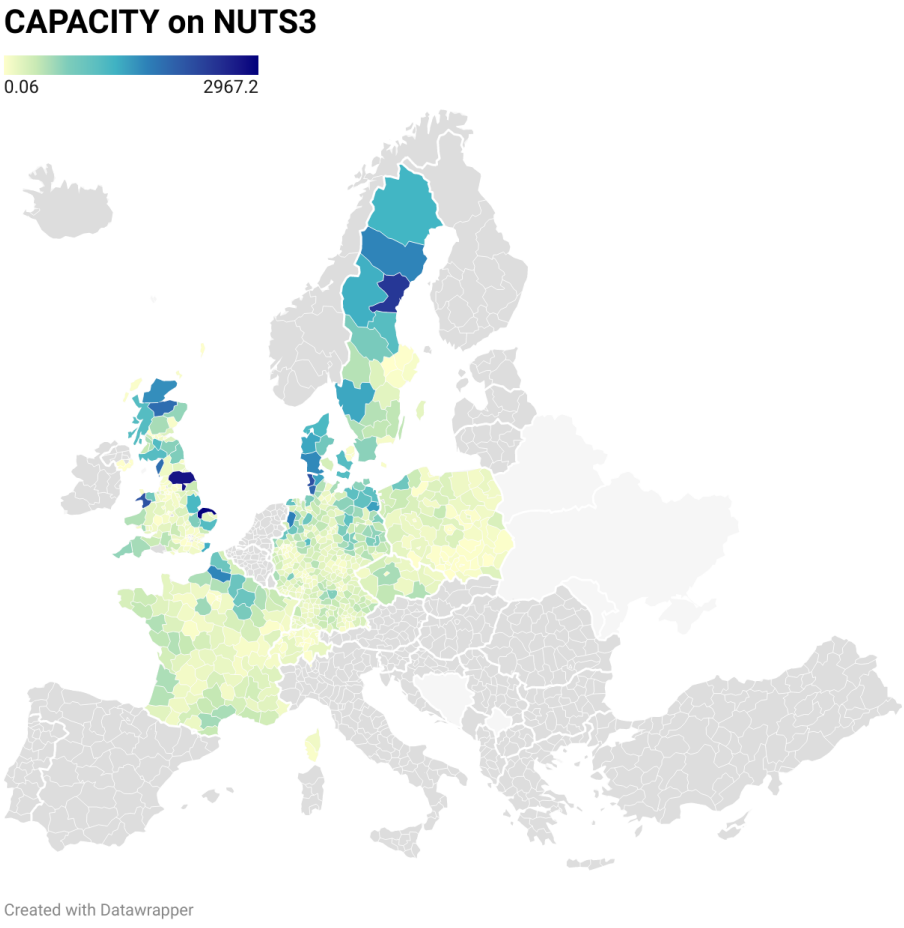



Fig. 7.26 Capacity distribution

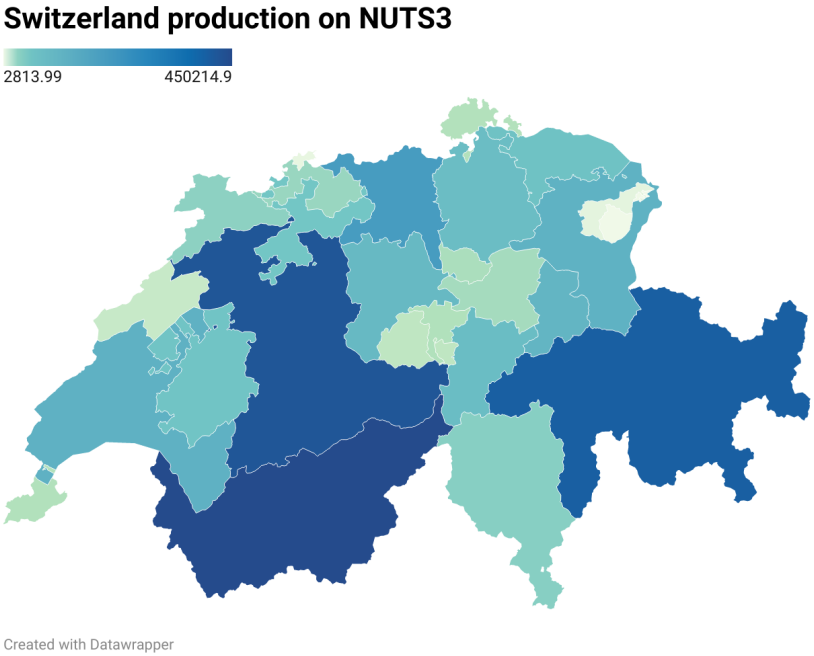



Fig. 7.27 Switzerland production distribution

The entire data visualization was created using python and in particular jupyter. The notebook with all the implementation details and all the views made (only some of them have been reported in the report) is available at the following link: https://drive.google.com/file/d/1pR0NuC11mo0fhILKaYB0CbUfEODMJNWn/view?usp=sharing