**Data Preparation:**

Data collection and preparation are the foundation for trusted ML models.

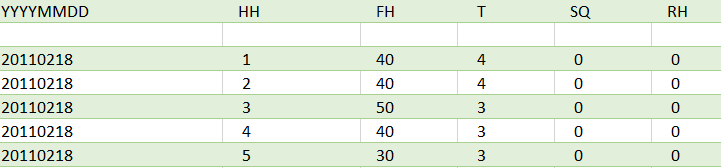
Once the data is collected, it’s time to assess the condition of it, including looking for trends, outliers, exceptions, incorrect, inconsistent, missing, or skewed information.

Because we have data from two different sources additional task should be implemented, such as: Find the optimal intersection or even the applicable one between the two datasets.

The data in the Energy consumption sheet was forms as following:



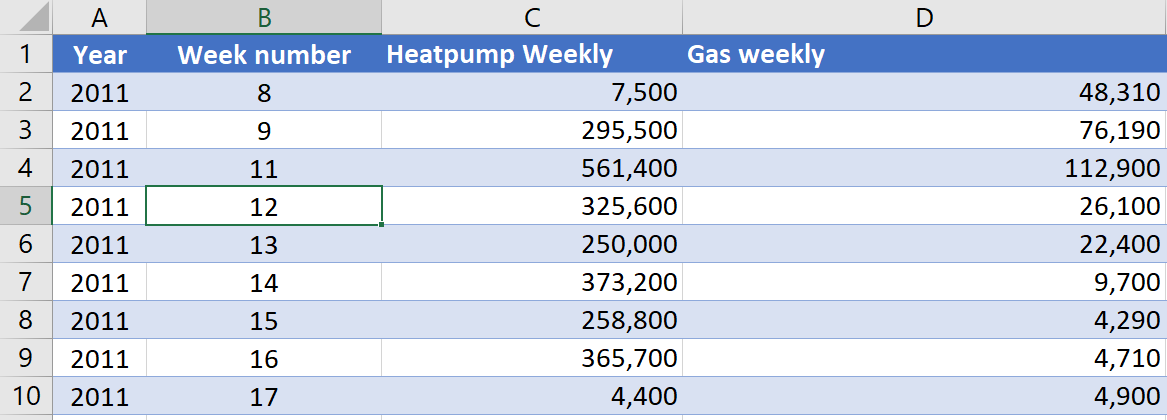
And the data derived from the KNMI site was formed as following:



Determine the starting date and the ending date of the measurements and the measurements intervals.

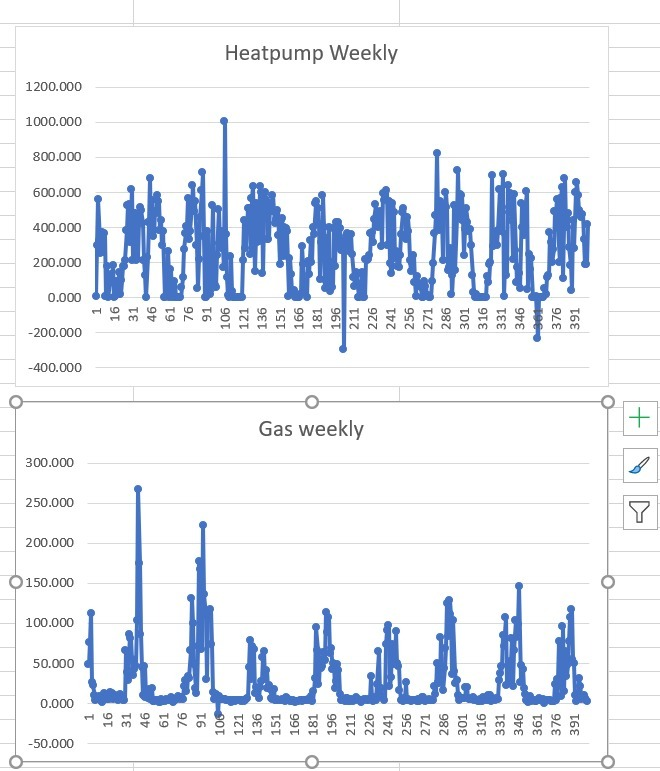
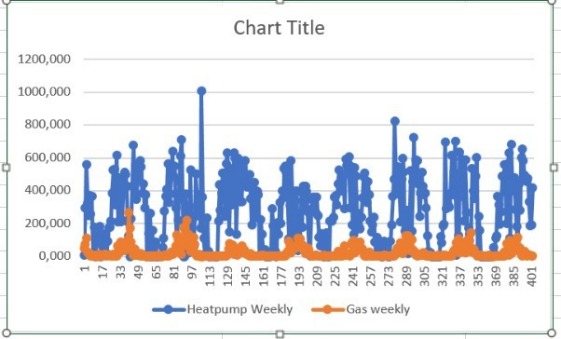
The Energy consumption interval measurements were not equal, but a weekly interval will be so most convenience. But still in the dataset there are some missing weeks and intersecting intervals, so a lot of manual correction was done.

The resulted data set is as following:

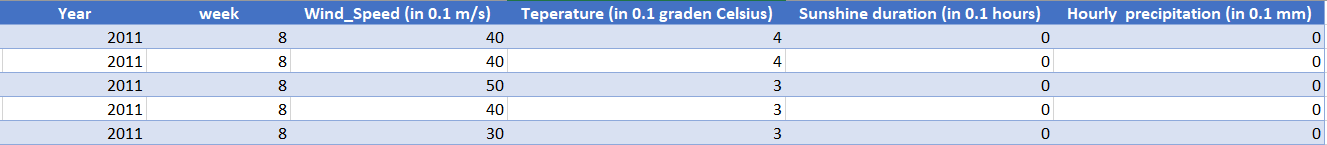


To prepare the KNMI dataset and fit it to the energy consumption dataset:

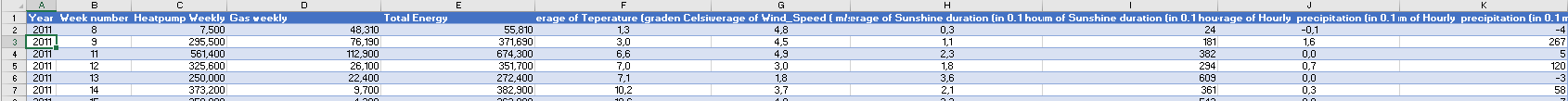
* We change the unit in windspeed and temperature columns
* We calculated the average and the sum on weekly bases for our four chosen columns.
* We omitted manually all the weeks that have no equivalent in the energy consumption table and correct all the intersecting intervals, either by emitting or by adding them to the right week number.
* We plotted the resulted dataset and omitted all the outliers’ values manually.



The preminateru resulted KNMI dataset is as following:



We combine the two tables together and created extra columns, such as total energy “Total energy” to get the following final table:



The weekly intervals for each Year.

Now our data is ready for the ML model.