Energy consumption in the Netherland

URL

<https://www.kaggle.com/hark99/energy-consumption-in-the-netherland>

CONTEXT

In Netherland, three companies are responsible for providing energy in terms of electricity and gas to the whole country. The three companies are Enexis, Liander and Stedin. Every company has shared data of ten years from 2009 to 2018. The total data is consisted of Six Million observations. The data is splitted into two categories, Electricity and Gas. Each year data is comprised of twelve features. Among features, five are providing statistical information. Our objective is to predict total Energy consumption for the ongoing year, 2019

# PRPROCESSING

# The preprocessing involves initial data exploration, extract transform load, data cleansing and feature creation.

# Initial Data Exploration

# The three companies’ data are available in csv format. There are almost twenty csv files of each company, showing electricity and gas data.

# ETL: Extract, Transform & Load

# ETL is an import tool to convert data into readable format. There are three phases. In first phase, we will extract each company's data. In in other two phases, we will first merge the data according to our requirements and then load the data frame. We will make three data frames, one for electricity data, the other for gas data and the last one consisting of all the data.

# Feature Creation

# Now it’s time for feature engineering. We will try two different methods to design our model to predict Annual Consumption, the target variable. First, we will select our features, which we call feature engineering. The features will be selected on the basis of correlation information. If it suitable we will proceed to model training. Otherwise we will create a new feature. and then predict the Annual Consumption. First we have tested every feature or variable against the target variable to see the dependency. It is clear after the check that the target variable dependency is very limited on the available features. So now we have create a new feature ‘Year’ on the basis of every year information regarding the yearly consumption.

# the basis of every year information. In this case we have only one featrue, the 'year'. First we tested the variables aganist

### The correlation is very weak among the features, so it won't work to train the model. So now we have to create a new feature, called 'Year' aganist the annual consumption.

### We will also sort the other data for visualization. This will help to see the shape of the data and therefore designing the model will be easy.