**client\_data.csv**

* id = client company identifier
* activity\_new = category of the company’s activity
* channel\_sales = code of the sales channel
* cons\_12m = electricity consumption of the past 12 months
* cons\_gas\_12m = gas consumption of the past 12 months
* cons\_last\_month = electricity consumption of the last month
* date\_activ = date of activation of the contract
* date\_end = registered date of the end of the contract
* date\_modif\_prod = date of the last modification of the product
* date\_renewal = date of the next contract renewal
* forecast\_cons\_12m = forecasted electricity consumption for next 12 months
* forecast\_cons\_year = forecasted electricity consumption for the next calendar year
* forecast\_discount\_energy = forecasted value of current discount
* forecast\_meter\_rent\_12m = forecasted bill of meter rental for the next 2 months
* forecast\_price\_energy\_off\_peak = forecasted energy price for 1st period (off peak)
* forecast\_price\_energy\_peak = forecasted energy price for 2nd period (peak)
* forecast\_price\_pow\_off\_peak = forecasted power price for 1st period (off peak)
* has\_gas = indicated if client is also a gas client
* imp\_cons = current paid consumption
* margin\_gross\_pow\_ele = gross margin on power subscription
* margin\_net\_pow\_ele = net margin on power subscription
* nb\_prod\_act = number of active products and services
* net\_margin = total net margin
* num\_years\_antig = antiquity of the client (in number of years)
* origin\_up = code of the electricity campaign the customer first subscribed to
* pow\_max = subscribed power
* churn = has the client churned over the next 3 months

**price\_data.csv**

* id = client company identifier
* price\_date = reference date
* price\_off\_peak\_var = price of energy for the 1st period (off peak)
* price\_peak\_var = price of energy for the 2nd period (peak)
* price\_mid\_peak\_var = price of energy for the 3rd period (mid peak)
* price\_off\_peak\_fix = price of power for the 1st period (off peak)
* price\_peak\_fix = price of power for the 2nd period (peak)
* price\_mid\_peak\_fix = price of power for the 3rd period (mid peak)

Note: some fields are hashed text strings. This preserves the privacy of the original data but the commercial meaning is retained and so they may have predictive power

Insights from the Exploratory Data Analysis

The features from the price dataset have low correlation to the target of our analysis, which is the churn rate. This means that the price of the product is not a major factor in the churn rate of the customer. This is a good thing because it means that the company can focus on other factors that are more important to the churn rate of the customer.

All consumption features are positively skewed, which means that the majority of the customers are consuming less than the average consumption. This is a good thing because it means that the company can focus on the customers that are consuming more than the average consumption.

There are no high correlation among independent features to target feature, the highest correlation are just 0.10 on margin feature. This means that the features are not highly correlated to each other, which is good because it means that the features are not redundant and can be used in the model.

Finally, the churn rate is 0.15, which means that 15% of the customers are churning. This is a good thing because it means that the company can focus on the 15% of the customers that are churning.