EDA

1.Points observed after just analysising the content of columns:

* The ususual date\_end for contract is 2016, and SME are churning, because they are not renewing the contract.
* The date\_modif\_prod is equal to date\_activ if the customer has not churned, if churned then date\_modif\_prod has churned on this date by modifying the contract.
* Forecast\_cons\_12m,forecast\_cons\_year,forecast\_discount\_energy for churned customer is 0.

2. Definition related to domain knowledge:

* Off\_peak: 10pm-7am weekdays, Cheaper avg 7 hours a day
* Peak:2pm-9pm weekdays,costly, avg 17 hr a day.
* Mid-peak: 7am to 2pm and 8pm to 10pm on weekdays, 7am to 10pm weekends

3.Numerical columns having maximum corelation with churn:

* Margin\_gross\_pow\_ele
* margin\_net\_pow\_ele
* num\_years\_antig

4.Lets define price sensitivity as difference between next month and present month divided by present month price.

5.The first row for every new customer are made to np.nan as there is no previous record available for such customer. Price\_sen dataframe is created.

6.The sum of prices for each price column in price dataframe is taken by grouping by client id, and a corelation check is done with churn column. Columns having highest corelation:

* **Total\_fix: Total sum of ‘each sum of peak, midpeak and off peak power prices’ for given customer.**
* **price\_mid\_peak\_var: sum of mid\_peak energy prices of given customer.**
* **price\_off\_peak\_fix: sum of off\_peak power prices of a given customer.**

**7.In the above designed price sensitivity dataframe:**

* **The off-peak price sensitivities have the most correlation with churn column.**

**8.Final dataframe contains following column, which was taken for modelling:**

