Summary and Findings:

Analysis of Client_Data:

I noticed that the number of Customers churned is significantly lesser than customers Retained. This shows that the data is imbalanced, and I have to balance the data before I proceed with modelling.

The data is mostly skewed to the left.

People who were not consumers of “gas”, definitely churned more than customers who were consumers of “gas” and “electricity”.

As per Correlation, **Consumption of electricity past 12 months, Consumption of gas past 12 months, Consumption of electricity the last month, gross margin, net margin, and paid subscription have been the most correlated to churn.**

At this point, I tried to verify and examine these features even more to check if they had a direct correlation with churn, and if features from the price datasheet also are responsible for churn.

But for this, I had to analyse the price data, append these datasets together, and then analyse.

Analysis of Price_data:

Price data gives insight that Retention has always been high and Churn has been low. This is an imbalance in data and will be rectified before proceeding with Modelling. The highest churn rates are observed in columns:

price_off_peak_fix, price_mid_peak_var, price_mid_peak_fix have high churn percentages.

price_off_peak_var and price_off_peak_fix seem to have the least percentage effects of Churn.

price_off_peak_var = least 4.19 (low)

price_peak_var = **8.97** (highest)

price_mid_peak_var = **8.65** (high)

price_off_peak_fix= 0.68 (least)

price_peak_fix= **8.64** (high)

price_mid_peak_fix = **8.64** (high)

We can notice that churn has been high during the periods where price was on its peak periods. However, after checking the data for price, we have noticed that price for energy and electricity has been significantly less during the peak periods compared to off peak periods.

however, we shall check how much price has been offered to clients who have churned and not churned to confirm price sensitivity.

Key Findings:

As we can see from the series, the average price offered to customers who churned in price_peak_var is 0.054 points higher than the price offered to customers who have not churned.

There is a slight price rise which might be the reason why customers have churned.

This is also true for price_mid_peak_var where the prices offered to retained customers has a difference of 0.302. And also for pice_peak_fix (difference of 2) and price_mid_peak_fix(1.35).

We can also verify this, after comparing to prices in during off-peak for both electricity and energy, yes, the given average price seem too high for off peak, however, customers have churned on those rates too.

Therefore, we can to an extent think, that price sensitivity exists. There is a slight variation in prices, and prices offered to customers could be leading to more churn than retention. Simply seeing the price rates during peak and off peak periods gives the impression that customers have been charged low, or given reasonable rates.

However, the real picture speaks when we examine deeper and check the actual rates offered to customers who have churned.

I will proceed with Pre-processing the data for Modelling purposes, and calculate price Sensitivity.