EDA Summary

Based on the exploratory data analysis performed on the PowerCo dataset, several key findings were observed.

- The dataset contains information on 15,458 customers and 26 variables related to their electricity usage and account information.
- The variables have a mix of data types, with some being numerical, categorical, and datetime.
- There are no missing values in the dataset, which is a good sign for data quality.
- Approximately 10% of customers have churned suggesting that the client base is relatively stable.
- The distribution of some variables, such as forecast_price_energy_p2, is skewed, indicating a potential need for normalization or transformation before modeling.
- There is a weak negative correlation between price sensitivity and churn, suggesting that customers who are less sensitive to price changes are more likely to churn.
- It was also observed that the distribution of the forecast prices for energy varied significantly across different times of the day and days of the week. This suggests that time could be an important factor in predicting energy prices.

Suggestions

- open-source datasets on energy consumption and weather patterns could be useful in predicting energy demand and prices. This could help in developing more accurate forecasting models for energy prices.
- Survey data from customers who have switched or are planning to switch providers on their reasons for switching, their satisfaction with the company, and their attitudes toward energy pricing and discounting strategies would also be useful.
- Competitor price data - perhaps a client is more likely to churn if a competitor has a good offer available?