PowerCo Customer Churn

EDA

Rajathadri A S

Summary of EDA

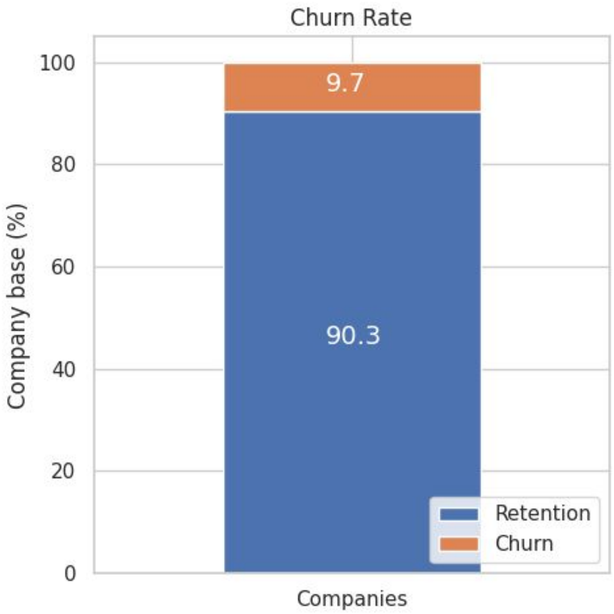
Two Datasets were provided to us:

1. Price_df : contains 193002 rows and 8 columns.
2. Client_df : contains 14606 rows and 26 columns.

There were no duplicate entries or null values in our datasets.

We initiated our EDA by merging the datasets and converting the datatypes of features in required formats.

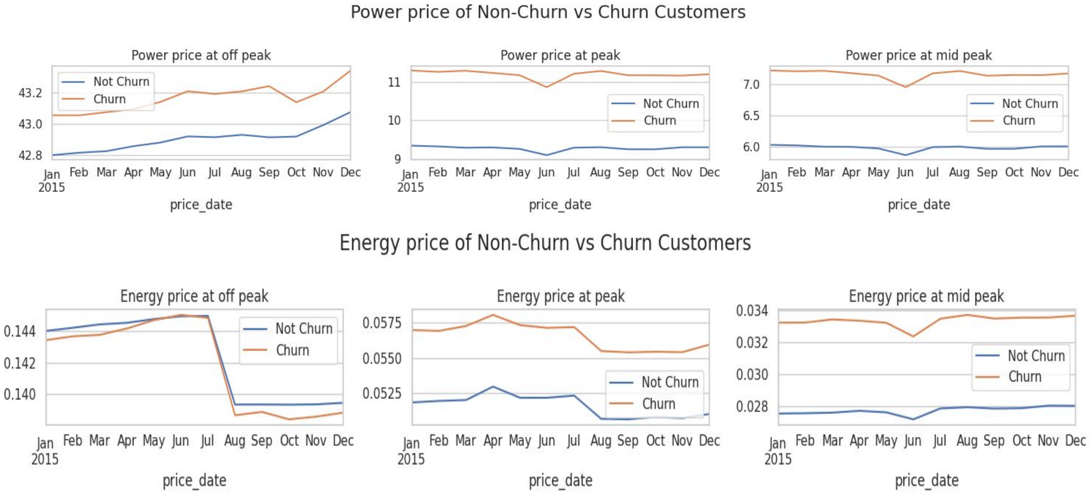
Churn Rate



Overall, about 9.7% of the customers of PowerCo are churn customers.

Energy vs Power

There is some variation of price between churn and un-churn clients. Churned clients have slightly low off-peak energy prices and high off-peak prices. But we can't confidently say this is the factor for churn. We need to analyze the data further to arrive at a concrete conclusion.



Client Price Sensitivity

The provided dataset doesn't allow us to use the actual formula to compute client price sensitivity.

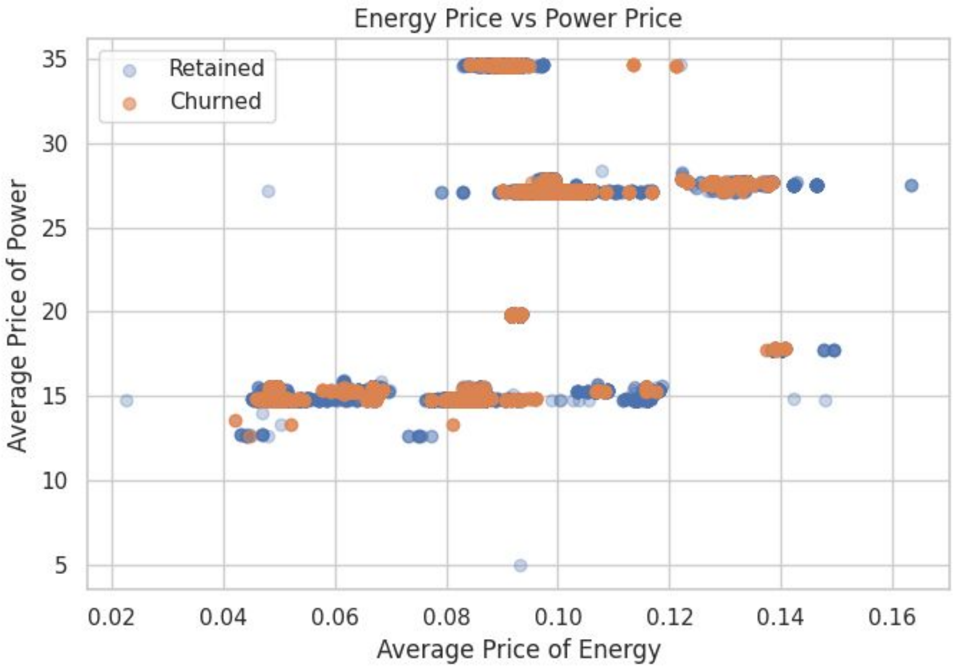
We were not able to determine whether or not price sensitivity and churn are correlated with the current data.

If PowerCo has panel data for their clients, it would be possible to compute the average client price sensitivity utilizing the real formula.

Client Price Sensitivity (contd.)

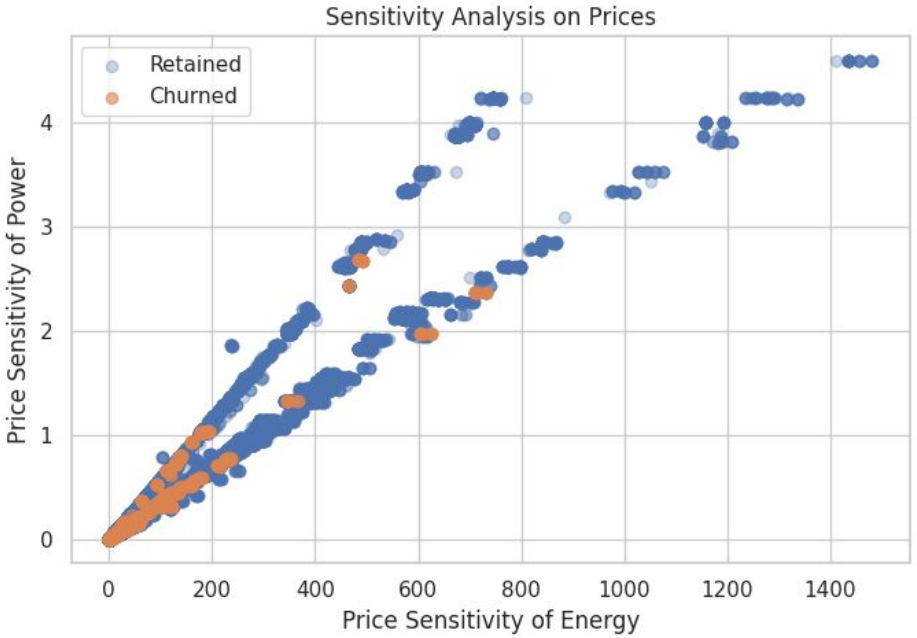
More clients seem to churn as a result of increases in the price of energy rather than power. In other words, PowerCo could likely get away with increasing power prices without any losing any clients.

Customers who churned have lower electricity and gas consumption levels than non-churners, except consumption is about even if the client does not use gas.



Client Price Sensitivity (contd.)

Customers who have a higher net margin seem to churn more. Perhaps customers are more inclined to churn if their net margin on power subscription is higher than the non-churn customer average? Whether or not the client uses gas does not differentiate the groups by much in this case.



Hypothesis Test

Null Hypothesis : Changes in price have a significant impact on customer churn, particularly in the SME segment of the gas and electricity market in Europe.

From the hypothesis test, we can observe that we fail to reject the Null Hypothesis since the p-value is less than the significance level.

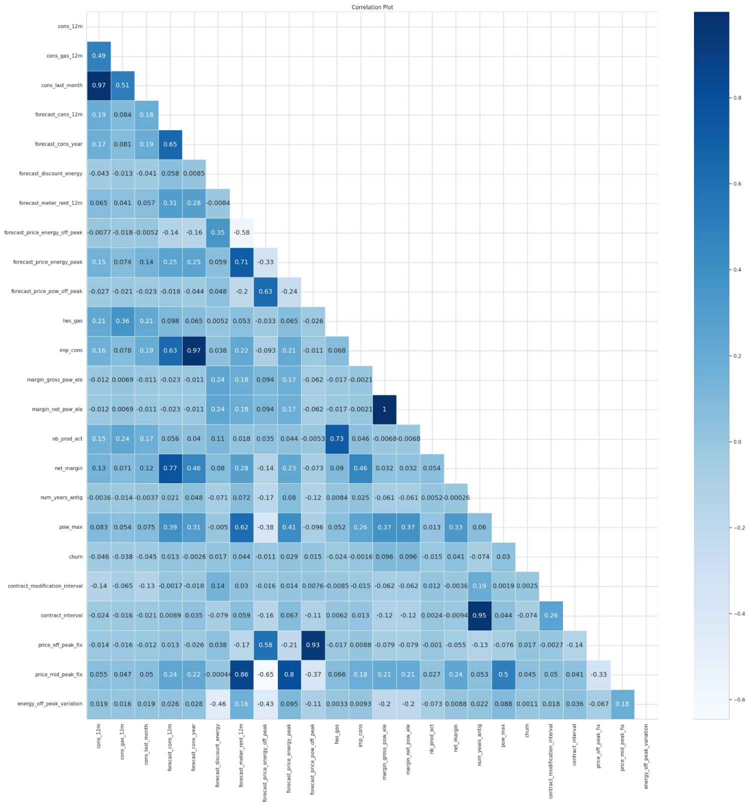
```python
# Importing ttest_ind library
from scipy.stats import ttest_ind


# Define the null hypothesis: changes in price do not significantly impact customer churn
null_hypothesis = 'Changes in price do not significantly impact customer churn'

# Define the alternative hypothesis: changes in price significantly impact customer churn
alternative_hypothesis = 'Changes in price significantly impact customer churn'


# Define the significance level
significance_level = 0.05

# Split the dataframe into two groups based on churn status
churn_yes = df[df['churn'] == 1]['net_margin']
churn_no = df[df['churn'] == 0]['net_margin']

# Perform a two-sample t-test on the two groups
t_stat, p_value = ttest_ind(churn_yes, churn_no)

# Determine whether to reject or fail to reject the null hypothesis based on the p-value
if p_value < significance_level:
    print(f'{alternative_hypothesis} (p-value = {p_value:.4f})')
else:
    print(f'{null_hypothesis} (p-value = {p_value:.4f})')
```

```
Changes in price significantly impact customer churn (p-value = 0.0000)
```

Correlation Heatmap

From the Heatmap,

We can observe that subscribed power and total net margin seems to be an influential factor for churn, with the former being slightly stronger.



End of EDA

Thank you.