Detailed Hypothesis:

- **Churning of customers:**

```
churn_counts = merged_data['churn'].value_counts()
print(churn_counts)
```

```
0    158146
1     17003
Name: churn, dtype: int64
```

The output indicates that there are 2 unique values in the “churn” column which are represented by 0 and 1. Out of 100% of data around 10% customer are churned.

There are 158146 occurrences of 0 (i.e., clients who have not churned) and 17003 occurrences of 1 (i.e., clients who have churned) in the “churn” column.

Therefore, the output suggests that out of all clients in the dataset, the vast majority have not churned over the next 3 months.

- **Mean average of churning of customers:**

```
cons_12m                          78943.054167
cons_gas_12m                       9248.511792
cons_last_month                    7213.902782
forecast_cons_12m                  1962.145197
forecast_cons_year                 1374.616950
forecast_discount_energy              1.233665
forecast_meter_rent_12m              72.002211
forecast_price_energy_off_peak        0.136484
forecast_price_energy_peak            0.054862
```

The variable with the highest average value for customers who have churned is ‘cons_12m’. This suggests that the customers who have a “high consumption of electricity in the past 12 months” are more likely to churn.

The variable with the highest average value for customers who have churned is ‘cons_12m’. This suggests that the customers who have a “high consumption of gas in the past 12 months” are more likely to churn.

- To determine which clients are mostly churned, need to analyze the features that have the stronger correlation with churn. The correlation coefficients range from -1 to 1 and indicates the strength and direction of the relationship between the features and target variable

```
margin_net_pow_ele              0.095820
margin_gross_pow_ele            0.095774
price_peak_fix                  0.046852
price_mid_peak_var              0.046120
price_mid_peak_fix              0.044446
forecast_meter_rent_12m         0.044257
net_margin                      0.041077
pow_max                         0.030414
price_peak_var                  0.029314
forecast_price_energy_peak      0.029254
forecast_discount_energy        0.017113
```

- A positive correlation coefficient between a feature and churn means that higher values of that feature are associated with a higher likelihood of churn. Negative correlation coefficient between a feature and churn means that higher values of that feature are associated with a lower likelihood of churn.
- Based on the data, it appears that "margin_net_pow_ele" and "margin_gross_pow_ele" have the strongest positive correlation with "churn" and "cons_12m", "cons_last_month", and "cons_gas_12m" have the strongest negative correlation with "churn"
- Its possible for a feature to have a strong negative correlation with the target variable (churn) and still have a higher average churn rate.
- Its possible that customers with 'high energy consumption' (i.e., high values of 'cons_12m' and 'cons_gas_12m') may also be more sensitive to changes in prices or other factors that may cause them to churn. Therefore, even though these customers are less likely to churn based on their energy consumption levels, they may still have a higher average churn rate due to other factors.

<u>Price Sensitivity:</u>

- Price sensitivity is basically how price change would influence the customer's decision
- Price sensitivity would have a significant impact on customer churn rate

```
print("Price sensitivity:", lr.coef_[0])
```

```
Price sensitivity: 0.8000000000000159
```

- The price sensitivity value of 0.80 indicates that customers in this dataset are somewhat price sensitive. If the price of a product or services increases by 1%, the demand for that product or service will decrease by approximately 0.80%
- 'A high price sensitivity rate suggests that customers are more likely to switch to a competitor if price increases'.
- This means company need to carefully balance pricing strategies with customer retention efforts to maintain their market position and profitability.

<u>Strategies to reduce churning:</u>

- **<u>Offering discount of 20%:</u>** Offering customers at high propensity to churn a 20% discount can be an effective strategy to retain, especially if pricing is a significant factor to their likelihood of churning. It's important to note, that offering a discount may not be enough to retain all customers at high risk of churning.

```
discounted_prices = merged_data['cons_12m'] * 0.8
merged_data.loc[:, 'discounted_prices'] = discounted_prices
high_churn_prob = merged_data[merged_data['churn_prob'] > 0.5]
```

- This code calculates the discounted prices for each customer in the dataset df by multiplying their annual consumption (cons_12m) by 0.80 (i.e. a 20% discount).

- This creates a new dataframe called high_churn_prob that includes only those **'customers whose churn probability is greater than 0.5, indicating a high likelihood of churning'**

- The output of the model can then be used to identify customers who are at high risk of churning and take proactive measures to retain them.

- **<u>Based on correlation factor coefficients:</u>**

```
margin_net_pow_ele                  0.095820
margin_gross_pow_ele                0.095774
forecast_meter_rent_12m             0.044257
net_margin                          0.041077
pow_max                             0.030414
forecast_price_energy_peak          0.029254
forecast_discount_energy            0.017113
forecast_price_pow_off_peak         0.014872
forecast_cons_12m                   0.012882
imp_cons                           -0.001552
forecast_cons_year                 -0.002540
forecast_price_energy_off_peak     -0.010703
nb_prod_act                        -0.014780
cons_gas_12m                       -0.037897
cons_last_month                    -0.045237
cons_12m                           -0.045918
num_years_antig                    -0.074033
```

- The correlation values between the features and churn indicate the strength and direction of their relationship with the target variable

- To identify possible churning of customers and stop them, we can you can focus in the features that have higher correlation with the target variable

- In this case 'gross margin on power subscription' and 'forecasted bill of meter rental for the next 2 months'and 'net margin on power subscription' have higher correlation coefficient.

<u>Interpretation from above results:</u>

- **This suggests that 'customers who have lower margin or higher forecasted meter rent are more likely to churn'**

- **Both 'gross margin on power subscription' and 'net margin on power subscription' has a positive but weak correlation with customer churn. This means margin increase, there is a slight increase in the likelihood of customer churn.**

- **The positive correlation coefficients between these variables and churn suggests that customers who have higher profit margins are more likely to churn**

- **While the negative correlation coefficients between other variables (such as “cons_12m” and “cons_gas_12m”) and churn suggest that customers who have higher energy consumption are less likely to churn.**

- **To reduce customer churn, it may be important to identify why the margins are increasing and if there are any actions that can be taken to reduce them without negatively impacting the customer experience**