PowerCo Customer Churn Analysis

About PowerCo and Problem Statement:

PowerCo is a prominent gas and electricity utility provider that caters to residential, corporate, and Small & Medium Enterprise (SME) customers. Due to the energy market's liberalization in Europe, there has been a significant customer churn, particularly in the SME segment. To diagnose the root cause of SME customer churn, PowerCo has partnered with BCG.

One plausible hypothesis is that changes in prices are responsible for customer churn. To address this, it would be useful to identify customers who are more likely to churn at their current price by building an effective predictive model.

Therefore, our hypothesis must align with the notion that price has a significant impact on customer churn.

Hypothesis Statement:

- **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.
- **Alternative Hypothesis** : Changes in price have no significant impact on customer churn, particularly in the SME segment of gas and electricity market in Europe.

Step By Step Procedure:

1. **Data Collection**: Collect data on customer churn, including information on customer demographics, usage patterns, and reasons for leaving. Additionally, collect data on pricing plans and discounts.
2. **Data Preprocessing:** Preprocess the data by removing any duplicate records, handling missing values, and dealing with outliers.
3. **Exploratory Data Analysis:** Conduct exploratory data analysis to understand the underlying patterns and trends in customer churn. This includes examining summary statistics, distribution plots, and correlation matrices.
4. **Feature Selection**: Select the most important features that influence customer churn. This involves using statistical techniques such as correlation analysis, chi-squared tests, and feature importance rankings.
5. **Model Selection:** Select the appropriate machine learning algorithm for predicting customer churn. This includes comparing various algorithms such as logistic regression, decision trees, and random forests, and selecting the one with the highest accuracy.
6. **Model Training:** Train the selected model on the preprocessed data. This involves splitting the data into training and testing sets, fitting the model to the training data, and evaluating the performance on the testing data.
7. **Model Tuning:** Optimize the selected model by tuning its hyperparameters using cross-validation techniques such as grid search or randomized search.
8. **Model Evaluation**: Evaluate the performance of the final model using various metrics such as accuracy, precision, recall, and F1-score.
9. **Prediction**: Finally, use the trained and optimized model to predict customer churn and identify those who are at risk of leaving. Additionally, use the model to identify the optimal discount that should be offered to incentivize customers to stay.