

Summary of Key Findings:

1. **Churn Distribution:** The distribution of churn indicators shows that there are both churned and retained customers in the dataset, allowing us to perform meaningful analysis.
2. **Variable Distributions:** The distributions of usage, forecasted usage, variable price, and fixed price show that there is variability in these parameters. This variability could be indicative of different customer behaviors and characteristics.
3. **Price Sensitivity and Churn:** The boxplot visualization indicates that customers who churn tend to have a higher median price change percentage compared to retained customers. This suggests a potential correlation between higher price sensitivity and churn.

Suggestions for Data Augmentation:

1. **Customer Demographics:** Including additional customer demographic data such as industry type, company size, and location could provide insights into how different customer segments respond to price changes.
2. **Competitor Pricing:** Obtaining data on pricing from competitors could help in assessing the competitiveness of our client's pricing strategy and its impact on churn.
3. **Economic Indicators:** Incorporating relevant economic indicators, such as inflation rates or overall market conditions, might provide a broader context for understanding customer behaviors.
4. **Promotions and Offers:** Data on past promotions, discounts, or offers provided to customers could help in analyzing the effectiveness of such strategies on customer retention.

Open Source Datasets:

1. **Energy Consumption Datasets:** Publicly available energy consumption datasets could provide additional insights into energy usage patterns across different customer segments.
2. **Consumer Price Index (CPI) Data:** CPI data could be used to analyze the impact of general price changes on customer churn and overall purchasing behavior.
3. **Online Retail Datasets:** Datasets from online retail platforms could offer insights into how consumers react to price changes in other industries, potentially providing valuable parallels.
4. **Customer Review Databases:** Sentiment analysis on customer reviews related to pricing and churn could shed light on customers' perceptions and emotions.

Conclusion:

The initial analysis suggests that price sensitivity is indeed correlated with customer churn. Customers with higher price change percentages are more likely to churn. To gain deeper insights and build a more accurate predictive model, the suggested data augmentation sources and open-source datasets could provide valuable context and additional features. By leveraging these resources, we can refine our analysis and model to better understand and address customer churn in the SME segment.

Kshitij Chaudhari