Sub-Task 3:

In summary, my exploratory data analysis showed that the dataset contains information on historical customer data, historical pricing data, and a churn indicator. The majority of the data was numerical in nature and showed no major outliers or missing values.

Regarding the hypothesis of price sensitivity being a driver of churn, the analysis showed that there was a weak negative correlation between price sensitivity and churn, but the correlation was not statistically significant. Therefore, while price sensitivity may be a factor in customer churn, it is not the only or the largest driver.

For data augmentation, I would suggest that the client provide data on customer demographics, such as age, gender, and location, as this may help to understand the customer base better and identify any specific groups that are more likely to churn. Additionally, data on customer satisfaction surveys could be useful in identifying specific pain points for customers and areas where improvements could be made to reduce churn.

Open source datasets that could be useful include energy consumption data by region, as this could help to identify any external factors, such as changes in the weather or economic conditions, that may be contributing to customer churn. Other useful datasets may include customer reviews and ratings for competing energy providers, which could help to identify areas where PowerCo could improve its services to reduce churn.

Overall, while some evidence suggests that price sensitivity may be a driver of churn, further analysis is needed to determine the full extent of its impact. By incorporating additional data sources and conducting more in-depth analysis, we can better understand customer churn behaviour and identify strategies to reduce churn and improve customer retention.