Hi AD,

I have two hypotheses on the customer churn in PowerCo, they include customer price sensitivity and black outs caused by collapse of electricity equipment used by PowerCo. Going on with the latter, we would need to model the total number of hours a customer spent in blackouts due degrading materials on churn rate. We would need the following data to confirm this hypothesis.

Data needed:

1. Power outage data: This should contain the total number of hours spent by customers in black outs
2. Material equipment age data : This should contain the year of the installed electrical infrastructures used by the customer or the customer vicinity.
3. Emergency rate data: The response rate of PowerCo to fixing of failed infrastructure.
4. Churn Data: Data indicating which customers stayed with PowerCo

Once the data is provided, we can

1. Provide EDA that shows that total hours of blackout is directly proportional to the churn rate of customers
2. Show that customers in vicinities with old electrical equipment have a higher tendency to churn.
3. Plot graphs showing that the older the materials the more possibility it is to collapse.
4. Engineer certain features for the data that can better aid prediction.
5. Choose highly predictive features for the models so as to prevent high feature space and avoid model overfitting
6. Fit a simple classification model to test our hypothesis and also ensure model interpretability.

To test the customer price sensitivity hypothesis. The following data will be need

Data needed:

1. Electricity Price Data; The price increase the customer has experienced over the years. This should also contain the average amount of electricity the customer purchases over the years.
2. Churn Data: Data indicating which customers stayed with PowerCo
3. Bills delay Data: This to show customers who had delayed in their bill payments over the years

Once the data is provided, we would proceed to:

1. Establishing a relationship with price increase and customers' who churned. We would plot several graphs to show the relationship between the price increase over the years and the customer who churned over the years.
2. We would also establish relationships between the percentage increase in customers' utility bill and their churn probability.

3. Establish a connection between customers who had delayed in bill payment and the probability for them to churn.
4. Engineer certain features for the data that can better aid prediction.
5. Choose highly predictive features for the models so as to prevent high feature space and avoid model overfitting
6. Fit a simple classification model to test our hypothesis and also ensure model interpretability.

Regards, Kolade.