

Hi AD,

I am contacting you for discussing the approach that we might take in order to explain customer churn and see how it relates to price change increases:

Dataset

In order to build a model for customer churn and use it for prediction, a dataset with the following columns will be useful:

- `sme_id` : name of the sme companies
- `sme_turnover` : turnover of the sme companies
- `sme_location` : one or more columns referring to the location of the sme companies
- `sme_industry` : the industry in which the sme is operating
- `date_of_subscription` : the first month in which the sme subscribed to Power.co services
- `average_consumption` : average monthly consumption of energy
- `percentage_change` : the percentage change between the average monthly cost that the company was paying until the current month (or before churn) and the expected price of the current month (or month of churn)
- `churn` : whether or not a given sme is a churn

Historical data could also prove to be useful, since we can use to build our custom dataset that we'll use for prediction

Exploratory & Modeling

We can start our exploratory analysis by looking at the different associations between the variables:

- What are locations with the most churns
- What are the industries with the most churns
- Who is more likely to churn, smes with higher or lower turnovers,
- Do old customers churn with the same rate as new customers

And we can then test the association between `percentage_change` and `churn`, by calculating the correlation ratio or performing a two sample t-test if the assumptions are satisfied.

Concerning the model building phase, we will need to have a binary classifier, where the target is the churn variable, and we'll use the other variables such as sme characteristics and price information as predictors. Depending on the nature of the relationship, we will look at the performance of both linear and non-linear models, we can also consider ensemble models if their metrics are significantly better.

For evaluating the results, since we definitely care about not losing customers, also we can't tolerate predicting customers as churn while they are not, since the effect is heavy on the company with the 20% discount, so the f1-score might be our go to metric.

This is how I thought about solving this problem, I would like to hear back from you as soon as possible to see what changes we can make to this approach.

Best regards,

NIJADI Mohamed