Introduction to Analytics Modeling – HW8

The power company's problem has a great deal of factors to consider when making their decision of whether or not to cut off a customer's power. These include cutting off the power of people who legitimately forgot to pay, those who legitimately cannot afford to pay, those who will eventually pay, and those who intend on never paying at all. Further than this, the power company must consider the cost of sending out an employee to shut off the power, the cost of sending them back to restore power if/when the customer picks the bill back up, the cost of delaying a non-paying heavy-user's power cut-off because they are far away, etc. With all of this in mind, I have come up with what I believe to be the most viable solution to this problem.

First, I recommend classifying the customers by their payment history/status. That is,

- Given a customer's:
  - current balance
  - number of missed payments
  - number of months since last payment
  - income
  - credit score
  - household size
- Use a support vector machine
- To determine which customers are:
  - purposefully not-paying
  - are paying (or have forgotten and/or qualify for assistance)

I suspect that purposeful non-payers would generally have more than one month's balance, a high number of missed payments, a handful of months since last payment, lower incomes, and lower credit scores. Whereas paying customers will have zero-balances and zero months since last payment, and forgetful/qualified assistance customers will have low balances, low number of missed payments, and a low number of months since last payment.

At this point, the power company should have a team work out a plan to reach out to forgetful customers (incentivize them with a small discount to enroll in autopay!) and/or work with those who need assistance paying their bill. This is outside of the scope of this analysis.

Then, for the purposeful non-payers,

- Given a purposeful non-payer's:
  - current balance
  - average monthly usage
  - distance from maintenance workers, which flows into the following calculations/factors:
    - person-hours
    - gas
    - vehicle wear
    - high possibility of having to repeat the process to *restore* their power
- Use linear regression
- To predict which customers will be the most costly

While linear regression is mostly useful for the determining the prediction-value of factors on the response, these calculations can be used to determine which customers will be most damaging to the company's bottom line. For example, if the linear regression shows that a customer's costliness is more heavily influenced by a customer's average monthly usage, then this can tell the power company to worry less about the customer's distance from the maintenance facility, or vice versa, etc. Understanding which factors have the highest influence on a customer's costliness gives the company an understanding of what to prioritize, or more importantly, gives the power company's analytics consultant (me!) a better idea of how to develop the optimization model.

Lastly,

- Given:
  - The earlier linear regression model's determination of costliest customers
  - Number of maintenance workers available
  - Traffic details for the day
- Use an optimization model
- To determine the optimal route for each driver to take.

Tracing this solution from the beginning to the end, we can see that this approach starts by identifying purposeful non-payers, determining which factors about a purposeful non-payer makes them most costly and thus which purposeful non-payers are the costliest, and finally, how to most efficiently utilize the power company's resources in cutting off the costliest purposeful non-payers.