**Power Plant Shut-Off Problem**

*A power plant is trying to determine which customers should have their power shut off without affecting customers that simply can’t pay. What’s the best way to approach this?*

I think the initial analysis that the students in question performed is a great start. By collecting a certain amount of information on the customers, a logistic regression model would allow the company to determine the probability that a customer may not pay his or her power bill. The data to be collected would be the tricky part, but an example of possibly beneficial data is outlined below in my *initial model outline*. I tried to focus on data that wouldn’t be too difficult to collect and would likely be information that the company would already have that just needed to be sorted.

**Initial Model Outline**

*{Given}* The following predictors:

1. Binary value stating whether the customer has reached out to the company regarding not being able to pay *for this month*.
2. Quantitative value stating how many times the customer has previously reached out to the power company regarding not being able to pay *for any month within the past 3 years*. (In this case, 0 is like the binary value above where the customer has not reached out).
3. Binary value stating whether the customer has purchased the home or is just renting
4. Categorical value stating how many incomes the household currently has (this could be presented in a questionnaire / optional survey where 0 represents the lack of an answer (I also realize this isn’t readily available information but could be pertinent)).
5. The cost of the power bill for this month
6. The amount that the power bill has changed over the past 3 years
7. How long the customer has been with the same power company
8. How many late payments the customer has had over the past 3 years
9. Total amount that the customer owes, if any

*{Use}*  Logistic regression

*{To}* Determine the likelihood that the customer will pay the power bill

Once the logistic regression model has been run, a confusion matrix would be beneficial to determine an appropriate p-value to set as a threshold. Working with the company to determine a rough estimate on how much it would cost for a false positive (turning off the power for somebody who will pay) compared to the amount of money the company would save for a true positive (turning off the power on somebody who was not going to pay).

After an appropriate p-value has been established, an ARIMA model can be used to determine the average power consumption over a certain amount of time based on the account history of the customer. Utilizing S = 12 would sufficiently account for natural seasons and the amount of power that would be needed to heat / cool the household respectively. This would help the company predict an average amount that could be saved by shutting off the power for a specific household.

*{Given}* The past account history of a specific household in question

*{Use}* An ARIMA model utilizing S = 12 to account for natural seasonality

*{To}* Calculate an estimated average for how much could be saved by turning off the power on that household

Once an estimated cost saving has been calculated based on the ARIMA model above, K-Means can be used to group customers into certain tiers of savings. This model would likely have to be run each month to account for differences in seasonality. For the purposes of this study, let’s assume the appropriate month has been selected and the average is used for the proceeding models.

*{Given}* The ARIMA-calculated averages

*{Use}* K-Means

*{To}* Group customers into equally “valuable” groups based on how much money could be saved by shutting off their power

Once the customers have been clustered into appropriate groups, placing them geographically on a map to visualize distance between households would be the next step. The power company could provide statistics based on how long it takes a worker to shut off the power to a single household. This information could be used to determine roughly how many households a group of workers could hit during the day. Naturally, travel time would have to be accounted for and it would need to be determined whether to hit fewer higher-value homes first or many lower-value homes. A simulation model can help us determine this as well as determine whether we would need more workers to reach optimal performance.

*{Given}* The average cost-savings per household, time needed to travel between households, and number of workers turning off power on any given day

*{Use}* A simulation model

*{To}* Determine how many workers would be needed to reach as many households as possible and determine whether it would be more profitable to target fewer high-value homes or more low-value homes.

One important consideration for the simulation would be to consider the bad publicity that would come from turning off the power on too many households at one time. If the company has received negative backlash regarding this sort of behavior before, then this could be entered in as a constraint on the simulation. If the company has not, and no other power company has, then the constraint could start off small and slowly increase over time while monitoring whether there is an increase in backlash.