**Question 18.1**

Describe analytics models and data that could be used to make good recommendations to the power company.

Here are some questions to consider:

* The bottom-line question is which shutoffs should be done each month, given the capacity constraints. One consideration is that some of the capacity – the workers’ time – is taken up by travel, so maybe the shutoffs can be scheduled in a way that increases the number of them that can be done.
* Not every shutoff is equal. Some shutoffs shouldn’t be done at all, because if the power is left on, those people are likely to pay the bill eventually. How can you identify which shutoffs should or shouldn’t be done? And among the ones to shut off, how should they be prioritized?

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

# You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}.

Have fun! Taking a real problem, and thinking through the modeling and data process to build a good solution framework, is my favorite part of analytics.

**Answer**

We split our analysis into two process: the first is to determine whether this person is never going to pay the bill. The second is to optimize the cut off process.

# Identify whether this person is never going to pay

Given {the past payment history of the clients who didn’t pay the bill last month, the contact history of such clients(whether they replied the message sent by power company), income data, the education level of the clients, employment}, we use {logistic regression model} to {predict if this person is never going to pay the bill}.

We use logistic regression model to do the prediction and classification. We use the past payment history data, the average income, the education level, the unemployment to train the logistic regression model. The response will be a probability that this client is going to pay the bill eventually.

In the logistic regression model, we can set up a probability P, as the classification threshold. If the probability client i will eventually pay the bill is less than this P, we can classify this client as someone who is never going to pay.

# Identify the expected value of cutting off a single client’s power

Given{the power using history, the economic condition, seasonal data, type of residential, response to communication}, We use{linear regression model} to {Identify the expected value of cutting off a single client’s power}

# Optimizing the cut off process.

Given {the set of clients that will never pay the bill, the expected value of cutting off a single client’s power, the location of each client who is classified never going to pay, the amount of resources(workers, trucks) we have, the time to do each cut off, the traffic patterns in this city, the gas price}, we use {optimization model} to determine {whether we want to cut off this clients’ power , how many workers do we need as well as the sequence and schedule of the cut off in order to maximize our benefit}.

The objective function is to maximize the total expected value, which is the expected benefits of cutting of the clients power – expected cost during the traveling

The decision variables are: whether we need to cut off this clients’ power(binary), how many workers do we need, the sequence and schedule of the cut off.

The constraints are: number of resources, the worker times available,