**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.

# First lets breakdown the problem statement, we are looking to:

* Reduce cost on non-paying customers
  + Deciding on whether Shutoffs should happen or not
    - If shutoffs happen, make sure they are not the one who will pay after that so as to reduce additional cost to switch on again
      * Identifying types of customers

**If shutoffs were decided:**

# Most efficient route & number of workers to hire

**Model Conceptualization**

# Customer identification:

* 3 types of customers:
  + Group 1 – Paying Customers
  + Group 2 – Customers who paid after non-payment
  + Group 3 – Customers who will never pay

# Since we are looking to identify customers within the non-paying groups (Group 2 & 3), paying customers (Group 1) and their usage patterns will not provide any significant insight. Therefore, our focus will be on Group 2 and 3.

* To begin, it is important to do exploratory data analyses on Group 2 customers. Given {historical payment data on Group 2 customers}, use {probability distribution} on the duration in terms of months it generally took for Group 2 customers to payback after reminders from the company. It will help to {understand what is the amount of leeway we can expect from customers who might pay us back after non-payment.}

# Given {the probability distribution from the analyzing Group 2 customers, how long has the current customer has not paid, together with tenure with the company, credit score, power usage history, payment methods, communications with the company’s customer service, housing type, family size as predictor variables}, use {k-nearest neighbor} as a supervised learning model on a monthly basis on customers that have not paid to {identify groups between non-paying customers that has a high likelihood to fall into Group 3.}

* Given {the likelihood of non-paying customers falling into Group 3}, we will be able to then use {confusion matrix, roc curve analysis} to {determine the probability threshold of acceptable false positive/false negative rates the company can accept.}
* Given {the false positive/negative rates}, we are able to {calculate cost at different threshold levels and optimize for the lowest trade-off between cost of power for a non-paying customer vs. cost to switch off.}

Worker optimization for Switch-offs:

# After identifying customers that are likely to fall into Group 3, we are able to use {k-Means clustering algorithm} to {identify clusters of potential Group 3 customers scheduled for switch-offs}. This is to determine the number of workers required according to number of clusters.

# Within each cluster, we will be able to then use {Dijkstra’s algorithm} to {identify the shortest path between each customer}, and use {optimization} to {find the shortest route to all locations}.