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.

Case Study

High Level steps to solve the power company's problem

1. Identify customers will not pay the bills.
2. Estimate the amount of power a potential shutoff customer will use in the next 3 months.
3. Prioritize shutoffs given the capacity constraints.
4. Find out the quickest route to shut power off.

Step 1: Identify customers will not pay the bills.

given:

- number of months delinquent
- have there been past delinquencies (binary variable)
- outstanding balance on the account
- type of service address (e.g. residential, business)
- credit score
- payment history (overdue payment)
- length of account
- method of payment (debit, credit, check, etc.)
- household income
- whether receive The Home Energy Assistance or not (binary variable)

use:

KNN (K-nearest neighbors)

to:

identify customers who will not pay

In the United States, it's illegal to use the information such as race, sex, age, zip code and other demographic factors in modeling. So, I would like to use the predictors listed above to build a classification model.

First, I will check the outliers and do some analysis to determine if an outlier should be removed or not. Then for any missing values, I would use imputation methods to replace the missing values with an estimate. And analyze the full dataset as if the imputed values were actual observed values.

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. In the power company case, I would classify customers into 3 categories: customer who paid on time, customer who forgot or got behind but eventually will pay, and customer not ever going to pay.

For customers in the last category who will not pay, the power company would have to consider shutting off their power.

Step 2: Estimate the amount of power a potential shutoff customer will use in the next 3 months.

given:

- power usage history
- weather condition
- Building Type
- Building construction date
- Number of members in the family

use:

Holt-Winters Exponential Smoothing: triple exponential smoothing including the trend and the cyclical components

to:

Estimate the amount of power a potential shutoff customer will use in the next 3 months

Once potential shutoff customers are identified, I am going to analyze their power usage history with Holt-Winters Exponential Smoothing Method including the trend and the cyclical components to predict the amount of power a potential shutoff customer will use in the next 3 months.

Exponential smoothing is a time series forecasting method for univariate data that can be extended to support data with a systematic trend or seasonal component.

Trend Analysis: It is necessary to estimate the trend factor from one month to the next since a customer's power usage can increase or decrease over time. So, it is important to identify trend components in the time series data.

Seasonality Analysis: Seasonality is another component of the time series pattern. The assumption is that data is cyclical. And in each cycle, certain periods tend to be consistently higher or lower than the average. Since the power usage is closely related to the weather, in this model, I would use 12 months in a yearly seasonal structure. The predicted power consumption for the next 3 months will be used as a factor in the next step to build an optimization model.

Step 3: Prioritize shutoffs given the capacity constraints.

In Step 3, I will use an optimization model to find out which shutoffs should be done, and the number of technicians needed. Then based on the results of Step 3, I will do a simulation and create a network model to determine the schedule and route of power shutoffs.

given:

- cost of predicted power consumption for the next 3 months from Step 2
- number of technicians
- technicians' salary
- cost of mileage / vehicle cost
- average mileage driven a day
- expected cost of turning power off
- amount of time needed to do a shutoff

use:

Optimization Model

to:

minimize the total cost and find out which shutoffs should be done and the number of technicians needed.

An optimization model has three main components: an objective function, a collection of decision variables, and a collection of constraints that restrict the values of the decision variables.

Some constraints of the optimization model for the power company are:

- The total hours a technician can work is 8 hours a day and 40 hours (about 1 and a half days) a week.
- A technician cannot drive more than 300 miles a day for safety concerns.
- The number of technicians must be greater than 0.
- Given a fixed budget for technicians' salary expenses.

The goal is to minimize the total cost. And using the optimization model to find out which shutoffs should be done and the number of technicians needed. Next, I am going to create a simulation and network model to find out the quickest route to shut power off.

Step 4: Find out the quickest route to shut power off.

given:

- customers' address got from Step 3
- number of technicians
- past traffic data
- GIS data
- the start time
- drive time
- the amount of time needed to shut off power

use:

Discrete-Event Stochastic Simulation and Network Model

to:

find out the quickest route to shut power off.

A simulation is the execution of a model, represented by a computer program that gives information about the system being investigated. Since I have got the information of which

shutoffs should be done and the number of technicians needed, now I am going to use DiscreteEvent Stochastic Simulation and Network Model to find out the quickest route to shut off power.

The reason I am looking for the quickest route instead of the shortest is that the technicians are paid hourly. Driving with the quickest route can save labor costs. On the other hand, the shortest route may take longer to drive to the destination due to the traffic conditions. The goal is to find out the most efficient way to shut off power and minimize the company's operational cost.