**Power Company – Case Study**

Our first case study is to work with our local power company and help them efficiently decide how to use their manpower to shut off delinquent customers' power. We will focus on two main approaches: Identifying those customers that will never pay and optimizing the crews' routes to maximize their ability to shut off customers power.

For the first issue, we have to be wary of protected data classes. We will not be using any data that could identify a protected class of data such as gender, race, age, but we will also not be using any data elements that could be highly correlated to a protected class such as zip code. Since the question is resolving around propensity to pay, I would think starting with a credit record would give us some good data to analyze our clients paying potential.

**Given** {Income, Credit History[Accounts delinquent, number of months missed payment, credit score], Payment history with Utility Company[Amounts Due, number of days late, defaults]}

**Use** {K Means Clustering}

**To** {Quickly identify any natural clusters that may exist in the data.}

With our clusters, we are hoping to see a breakout of groups that pay, would be likely to pay in the future, will not pay and are unable to pay, and able to pay but will not pay. If we find some logical clusters existing, we can take our group that always pays and eliminate them and our group that is not paying and is unable to pay and setup a program to mail them a post card explaining charity programs.

Now we have two groups, both have not been paying and both seem like they could afford to pay, so now we have to decide how likely they are to never pay us.

**Given** {Our two clusters from the above results that are not paying, but should be able to}

**Use** {Logistic Regression}

**To** {Identify a propensity to pay based on original factors}

We will need to analyze our results and adjust our threshold to find the most efficient customers to have their power shut off. We want to exclude those customers who we believe will be paying the bill in the coming months.

We can even go farther on our customers and try to analyze how much power they will use and highlight the cost to the Power Company of leaving their power on. Taking our two groups of customers who may not pay, we can split them into groups with long account histories and newer customers.

**Given** {Customer Credit History, Account Payment History, past usage}

**Use** {ARIMA}

**To** {Adjusting for seasonality to estimate the amount of power they will use next month}

For newer customers, with shorter account history

**Given** {Customer Credit History, Account Payment History, past usage, seasonal averages for customer base}

**Use** {Linear Regression}

**To** {Using the overall customer averages to adjust for seasonality, the expected power usage next month}

Now we should have a pool of customers with a calculated probability that they will pay, as well as their potential cost in unpaid power that they will accrue in the coming month. Combining these two elements will give us a score of our high target customers to have their power shut off. We can map these customers to see if any natural clusters exist. Knowing our number of drivers, we could use that as our nodes, so each driver would have a compact geographical area.

**Given** {Customers identified Previously, Customers Address}

**Use** {k means clustering for K drivers}

**To** {Identify the geographically close customers to assign to drivers}

Then for each cluster we could run an optimization model to see which clients should have their power shut off.

**Given** {Calculated customer propensity to pay, Customers Address, average drive time, average work time}

**Use** {Optimization with binary customer variable and constraints to hours of work in a day and drive and work times}

**To** {Identify the customers that will be assigned for power shut off}

This is probably already more work than one team would want to tackle, but if we were still looking to add more value for the Power Company, we could run some simulations to see the potential efficiency improvements of adding more crews.

**Given** {# of Crews, cost of crew, average drive time, average work time, # of monthly customers to shut off, Loss of income from not shutting off customer}

**Use** {Simulation to see effect on adding more crews}

**To** {Identify the potential improvements of adding more work crews}

This is probably an over designed impractical solution, but this is just homework so I was trying to have fun with covering all possibilities. I believe I have given the Power Company some thoughts on how they can pursue their problem and look forward to see what my fellow students have developed.