

Question 18.1

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

Summary:

Power Company facing problem about some of the customer don't pay their bills.

**Not the group aren't able to pay, since power company has fund to support those who aren't able to pay.*

Main Concern: People who could pay but don't.

Goal:

1. Shut off power for those who are not ever going to pay.

**Exclude for who just miss payment or whoever eventually pay.*

2. Prioritized and be efficient when decide the shut off location due to the shut off had to be done manually and limit of resource of people doing it.

I. Analytic model

- Classification Model:

Logistic Regression: Identify if a customer able to pay

Logistic Regression: Out of those who able to pay, identify those never pay

- Regression Model:

Linear Regression: Determine cost leave power on for the next quarter

- Optimization:

Determine the daily schedule for power shut off task for available personnel based on object function.

II. Data need

- To identify if a customer able to pay never pay:

Data of the Payment history, credit score, household income, geolocation, Age, Number of people in as a family, Current employment status, criminal record (Company like comScore or Experian will have even more detail information regarding a person's information, including if you own a pet or if you are a wine lover, so data is accessible in reality.)

But the most straight forward data and most weights on should be the payment history. Those who never pay will have a long history with always past due amount attached with the account and no banking information attached as well, should be easy to find these cases. I would use historical data about customers who paid/did not pay to train a model to make predictions.

- To determine the cost leave power on for the next quarter:

Data of median of power usage for the similar area with same size home and number of family members, next quarter seasonality median power usage, if owns an electric car or not, potential cost of abuse electric usage due to not shut off

- Determine the daily schedule for power shut off task for available personnel based on object function

Data of cost of leave power on from previous step, cost of sending someone to shutoff the power, treat geolocation cluster as non-pay group, and based on the distance between electricity box in the cluster, we will have time of travel, time stop for shut down, weather, best route (For example, UPS is design the delivery truck route that avoid as much left turn as possible) Cost of hire new personnel/Training for shutdown(Less likely since technology is allow you shutdown remotely which will reduce most of the cost of shutdown.) Also per shutdown, the cost of turn back on when a customer realized and pay off all the dues. (Assuming that situation is rare since they are in the category who able to pay but never pay for the entire time)

III. Analytic Process

On the high level, this involve two part of the solution for the Power Company.

First part is to identify target customer need to be shut off the power. My approach is dividing the first part into two steps, first step is identify those who able to pay the bill either by themselves or by government or third-party. These are all good retention customers regardless how they paid for the bill, as long as they paid the due they will be in this group. For those unable to paid, I will simply exclude them since they are not in the shutdown consideration this time. Power company has program to support them. Now out of those able to paid, further classify which are never paid at all. This is also a logistic regression, also two group in this step: who never pay vs. who paid at least 1 time. The former are those the company should bring into for the next step.

Second part is to determined shutdown priority. Based on the output from part 1, my understanding was whichever increase the most business lost after minus the sum of total cost of shutdown should be consider to shutdown first with higher priority. For these group of people, I suggest we should used group total value, group by geolocation for minimum travelling purpose and total value of business lost when leave the power on for the next quarter. Since the situation indicating we don't have the estimate usage by those group. We need to use regression model to predict the usage based on factors like season or property type to get the usage to determine the cost. Also, the other costs which including hiring someone to do the work, optimization of split the time into a proper schedule to shut off.

IV. Conclusion

- **Given**

Payment history, credit score, household income, geolocation, Age, Number of people in as a family, Current employment status, criminal record.

power usage for the similar area with same size home and number of family members, next quarter seasonality median power usage, if owns an electric car or not, potential cost of abuse electric usage due to not shut off.

cost of leave power on from previous step, cost of sending someone to shutoff the power

Cost of hire new personnel/Training for shutdown

cost of turn back on, Geolocation

- **Use**

Tree model with branch of classification customer into group, use logistic regression model to identify the probability if a customer is fall into a specific group. Turn the output as binary variable for further analysis.

Use Linear regression to estimate cost function and use optimization on the function.

- **To**

Identify the probability if a customer is fall into a specific group, if a person falls into target group which can pay but not pay.

Estimate cost of leave the power on, cost of sending someone to the field do the shut off work and ending someone to turn back on. Calculated the total cost of not shut off vs. shut off, based on the current business loss to the determined and shut off priority by those group will generate the highest business loss.