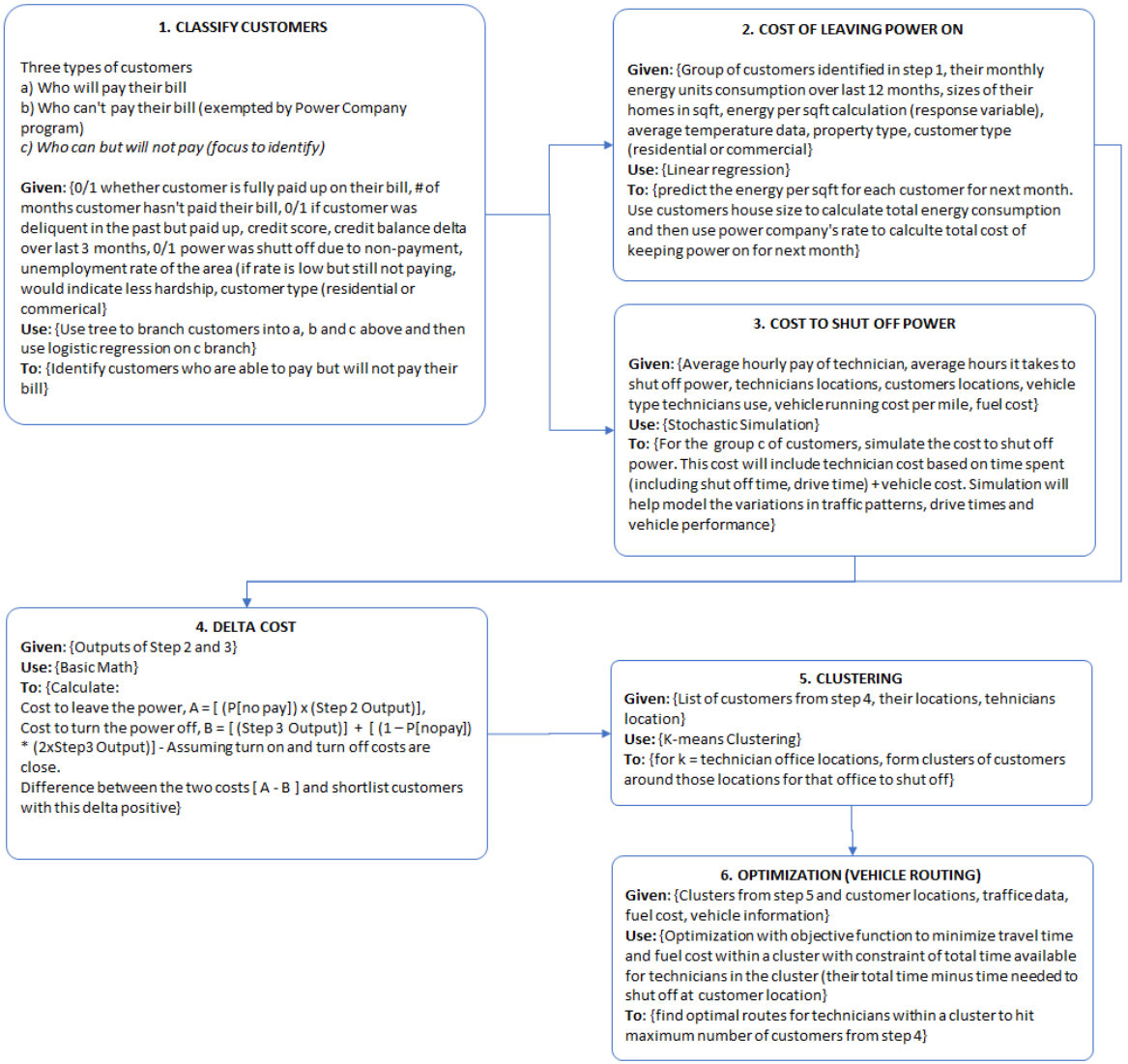
Week 13 Assignment – Power Company Case

Omer Farooq (EDx ID: mfarooq4)

04/08/2020

Table of Contents

SUMMARY


1. CLASSIFY CUSTOMERS
Three types of customers
a) Who will pay their bill
b) Who can't pay their bill (exempted by Power Company program)
c) Who can but will not pay (focus to identify)
Given: {0/1 whether customer is fully paid up on their bill, # of months customer hasn't paid their bill, 0/1 if customer was deliquent in the past but paid up, credit score, credit balance delta over last 3 months, 0/1 power was shutt off due to non-payment, unemployment rate of the area (if rate is low but still not paying, would indicate less hardship, customer type (residential or commerical}
Use: {Use tree to branch customers into a, b and c above and then use logistic regression on c branch}
To: {Identify customers who are able to pay but will not pay their bill}
2. COST OF LEAVING POWER ON
Given: {Group of customers identified in step 1, their monthly energy units consumption over last 12 months, sizes of their homes in sqft, energy per sqft calculation (response variable), average temperature data, property type, customer type (residential or commercial}
Use: {Linear regression}
To: {predict the energy per sqft for each customer for next month. Use customers house size to calculate total energy consumption and then use power company's rate to calculte total cost of keeping power on for next month}
3. COST TO SHUT OFF POWER
Given: {Average hourly pay of technician, average hours it takes to shut off power, technicians locations, customers locations, vehicle type technicians use, vehicle running cost per mile, fuel cost}
Use: {Stochastic Simulation}
To: {For the group c of customers, simulate the cost to shut off power. This cost will include technician cost based on time spent (including shut off time, drive time) + vehicle cost. Simulation will help model the variations in traffic patterns, drive times and vehicle performance}
4. DELTA COST
Given: {Outputs of Step 2 and 3}
Use: {Basic Math}
To: {Calculate:
Cost to leave the power, A = [(P[no pay]) x (Step 2 Output)],
Cost to turn the power off, B = [(Step 3 Output)] + [(1 – P[nopay]) * (2xStep3 Output)] - Assuming turn on and turn off costs are close.
Difference between the two costs [A - B] and shortlist customers with this delta positive}
5. CLUSTERING
Given: {List of customers from step 4, their locations, tehnicians location}
Use: {K-means Clustering}
To: {for k = technician office locations, form clusters of customers around those locations for that office to shut off}
6. OPTIMIZATION (VEHICLE ROUTING)
Given: {Clusters from step 5 and customer locations, traffice data, fuel cost, vehicle information}
Use: {Optimization with objective function to minimize travel time and fuel cost within a cluster with constraint of total time available for technicians in the cluster (their total time minus time needed to shut off at customer location}
To: {find optimal routes for technicians within a cluster to hit maximum number of customers from step 4}

PROBLEM STATEMENT

The Power Company wants to manage their revenue and costs by effectively shutting off power for the customers who are not paying their bills. But they do not want to create more hardship for the customers who legitimately can not pay their bills due to financial stress. They have a program for such customers. But the ones who can pay but are intentionally not paying their bill, the Power Company wants to analyze the option of shutting off their power as long as the cost to shut off is not higher than the expected cost to keep the power on for that customer for a month.

Problem has following components:

1. Who are the customers who can pay but are not paying their bill?
2. What is the cost of leaving their power on versus cost of shutting their power off?
3. What is the subset of these customers for which shutting off power makes sense?
4. Operationally, what is the best solution to shut off the power for the identified customers?

STEP 1 – CUSTOMER CLASSIFICATION

There are 3 types of customers here,

a) Customer will pay their bills on time

b) Customers who will not pay and can't pay

c) Customers who will not pay but can pay.

Types a and b are not the focus in this problem b/c the power company has a program for type b customers. Our goal in this step is to identify the customers of type c.

Data:

We will need following data at minimum for this step:

- 0/1 whether customer is fully paid up on their bill
- # of months customer hasn't paid their bill
- 0/1 if customer was delinquent in the past but paid up
- Credit score
- Credit balance delta over last 3 months. This would be a good indicator of economic hardship as well. If the balance went up significant over last 3 months, it would

indicate that customer might fit in type b. If balance was unchanged but customer still hasn't paid, this would indicate that customer might belong to type c.

- 0/1 if power was shut off due to non-payment in the past 12 months
- Unemployment rate of the area (if rate is low but still not paying, would indicate less hardship,
- customer type (residential or commercial
- A 0/1 binary response factor based on historical data. If a customer was not in Power Company program to support their bill and has been paying up their bill on time that customer is 0, otherwise 1.

Use:

Use a tree-based model to branch customers into a, b and c above and then use logistic regression on c branch.

To

Identify customers who can pay but will not pay their bill. Model will provide probabilities which will need a threshold analysis to determine 0 and 1 (our target group). This also allows to manage the number of customers in group c but tightening the threshold.

STEP 2 – COST OF LEAVING POWER ON

In this step, focus is to predict the cost of leaving the power on for the type c customers from step 1. This cost will be:

Cost of leaving power on = Energy units used by type c customers in next month (units)

x Avg. cost per unit ($/unit)

Given:

- Group of type c customers identified in step 1
- Their monthly energy units consumption over last 12 months
- Sizes of their homes in square feet
- Energy per sq-ft calculation (response variable)
- Average temperature data
- property type (single family, multi-family etc.)
- Customer type (residential or commercial)

Use:

Linear regression

To:

- Predict the energy per sq-ft for each customer for next month.
- Use customers house size to calculate total energy consumption
- Then use power company's rate to calculate total cost of keeping power on for next month

STEP 3 – COST TO SHUT OFF POWER

Focus in this step is to identify the cost of shutting off power for all type c customers identified in step 1. This cost is function of multiple factors.

Given:

- Average hourly pay of technician
- Average hours it takes to shut off power (this would be the time spent at customer location)
- Technicians locations
- Customers locations
- Vehicle type technicians use
- Vehicle running cost per mile
- Vehicle maintenance cost per mile
- Traffic data for the area with a distribution for variation
- Average fuel cost

Use:

Stochastic Simulation

To:

For the group c of customers from step 1, simulate the cost to shut off power. This cost will include:

Total technician cost = [Avg $/hour

x (Shut off time at customer location + drive time)]

+ Vehicle fuel cost + Vehicle maintenance cost

Simulation will help model the variations in traffic patterns, drive times and vehicle performance.

STEP 4 – DELTA COST

We need a way to identify the subset of type c customers, for which shutting off Power makes sense. This step focuses on that.

Given:

- Step 2 output: Total cost of leaving power on
- Step 3 output: Total cost of shutting power off

Use:

Basic Math

To:

Calculate probabilistic costs:

Cost to leave the power, A = [(P[no pay]) x (Step 2 Output)]

Cost to turn the power off, B = [(Step 3 Output)] + [(1 – P[no pay]) * (2 x Step3 Output)]
Assuming turn on and turn off costs are close, thus similar.

Calculate the difference between the two costs [A - B] and shortlist customers with positive delta. This would indicate that for these customers, the cost to leave power on is higher even after deducting the cost to shut off the power.

STEP 5 – CLUSTERING

The goal of this step is to take the subset of type c customers from step 4 and identify the right groupings of these customers based on the locations of Power Company's technicians.

Given:

- Locations of customers from step 4
- Technician offices locations.

Use:

K-means Clustering

To:

For k = technician office locations, form clusters of customers around those locations. This would give a pool of customers to each office location to handle shut offs.

STEP 6 – VEHICLE ROUTING OPTIMIZATION

Now that we know the customers, we want to shut off power for and have logical groups of these customers with technician offices, last step is to provide these technician offices with optimum routes for an optimum list of customers.

Given:

- Clusters from step 5
- Deltas from step 4
- Customer locations
- traffic data
- Fuel cost
- Vehicle information

Use:

Optimization with objective function to minimize a combined function of

- Travel time within a cluster (minimizing time will automatically minimize cost)
- Total delta from step 4

This function would look something like this:

Obj Function = Min [Travel Time + α x Delta from step 4]

This function would allow to balance b/w the total time available and hitting customers with maximum deltas.

Constraint would be of total travel time available for technicians in the cluster (their total time minus time needed to shut off at customer location).

To:

Find optimal routes for technicians within a cluster to hit maximum number of customers from step 4 with maximum deltas.