

Status Report for NOVEC Team

Anita Ahn Meselework Aytenifsu Randall Barfield Daniel Kim

October 27th 2016

Group meeting with Professor: Tuesday 18th

Questions that were brought up during the meeting:

- How will NOVEC use these customer segments to predict total system's energy usage?
- How can we validate that the number of customer segments should be 6 or 8?
- How are the regions distinguished from one another? What do the different customer clusters mean?
- How can we validate that the stratified sampling data that we have from NOVEC is actually a good representation of the entire population's data? Does it capture variability?
 - May need to follow up with Statistics Department professor to verify
- What is the course of action if other months show different customer segments?
- How can we prove that the customer segments is a good way to cluster NOVEC's customer population?

In progress:

- Scheduled meeting with Statistics Department Professor Xu on Monday 31st to discuss the sampling techniques that NOVEC used for its stratified sampling
- Validating customer segments through using weighted clustering technique to find total NOVEC's system's energy usage. Client has identified that they would want 5% Margin of Error for these clusters
- Fix goals and presentation slides to reflect NOVEC's purpose in customer segmentation

Completed:

- Emailed Professor with presentation slides that explain NOVEC's background, weighted clustering method, and how the team's analysis will be used to represent total energy usage
- Created an online website for team NOVEC
- Scheduled meeting with Professor Xu

After the meeting with the professor on Tuesday, October 18th, the NOVEC team identified several questions that should be addressed by the final presentation. Through using metrics like Load Factor, Demand Factor, and Coincident Peak Factor, the NOVEC team had come up with 8 different clusters of customer segments. The size of the clusters was determined by using K means square method and 8 clusters was the number that decreased the sum of square errors significantly. However, in order to validate that the customer segments are representative of the total population, the team had to come up with a metric that measures the effectiveness of the clusters. It was important to understand how NOVEC will use this customer segmentation internally so that the team will be able to measure whether or not the segments prove useful for the company. Without this understanding, the team would not know whether or not the current analysis and clusters is correct, since there is not a clear distinction between different customer segments currently represented by our clustering technique.

The key questions that were asked by the professor were: How can NOVEC use the clusters that the team has made? What is dividing one region of clusters from another? How does this customer segmentation relate to other information that NOVEC has, to be useful? Why is this clustering better than other clusters? Does the stratified sampling technique provide good data that accurately represents the population of NOVEC's customers? In order to answer these questions, the team has to understand NOVEC's system and how the company operates, to understand how the customer segments will be used to achieve its intended purpose.

NOVEC is an electricity provider that has a collection of circuits (or power lines) that comprise of different customers. NOVEC plans system construction based on the forecasts of future capacity for these circuits, where capacity is defined to be the maximum amount of electricity customers will use for one hour. The reason why NOVEC needs to segment customers is because the company does not have good prediction down to individual circuit level of detail although the aggregate level of prediction is relatively accurate. If the project team is able to identify good customer segments, it will help NOVEC predict clients' electricity usage by circuits, therefore increasing their capability to plan and construct at circuit level of detail.

With the provided sample of customer's electricity usage, NOVEC wants to identify clusters that contain customers of similar usage pattern that can explain the group's contribution towards NOVEC's system peak and total energy usage. In order to tie the customer segments to NOVEC's total system energy usage, the project team will use the Weighting Mechanic. If a circuit is identified to consist of customers from clusters 1,2,3, and 4, the project group would first find out the weight that each cluster has on the circuit, determined by the number of customers. Each cluster will be weighted by the number of customers it has to produce a "weighted electricity usage" and when you sum up all these numbers for each cluster in the circuit, it will tell you the "total circuit energy usage" for the respective circuit. In order to validate whether or not the current clusters provide a good customer segmentation, the team will sum up all the "total circuit energy usage" for each circuit and compare that number with NOVEC's historical data on total system's energy usage. A good clustering will be able to produce a "total system energy usage" number that is within 5% margin of error when compared to historical records.

Typically, the circuit capacity is reached during the month of July, which is why the project team has decided to segment the customers for July before doing other months. However, if the other months show a customer segmentation that is very different from the month of July, the team will have to conclude that the customer segmentation is not consistent over the months.

The team will continue its efforts on validating the current customer segments through this weighted methodology and discuss the stratified clustering technique with Professor Xu from the Statistics department to ensure that the data received from NOVEC will be able to represent the population well for our analysis to be considered valid.