

Applying Data Analytics to Improve Sustainability of Public Utilities

CEJS Mid-Year Report

Gareth P. Green
Department of Economics

The primary objective of this research project is to improve the environmental and social sustainability of public utilities. The method is to first learn more about customer behavior by applying data analytics to a variety of data sets collected by Seattle Public Utilities (SPU), then examine policy to promote sustainable behavior in a just manner. A secondary objective is to develop course materials for the new and existing courses. The third objective is to introduce techniques from data analytics to understanding sustainability, environmental justice, economics, and developing public policy.

Activities and Progress to date:

Spring 2015: Acquired data sets from SPU and met to discuss questions of interest. SPU is particularly interested in additional analysis on race regarding quality of service and environmental behavior. Began exploratory analysis of the data sets. I found it is going to be more difficult to merge the data sets than I initially thought because there is not one “best” variable in common across the different types of data sets. For example, I could merge the data sets based on neighborhood but that could reduce variation within neighborhoods. Instead I have been working within given data sets to better understand them. I may try merging the data sets based on different common variables and see how the results differ. As such, I have made steps toward fulfilling the first project objective, but still have a lot of analysis to complete.

Summer 2015: I attended the American Environmental and Resource Economics conference June 2-5. In one sense the conference was a disappointment because it was advertised as having a primary theme of “the use of big data in environmental and natural resource applications”; however, there was only one session that applied data analytics to environmental and natural resource issues. On the other hand, this is good because it highlights how the use of data analytics is starkly missing from this field so the area may be ripe for publication of results from this project, which relates to the third project objective.

Fall 2015: I was not able to make much progress during fall quarter due to teaching two intensive courses and having to tend to a heavy load of chair duties. I did develop an ANOVA assignment for my ECON 3100 Quantitative Methods and Applications class examining customer service ratings by race and neighborhood, which relates to the second project objective.

Winter 2016: I continued data analysis using a variety of data mining tools in R, including nearest neighbors, naïve Bayes, decision trees, support vector machines, several clustering, and association rules. Some of these methods were not appropriate for this project, but my goal is to learn by application. I also designed a policy analysis for intermediate microeconomics.

Plans for the remainder of the year

Unfortunately I have a long way to go and am not as far along on this project as I had hoped. However, spring quarter is when I have my course release so I hope to get a lot accomplished. My long to do list includes: 1) finalizing whether it is beneficial to merge the different data sets; 2) complete analysis

based on race and environmental behavior; 3) write up project research for presentation and publication; 4) complete and submit final report.