Reducing Household Carbon Emissions using Machine Learning

Stepheny Perez, Omid Jafari, Tanya Olivas Rico

Fall 2018

1 Introduction

Our team found an analytic challenge appropriate for this project. This is a real-world example with a tangible reward other than the experience we will acquire using techniques learned during this course. The dataset provided contains information about households and their power consumption and associated carbon footprint. In addition to consumption, quality of life importance (or QoLI) is provided for each household, which indicates how important that particular activity is to that household. Using this information, we want to suggest consumption modifications to households to lower their carbon footprint without reducing their quality of life.

By analyzing said data, the team will be able to describe how the data product succeeds mathematically in minimizing an individual's carbon footprint with minimal negative impact on their utility. The end product is implemented operational code that can be run with any data set provided by the client in order to make accurate predictions for said read-world problem.

2 Motivation

Environmental sustainability is an important goal for our generation. In this project, we hope to show how small changes in a household can make a large difference in overall carbon footprint. Another incentive is winning $2200 per person and a 3-day and 2-night trip to San Francisco, CA in March 2019, should we be selected by Wells Fargo after completing this project.

3 Methodology

3.1 Dataset

We are given a data set describing 1000 individuals' daily activities; there are 27 activities which include units of consumption and are given a Quality of Life factor ranging 1-100 each. There is also a list of equipment each individual might use for each activity. We are also given the Carbon footprint of each equipment for each activity.

Below is a simplified representation of the dataset provided to us for this challenge.

household1	a1 consumption	a1 QoLI	a1 CF
household1	a1 consumption	a1 QoLI	a2 CF
household1	a3 consumption	a3 QoLI	a2 CF
...	...	...	...
household2	a1 consumption	a1 QoLI	a1 CF
household2	a1 consumption	a1 QoLI	a2 CF
household2	a3 consumption	a3 QoLI	a2 CF
...	...	...	...

3.2 Preprocessing

The first step is preprocessing the data. The data was provided in an Excel file with two sheets, where the second sheet is simply a lookup table containing the carbon footprint of certain activities. We combine this data with the full dataset so the carbon footprint of an activity is readily available for analysis.

The second step is to replace all empty cells with '0' (after concluding this was the intention of an empty cell), and to normalize the data.

The third step is to prepare for clustering. As it currently stands, each household has a row for each activity in the dataset. We combine and rearrange these values so each household has only one row in the dataset, along with consumption for each activity and QoLI for each activity.

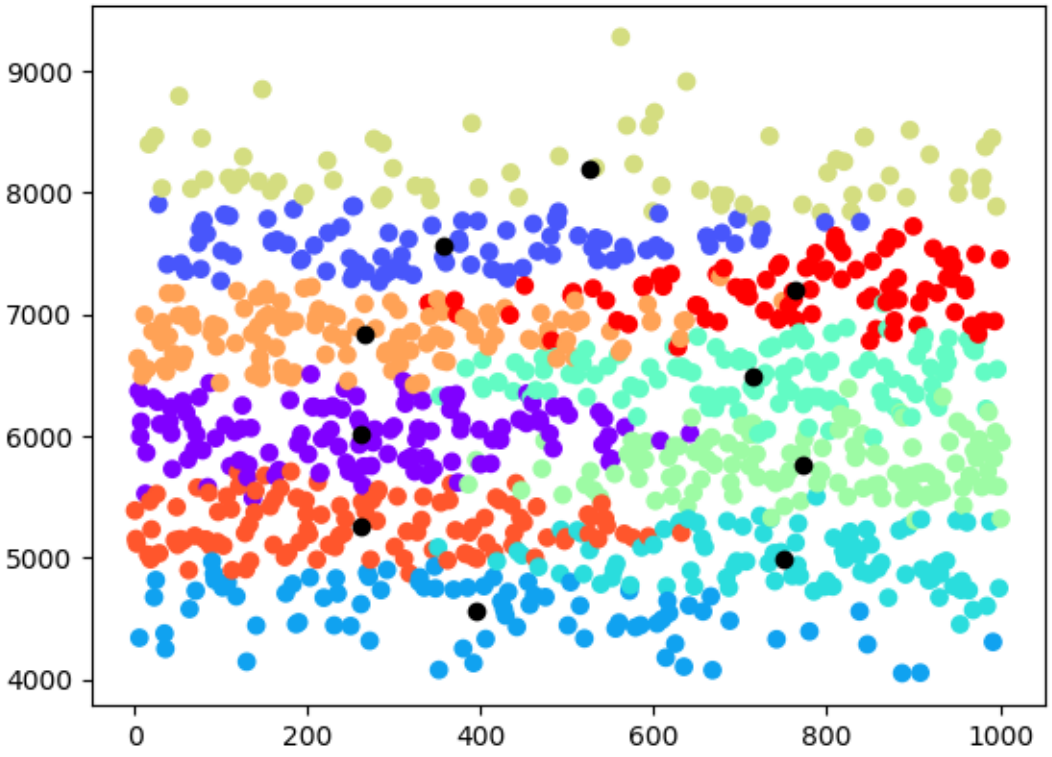
3.3 Clustering

The second step is to cluster the rearranged data. This should give us clusters of households with similar usage patterns. We will be using the k-means clustering method, though this may change in the future if we learn new clustering methods that are more appropriate. We will determine our k value by using the elbow method, because we don't know exactly how many clusters will be needed to represent similar households.

3.4 Consumption Suggestions

Next, we will suggest consumption modifications to each cluster of households. Because each cluster represents similar households in terms of consumption and QoLI, we can assume that any consumption modifications will similarly affect each household in that cluster and therefore be appropriate for each household in that cluster. We have not yet implemented this step.

4 Preliminary Results



As you can see above, we have done clustering to group each household together with other households that have similar behavioral and preference patterns. Now, in the next steps, we can suggest consumption modifications to each group dependent upon their consumption and preferences.