Motivation:

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.

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.

Problem Definition:

Using machine learning, we will create a data product to help individuals optimize the balance between their carbon footprint and 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.

Currently, we cannot predict how complicated this project will present itself for our team; we can assume it will be no easy task given the fact that it is an open challenge to individuals currently attending college for Computer Science.

Data Analysis:

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.

Proposed Solution:

We will analyze data in order to recommend environmentally-friendly changes to everyday actions without lessening individuals' quality of life.

By creating a Machine Learning algorithm, we will find substitutes for these high-carbon-emitting activities per person, without compromising the quality of life coefficient given to each activity. We are to note that the data is not real, and our final solution will work with any data set that is fed to it.

Our data preprocessing includes substituting empty cells of the first sheet (individuals information) with zeros, as well as substituting the empty cells of the second sheet (Carbon footprints) with numbers as close to zero as possible in order to reduce the error in our calculations.

Each activity is given either a zero or one under equipment used. We take the product of the equipment and the values in the second sheet and average them. A real data set will not have zeroes in all activities, therefore we can disregard the all zero activities in this data set for now. We created a new variable with said average, so we can keep track of it. This calculation is saved per individual, showing their total quality of life as an average of all the activities performed.

As a goal, once the algorithm is run, we will make suggestions on activity changes that drop carbon-footprint without reducing the individual's quality of life, thus promoting environmental sustainability, which includes accelerating the transition to a low-carbon economy. The challenge asks us to provide these suggestions in taking individual actions in order to encourage collective responsibility to help achieve said environmental sustainability.