Hello everyone, for my project, I partnered with ezhome, which is a tech start up that provides lawn services to homeowners in the bay area. ezhome specifically wanted me to find features that predict ratings, 1 to 5 stars, given by the new users after each visit.

First, I sat down with ezhome to come up with the relevant features, which I then extracted from their Amazon redshift data warehouse using SQL queries. After manipulating the data, I used a random forest classifier model, because it gave me higher accuracy. And also because it is a flexible method that scores features on their importance, which is exactly what ezhome wanted.

So, here are the important features that predict user ratings. The very top one, user engagement, measures how often users log in, either using their phone app or the website. The second one, gardener consistency, is a measure of how often homes get the same gardeners. Basically, the number of different gardeners divided by the number of visits. So, my task was done, I had what ezhome wanted. Then, I started thinking if there is something that could be even more useful to them, because what they are really looking for is a lever they can pull to improve user ratings. If I could quantify the impact of one of the easier-to-change features using a causal model, that could be really useful.

Specifically, I was interested in this feature or variable, gardener consistency. So, why was this feature included in the first place?

Before delving deep into the project, I had started by reading feedback comments left by users, because I thought that will give me valuable insights into what the users were thinking precisely when they gave these ratings. I noticed that the highest rated reviews tend to use the first names of the gardeners while the lowest reviews used more generic nouns, which led me to think that knowing your gardener or having the same gardener consistently could probably result in higher ratings. I could think of two possible reasons for this happening: first, familiarity breeds fondness towards the gardener. Second, and perhaps more important, the gardener knows more about the landscape of the home as well as the owner’s preferences and does a better job. Either way, it is interesting and could be useful.

To explore this further, I focused the first two visits. I noticed that, for the second visit, some homes get the same gardener from the first visit, while the others do not. I dug around to figure out how the gardener-home assignment works.

Turns out, it is randomly assigned! In other words, ezhome was unknowingly running an experiment! Now to estimate the impact of sending the same gardener consistently, I could exploit this natural experiment and compare the ratings between these two groups, which I did by running an ordinal logistic regression. And the results were statistically significant.

To summarize, ezhome has at least 10% higher chance of getting a 5-star rating if they send a familiar gardener, no matter what their rating was in the first visit.

The flip-side is also true. The probability of getting a bad rating, 1-star rating, is also lowered if ezhome sends a familiar gardener.

So, my recommendation for ezhome is to consider a more consistent assignment. But, of course, benefits of doing so, has to be balanced against the costs. And finally, if feasible, run this experiment on subsequent visits to get more relevant estimates.