1. For the key measure of success, I would choose the increase in number of times drivers cross the bridge. This metric should be particularly useful in comparing the effectiveness of the proposed change because it is likely that it has some relatively stable average prior to implementing the change. Any significant deviations from this average after implementing the reimbursed toll costs should signify that the change is successful.
   1. The experiment would be implemented by continuing to record the number of times drivers cross the toll bridge. If Ultimate was not already recording this, then there would be a trial period of anywhere from 2-4 weeks (assuming no holidays or other reason to suspect irregular driving behavior) so that baseline statistics about normal driver tendencies to cross the toll bridge could be understood. Then another testing period would be put in place once drivers were informed they would be being reimbursed for tolls. The testing period here would most likely be another 2-4 weeks, to make sure that enough data is acquired (and to incorporate the difference in driving behavior between the 2 cities during the weekdays), but also completed in a timely manner. Testing periods could be modified based on considerations and preferences of management. Ideally, the data on drivers crossing the toll bridge without the reimbursed tolls would already have been being collected for a while so we would have more stable statistics about exactly how much drivers cross the toll bridge without tolls being reimbursed.
   2. Assuming there is an increase in traffic through the tollway/bridge, it is important to make sure that the increase is actually due to the policy change, as opposed to just an increase by chance. A frequentist statistical approach should be sufficient to confirm if the increase happened by chance or is statistically significant (which would suggest that the increase was indeed due to the reimbursement of tolls). This could be performed by looking at summary statistics (mean and standard deviation) of how many times drivers cross the bridge. Once we have some preliminary data after the tolls have started to get reimbursed, we can then compare the new data on how many times drivers are crossing the toll bridge. We can quantify the likelihood of seeing the new statistics, if the true underlying proportions were what they were before the tolls were being reimbursed. Then, we could declare whether of not the new toll bridge crossing behavior was indeed statistically significant. The necessary level of statistical significance would be set ahead of the experiment (i.e. is management comfortable with 95% confidence of rejecting the null hypothesis that driver behavior has remained the same or would they want more certainty).
   3. The results would be presented in terms of statistical significance and rejecting a null hypothesis. Following on to the framework laid out in part b, the null hypothesis would be that driver behavior has not changed. After the test, if the results were statistically significant, I would say that we can be 95% confident that we can reject the notion that driver behavior has not changed. I would state that while we can never be certain, that the results suggest that drivers indeed responded to reimbursement of tolls by driving across the toll bridge more frequently. Assuming that the cost-benefit analysis has been done in terms of this being a profitable move for the company, I would go ahead with recommending the reimbursement of tolls. If I were presenting to non-technical stakeholders, I would take out words like null hypothesis and statistically significant to be more accommodative of my audience.