Ultimate Challenge Question 2:

In this thought experiment, there are two adjacent cities separated by a two-way toll bridge that discourages drivers to service both cities rather than just their own. A good metric to measure this would be the percentage of toll crossings by drivers compared to all rides taken. Additionally, this would need to be combined with the percentage of drivers with toll crossings compared to those with none. Alone, each of these metrics captures one aspect of the problem — the first shows how often crossings are made, but doesn't show how many drivers are participating in this activity, while the latter shows the opposite. Together, they can flesh out a more complete picture of this phenomenon and whether it has improved over time.

To experiment with a new system, we should utilize an A/B test to determine its effectiveness. In this system, the tolls would be reimbursed; the hypothesis is that this reimbursement program will encourage drivers to service both cities. To test this, you would randomly select a large chunk of drivers who are representative of both cities and a variety of other factors, who will get put into the new program, while the rest of the drivers are kept in the old system. You then track over time the aforementioned metrics and see if there are differences in them over time. In particular, you need to make sure that there is data for this program during the week, during relevant rush hour times, and on weekends. Given the circadian rhythm of city driving described in the problem, one might expect that the toll reimbursements may have a larger effect on weekends than weekdays, given that's when there's the most overlap in activity (or the least, conversely, if there's so much service that crossings are less necessary).

To see if these metrics change, you could do a z-test for a statistically significant difference in the distributions. Additionally, you'd look to see the magnitude of the change. It would also be instructive to run these tests for individual timeframes (for example, on weekdays, weekends, rush hour times, etc.) and see if any other patterns noticed are statistically significant.

If the program is a success, then you'd see statistically significant, large changes in the observed metrics in the experiment group compared to the control group. That is to say, more drivers would be crossing the toll bridges more often in the experimental group. Additionally, the time insights can be a useful tool in tailoring the program to be more fine-tuned. For example, by reimbursing tolls, there is a cost incurred on the companies, and so if it's very effective in increasing crossings on weekends, but not on weekdays, then only implementing the policy on weekends could save some money for the company while still being an effective policy. The same could be true of rush hour times vs. off-peak. If the policy is very effective, it would be worth noting the extra cost of paying for those tolls compared to the increase in ridership happiness and usage, and seeing whether these costs are offset and whether or not it's considered worthwhile. Extra analysis would be necessary for this, and could be done using things like sentiment analysis, surveys, and models of overall ridership in the experiment group.