Part 2: Experiment and metrics design

The neighboring cities of Gotham and Metropolis have complementary circadian rhythms: on weekdays, Ultimate Gotham is most active at night, and Ultimate Metropolis is most active during the day. On weekends, there is reasonable activity in both cities.

However, a toll bridge, with a two way toll, between the two cities causes driver partners to tend to be exclusive to each city. The Ultimate managers of city operations for the two cities have proposed an experiment to encourage driver partners to be available in both cities, by reimbursing all toll costs.

1. What would you choose as the key measure of success of this experiment in encouraging driver partners to serve both cities, and why would you choose this metric?
2. Describe a practical experiment you would design to compare the effectiveness of the proposed change in relation to the key measure of success. Please provide details on:
    - how you will implement the experiment.
    - what statistical test(s) you will conduct to verify the significance of the observation.
    - how you would interpret the results and provide recommendations to the city operations team along with any caveats.

Response:

1. The average daily/weekly/monthly/yearly percentage of drivers crossing the toll bridge (being present in another city) before and after the toll reimbursements. I might also want to implement experiments separately for weekdays and weekends, because the problem statement tells us that the these two cities have different activity patterns on weekdays vs weekends.
    - First, The metric is very simple to measure and calculate: All we need is to take the number of drivers crossing the bridge divided by total number of drivers in both cities.
    - Second, it's highly reflective of the issue we want to investigate, the percentage is higher, then there are more drivers present in both cities.
    - Third, it's a suitable metric for A/B testing.

2. As we mentioned earlier, we are doing a A/B testing.
    - Suppose we have total drivers z in Gotham and Metropolis. And average drivers who cross the bridge is z1 during weekday, and z2 during weekends before toll imbursement, and z11 and z22 respectively after toll imbursement (for the period

of concern, whether daily/weekly/monthly/yearly average). Then for weekdays for example, first we calculate the percentage before: z1/z , and after: z11/x, and difference: z11/z -z1/z. Then we do a permutation test, and calculate p-value that we get observed or more extreme percentages difference (larger than observed difference). The test is similar for weekends.

- Therefore the null hypothesis Ho: the daily average percentage of driver crossing the bridge is not affected by the toll reimbursement. Ha: affected.
  Then we need to specify a significance level to test on, for example, most common choice is 0.05.
- If we get a calculated p-value less than our chosen significance level (<0.05), then we can reject the null hypothesis that the improvement is due to chance, in other words, the toll imbursement is effective in encouraging drivers to serve both cities. And smaller the p-value, more confident we're about our rejection of null hypothesis, which means we are more confident that the reimbursement policy is making a difference.

However we still need to keep one thing in mind that our A/B testing shows whether the reimbursement is affecting the driver presences in both cities on average, but there might be some underlying differences on these two cities, for example: Metropolis is a just an usual city, while Gotham is famous for casinos which is why it is most active at nights. So even if the reimbursement doesn't work on average, it might work for one of them or there could be different policies that work better. In other words, reimbursement might be one but not the only option to encouraging driver presence in both.