*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:*

*a. how you will implement the experiment*

*b. what statistical test(s) you will conduct to verify the significance of the*

*observation*

*c. how you would interpret the results and provide recommendations to the city*

*operations team along with any caveats.*

1. I would measure the fraction of drivers that service both cities from the total number of drivers in the two groups with and without toll reimbursement. If the fraction of drivers servicing both cities increases when drivers get reimbursed for tolls, then toll reimbursement is a good idea and serves the purpose of encouraging the drivers to be available in both cities.
2. a) The drivers in both cities will be split into 2 groups, 1 group will get reimbursed for tolls, the other wouldn’t:

|  |  |  |
| --- | --- | --- |
| City/Group | A | B |
| Gotham | reimbursed | not reimbursed |
| Metropolis | reimbursed | not reimbursed |

To save the company money we need to determine the smallest number of drivers to be assigned to group A (toll reimbursement). The same number will be assigned to group B.

Power analysis to determine sample size:

* + 1. Depending on the budget pick desired power 0.8-0.9, which is probability of detecting the effect if it is really there
    2. Estimate the effect size – this one depends on the business needs of Ultimate Technologies, for example is a 1% increase in the fraction of drivers servicing both cities worth the investment into reimbursements. It might be if a larger driver network is more resilient and reduces passenger wait times. Let’s say we want to see at least a 10% increase We will use the current mean and standard deviation to estimate this: (1.1\*mean-mean)/st dev
    3. Significance level 0.01-0.05 (probability of rejecting the bull hypothesis when it is true)
    4. Finally, once the three parameters have been fixed calculate the required sample size.

Group assignment randomization will have to be carefully considered to include drivers with different demographics. I recommend to run the experiment with both Gotham and Metropolis drivers and analyze the results separately because there might be a e.g. a cultural difference that would make one city drivers more likely to change their behavior.

Finally, the experiment should last for at least 1 week to exclude confounding daily fluctuations, i.e. different behavior on weekend and weekdays.

1. I would conduct a two-tailed t-test with the following hypotheses:

Null: the drivers remain exclusive to their cities and the incentive does not work.

Alternative: The incentive is working and more drivers serve both cities.

If the proportions of the groups with and without the incentive are statistically significantly different at 0.01-0.05 significance level, I would recommend adopting the reimbursement program.

If only one city has a significant difference, I would recommend adopting the program if the change is sufficient to justify cost.

Finally, if the deference for both states is not significantly different, I would recommend to abandon the reimbursement program and consider alternative incentives, e.g. a year-end bonus for drivers serving both cities.

Once the program is rolled out full scale, I recommend to keep monitoring the proportion metric and try to estimate the economic value the program generates to confirm our assumptions and general applicability of the conclusions. For example, there is always a chance that the groups were not properly randomized or an unknown confounding effect was responsible for the observed difference or the assumed economic value is not realized after all.