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) Important assumption: the toll bridge is the only way to move between the two cities.
   To make sure the drivers are not abusing the system, they will only be reimbursed when crossing the bridge because of a ride. There are two possibilities:
   - (A) Without a rider: Driver is in Gotham and accepted a ride on Metropolis (or vice-versa)
   - (B) With a rider: Rider is in Gotham and requested a ride to Metropolis (or vice-versa)

   I would keep track of Indicator = [((A) + (B))/total rides] over time, aggregated by day.

   Why?

   - Because it is simple to gather this information. All we need is the driver's current city when a ride is accepted, city where the ride starts and city where the ride ends.
   - It is a simple to understand, thus it is easy to communicate to different business departments.
   - Total number of rides varies over time. In addition, this measure may increase the total number of rides. For that reason, we need to divide by the total rides in a day, so we can investigate the proportion.

2) Describe a practical experiment
   a) How you will implement the experiment
   Ultimate keeps track of all rides' origins and destinations, as well as the location of every driver currently active. If that were not true, the Ultimate's algorithm would not know to which driver they should offer an upcoming ride. This means that all the data I need is already in the database.
   If the current driver's location is not already easily available in a table, here is what I would do: I am guessing there is a table that records all events that happens over time, like the moment a ride is requested by a user, the moment a ride is accepted by a driver, etc. It suffices to search for the event "ride accepted", retrieve the "driver_ID" and the "time", then search for the tuple ("driver_ID", "time") at the Driver's location table, retrieving the city.

Now we are able to create two flags:

- FlagA – 1 if driver's current City is different than ride's origin City, 0 otherwise
- FlagB – 1 if ride's origin City is different than ride's destination City, 0 otherwise

Once that is done, I would separate that data into daily bins, aggregating by sum(). Each day would become one record, composed of sum(flagA) and sum(flagB). Add these two quantities and divide by the total number of rides in that day and we have the indicator.

One final note, verifying how that indicator behave along three months (one month before and two months after the implementation of the measure) would be enough to attest the efficacity of the experiment.

b) What statistical test(s) you will conduct to verify the significance of the observation?

A three months plot, as described above, should be enough to graphically spot a change in the behavior. However, we can always double-check with a hypothesis test two compare the two means. The test would go like this:
Distribution 1: data before the event
Distribution 2: data after the event

Hypothesis H0: mean1 = mean2
Hypothesis H1: mean1 < mean2

Then calculate the Z-statistic:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2}}$$

X bar is the mean and
Sigma is the is the standard deviation of the sample divided by the square root of the number of data points

Once calculated, define the level of significance (usually 95%) and compare with the Z table.

c) How you would interpret the results and provide recommendations to the city operations team along with any caveats

If the result is NOT statistically relevant (meaning we do NOT reject H0), I would recommend them to abandon the measure.
However, it we do reject H0, I would probably put someone from the business in the loop, to verify how this increase in the indicator reflects on the financial side. It needs to, at least, compensate all the reimbursements.
Measuring all the effects is hard. Maybe the riders' satisfaction will increase due to less waiting time. Maybe the possibility of operating in both cities without paying for the toll will attract more drivers to work for Ultimate. All of those would take time to evaluate.
In conclusion, if we do notice an increase in the indicator, I would recommend keeping the measure while the business side (or the data side) tries to turn some of the intangibles into tangibles.