

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?

First, I am going to make two assumptions about how the Ultimate cities are going to be conducting the reimbursements. The first assumption is that they will be advertising this chance for reimbursement to potential customers in both cities. If they don't then they only serve to bring new traffic on the toll bridge only by word of mouth, or by the chance that a person finds themselves required to pay the toll and only realizing then that they could be reimbursed. The second assumption, is that the toll booth uses a mail-in rebate system in order to reimburse their customers for the cost. This may be effective for the experiment because people tend to forget to mail in the rebate. Therefore, the experiment would incur less costs while still bringing in the drive partners to use the toll bridge. It would also allow an easy way to track if there is an increase in purchase rate over the duration of the experiment compared to baseline.

However, the other possibility is that the toll reimbursement the person reaches the toll booth they are given a ticket with the cost of zero. This would still allow for the transaction to be recorded while allowing for the easiest customer experience since customers generally hate mail-in rebate systems. Although I will be assuming the mail-in rebate system, I believe that this would lead to the highest rate of conversion for driver partners to begin to transverse between the two cities.

The key measure of success for this experiment would be an increase in the toll booth transactions. I would choose this metric because it is readily available. We should already have baseline data from all of the previous months of the toll to compare what the trends are in the amount of traffic that happens over the toll bridge on a regular basis. Then we can easily compare that transaction data to the transaction data that comes during the experiment. I would be looking for a significant increase in transactions during the experiment vs. control.

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

The experiment that I would design I have alluded to above, but I would implement an experiment over the course of 6 months to a year depending on the budget of the two Ultimate cities. A year would be ideal because we would be able to see how the reimbursement effects the traffic across the toll bridge across an entire cycle of transaction data to compare to the years of previous data to see if the experiment increased traffic across the bridge. However, it is more likely for these governments not to have unlimited budgets to be able to conduct the experiment for a year. It would be more informative to first look at the weekly and monthly trends in the transaction data for the toll booth to tell what the previous trends in traffic across the toll bridge is in both directions. If the previous data shows consistency across all months, it may only truly need to be done for two months, a month for the people in both cities to realize that it is actually happening and a month to receive the data for the effect of the experiment. However, most likely there are general trends and seasonality to traveling across a toll bridge; therefore, the minimum that I would recommend is 3 months. Once again, a month for the people to realize the reimbursement is occurring and 2 months to compare the effect of the experiment. I extend the length for seasonality effects because it could be that in a given month the offer of reimbursement may be more of an effective motivator for a person to choose to travel across the bridge than in another.

If the toll bridge has been in place for several years, we would not need to provide a control group during the experiment since we have several years of transaction data to compare to. If the cities wanted to compare the effectiveness of the mail-in rebate vs. the zero charge, the cities could conduct the mail-in rebate for 3 months then do the zero charge for at least 2. I would aim for 3 months during the months that show the highest amount of traffic over the bridge so that we could see the effect based upon the greatest amount of possible data. Most likely this would be in summer months when people are doing the most traveling and in most northern locations the months when traveling across the toll bridge is actually possible. Although an argument could be made to conduct the experiment over moderate travel months because there you may be able to detect a difference better than the highly traveled months.

Now we also must consider the two cities' goal of reimbursement. If the cities are looking to just reimburse during the experiment and then return back to their original

prices, then this would attract different customers to choose to use the toll bridge then if they were thinking about removing the toll altogether if the experiment is effective. If they are going to remove the toll altogether than this may attract businesses to begin to use the toll bridge where they wouldn't before and could introduce more commerce exchange between the two cities. However, if it is not in the budget to remove the toll altogether, than it may be more effective to design an experiment that would look at a range of reimbursement amounts chosen randomly per customer to determine the optimal rate to charge per customer to create the most traffic while maintaining the revenue.

Assuming they have considered the full versus discount reimbursement and are still considering the full reimbursement. The experiment would last for 3 months gathering transaction data with details about the transaction as well (i.e. time of transaction, personal or commercial vehicle, vehicle's size).

b. what statistical test(s) you will conduct to verify the significance of the observation
Assuming that the cities only are looking to determine if there is a difference in traffic during the experiment compared to before, I would conduct an ANOVA to determine if there was a significant difference between any of the treatment months versus the previous 3 months and the same 3 months in the previous year. For example, if the experiment took place in May, June, and July of 2019, then I would compare the average transaction amounts from those months both monthly and as a whole with transaction averages from February, March, and April of 2019 (monthly and as a whole). A separate ANOVA similar comparison would be made with May, June, and July of 2018. I may also look into the averages for all May, June, and July data to see if there is any further information in the comparison between that data and the treatment averages. I would also separate the data into separate ANOVA comparisons with just demographic data as well (i.e. personal vs commercial, size) to see where the differences in the averages are coming from and who the experiment targeted. I would also conduct a time-series analysis to determine if there are certain times in the day where the free toll was the most effective in generating more traffic.

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

I would look into the results of the ANOVA for overall significance and then into which of the month comparisons are significant using a post-hoc comparison such as Tukey's HSD comparison. These comparisons would allow me to determine whether there is a significant difference in the months due to the treatment. We would be able to see the

effect while controlling for seasonal differences. We would also be able to see where the differences were coming from and whether this was the intended target for the cities. Depending on the size of the difference between the groups, we may be able to tell where the differences were coming from and who the free toll was benefitting the most and whether this group is important influence for the cities. Thus, I would make a recommendation that they may want to decrease their toll on a certain group to maintain the traffic that we saw during the experiment. Or I may be able to provide the suggestion that they then conduct another experiment that investigates the optimal rate to charge in order to find out what they should charge each type of customer.

With the time-series analysis, I would be able to make recommendations on when they should charge a toll and when they shouldn't. Based on the data we should be able to find times when it didn't matter what the toll was because there was no increase or decrease in the amount of traffic. This may be due to some sort of constraint where during this time people are going to travel across the bridge no matter whether there is a toll or not. Maybe there is a seasonal event that happens, or people are traveling to Gotham to enjoy the night life while people travel to Metropolis in the daytime to work. During these times it would be effective to implement the toll because people have to go no matter whether there is a toll or not so they will pay it. However, to increase the amount during the other times they may want to implement a decreased or free toll so that we can see the increase that we saw during the experiment.