Question 2: Experiment and Metrics Design – Ultimate Data Challenge

In this problem, there is a toll that is assumed to be a barrier to having driving partners from Gotham make themselves available in Metropolis and vice versa. In order to accurately assess the impact of the proposed change to the toll structure, it is necessary to measure the impact on the three basic types of customer ride: completely in the home city, completely in the other city, and cross bridge. Additionally, we must account for empty rides across the bridge.

To test this toll change impact, a good approach would be an AB test. I would structure the test as follows:

Choose Sample Population:

Randomly choose a large sample population for the test. The chosen population should have as many of the “average” characteristics as possible (ie: similar distributions on hours worked, fares earned, number of rides per week, etc.) to minimize sample bias. Effort should be made to remove “outliers” from the driving population (ie: drivers that only pick up passengers on their way to and from work, or drivers that only drive once per week in a specific location etc.)

From this population, randomly choose two equal size populations one from Metropolis and one from Gotham. Break each population into equal control and test groups.

Test structure:

For each city, the control group will maintain status quo toll structure. The test group will have their tolls refunded per the planners’ proposal. The time period for the test must be set to cover all possible traffic patterns (ie: rush hour, weekends, etc.) and preferably would allow multiple cycles through those traffics patterns. Given the difference in in circadian rhythms of the cities, it is obviously important to capture continuous 24 hour data. From a practical standpoint, it won’t be possible to run the test for several years, to capture all situations, test time should be maximized.

Data Collection:

For each driver, as described above three fare types are possible, as well as empty car bridge tolls. That condition requires extracting the following information from the app database:

Driver Group	Total Rides	In Zone	Cross Zone	Ext Zone	Total Fare	Fare InZone	Fare Cross Zn	Fare Ext Zone	Toll	Toll/ FareExtZn

Note: if the proposal goes forward, individual driver data can be examined to determine if a more complex toll structure (ie demand driven refunds) is warranted. But in terms of assessing the idea, this data will suffice.

Evaluation:

If the policy change is effective, the following things should true:

1) The test groups should see an increase in both raw numbers and percentage of cross bridge trips compared to the control groups.

2) The percentage of external rides should be higher for the test groups than the control groups.

These two criteria are the basic must haves in the evaluation and should be shown to be statistically significant (show changes in means that are statistically significant to a high confidence level through the use of an appropriate statistic for the distribution. Assuming a normal distribution with reasonable numbers of participants, a Ztest would be appropriate using a null hypothesis that the means do not change.)

It is also necessary to judge impact of the toll change from a cost benefit standpoint. Presumably, the driving force behind this proposal is some level of localized shortages of drivers. If that is true, the overall number of rides should increase within each city (it would be no net benefit if the same number of rides were provided by different drivers. The overall increase in rides must be measured and again shown to be statistically significant.

Finally, it is necessary to judge the financial impact to the change. From the cities' perspective, loss of toll revenue should be compared to the increase in number of total rides in each city. The impact on a driver's overall fares should be calculated as a percentage. The changes should be compared both to the control group and the test group from the other city to judge the efficacy of the policy.

Recommendations:

The evaluation results above would drive the recommendations to the city operations team. While there are a large number of possible outcomes, the most likely are:

1) The toll reduction drives a statistically significant increase in both numbers of rides and rides provided by out of town drivers. The lost revenue for the cities is judged to be acceptable relative to the improvement. Recommendation: Proceed with toll change.

2) The toll can not be shown to statistically improve the total rides and driver distribution. Recommendation: stop test, do not implement toll change.

3) If there is an improvement, but it is too expensive from the cities' perspective, propose modifications to the proposal. Perhaps the tolls could be relaxed at peak demand times. These recommendations would require a more sophisticated analysis that included time series and varying load rates and is not described here.