1. The metric I would use for measuring the success of the experiment is the total number of toll crossings on a weekly basis.
    a. First, I need a way to measure the number of times a driver actually crosses the toll bridge - using the geolocation to measure bridge crossings is perfect.
    b. Secondly, I need a way to account for both the weekday opposite rhythms of both cities as well as the smooth activity over the weekends. Taking the total number of toll crossings over the course of a week will capture this information at a macro level and allow us to identify if the toll-reimbursement experiment is actually incentivizing drivers to be available in both cities.
    c. Lastly, this metric is composed of daily metrics, and can be used to easily break down and analyze weekday vs weekend trends, the daily average, and will work with a variety of other weekly metrics.
2. I would initiate the experiment by sending a push notification to all drivers in both cities that we will be reimbursing them for toll fees in order to ensure that they have access to both markets since the time for peak demand in both cities alternates in order to give them maximum flexibility in their schedules.
    a. This experiment would run for a minimum of four weeks in order to give several time periods to be examined, and would need a four week 'blank' period prior to collect a base level of metrics.
    b. Prior to the start of the experiment I would label drivers as primarily Gotham or Metropolis
    c. I would set up a weekly dashboard to track: total refunds per unique driver, total refunds per Gotham driver, and total refunds per Metropolis driver, and total refunds. This would be used in weekly meetings to monitor the progress of the experiment.
    d. I would then conduct several statistical tests. The first would be a one-sample T-test on each week of metrics compared to a baseline normal for that metric collected in the previous 4 week 'blank' period. I would then conduct 6 dependent samples T-tests in order to assess if the difference between weekly results were statistically significant
        i. If there was a lack of adoption in the first week we could also change the messaging to drivers to see if we can elicit a change in adoption. The dependent sample tests would need to be replaced by independent sample tests.
    e. There are two things that I'm looking for in success for this experiment.
        The first is if the toll reimbursements INCREASED toll crossings when compared to the 'blank' period.
        The second is if any change in toll crossings is statistically significant from the baseline established in the 'blank' period.

        I would consider the experiment a success IF at least two weeks of the four week experiment there was an increase in toll crossings AND they were statistically significant.

i. If there was a measurable increase in the toll crossings but it was not statistically significant I would recommend that we attempt the experiment again as long as the cost would be within budget.
ii. If there was no measurable increase in the toll crossings then I would recommend that we try alternate methods to increase driver availability in both cities.
iii. If there was a measurable increase in only 2 weeks I would continue the experiment, but if in 3 to 4 weeks there was an increase I would recommend we launch toll reimbursements permanently.