Take Home Challenge Part 2

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?

Looking at how the **proportion of drivers making pickups/drop-offs in both cities** changes for reimbursed vs. non-reimbursed drivers could serve as the key measure of success for this experiment. This metric was selected because the purpose of the experiment was to encourage drivers to serve both cities. If a higher proportion of drivers operate in both cities when toll costs are reimbursed, then the new strategy could be considered effective (assuming the results are statistically significant).

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.

We can begin by defining our null and alternative hypotheses. The **null hypothesis** is: the proportion of reimbursed drivers operating in both cities is less than, or equal to, the proportion of non-reimbursed drivers operating in both cities. The **alternative hypothesis** is: the proportion of reimbursed drivers operating in both cites is greater than the proportion of non-reimbursed drivers operating in both cities

As stated in question 1, our **test statistic** is the difference between the proportion of reimbursed drivers operating in both cities and non-reimbursed drivers operating in both cities. Since we do not have access to the city data, we cannot make a determination about the data's distribution. Therefore, a **non-parametric test** should be selected. In this case, a **permutation test** is appropriate to test for significance.

In order to carry out the experiment, the drivers should be randomly assigned to a **control group** (non-reimbursed) and an **experimental group** (reimbursed). The random assignment is an effort to minimize bias in the groups. The **duration** of the experiment will depend on available time and resources. It is advisable to gather as much data as possible.

During the experimental period, we will collect data on the location of pickups and drop-offs for each driver. We can track whether drivers operated in both cities using an "operated in both cities" flag. Depending on the specific goal of the experiment, we could define a minimum threshold before considering a driver to be operating in both cities. For example, we may define "operating in both cities" as a pickup/drop-off in each city daily, once per week, etc. At the conclusion of the experiment we will calculate the test statistic: the difference between the proportion of reimbursed drivers operating in both cities and the proportion of non-reimbursed drivers operating in both cities. This will be the **observed difference**.

After calculating the test statistic, a permutation test will be carried out to test for significance. In order to carry out the permutation test, we will keep the driver groups (treatment vs. control) constant but shuffle the results (operated in both cities vs. did not operate in both cities). The test statistic will then be calculated on this synthetic data. This process will be repeated 10,000 times. We can then calculate a **p-value** by looking at the proportion of the synthetic datasets where the test statistic is greater than or equal to the observed difference. When selecting a **significance level** the business context would have to be taken into account. For example, the costs/risks of implementing the new policy would need to be considered. However, to provide an example, if the commonly used significance level of 0.05 is selected then the results would be considered significant if the p-value is less than, or equal to, 0.05. In this case the recommendation to the city would be to go ahead with the new driver reimbursement policy.

There are some caveats associated with this experimental approach. Despite significance testing, it is always possible that results labeled significant actually occurred due to chance. For example, if the p-value is 0.02 then there is still a 2% probability of witnessing the observed difference when the null hypothesis is true. Additionally, there may be unknown biases in the experimental design that influence the test results.