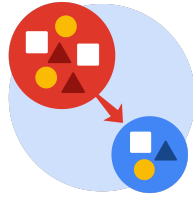


Course Four

From Data to Insight: The Power of Statistics



Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. As a reminder, this document is a resource that you can reference in the future, and a guide to help you consider responses and reflections posed at various points throughout projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- ☐ Complete the questions in the Course 4 PACE strategy document
- ☐ Answer the questions in the Jupyter notebook project file
- ☐ Compute descriptive statistics
- ☐ Conduct a hypothesis test
- ☐ Create an executive summary for external stakeholders

Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

- How would you explain an A/B test to stakeholders who may not be familiar with analytics?
- If you had access to company performance data, what statistical tests might be useful to help understand performance?
- What considerations would you think about when presenting results to make sure they have an impact or have achieved the desired results?
- What are some effective ways to communicate statistical concepts/methods to a non-technical audience?
- In your own words, explain the factors that go into an experimental design for designs such as A/B tests.



Reference Guide

This project has four tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



Data Project Questions & Considerations



PACE: Plan Stage

- What is the main purpose of this project?

The main purpose of this project is to perform hypothesis testing to determine if there's a statistically significant in the average amount of rides between iPhone and Android users.

- What is your research question for this project?

Is there a statistically significant in the average amount of rides between iPhone and Android users?

- What is the importance of random sampling?

Random sampling is important because we want to get a sample that represents the characteristics of the whole population. In other words, every element of the population has an equal chance to be included in the sample data.



- Give an example of sampling bias that might occur if you didn't use random sampling.

About nonresponse bias, which is certain people are unrepresentative.



PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful in order to have a general and specific understanding of the whole dataset.

- How did computing descriptive statistics help you analyze your data?

With different methods such as measures of central tendency, dispersion, and position. In each of them, we can know how the variables of our data are being centered (mean, median), how spread it is (standard deviation), and the percentiles respectively (quartile).

- In hypothesis testing, what is the difference between the null hypothesis and the alternative hypothesis?

The null hypothesis is the one we first want to assume is true until we get the evidence to reject or not. The alternative hypothesis is the one we favor to achieve when we have sufficient evidence.

- How did you formulate your null hypothesis and alternative hypothesis?

First, we set the null hypothesis. Usually, we say that there's no difference between our sample mean/proportion. Then, we set the alternative hypothesis which is there's a significant difference between our sample mean/proportion.



- What conclusion can be drawn from the hypothesis test?

The p-value is the one we focus on. In the end, we might choose to reject or fail to reject the null hypothesis if the p-value is less than or greater than the significance level.



PACE: Execute Stage

- What key business or organizational insight(s) emerged from your A/B test?

The descriptive statistics show the average amount of rides from iPhone users is higher than from Android users. However, the result of the A/B test shows that there's no difference in the average amount of rides between iPhone and Android users.

In other words, these users have the same average amount of rides no matter the device they use.

- What recommendations do you propose based on your results?

From the A/B test above that showed an insignificant difference, we might choose to perform another hypothesis testing based on the different variables, such as comparing both retained and churned users. Also, we might address the insufficiency of data values to accurately perform the test.