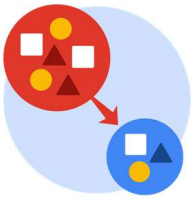
Course Four

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

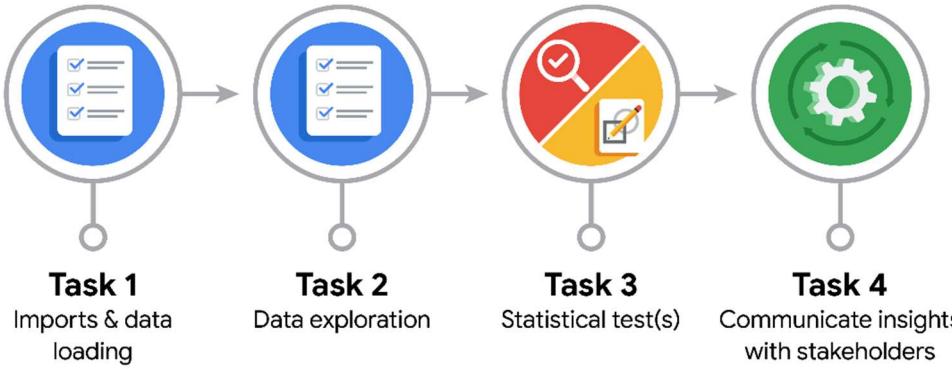
Relevant Interview Questions

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 primary objective of this project is to investigate the reasons why users stop using the Waze app, a phenomenon known as churn. Once we understand the underlying causes, we can implement strategies to identify users at high risk of churning and take proactive steps to prevent it.

- What is your research question for this project?

The overarching research question for the project is how we can reduce user churn. For this specific phase, we aim to determine if there is a statistically significant difference in the average number of drives between iPhone and Android users.

- What is the importance of random sampling?

Random sampling helps to minimize the risk of bias and ensures that the sample data accurately reflects the population being studied. Moreover, it enables statistical inference, allowing researchers to draw more reliable conclusions from a representative sample.

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

In this specific project, we observed that the variables related to driving were significantly higher than what would be expected from the average driver. This suggests a potential bias toward a particular group of users, such as rideshare or long-distance commuting drivers, who may not represent the overall population. It's unknown whether random sampling was used to collect this data, but if it wasn't, it may be an explanation for the unusually high values.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics provide a valuable overview of the variables within a dataset. By examining distributions, outliers, missing data, and key metrics like mean, median, mode, standard deviation, variance, and IQR, analysts can gain a general understanding of the data's characteristics. This information helps identify areas of interest and focus subsequent analysis.

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

By calculating descriptive statistics for device type and number of drives, I was able to examine the mean, standard deviation, minimum, and maximum values. These results suggested that there might not be a significant difference in the average number of drives between the two device types. To confirm this hypothesis, I conducted an independent two-sample t-test.

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

The null hypothesis states that there is no effect, difference, or relationship between variables. In contrast, the alternative hypothesis posits that a statistically significant difference, effect, or relationship exists between two variables.

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

Descriptive statistics revealed no apparent significant difference in the average number of drives between device types. To verify this observation, a hypothesis test was conducted. Based on the available information, the null hypothesis was that there was no statistically significant difference

between average drives and device type. Conversely, the alternative hypothesis posited a statistically significant difference.

- What conclusion can be drawn from the hypothesis test?

The conclusion is that there is no significant difference in average drives between device types. The resulting p-value of 0.14 was greater than the significance level (alpha), leading to the failure to reject the null hypothesis.

PACE: Execute Stage

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

The independent two-sample t-test revealed no significant difference in the average number of drives between users of iPhone and Android devices.

- What recommendations do you propose based on your results?

Given the findings, resources can be reallocated to explore other variables within the data that may be more relevant to addressing user churn.