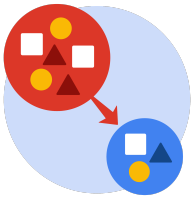
Course Four

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

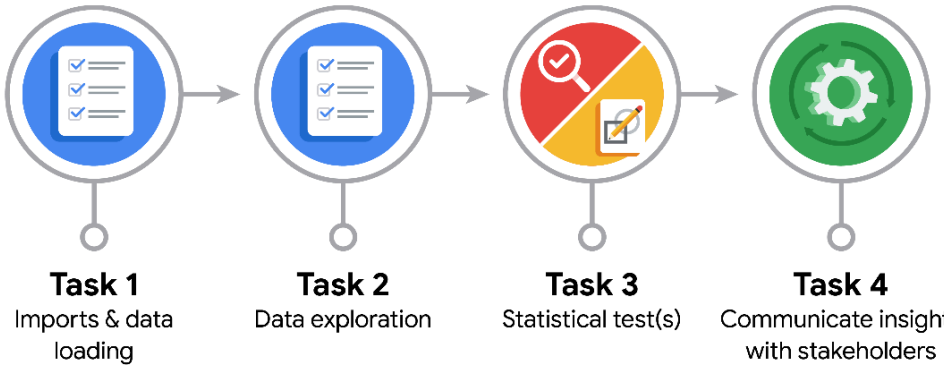
Relevant Interview Questions

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 purpose of the project is to create a machine learning model that will distinguish between claims and opinions.

- What is your research question for this project?

For this part we are checking if there is a significant statistical difference between the view counts of verified and unverified users.

- What is the importance of random sampling?

It ensures that the population is well represented and avoids bias when choosing the sample.

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

A teacher may only choose high scoring students in her class when sampling to make her average grade higher so she seems like a better teacher.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

They give information about distribution and concentration of various variables allowing for statistical inference to happen.

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

It made it easier to see which values had outliers and which columns had missing data.

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

The null hypothesis indicates no difference or effect in changing something while the alternative hypothesis shows there is a statistically significant difference between the options.

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

I considered the problem statement as trying to prove one of two options and then chose them as null and alternative hypothesis.

- What conclusion can be drawn from the hypothesis test?

The conclusion is that there is a statistically significant difference between verified and unverified views.

PACE: Execute Stage

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

Verified users get much less views than unverified ones.

- What recommendations do you propose based on your results?

Change the recommendation algorithm to push verified users more, or convince unverified users to verify their accounts when they reach certain engagement metrics.