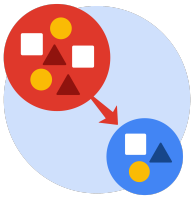
Course Four

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

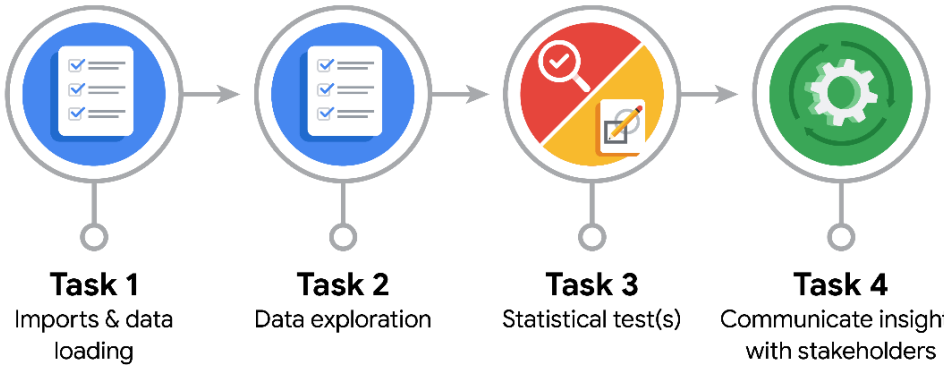
Relevant Interview Questions

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?

To find out if there is a significant difference in view count of videos with an author that is verified or not.

- What is your research question for this project?

Is there a relationship between view count and the verification status of the video's author?

- What is the importance of random sampling?

Random sampling helps with ensuring that the sample is representative. If you do not have a representative sample, then your results will likely be biased.

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

Self-selection bias might occur if sampling was done voluntarily. This bias happens when a certain type of person is more likely to take a survey or be part of the experiment.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics allow the analyst to have a better understanding of the data. Using descriptive statistics, you are able to interpret the data in a simple way before conducting further analysis.

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

I was able to identify and remove rows with missing data and get an initial understanding of the mean view count for both verified and unverified videos.

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

The null hypothesis is the default, it's what we assume happens by chance. The alternative hypothesis is what we are testing for, it's what we assume does not happen by chance.

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

I based it off of the research questions and set the null to be = and the alternative to be not equals.

- What conclusion can be drawn from the hypothesis test?

The p-value is a very small number which is obviously less than the 5% significance level. Because of this, we reject the null hypothesis in favor of the alternative hypothesis. This means that there is a statistically significant difference in view count between the verified statuses.

PACE: Execute Stage

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

I found that there is a statistically significant difference in the average view count of verified and unverified users. We can intuit from this that there may be other behavioral differences between the two different verified statuses.

- What recommendations do you propose based on your results?

We could do further research into why this difference exists. We can also build a regression model on verification status in order to further our analysis on the behavior of the two groups.