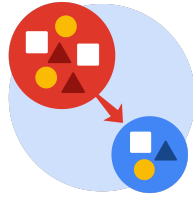


Course Four

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 investigate whether there are any significant differences between the data for verified vs. unverified TikTok accounts.

- What is your research question for this project?

Is there a statistically significant difference between video view counts for verified accounts compared to unverified accounts?

- What is the importance of random sampling?

Random sampling gives a smaller and easier to work with dataset that is representative of the population dataset.

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

The ratio of unverified to verified accounts might be different than that of the population, this would affect the result and conclusions drawn from the hypothesis test.



PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful to get an initial idea of mean and percentile values of the variables in the dataset.

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

Descriptive statistics provided the mean values of verified and unverified accounts as an initial result.

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

Null hypothesis typically suggests that there is no significant difference between the values of the variables being tested. Alternative hypothesis suggests the opposite, i.e., there is a significant difference.

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

Null hypothesis : There is no statistically significant difference between the video view counts of verified and unverified accounts.

Alternative hypothesis : There is a statistically significant difference between the video view counts of verified and unverified accounts.

- What conclusion can be drawn from the hypothesis test?

The p-value is less than the significance level of 5% so the null hypothesis can be rejected. This means that there is a statistically significant difference between the video view counts of verified and unverified accounts.



PACE: Execute Stage

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

It can be concluded that the video view counts of unverified accounts were significantly higher than that of verified accounts. This is worth investigating from a business perspective as it could indicate more click-bait videos or spam bot associations for unverified accounts.

- What recommendations do you propose based on your results?

The next step could be to analyse the variable 'video_view_counts' based on the verified status of the users. This could be a key variable for the claims classification project and could be useful to build a regression model.