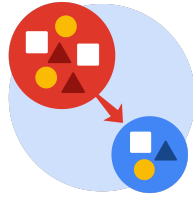


Course Four

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?

Find the relationship between the number of views for TikTok videos posted by verified accounts versus unverified accounts.
Test the hypothesis of the number of views for TikTok videos posted by verified accounts versus unverified accounts.

- What is your research question for this project?

Is there a relationship between the number of views for TikTok videos posted by verified accounts versus unverified accounts.



- What is the importance of random sampling?

Random selection means that every member in the population has an equal chance of being chosen for the sample.
Reduce bias in data.

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

Some members of the population are more likely to be selected than other members, it will lead to bias and skewed result.



PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful because they let you quickly explore and understand large amounts of data.

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

Computing descriptive statistics helps you quickly compare the number of views for TikTok videos posted by verified accounts versus unverified accounts.

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

Null hypothesis (H_0): A statement that is assumed to be true unless there is convincing evidence to the contrary.
Alternative hypothesis (H_a): A statement that contradicts the null hypothesis and is accepted as true only if there is convincing evidence for it.



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

Ho : There is no difference in the number of views for TikTok videos posted by verified accounts and unverified accounts.
Ha : There is difference in the number of views for TikTok videos posted by verified accounts and unverified accounts.

- What conclusion can be drawn from the hypothesis test?

If $p\text{-value} < 5\% (0.05)$: Reject the null hypothesis.
If $p\text{-value} > 5\% (0.05)$: Do not reject the null hypothesis.



PAC: Execute Stage

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

It would be interesting to investigate to root cause of this behavioral difference. For example, do unverified accounts tend to post more clickbait video ? Or are unverified accounts associate with spam bots that help inflate view counts ?

- What recommendations do you propose based on your results?

The next step will be to build a regression model on `verified_status` to make predictions on claim status.