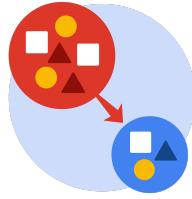# Course Four
From Data to Insight: The Power of Statistics

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.

Task 1
Imports & data loading

Task 2
Data exploration

Task 3
Statistical test(s)

Task 4
Communicate insights with stakeholders

Data Project Questions & Considerations

PACE: Plan Stage

- What is the main purpose of this project?

The main purpose of this project is to use statistical methods to analyze and interpret TikTok video performance data, specifically to determine whether verification status has a significant impact on video view counts. This analysis will help the team understand engagement trends and provide insights for improving content visibility on the platform.

- What is your research question for this project?

Is there a statistically significant difference in the average video view counts between verified and unverified TikTok accounts?

- What is the importance of random sampling?

Random sampling ensures that every data point in the population has an equal chance of being selected. This reduces bias and helps make the sample representative of the entire population, allowing results and conclusions to be generalized more accurately.

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

  If only videos from verified TikTok accounts were chosen for analysis, the sample would overrepresent popular or high-engagement accounts. This would create a bias that makes it seem like all TikTok videos receive high view counts, even though that may not be true for unverified users.
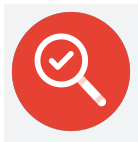
PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics summarize and organize large datasets, making it easier to understand key patterns and characteristics. They provide measures such as mean, median, and standard deviation that help describe data distribution and identify trends before deeper analysis.

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

  Computing descriptive statistics helped reveal basic patterns in the TikTok dataset, such as the average video views, likes, and shares. It allowed me to compare engagement metrics across verified and unverified accounts and identify differences that might be worth testing statistically.

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

  The **null hypothesis (H_0)** assumes there is no significant difference or effect — in this case, that verified and unverified TikTok accounts have the same average video view counts.

The **alternative hypothesis (H₁)** proposes that there is a significant difference — meaning verified accounts have different (likely higher) average view counts than unverified ones.

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

The null and alternative hypotheses were based on the question of whether verification status affects video view counts.

Null hypothesis (H₀): There is no significant difference in average video view counts between verified and unverified TikTok accounts.

Alternative hypothesis (H₁): There is a significant difference in average video view counts between verified and unverified TikTok accounts.

These hypotheses were formulated to test whether verification status influences engagement on the platform.

- What conclusion can be drawn from the hypothesis test?

The t-test results showed a **t-statistic of -25.49** and a **p-value of 2.6×10^{-120}**, which is far below the 0.05 significance level.

Therefore, we **reject the null hypothesis** and conclude that there is a **statistically significant difference** in video view counts between verified and unverified accounts. Verified accounts, on average, receive substantially more views than unverified ones.

PACE: Execute Stage

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

The analysis revealed that **verified TikTok accounts consistently achieve significantly higher video view counts** than unverified accounts. This suggests that verification status is associated with greater visibility and engagement on the platform, possibly due to algorithmic boosts, increased credibility, or audience trust.

- What recommendations do you propose based on your results?

Encourage more creators to pursue account verification, as it appears to enhance reach and engagement potential.

Provide guidance or incentives to help unverified users meet verification requirements to improve their visibility.

Further analyze contributing factors—such as content quality, posting frequency, and follower count—to better understand what drives engagement beyond verification status.

Develop predictive models to forecast engagement outcomes and identify creators who could benefit most from verification.