**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 this project is to demonstrate knowledge of how to prepare, create, and analyze hypothesis tests.

* What is your research question for this project?

There are a few possible ways to frame the research question.

1) Do videos from verified accounts and videos from unverified accounts have different average view counts?

2) Is there a relationship between the account being verified and the associated videos' view counts?

* What is the importance of random sampling?

Random sampling is important as it ensures that the sample data accurately represents a larger population.

It minimizes bias and increases the likelihood that the results of the analysis can be generalized to the entire user base.

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

If the sample only included popular influencers or accounts with millions of followers, the data would be skewed toward higher video view counts, which would not represent the average TikTok user.

This would lead to incorrect conclusions about the impact of verification status on view counts.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics summarize and organize data in a meaningful way, allowing analysts to better understand patterns and trends.

They help identify central tendencies, variability, and distributions within each dataset.

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

Computing the mean video view counts for both verified and unverified accounts helped identify an initial pattern.

This pattern was that verified accounts tend to have higher average views.

This observation was integral in guiding the direction of the hypothesis testing.

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

The null hypothesis represents the assumption that there is no effect or no difference between groups.

The alternative hypothesis proposes that there is a significant effect or difference that is present.

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

The Null Hypothesis formulated states that there is no difference in the average video view counts between verified and unverified TikTok accounts.

The Alternative Hypothesis formulated states there is a significant difference in the average video view counts between verified and unverified TikTok accounts.

* What conclusion can be drawn from the hypothesis test?

The two-sample hypothesis test yielded a p-value smaller than the significance level (e.g., 0.05), which means that the null hypothesis should be rejected.

Thus, there is a statistically significant difference in the average video view counts between verified and unverified accounts.

**PACE: Execute Stage**

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

Verified TikTok accounts have significantly higher average video view counts compared to unverified accounts.

With this information, it could be indicated that verification increases visibility or that more popular accounts are more likely to be verified.

* What recommendations do you propose based on your results?

Based on the results, the following recommendations can be made:

-Encourage high-performing users to apply for verification, as it may help boost their content visibility.

-Investigate the underlying mechanisms driving higher views for verified accounts to better understand platform engagement.

-Consider integrating verification status as a variable in TikTok’s content recommendation algorithms.