**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 main purpose of this project is to develop a predictive model capable of identifying claims from opinions in user reports.

* What is your research question for this project?

Is there a statistically significant difference in the number of views for TikTok videos posted by verified accounts versus unverified accounts?

* What is the importance of random sampling?

Random sampling is important because it helps researchers to reduce biases.

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

Undercoverage bias.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics are useful particularly because they provide a summary of our data.

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

Computing descriptive statistics has helped me analyze my data by letting me have a thorough understanding of the data. It has helped me in picturing my data that way I can adequately analyze the data however large the data may be.

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

The null hypothesis is what is believed to be true and accepted, universally-true if you will. The alternative hypothesis however is what we aim to prove with the results or outcome of our experiment. The alternative hypothesis provides us with evidence enough for us to safely reject the null hypothesis.

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

The null hypothesis is typically formed with these keywords: ‘no difference’, ‘greater than or equal to’, ‘less than or equal to’, ‘is equal to’. The alternative hypothesis typically uses keywords like ‘not equal to’, ‘less than’ or ‘greater than’

* What conclusion can be drawn from the hypothesis test?

There are two possible conclusions drawn from the hypothesis test. We can reject the null hypothesis or fail to reject the null hypothesis.

**PACE: Execute Stage**

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

The difference in video view count among verified and unverified users is statistically significant

and most likely not due to chance.

* What recommendations do you propose based on your results?

Conduct further analysis to understand why there is a difference in video view counts. Are verified users more active, do they post more engaging content, or is there another factor at play?

Use the observed difference in view counts as a feature in the predictive model after conducting regression analysis using this variable. This could help improve the model's ability to distinguish between verified and unverified users.