**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?

Conduct a descriptive statistic test and hypothesis test to determine there is a significant difference in a variable of interest.

* What is your research question for this project?

1) Do videos from verified accounts and videos 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?

To ensure data collected is independent and not biased.

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



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

They provide a quick overview on the distribution of the dataset.

* How did computing descriptive statistics help you analyze your data?
* In hypothesis testing, what is the difference between the null hypothesis and the alternative hypothesis?

Null hypothesis – no change in status quo

Alternative hypothesis – Change in status quo due to significant difference.

* How did you formulate your null hypothesis and alternative hypothesis?
* What conclusion can be drawn from the hypothesis test?

The analysis shows that there is a statistically significant difference in the average view counts between videos from verified accounts and videos from unverified accounts. This suggests there might be differences between these two groups of accounts.

**PACE: Execute Stage**

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

It would be interesting to investigate the root cause of this difference. For example, do unverified accounts tend to post more clickbait-y videos? Or are unverified accounts associated 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. A regression model is the natural next step because the end goal is to make predictions on claim status. A regression model for verified\_status can help analyze user behavior in this group of verified users. Technical note to prepare regression model: because the data is skewed, and there is a significant difference in account types, it will be key to build a logistic regression model.