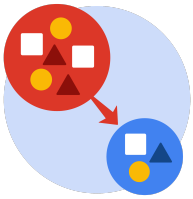
Course Four

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~~

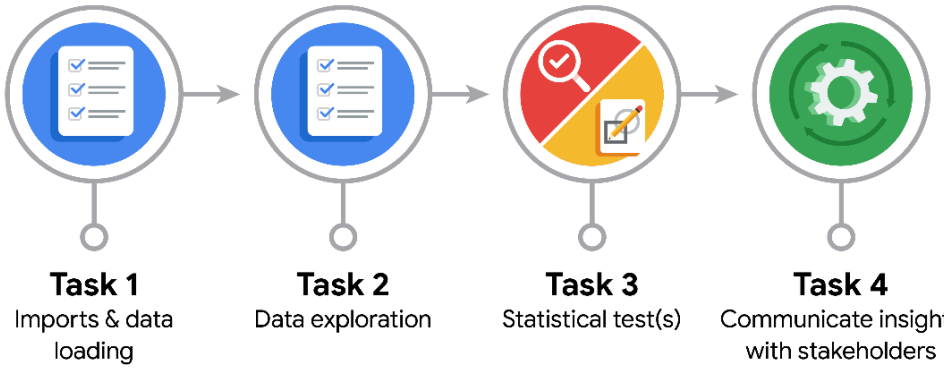
Relevant Interview Questions

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 conduct exploratory data analysis and hypothesis testing on TikTok user data to investigate the relationship between verified account status and video view counts. The goal is to uncover insights that can support the development of a machine learning model for claims classification.

- What is your research question for this project?

Does the verified status of a TikTok account have a statistically significant impact on the number of views that a video receives?

- What is the importance of random sampling?

Random sampling is essential because it ensures that every individual or item in the population has an equal chance of being selected. This helps to avoid bias and increases the likelihood that the sample

data accurately represents the population, which improves the reliability and generalizability of the analysis.

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

If we only selected data from popular TikTok influencers or videos with high engagement, the results would likely overestimate average view counts and skew the analysis toward verified accounts, which could misrepresent the broader TikTok user base.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics provide a summary of the main features of a dataset, such as measures of central tendency (mean, median) and variability (standard deviation). They help analysts understand the basic structure and characteristics of the data before conducting more advanced statistical analyses.

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

By computing the mean video view counts for both verified and unverified accounts, we identified an initial observation that unverified accounts had a higher average view count. This insight guided the next step—hypothesis testing—to determine whether this observed difference was statistically significant.

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

The null hypothesis (H_0) assumes that there is no effect or difference between groups—in this case, that verified status does not affect video view count.

The alternative hypothesis (H_1) suggests that there is a significant difference—in this case, that verified status does affect video view count.

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

Null Hypothesis (H_0): There is no difference in average video view counts between verified and unverified TikTok accounts.

Alternative Hypothesis (H_1): There is a difference in average video view counts between verified and unverified TikTok accounts.

- What conclusion can be drawn from the hypothesis test?

The results of the hypothesis test indicated a statistically significant difference in video view counts between verified and unverified accounts. This supports rejecting the null hypothesis and accepting the alternative hypothesis, suggesting that verified status does impact view counts.

PACE: Execute Stage

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

The test revealed that unverified accounts, on average, receive significantly more video views than verified accounts. This insight raises important questions about user behavior and content engagement, and it may highlight potential spam or bot-related activity among unverified users.

- What recommendations do you propose based on your results?

Investigate further why unverified accounts have higher view counts—are these views organic, or are they artificially inflated by bots or spam behavior?

Consider including verified_status as a feature in the upcoming claims classification model to improve prediction accuracy.

Conduct follow-up analysis using regression modeling to assess the influence of other variables, such as engagement metrics (likes, comments) on view counts.

Implement content monitoring systems to detect possible spam trends among unverified accounts.