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

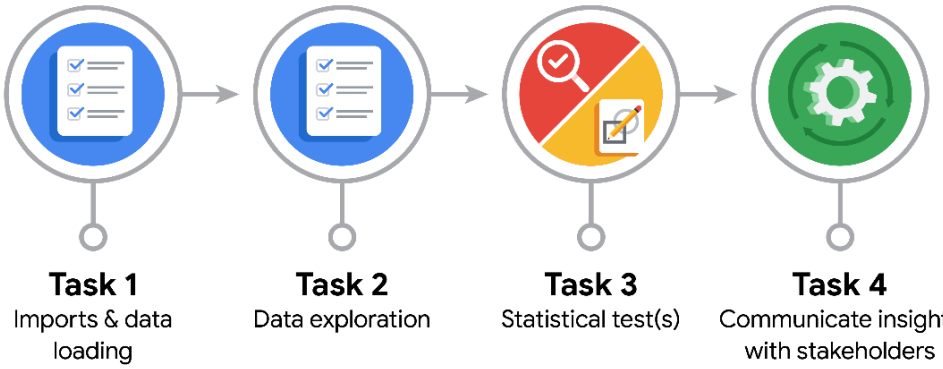
Relevant Interview Questions

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?

To find if there is a statistically significant difference in video view count between verified and unverified accounts on TikTok.

- What is your research question for this project?

Is there a statistically significant difference in video view count between accounts that are verified and unverified?

- What is the importance of random sampling?

Random sampling helps ensure we get a representative sample so we can have more actionable insights and data that isn't affected with bias.

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

An example of sampling bias that might occur is Undercoverage Bias which comes from Convenience Sampling. Convenience Sampling is a Non-Probability Sampling Method where you choose members of a population that are easy to contact or reach. It involves collecting a Sample from somewhere convenient to you such as your workplace, a local school or a public park. Undercoverage Bias is when some members of a population are inadequately represented in a Sample.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful because they help summarize and organize large amounts of data in a clear and understandable way. They provide key insights such as the average, spread, and distribution of values, making it easier to identify trends, patterns, and outliers.

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

Computing descriptive statistics helped identify the central tendency and variability of the data, such as the average view counts and standard deviation. This gave a quick overview of how views differed across videos and informed further analysis, including comparisons between verified and unverified accounts.

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

The null hypothesis (H_0) represents the default assumption that there is no effect or difference—in this case, no difference in view counts between verified and unverified accounts. The alternative hypothesis (H_a) suggests that there is a significant effect or difference. Hypothesis testing helps determine whether the observed data provides enough evidence to reject the null hypothesis in favor of the alternative.

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

I formulated the hypotheses based on the research question: whether verification status affects view counts. The **null hypothesis (H_0)** assumes there is no difference in mean view counts between verified and unverified accounts, and any observed difference in the sample data is due to chance or sampling variability. The **alternative hypothesis (H_a)** assumes there is a statistically significant difference in view counts between the two groups, and any observed difference in the sample data is due to an actual difference in the corresponding population means. This framework allows for a formal statistical test of the observed data.

- What conclusion can be drawn from the hypothesis test?

The hypothesis test resulted in a very small p-value (2.61×10^{-120}), which is much lower than the significance level of 0.05. Therefore, we **reject the null hypothesis** and conclude that there is a **statistically significant difference** in view counts between verified and unverified TikTok accounts.

PACE: Execute Stage

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

The A/B test revealed a surprising insight**: unverified TikTok accounts receive significantly higher average view counts** than verified accounts. This suggests that verification status alone may not drive user engagement, and that other factors—such as content type, posting frequency, or algorithm exposure—may play a larger role in video visibility.

- What recommendations do you propose based on your results?

Based on the results, I recommend that TikTok further investigate **the factors contributing to high view counts among unverified users**. Additionally, platform strategies should not rely solely on verification status to boost content visibility. Enhancing the algorithm to promote quality and engaging content—regardless of verification—may lead to better overall user engagement and satisfaction.