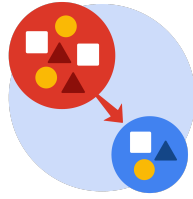


Course Four

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 primary purpose of this project is to analyze and interpret data related to the claims classification project at TikTok. The data analytics team is tasked with conducting descriptive statistics, hypothesis testing, and creating an executive summary to determine if there is a statistically significant difference in the number of views for TikTok videos posted by verified accounts versus unverified accounts. This analysis aims to provide valuable insights into user engagement based on account verification status, contributing to TikTok's mission to inspire creativity and enhance the efficiency of content moderation processes. The project also serves as a practical application of statistical methods and communication skills for the team members.

- What is your research question for this project?

The research question for this project is: "Is there a statistically significant difference in the number of views for TikTok videos posted by verified accounts compared to unverified accounts?" The team aims to investigate whether the verification status of an account has a measurable impact on the

viewership of videos on the platform. This question guides the statistical analysis and hypothesis testing that the data analytics team will conduct as part of the claims classification project at TikTok.

- What is the importance of random sampling?

Random sampling is vital as it ensures a fair representation of the entire population, reducing biases and increasing the likelihood of generalizable results. Providing each member an equal chance of selection enables statistical inference, making findings applicable beyond the sampled group. This method enhances external validity, simplifies analysis through probability theory, and contributes to the overall reliability of research outcomes. Random sampling is a foundational principle for obtaining accurate and meaningful insights in statistical studies.

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

An example of sampling bias without random sampling could be political polling conducted only through online surveys. If the survey is administered solely online, it may exclude individuals without internet access, leading to a bias in the sample towards a more tech-savvy population. This bias could result in inaccurate predictions or insights, as it needs to pay attention to the perspectives of those not represented in the online sample, such as older individuals or those with limited digital connectivity.



PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful because they provide a concise and meaningful summary of the main features of a dataset. They help organize and simplify large amounts of data, making it more understandable and interpretable. Descriptive statistics offer insights into central tendencies (mean, median, mode), variability (range, standard deviation), and data distribution, allowing researchers and decision-makers to grasp the key characteristics and trends without delving into complex details. This aids in effective communication of data patterns and supports informed decision-making processes.



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

Computing descriptive statistics played a crucial role in data analysis by offering a succinct summary of key features. It facilitated a quick dataset exploration, allowing a better understanding of central tendencies and variability. This process helped identify patterns and outliers, laying the groundwork for informed decision-making and guiding subsequent hypothesis testing. Overall, descriptive statistics provided valuable insights into the distribution of video views for verified and unverified accounts in the TikTok project.

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

In hypothesis testing, the null hypothesis (H_0) represents a statement of no effect or no difference, suggesting that any observed results are due to random chance. It serves as a baseline assumption that no significant relationship or effect exists in the population being studied. On the other hand, the alternative hypothesis (H_1 or H_a) contradicts the null hypothesis, asserting that there is a genuine and statistically significant effect, relationship, or difference in the population. Hypothesis testing aims to assess whether there is enough evidence in the sample data to reject the null hypothesis in favor of the alternative hypothesis. The choice between the null and alternative hypotheses defines the direction of the statistical test and guides the evaluation of the research question.

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

In formulating hypotheses for the TikTok project, the null hypothesis posits no significant difference in video views between verified and unverified accounts, serving as the default assumption. The alternative hypothesis contends a statistically significant difference in views, presenting the researcher's proposed effect. This formulation guides the subsequent statistical analysis, aiming to either accept the null hypothesis or reject it in favor of the alternative based on observed data patterns. The hypotheses address the inquiry and provide a foundation for hypothesis testing.

- What conclusion can be drawn from the hypothesis test?

The conclusion drawn from the hypothesis test depends on the results obtained. If the test leads to rejecting the null hypothesis, it suggests a statistically significant difference in the number of views between verified and unverified accounts on TikTok. On the other hand, if the test does not provide enough evidence to reject the null hypothesis, it implies no substantial statistical difference in video views between the two account categories. The conclusion informs whether there is support for the proposed effect (alternative hypothesis) or if the data does not provide convincing evidence to deviate from the null hypothesis.



PACE: Execute Stage

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

In the context of an A/B test for TikTok's project on video views for verified and unverified accounts, potential key insights include whether there is a statistically significant difference in viewer engagement based on account verification status. Suppose the A/B test rejects the null hypothesis. In that case, it may suggest that verification status influences the number of views, providing valuable insights for content moderation, user engagement strategies, and overall business decision-making at TikTok.

- What recommendations do you propose based on your results?

Suppose the A/B test indicates a significant difference in video views between verified and unverified accounts. In that case, recommendations may include prioritizing content from verified accounts for visibility or implementing targeted strategies to boost engagement for unverified accounts. Conversely, a more uniform approach to content promotion could be considered if no significant difference is observed. These recommendations aim to optimize user experience, content moderation, and overall platform performance based on the insights gained from the A/B test results.