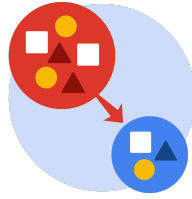


Course Three

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 3 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 demonstrate knowledge of how to prepare, create, and analyze hypothesis tests.

- What is your research question for this project?

To determine whether there is a statistically significant difference in the average number of views for TikTok videos posted by verified accounts versus unverified accounts.

- What is the importance of random sampling?

Random sampling is representative and avoids bias.

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

Inadequate representation of some members of a population, it can introduce subjectivity into the sample selection process.



PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are used to explore and summarize data.

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

Descriptive statistics help to describe and explain the features of a dataset by giving short summaries about the data.

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

The null hypothesis is a statement that is assumed to be true unless there is convincing evidence to the contrary. However, the alternative hypothesis is a statement that contradicts the null hypothesis and is accepted as true only if there is convincing evidence for it.

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

H_0 : There is no difference in the average video view count for TikTok videos posted by verified accounts versus unverified accounts.

H_A : There is a difference in the average video view count for TikTok videos posted by verified accounts versus unverified accounts.

- What conclusion can be drawn from the hypothesis test?

With a p-value ($2.6e-120$) being less than 0.05 (as the significance level is 5%), we reject the null hypothesis in favor of the alternative hypothesis.



Therefore, we can conclude at the 5% significance level that there is a statistically significant difference in the average video view count for TikTok videos posted by verified accounts versus unverified accounts.



PACE: Execute Stage

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

We can conclude at the 5% significance level that non-verified users published more TikTok videos in which a claim/opinion has been made.

- What recommendations do you propose based on your results?

It would be interesting to investigate the main cause of this difference. Also, it seems that encouraging customers to verify their accounts can decrease the amount of content that needs to be reviewed by moderators.

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.