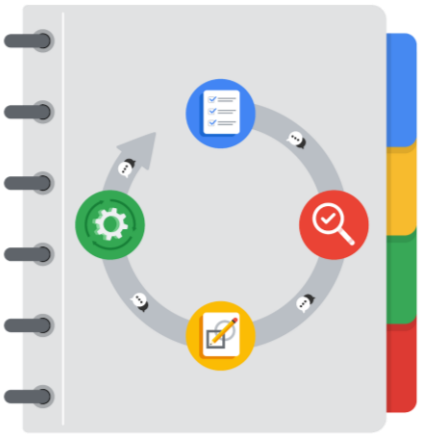
TikTok PACE Strategy Document IV

Statistical Review

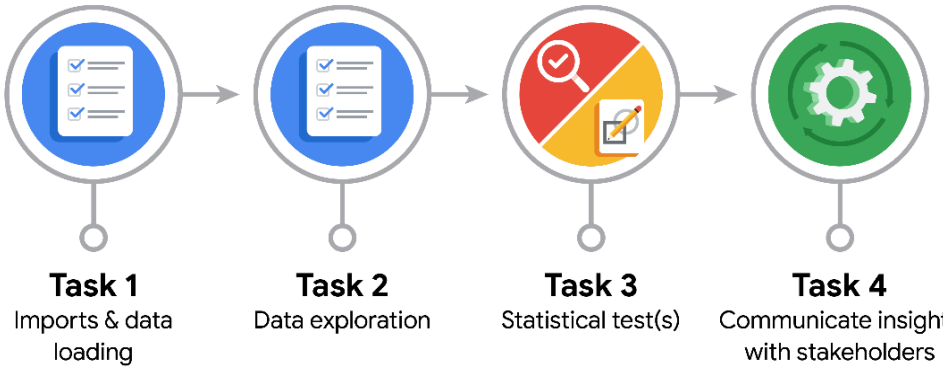
Introduction



PACE stands for Plan, Analyze, Construct and Execute. It is a framework that illustrates the foundation and structure for data analysis projects and each letter represents an actionable stage in a project. The stage "Plan" involves the definition of the project scope, the research of business data and the workflow development. The stage "Analyze" involves data scrubbing, data conversion and database formatting. The stage "Construct" involves building models and machine learning algorithms and selecting a modeling approach. The stage "Execute" involves the presentation of results to decision-makers, stakeholders and others in order to receive feedback. This framework is built upon an iterative cycle where each stage may reveal new insights, requiring the return to earlier stages. A PACE strategy document is used to record decisions and reflections at different stages of the data analytical process. It typically includes the definitions of roles and actions to ensure clarity and accountability.

Purpose

TikTok users have the ability to report videos and comments that contain user claims. These reports identify content that needs to be reviewed by moderators. This process generates a large number of user reports that are difficult to address quickly. TikTok is working on the development of a predictive model that can determine whether a video contains a claim or offers an opinion. With a successful prediction model, TikTok can reduce the backlog of user reports and prioritize them more efficiently. For this stage of the project, we identify four main tasks that are presented in the following visual.



Considerations

PACE: Plan Stage

- What is the main purpose of this part of the project?

  The goal is to analyze the relationship between account verification status and video view counts through hypothesis testing.

- What specific question are we trying to answer with our data?

  The question we are trying to answer is whether there is a difference in average view counts between videos posted by verified and unverified TikTok accounts.

- What is the significance of random sampling?

  Random sampling ensures that the sample data represents the entire population, reducing bias and improving the validity of the statistical conclusions.

- Provide an example of sampling bias that might occur if we do not use random sampling.

  If only videos from verified accounts with high follower counts are selected, the results might overestimate the average view counts for verified accounts, leading to inaccurate conclusions.

PACE: Analyze & Construct Stages

- How do descriptive statistics help us understand data?

  They provide a quick summary of key characteristics like averages, ranges, and distributions, making it easier to identify patterns or differences between groups.

- How did calculating descriptive statistics help us make sense of our data?

It allowed for a straightforward comparison of average view counts between verified and unverified accounts, highlighting significant differences between them.

- What is the difference between the null hypothesis and the alternative hypothesis in hypothesis testing?

The null hypothesis is a statement that is assumed to be true unless there is convincing evidence to the contrary. The alternative hypothesis is a statement that contradicts the null hypothesis and is accepted as true only if there is convincing evidence of it. In our case, the null hypothesis assumes no difference between groups, while the alternative hypothesis assumes a difference exists.

- How did we establish our null hypothesis and alternative hypothesis?

We assume that there is no difference in view counts between verified and unverified accounts (null) and then we test for evidence of a significant difference (alternative).

- What does the hypothesis test reveal about our hypotheses and what conclusions can be drawn?

The test showed a statistically significant difference, leading to the rejection of the null hypothesis and supporting the alternative hypothesis that verified and unverified accounts have different view counts.

PACE: Execute Stage

- What key findings did we uncover from our A/B test?

There is a significant difference in the average view counts between verified and unverified accounts, suggesting behavioral differences.

- What recommendations would we provide to the organization based on our findings?

We should investigate further into why unverified accounts have higher average view counts, explore user behavior patterns among these groups and proceed with building a regression model to predict claim status.