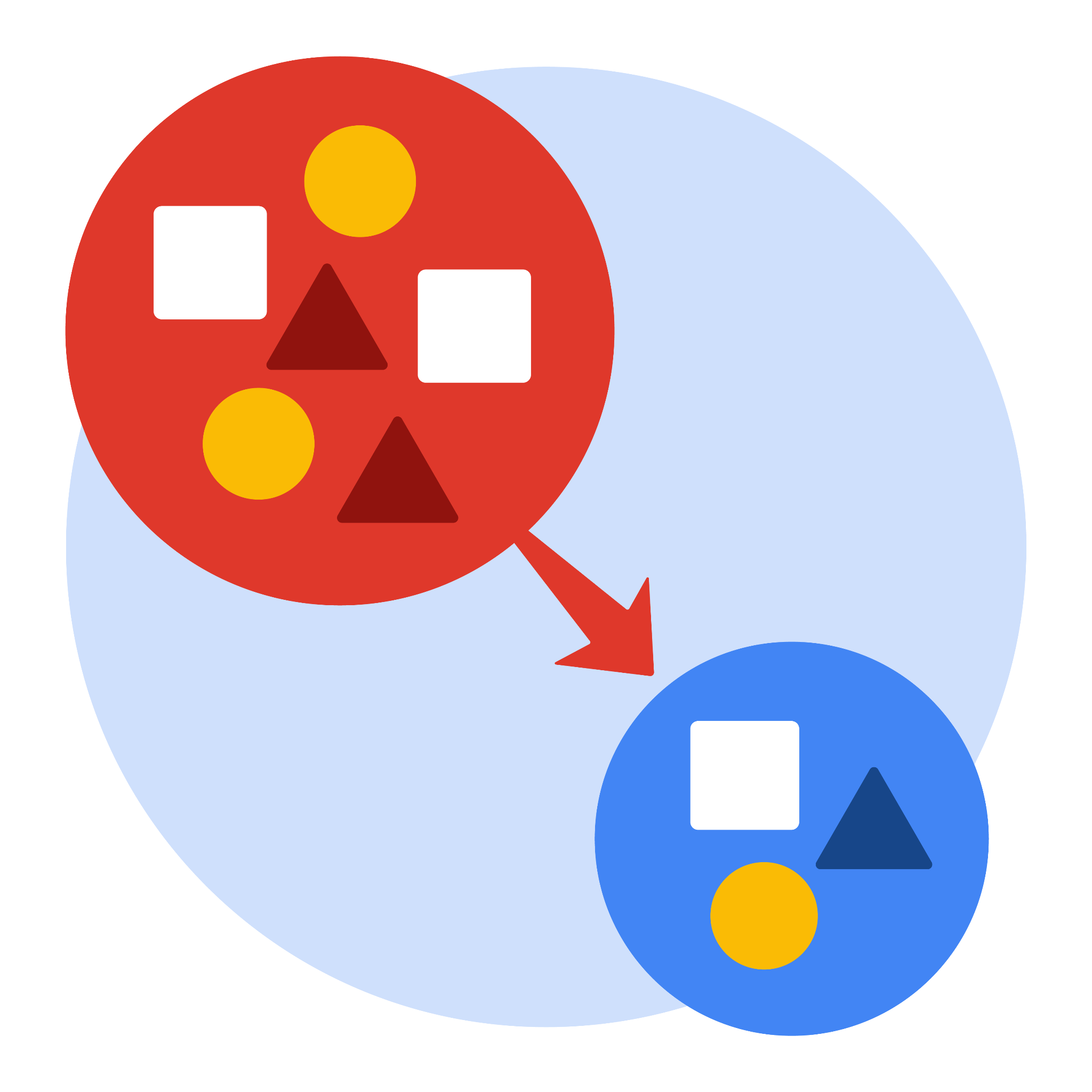
**Course Four**

# 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

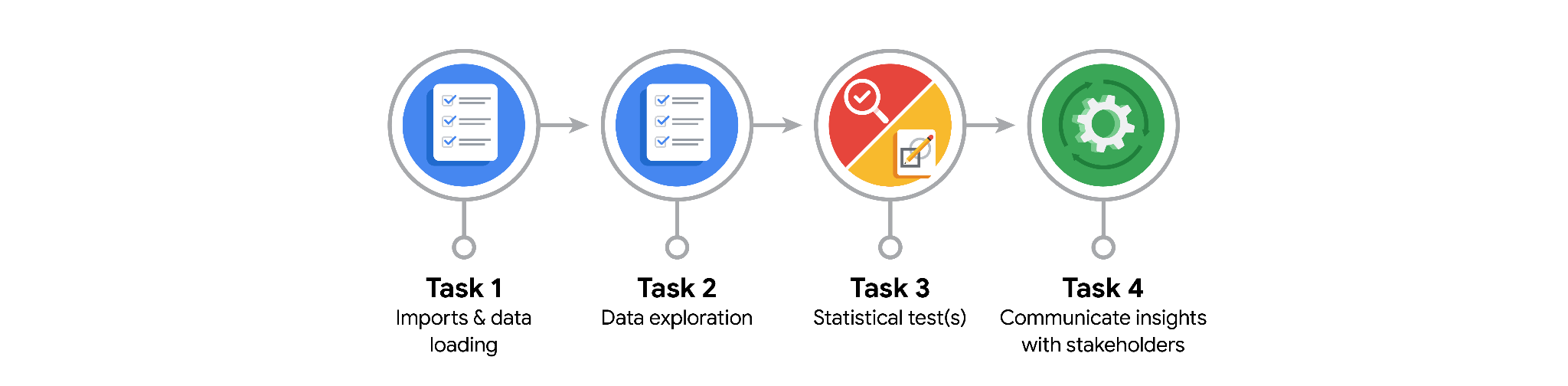
# Relevant Interview Questions

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 develop a machine learning model that can be used to predict the claim status of videos.

* What is your research question for this project?

Is there a statistically significant difference between the view count of videos from verified and unverified creators.

* What is the importance of random sampling?

Random sampling ensures that bias is reduced in the statistical test.

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

Sampling bias might occur if I used a convenience sample of family and friends since their use of the platform is highly unusual.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics are useful because they allow you to get a general understanding of the data and potential trends that might exist.

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

By looking at descriptive statistics I was able to see that there were many more views by unverified users.

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

The null hypothesis states that there is no effect or difference in the two populations but the alternative hypothesis states that there is a difference.

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

Null Hypothesis: There is no difference in view count between verified and unverified users.  
Alternative Hypothesis: Unverified users receive more views on average compared to verified users.

* What conclusion can be drawn from the hypothesis test?

Unverified users receive more views on average.

**PACE: Execute Stage**

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

The verified status of the author and view count can be used in the machine learning model.

* What recommendations do you propose based on your results?

The machine learning model should flag unverified users as most likely to post claims.