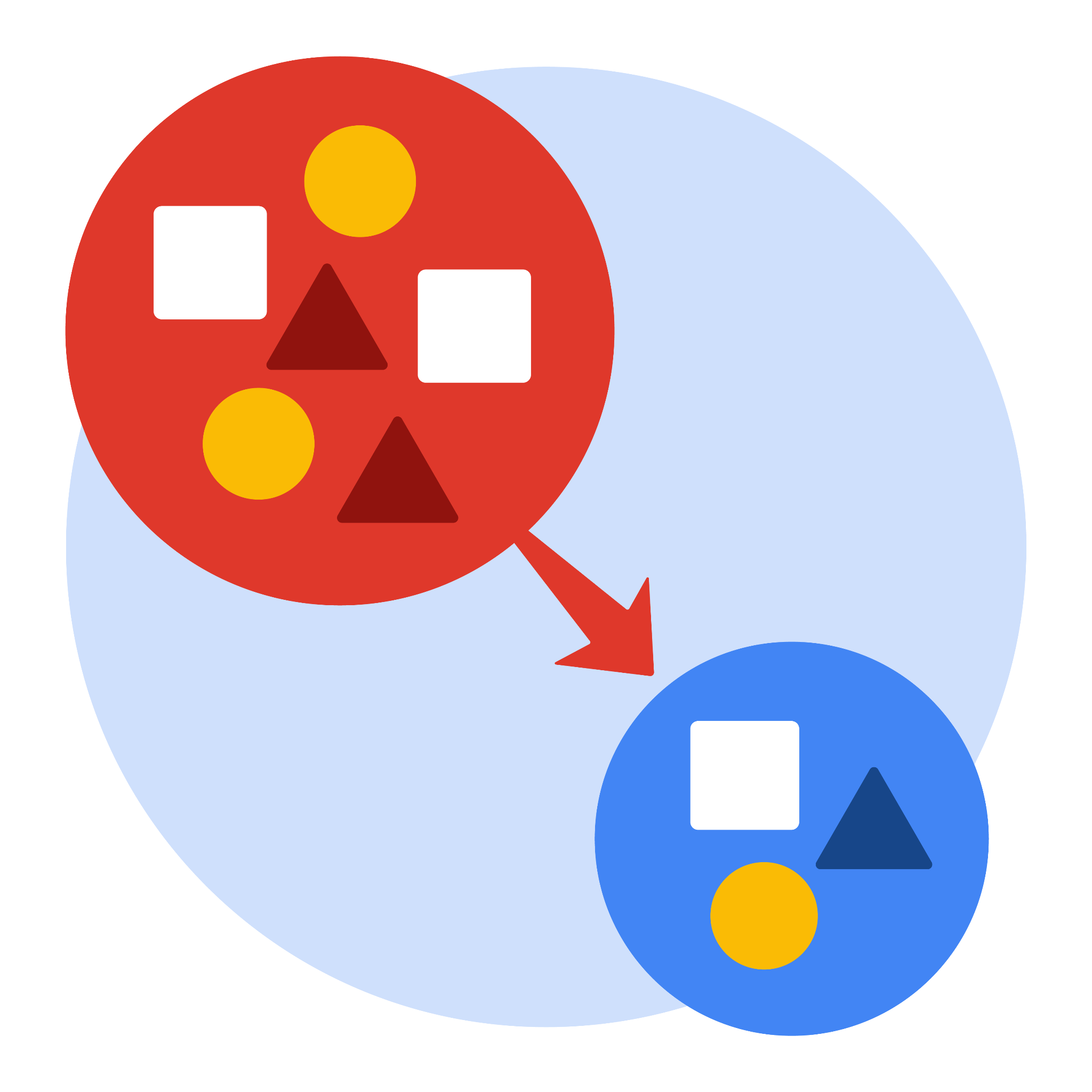
**Course Four**

# 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

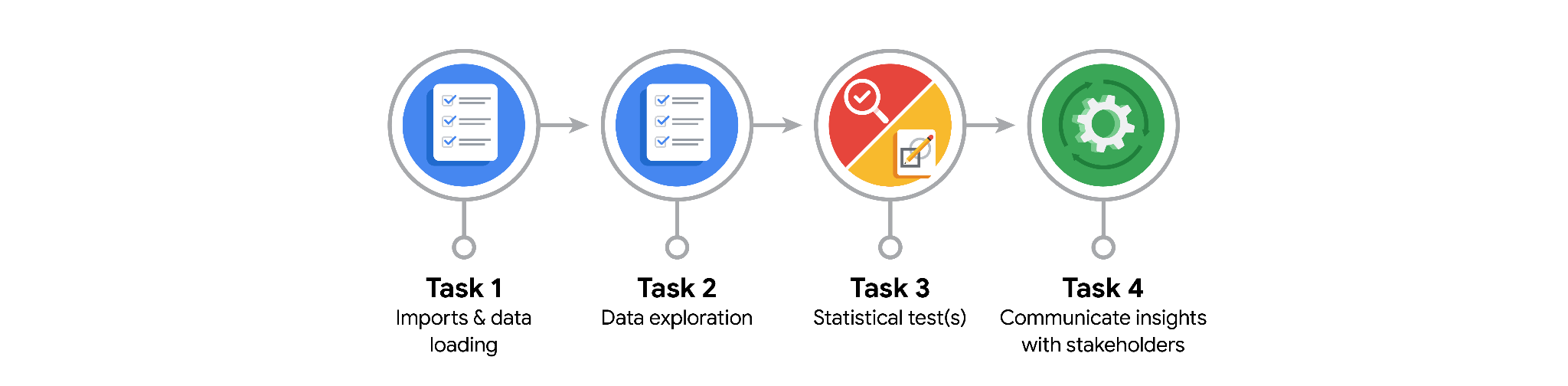
# Relevant Interview Questions

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 demonstrate knowledge of how to prepare, create, and analyze A/B tests. Your A/B test results should aim to find ways to generate more revenue for taxi cab drivers. (Automatidata)

To indicate which variable would be most insightful to test in terms of verified and unverified accounts? (TikTok)

To determine whether there is a statistically difference between mean number of rides and device type. (Wave)

* What is your research question for this project?

Apply descriptive statistics and hypothesis testing in Python. The goal for this A/B test is to sample data and analyze whether there is a relationship between payment type and fare amount. For example: discover if customers who use credit cards pay higher fare amounts than customers who use cash. (Automatidata)

Apply descriptive statistics and hypothesis testing in Python. The goal for this A/B test is to sample data and analyze whether there is a statistical difference in the data between verified and unverified accounts. To determine user behavior trends based on verification type. (TikTok)

Apply descriptive statistics and hypothesis testing in Python. The goal for this A/B test is to sample data and analyze whether there is a statistical difference in the data between Android and iPhone devices impacts the number of drives. (Wave)

* What is the importance of random sampling?

Helps removes or does not reinforce bias when collecting our data for analysis and our conclusions.

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

Conducting research with a group of participants that do not accurately represent the population



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics allow for the ease of data visualization. It allows for data to be presented in a meaningful and understandable way.

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

Help clarify what are the variables for any violation of assumptions that underlie statistical techniques, that addresses specific research questions.

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

A null hypothesis states that is assumed to be true unless there is convincing evidence to the contrary. Whereas alternative hypothesis contradicts the null hypothesis.

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

For the null hypothesis,

We compare the only one true mean (=)

For the alternative hypothesis,

We contradict to the null hypothesis (< or > or !=)

Can be applied in one or two sample testing

* What conclusion can be drawn from the hypothesis test?

According to the Automatidata lab, we reject the null hypothesis since our p-value is less than 5%.

Similar for the Tiktok lab, we reject the null hypothesis since our p-value is less than 5%.

Wave, we fail to reject the null hypothesis, since our p-value is greater than 5%.

**PACE: Execute Stage**

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

The data was not collected this way; so, an assumption had to be made to randomly group data entries to perform an A/B test. This dataset does not account for other likely explanations. For example, riders might not carry lots of cash, so it's easier to pay for longer/farther trips with a credit card. In other words, it's far more likely that fare amount determines payment type, rather than vice versa. (Automatidata)

* What recommendations do you propose based on your results?

We conclude that the hypothesis test that there is a difference in the average fare amount between customers who use credit cards and customers who use credit to generate more revenue. This project requires an assumption that passengers were forced to pay one way or the other, and that once informed of this requirement, they always complied with it. (Automatidata)

We conclude that the hypothesis test that it is statistically significant that there is a mean difference between verified users and non-verified users. For example, do unverified accounts tend to post more clickbait-y videos? Or are unverified accounts associated with spam bots that help inflate view counts? (Tik-tok)

Conclude, that the t-test is not statistically significant for the mean number of drives can impact the different between android and iphone devices. Makes it harder to decide on which device customers should use on taking (i.e. more number of drives) using Waze, further analysis is required to understand their customer behaviour.