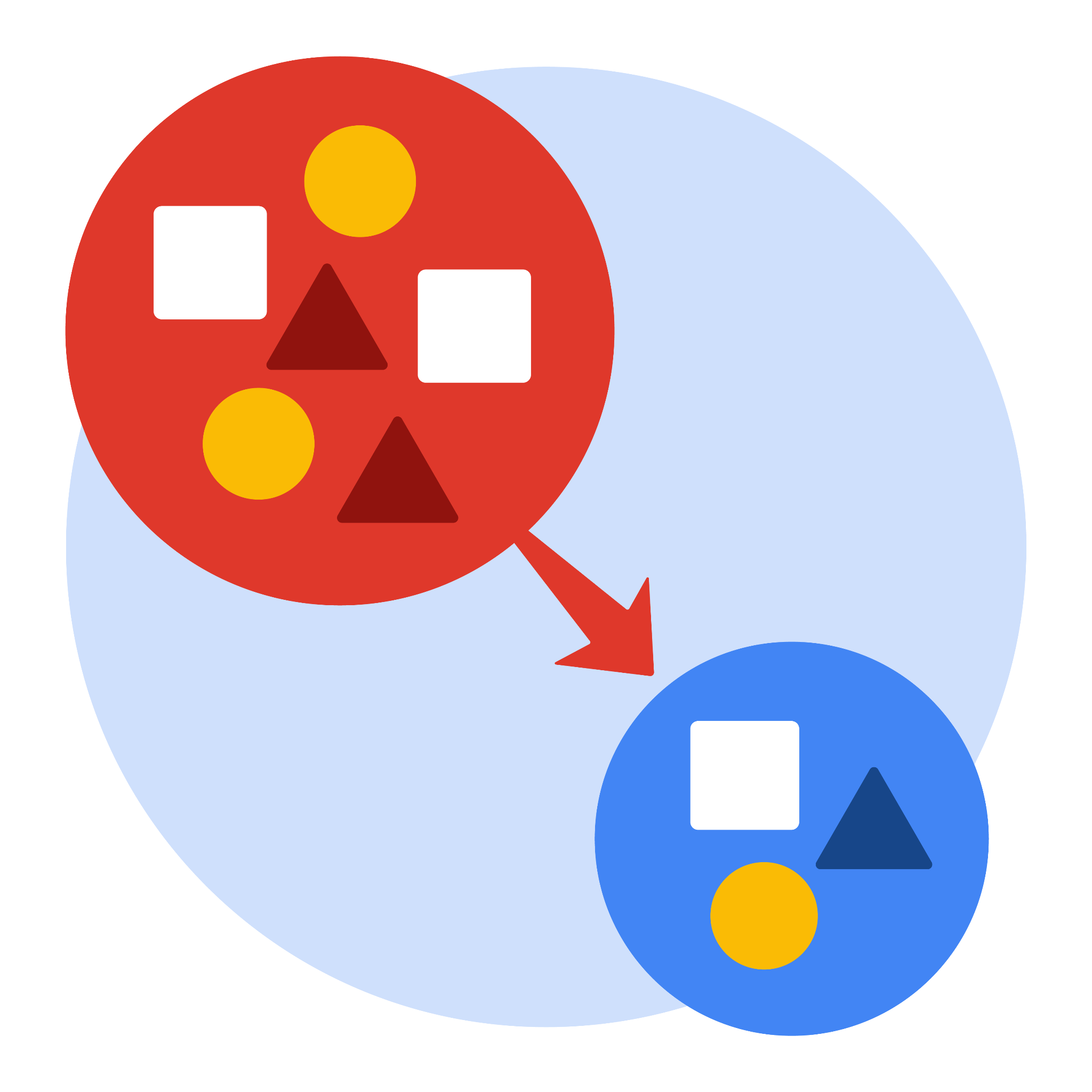
**Course Four**

# 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

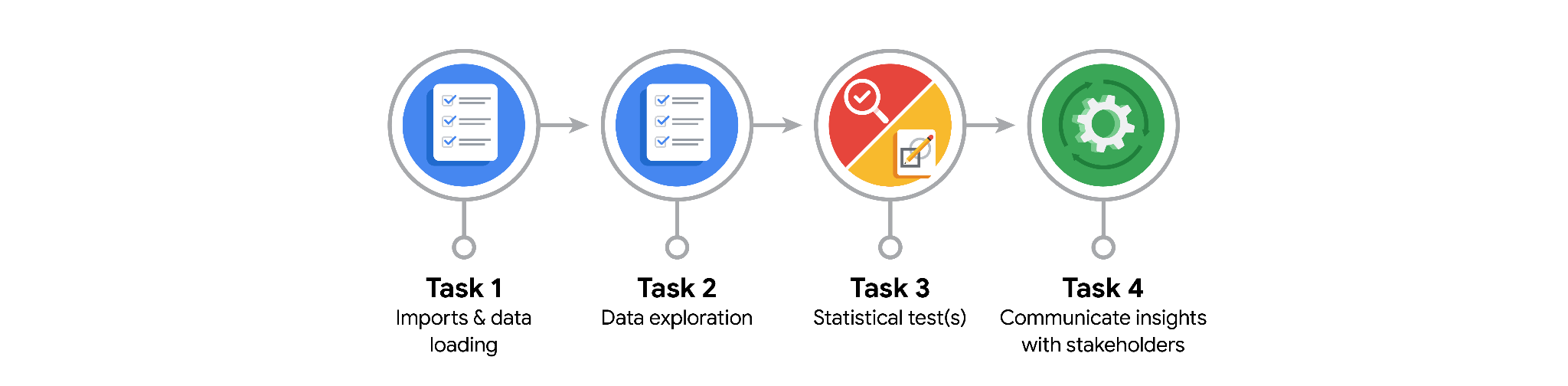
# Relevant Interview Questions

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 evaluate whether there is a statistically significant difference in fare amounts based on payment method (cash vs. credit card), and to apply statistical methods such as descriptive statistics and hypothesis testing in a real-world business context.

* What is your research question for this project?

Is there a significant difference in the average fare amount between taxi riders who pay by credit card and those who pay by cash?

* What is the importance of random sampling?

Random sampling ensures that every individual in the population has an equal chance of being selected. This reduces bias, improves the representativeness of the sample, and increases the generalizability of the results.

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

If the data only includes rides from a high-income neighborhood where riders are more likely to pay with credit cards, the average fare amount for credit card users may appear inflated, misrepresenting the true population.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics provide a quick and informative summary of the dataset, such as averages, ranges, and standard deviations. They help identify central tendencies, variability, and potential outliers that guide further analysis.

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

By computing the mean, median, and standard deviation for fare amounts across payment types, I was able to identify early patterns and variations in the data, which informed the development of my hypothesis and testing approach.

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

The **null hypothesis (H₀)** assumes there is no difference or effect—in this case, that average fare amounts are the same for cash and credit card payments. The **alternative hypothesis (H₁)** proposes that there is a meaningful difference between the two groups.

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

**Null Hypothesis (H₀):** There is no difference in the average fare between credit card and cash payments.

**Alternative Hypothesis (H₁):** The average fare amount for credit card payments is significantly different from that of cash payments.

* What conclusion can be drawn from the hypothesis test?

The hypothesis test revealed a statistically significant difference in average fare amount between credit card and cash payments. The results suggest that credit card users tend to pay higher fares, and we can reject the null hypothesis in favor of the alternative.

**PACE: Execute Stage**

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

The analysis indicates that encouraging credit card payments may result in higher fare revenue. This could influence business strategies such as offering discounts, rewards, or preferred services for credit card users.

* What recommendations do you propose based on your results?

The NYC TLC or taxi operators could **promote credit card payments** to potentially increase fare revenue.

Additional research should control for confounding variables (e.g., trip distance, location, time of day) to better isolate the effect of payment type.

**Clean and filter extreme outliers** in fare data before drawing firm conclusions, as these may skew averages and affect statistical accuracy.