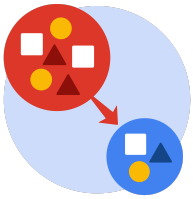
# Course Four

## 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

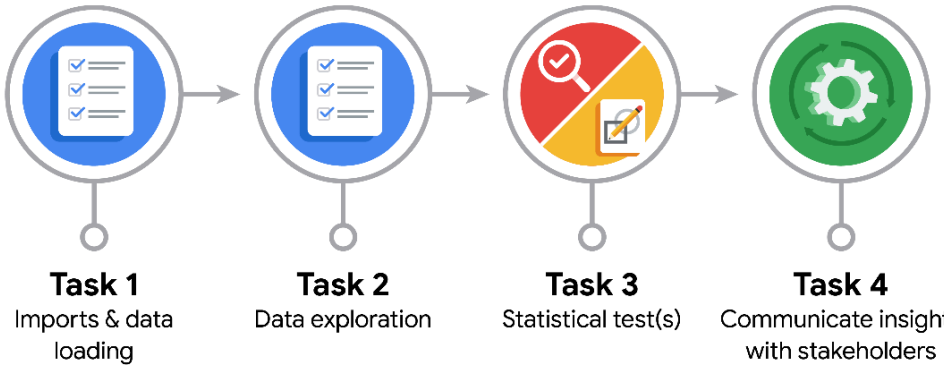
### Relevant Interview Questions

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 regression model for the New York City Taxi and Limousine Commission (TLC) to estimate taxi fares before a ride, based on data gathered from TLC's extensive trip records. Additionally, the project aims to analyze the relationship between fare amount and payment type and to determine if customers who use credit cards pay higher fare amounts than those who use cash.

- What is your research question for this project?

The primary research question is: "Is there a statistically significant relationship between the fare amount and the payment type used by customers, specifically between credit card payments and cash payments?"

- What is the importance of random sampling?

Random sampling is crucial because it ensures that every member of the population has an equal chance of being selected. This minimizes bias and makes the sample more representative of the population, leading to more reliable and generalizable results.

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

If random sampling is not used, sampling bias can occur. For example, if the sample only includes taxi rides from a specific part of the city or during certain times of the day, it may not accurately represent the overall population of taxi rides. This could skew the results and lead to incorrect conclusions about fare amounts and payment types.

## PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful because they summarize and provide an overview of the main features of a dataset. They help to understand the distribution, central tendency, and variability of the data, making it easier to identify patterns and insights.

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

Computing descriptive statistics helped by providing a clear summary of the data, such as the mean, median, standard deviation, and range of fare amounts. This made it easier to identify any anomalies or outliers and to understand the general trends in the data.

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

The null hypothesis (H0) is a statement that there is no effect or no difference, and it serves as the default assumption that any observed differences are due to random chance. The alternative hypothesis (H1) is a statement that there is an effect or a difference, and it represents what the researcher aims to prove.

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

For this project, the null hypothesis (H0) was formulated as: "There is no difference in fare amounts between customers who pay with credit cards and those who pay with cash." The alternative hypothesis (H1) was: "Customers who pay with credit cards have higher fare amounts than those who pay with cash."

- What conclusion can be drawn from the hypothesis test?

The conclusion from the hypothesis test will determine whether there is enough evidence to reject the null hypothesis in favor of the alternative hypothesis. If the p-value is less than the significance level (typically 0.05), we reject the null hypothesis and conclude that there is a statistically significant difference in fare amounts based on payment type.

## PACE: Execute Stage

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

Key insights might include whether customers who pay with credit cards tend to have higher fare amounts, which could inform the TLC's strategies for payment methods and customer engagement.

- What recommendations do you propose based on your results?

Based on the results, if it is found that credit card payments are associated with higher fare amounts, the recommendation might be to encourage the use of credit cards through incentives or promotions. This could increase overall fare revenue for taxi drivers and the TLC. Additionally, providing a seamless and secure payment experience could enhance customer satisfaction and loyalty.