**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 purposeis to demonstrate knowledge of how to prepare, create, and analyze A/B tests to make inferences about the relationship between payment type and fare amount.

* What is your research question for this project?

Is there any relationship between payment type and fare amount?

* What is the importance of random sampling?

Random sampling is a way to deter bias. It makes sure that every element of the population has an equal chance of being in the sample under consideration.

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

Suppose a company wants to measure customer satisfaction and only surveys people who shop online. This sampling method excludes people who prefer shopping in-store.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics are essential because they provide a straightforward summary of data, making complex data sets understandable and accessible. They help in identifying patterns, trends, and potential outliers, giving an initial sense of what the data reveals without performing advanced analyses.

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

Summarized data gave insights to support data-informed decisions. For example, knowing the averages of fare amounts against each payment type helped us to understand if credit card users pay more than riders who pay by cash.

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

The null hypothesis is a statement of no effect, no difference, or no association. It assumes that any observed effect in the data is due to random chance or sampling variability rather than a true effect. The alternative hypothesis contradicts the null hypothesis and represents the presence of an effect, difference, or association. It suggests that any observed effect is real and not due to random chance.

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

𝐻0: There is no difference in the average fare amount between customers who use credit cards and customers who use cash.

𝐻𝐴: There is a difference in the average fare amount between customers who use credit cards and customers who use cash.

* What conclusion can be drawn from the hypothesis test?

The high t-statistic and extremely low p-value together indicates that riders who pay through credit cards pay a higher fare amount as compared to riders who pay by cash.

**PACE: Execute Stage**

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

The key business insight is that encouraging customers to pay with credit cards can generate more revenue for taxi cab drivers

* What recommendations do you propose based on your results?

The Automatidata data team recommends that the New York City TLC encourages customers to pay with credit cards, and create strategies to promote credit card payments. For example, the New York City TLC can install signs that read “Credit card payments are preferred” in their cabs, and implement a protocol that requires cab drivers to verbally inform customers that credit card payments are preferred.