**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 purpose of this project is to predict taxi cab fares before each ride. At this point, this project’s focus is to find ways to generate more revenue for New York City taxi cab drivers. This part of the project examines the relationship between total fare amount and payment type.

* What is your research question for this project?

To predict taxi cab fares before each ride.

* What is the importance of random sampling?

It provides each individual or member of a population with an equal and fair probability of being chosen. So, there will be no bias.

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

Presidential election voters.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

In general, descriptive statistics are useful because they let you quickly explore and understand large amounts of data.

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

Use descriptive statistics to conduct Exploratory Data Analysis (EDA).

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

A null hypothesis states that there is no significant relationship or difference between variables, while an alternative hypothesis suggests that there is a significant relationship or difference.

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

The null and alternative hypotheses are two competing claims that researchers weigh evidence for and against using a statistical test:

Null hypothesis (H0): There's no effect in the population.

Alternative hypothesis (Ha or H1): There's an effect in the population.

* What conclusion can be drawn from the hypothesis test?

We use t-test and p-value to compare it with significance level. If p-value is less than significance level that then we reject the null hypothesis and vice versa.

**PACE: Execute Stage**

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

There is a statistically significant difference in the average total fare between customers who use credit cards and customers who use cash. Customers who used credit cards showed a higher total amount compared to cash.

* 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