**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

# 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 conduct an A/B test to discover whether or not customers paying with credit cards pay higher rates than customers paying with cash.

* What is your research question for this project?

Are the relations between fare amounts and payment types equal?

* What is the importance of random sampling?

Random sampling is important because it minimizes the amount of bias in the analysis of a sample of a population.

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

In the case of different payment types, non-random sampling may involve snowball sampling, where certain customers are asked to recommend people to join the study, or convenience sampling, where people are selected at the sampler’s ease, with no regard for factors that might bias the result.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Descriptive statistics can provide information such as the central tendency and distribution of a dataset. This helps identify relationships and patterns between variables.

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

Computing descriptive statistics helped determine the average fare amount for each payment type needed for the A/B test.

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

In hypothesis testing, the null hypothesis is the starting assumption about the state of a dataset. The alternative hypothesis is the updated view of the dataset, which will be accepted if the hypothesis test confirms it.

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

The question that needed to be answered was whether or not there was a connection between fare amount and payment type. Therefore, the null hypothesis was that the fare amount for each payment type is equal, and the alternative hypothesis was that the fare amount for one type of payment was higher than the fare amount for the other type.

* What conclusion can be drawn from the hypothesis test?

The hypothesis test revealed that, on average, customers who pay with credit cards pay more than customers who pay with cash.

**PACE: Execute Stage**

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

The key business insight that emerged was that encouraging customers to pay with credit cards will improve profits.

* What recommendations do you propose based on your results?

I would recommend a more in-depth look at the distance travelled during trips paid for with credit cards, to confirm whether or not the higher average was due to payment type or trip distance.