**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:

* Understand and assess the proposed scenario
* Demonstrate knowledge of how to prepare, create, and analyze an A/B test using statistics
* Apply descriptive and inferential statistics, probability distributions, confidence intervals, and hypothesis testing in Python
* Articulate findings in 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: Planning Stage**

* What is the main purpose of this project?

To show the application of the A/B tests and its processes. It aims mainly to find ways to generate more revenue for taxi cab drivers.

* What is your research question for this project?

What are the variables that have more influence in determining the total taxi fare amount?

e.g. are the two variables total\_amount and payment\_type correlated?

* What is the importance of random sampling?

It can help ensure that your sample is representative of the population and that the sample is high-quality. Once this simulation is done, the date is ready to have hypothesis test conducted.

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

A population in the sample might be more representative than others and so wrong conclusions might be drawn thus misrepresenting the entire population of the dataset that was surveyed.



**PACE: Analyzing & Constructing Stages**

* In general, why are descriptive statistics useful?

Because it helps to quickly explore, understand and compare between variables from a large dataset with literally one line of code if done with computers.

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

By providing key insights and summarizing important characteristics of the data set. For example, helping to understand the central tendency, data’s variability, shape of the data distribution, identifying outliers, discovering correlations between variables and making informed data comparisons.

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

The null hypothesis basically states that the observed event or result in the population is due by chance and thus is random. In the other hand, the alternative hypothesis states the opposite and thus that there is something interesting or significant in the event or result of the population that we are interested that is having an effect or relationship on it and thus it is not random.

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

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

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

* What conclusion can be drawn from the hypothesis test?

The p-value is significantly smaller than the significance level of 5%, thus I reject the null hypothesis.

I conclude that there is a statistically significant difference in the average total fare amount between customers who use credit cards and customers who use cash.

**PACE: Execute Stage**

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

- that motivating customers to pay with credit cards can generate more revenue for taxi cab drivers.

* What recommendations do you propose based on your results?

Similar to the above answer. To find methods to encourage customers to choose primarily credit cards as a payment option than cash.