**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 prepare, create, and conduct A/B testing to help find ways to generate higher revenues for taxi cab drivers

* What is your research question for this project?

Do the customers who use a credit card pay higher fare amounts than those who use cash?

In this statistical analysis more concrete and based on the task given the following:

Is there a relationship between total fare amount (total\_amount field) and payment type?

* What is the importance of random sampling?

Sample groups of customers in this project analysis:

1. required to pay with credit card
2. required to pay with cash

For this the assumption is made that customers have been randomly selected and divided into both groups respectively (disregarding of the reason for selecting this payment type and assuming the customers are completely paying the invoice with the respective payment type as per their group)

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

Age and trip distance may affect payment type. Non random sampling may focus only on young people who may prefer credit cards over cash as opposed to older customers who may be used to pay in cash.

Also customers driving longer distances may use credit cards as they do not have higher cash amounts readily at hand so they would not pay in cash. Focusing only on these would bias the results in favor of credit cards.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

To understand the data working with and to provide collaborators with a summary of the relative location of values in the data, as well an information about its spread

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

By understanding the characteristics and quality of the data one can assess how to deal with it to calculate summary statistics and to perform hypothesis testing (A/B)

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

- A null hypothesis (H0) always predicts no true effect, no relationship between variables, or no difference between the groups

- An alternative hypothesis (Ha or H1) states your main prediction of a true effect, a relationship between variables, or a difference between groups

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

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

𝐻𝐴: 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?

As the p-value is significantly smaller than the significance level of 5%, reject the null hypothesis

**PACE: Execute Stage**

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

Customers paying with credit cards generate higher revenues for the taxi drivers and as such should be encouraged paying with this payment type

* What recommendations do you propose based on your results?

As per assumption made (see jupyter notebook) both customer groups were generated based on a random experiment to perform A/B testing. Assuming both customer types were required to pay with the respective payment type and that they also actually did so (no bias).

As per data dictionary the tip amount (field) is automatically populated for credit card tips. But Cash tips are not included. So we would need to account for this and do the A/B testing based on the fare\_amount field and not as now conducted on the total\_amount field. This means that the cash paying sample group got undervalued as their cash tips are not included in the Total\_amount field values.

Third, longer trip distances may explain the preference for credit cards, as the total amount in cash would not be readily available to many customers.

**Recommendations:** While encouraging customers to pay with credit cards, and creating strategies to promote credit card payments – one should conduct further A/B testing with following analysis: use fare\_amount field instead of total\_amount field and choose appropriate sample groups that account for trip distance and reason for payment type choice etc.