**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?

To Conduct the A/B test among two randomly sampled groups, and discover if customers who use credit cards pay higher fare amounts than customers than who use cash.

* What is your research question for this project?

Is there a relationship between total fare amount and payment type?

* What is the importance of random sampling?

The importance of random sampling is that it reduces the chance of bias in sampling and blindly select units, thus no certain groups are favoured in this method.

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

For example, if we are collecting samples for the poll election from an area to see who will win in election, but in that area of city the most voters reside supporting the opposite party. This may include bias because we did not randomly select people from the whole city areas. Since the election in being conducted for the city, and the poll might include more supporters for opposite party.



 **PACE: Analyze & Construct Stages**

* In general, why are descriptive statistics useful?

Because they give you the idea of the spread of each column in dataset. You might conclude the

Values from the dataset.

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

If the column is continuous, you might have idea of values and spread of values in that column and find extreme values. You may infer the average and standard deviation. If the column is categorial, doing value\_counts() will give you the idea which values are mostly occurred in the dataset.

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

The null hypothesis is the claim that meant to be true initially. The Alternative hypothesis is the claim against the null hypothesis which forces to reject the null hypothesis based on some evidence. The null hypothesis states that the difference in alternative hypothesis is due to random chance.

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

$H\_0$: There is no difference in the average total fare amount between customers who use credit cards and customers who use cash.

$H\_A$: 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?

Since The p-value is more than the significal level, therefore we FAILED TO REJECT the NULL hypothesis.

Therefore, There is NO 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?

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 difference between average card payment fare and cash fare is inflated, because we use the total amount as the comparing variable. A possible reason for this occurance is because cash tips aren't declared. In turn, this means that we capture tips in one group but not in the other. Instead, one could compare the `fare\_amount` column.