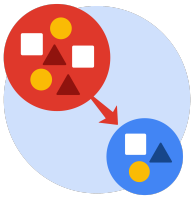
# Course Four

## 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

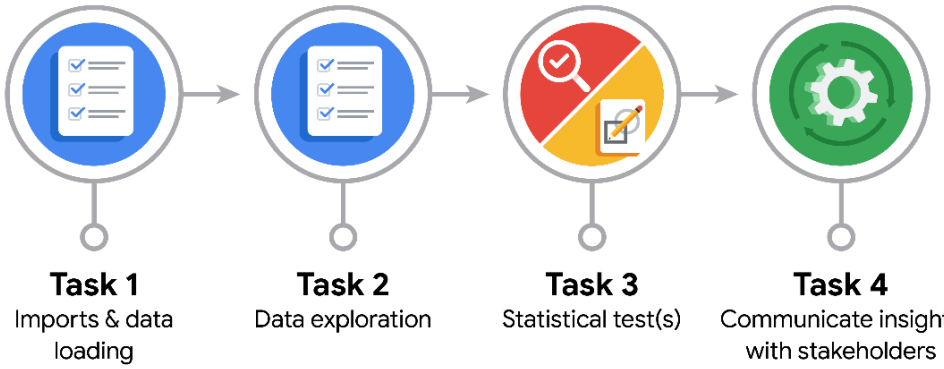
### Relevant Interview Questions

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 main purpose of this project is to prepare, create, conduct and analyze A/B tests to analyze the relationship between fare amount and payment type. The A/B test should aim to find ways to generate more revenue for taxi cab drivers |

- What is your research question for this project?

| |
|---|
| Do customers who use credit cards pay higher fare amounts than those who use cash?That is the relationship between fare amount and payment method? |

- What is the importance of random sampling?

It allows for unbiased data collection, which lets studies arrive at unbiased conclusions. There are no special skills involved in using this method, which can result in fairly reliable results. It also require little to no special method.

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

**Here are a few examples of of sampling bias:**

**Selection bias:** Survivors bias is an example of this kind of bias. This type of bias ignores those subjects that did not make it past a certain point in the selection process and only focuses on the subjects that "survived." This can lead to false conclusions

**Recall bias:** Memory is imperfect, and when survey participants can't recall correctly, it results in such bias.

**Observer bias:** When researchers consciously and subconsciously influence the interpretation of a data. It may take the form of focusing on certain datasets or influencing participants during the survey

Exclusion bias: This involves deliberately excluding specific participants from a study. This can affect the integrity of the research

**Non-response bias**: Non-response, or participation bias, occurs when a group of respondents refuses to participate in a study or drop out during the study period. This could be due to the length of the survey, the structure of the questions, or sensitive topics at hand.

## PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Data is summarized and explained in descriptive statistics. It allows us to display data in a more useful manner. During data collection, a sample population is considered. Here too descriptive statistics are useful in explaining the sample data.

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

There are several benefits of computing descriptive statistics. It helps understand the central tendencies of the data such as mean, median and mode. In addition, descriptive statistics provides a measure of variability in the data. Measures such as range, variance,standard deviation are vital in assessing risk, and predicting future trends.

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

The following are some differences between Null and alternative hypothesis:

| | Null Hypothesis | Alternative Hypothesis |
|---|---|---|
| Notation | $H_0$ | $H_a$ |
| Definition | Is a statement that is assumed to be true unless there is convincing evidence to the contrary | A statement that contradicts the null hypothesis and is accepted as true only of there is convincing evidence. |
| What is it? | Generally, researcher try to disprove or reject it | Generally, researchers try to approve or accept it |
| Testing process | Indirect | Direct |
| p-value | It is rejected if p-value is less than significance level, otherwise, it is accepted | It is accepted if p-value is less than significance levels, otherwise it is rejected |

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

1. There is no difference in the average fare amount between customers who use credit cards and customers who use cash.

2. There is a difference in the average 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 significantly smaller than the significance level of 5%, I reject the null hypothesis. There is a statistically significant difference in the average 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 following are two business insights that emerged from the A/B test:<br><br>1. Encouraging customers to pay with credit card can generate more revenue for taxi cab drivers and everyone involved. As the test suggests, customers using credit cards are either more willing to spend more, perhaps due to the psychological effect of not paying with physical cash, or they opt for longer or more expensive rides. |

- What recommendations do you propose based on your results?

| |
|---|
| Promotions or discounts could be tailored to encourage higher spending from the group that typically spends less. Alternatively, loyalty programs could be designed to reward frequent users, irrespective of their payment method.<br><br>**Enhancing Customer Experience:** This insight might suggest a need to streamline the payment process for both groups to enhance customer satisfaction. For example, improving the credit card payment system to be more user-friendly or offering incentives for cash users to switch to card payments due to its higher revenue potential. |