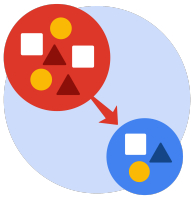
Course Four

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

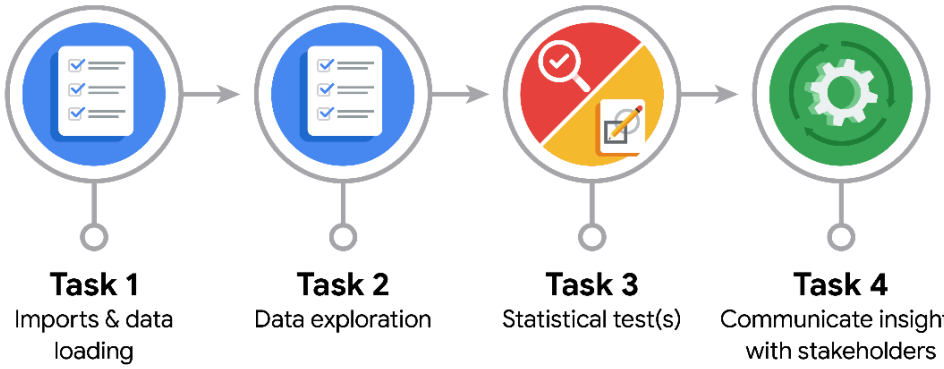
Relevant Interview Questions

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 analyze the relationship between fare amount and payment type within the New York City Taxi & Limousine Commission (New York City TLC) dataset, specifically to determine if there's a statistically significant difference in fare amounts between credit card and cash payments. Also to demonstrate knowledge of how to prepare, create, and analyze A/B tests.

- What is your research question for this project?

Is there a relationship between total fare amount and payment type?” More specifically, "Is there a difference in the average fare amount between customers who use credit cards and customers who use cash?"

- What is the importance of random sampling?

Random sampling is important because it minimizes bias, ensures that the sample is representative of the entire population, and allows for valid statistical inferences. It helps to ensure that any observed differences are likely due to actual variations in the population rather than systematic differences in the sample.

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

If you didn't use random sampling and instead, for example, sampled only taxi trips from business districts during rush hour, you might overrepresent credit card payments (as business travelers often use cards). This would lead to a biased conclusion that credit card users always pay higher fares, when in reality, it might just be a reflection of the specific time and location of the sample.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful for summarizing and understanding the main features of a dataset. They provide measures of central tendency (mean, median), dispersion (standard deviation), and shape (distribution), which help in identifying patterns, outliers, and potential relationships within the data.

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

Computing descriptive statistics helped by providing an overview of the fare amounts and payment types. It showed the average fare amounts for each payment type, revealing a potential difference between credit card and cash payments. It also helped to identify any potential outliers within the data.

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

The null hypothesis (H0) states that there is no significant difference or relationship between variables. The alternative hypothesis (HA) states that there is a significant difference or relationship.

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

• **H0:** There is no difference in the average fare amount between customers who use credit cards and customers who use cash. • **HA:** 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?

The hypothesis test resulted in a very small p-value (6.797387473030518e-12), which is significantly less than the chosen significance level (0.05). Therefore, we reject the null hypothesis. There is a statistically significant difference in the average fare amount between customers who use credit cards and those who use cash.

PACE: Execute Stage

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

The key insight is that customers who pay with credit cards tend to have higher average fare amounts than those who pay with cash. This suggests that encouraging credit card usage could potentially increase revenue for taxi cab drivers.

- What recommendations do you propose based on your results?

• Consider implementing strategies to encourage credit card payments, such as offering incentives or promoting the convenience of card payments. • Further investigate the factors that contribute to higher fares with credit card payments (e.g., longer trips, different customer demographics). • Recognize that the test was conducted with the assumption of random assignment of payment type, which is not how real world taxi rides occur. Therefore, consider these results as directional, and not definitive proof of cause and effect.