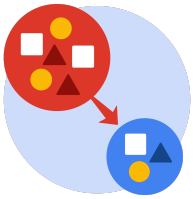
Course Four

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

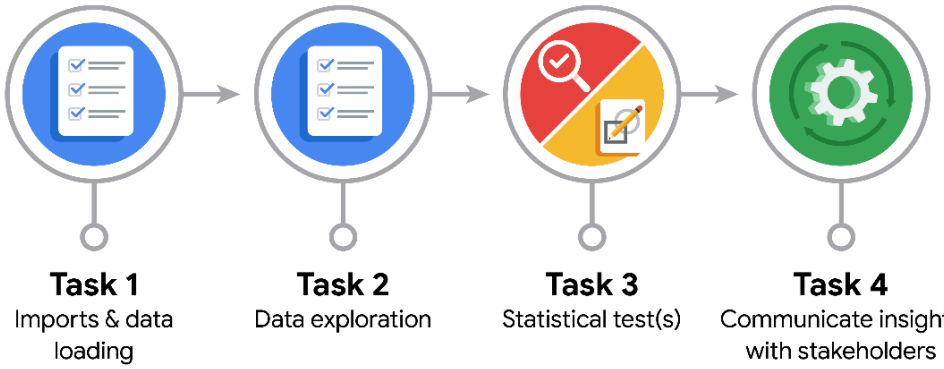
Relevant Interview Questions

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 the total fare amount of taxi rides and the payment type used. Ultimately, the goal is to identify potential strategies that could lead to increased revenue for New York City taxi cab drivers based on these payment patterns.

- 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?

  Random sampling is crucial because it helps ensure that the subset of data we analyze (our sample) is representative of the entire population of taxi rides. By giving every ride an equal chance of being included in our sample, we minimize the risk of introducing bias. This allows us to make more reliable inferences about the entire population based on our findings from the sample.

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

If we only analyzed taxi trips that occurred during weekday rush hours, our sample would be biased towards business travelers and potentially longer, more expensive trips to commercial areas. This would not accurately reflect the fare amounts and payment types of all taxi rides, including those taken during evenings, weekends, or in residential areas. Our conclusions about the relationship between fare amount and payment type might be skewed by this non-random selection of data.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics provide a concise summary of the main features of a dataset. They help us understand the central tendency (e.g., mean, median), the spread or variability (e.g., standard deviation, range), and the shape of the data distribution for individual variables. This initial understanding is essential for identifying patterns, potential outliers, and formulating hypotheses for further statistical testing.

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

Computing descriptive statistics for the 'Total Amount' variable, broken down by 'Payment Type', allowed me to see the average fare, the variability in fares, and the range of fares for each payment method (e.g., credit card, cash). This initial exploration helped me observe if there were apparent differences in the typical fare amounts associated with different payment types, which then informed the formulation of my null and alternative hypotheses for the formal hypothesis test.

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

The null hypothesis (H0) is a statement of no effect or no difference. It represents the status quo or a commonly accepted belief. In our case, the null hypothesis would state that there is no significant difference in the average total fare amount across different payment types.

The alternative hypothesis (HA) is a statement that contradicts the null hypothesis. It proposes that there is a real effect or a significant difference. In our case, the alternative hypothesis would state that there is a statistically significant difference in the average total fare amount across different payment types.

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

Based on the project goal of finding ways to generate more revenue and my initial observation of descriptive statistics, I formulated the hypotheses as follows:

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

Alternative Hypothesis 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?

Since the p-value (almost 0%) is significantly smaller than the significance level of 5%, I reject the null hypothesis.

I conclude 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 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?

Consider strategies to encourage credit card payments, such as highlighting the convenience or offering small incentives (if permissible and cost-effective).

Further investigate the reasons behind the higher average fares for credit card users. Are they taking longer trips, traveling during different times, or using different rate codes? Understanding these underlying factors can inform more targeted strategies.

Explore partnerships with credit card companies for potential promotions or loyalty programs that could further incentivize credit card use and potentially increase overall spending.