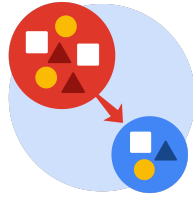


## Course Four

### 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

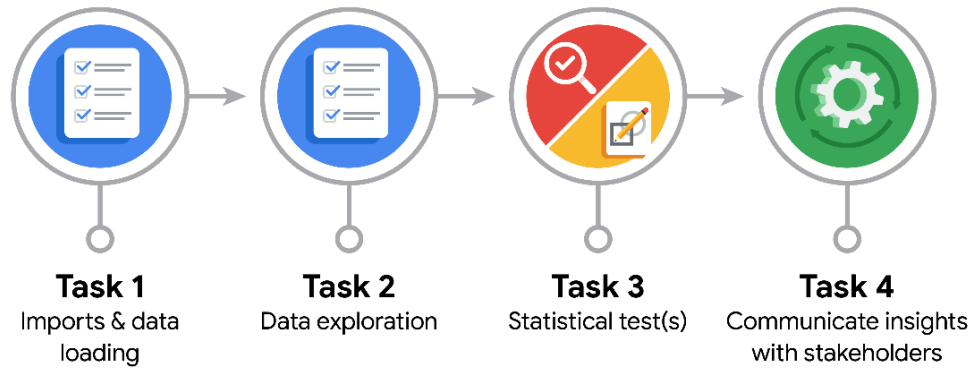
#### Relevant Interview Questions

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?

This project aims to conduct a hypothesis testing between payment type & fare amount, and determine the statistical significance of whether people who pay by credit card are higher in the fare rides amount than those who pay with cash.

- What is your research question for this project?

Is there a statistically significant difference between fare amount and payment type, which those who pay by credit card are higher in fare amount than those who pay by cash?

- What is the importance of random sampling?



Random sampling helps to ensure that choosing members in a specific experiment represents the whole population. More importantly, it comes with choosing randomly (or probability sampling), rather than choosing with intended purposes (non-probability sampling).

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

The under-coverage bias comes from convenience sampling. For instance, researchers conduct opinion polls/interviews with customers by standing at the mall to ask them about their experiences. The bias shown here is the people who aren't interviewed (either they don't come to the mall, they don't like it, or go to different malls).



### **PACE: Analyze & Construct Stages**

- In general, why are descriptive statistics useful?

Performing descriptive statistics helps to understand the summarization of the whole dataset quickly and easily.

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

Descriptive statistics come from many measures to compute such as central tendency (mean or median), dispersion/spread (standard deviation), and position (quartile) for variables which include numerics and objects.

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

The null hypothesis is the one we first assume to be true unless there's evidence for the contrary. The alternative hypothesis is the one we favor, which is the contrary of the null hypothesis, and also has to be supported by the evidence.



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

In this project, the null and alternative hypotheses are:

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

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?

We'll choose to reject or fail to reject the null hypothesis based on the test result. Specifically, the p-value (or probability of observed difference between those two means as or more extreme if the null hypothesis is true) and significance level. If the p-value < significance level, we reject the null hypothesis. Vice versa, if the p-value > significance level, we fail to reject the null hypothesis.



### **PAC: Execute Stage**

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

The p-value of this test is 0.00...679 (6.79...e-12), which is smaller than the significance level of 0.05. By that, we can reject the null hypothesis which is there is no difference in the average fare amount between customers who use credit cards and customers who use cash.

- What recommendations do you propose based on your results?

Based on the results, we might encourage more customers to use credit card as primary payment type, and advantages and conveniences, and special offers.

