Course Four
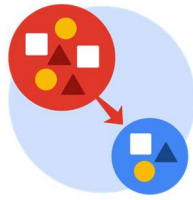**From Data to Insight: The Power of Statistics**

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.

- An A/B test is a method used to compare two different versions of a model, product, or set of data to determine which one performs better in the context of the business question.
- When analyzing employee performance metrics, a one-sided Z-Test may be useful to determine the percentile employees fall into regarding specific data. A two-sided T-Test may be beneficial for determining
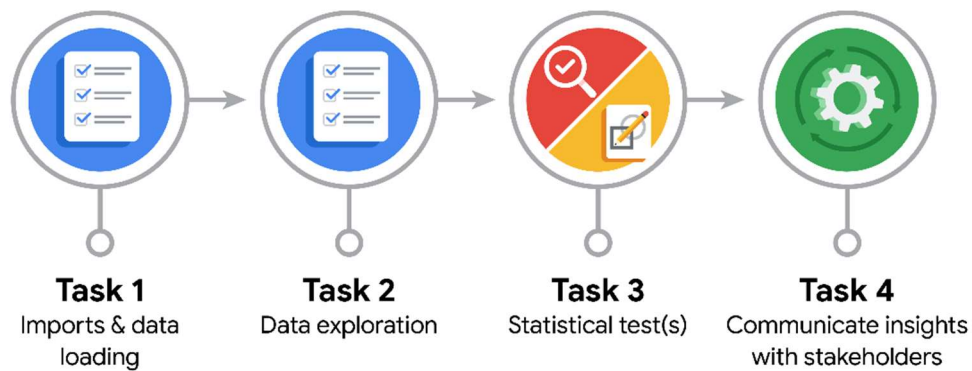
better versions of the company website that will promote more sales, click rate, and ultimately time spent on the website before leaving.

- Ensure that the results presented to stakeholders are clearly and concisely explained using non-technical jargon, depending on their level of knowledge. Additionally, utilizing plots can aid in retaining attention and making conclusions easier to digest.
- Plots can make communicating technical information to a non-technical audience much easier. Additionally, descriptive statistic summaries should be formatted in a way that produces only relevant points of information.
- Key components of an A/B test include formulating a hypothesis, choosing a variable to test, and employing random sampling while breaking up the audience into two groups (control and variation).

Reference Guide

This project has four tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.

Task 1
Imports & data
loading

Task 2
Data exploration

Task 3
Statistical test(s)

Task 4
Communicate insights
with stakeholders

Data Project Questions & Considerations

PACE: Plan Stage

● What is the main purpose of this project?

> The main goal of the project is to create a reliable and precise model that can predict taxi fares before the ride takes place. For this specifics task, we will use descriptive statistics and hypothesis testing to investigate if there is any connection between the payment method used and the fare amount

● What is your research question for this project?

> Is there a relationship between the payment method used and the amount paid for the fare?

● What is the importance of random sampling?

> Ensures independence of observations and that all parties involved have an equal opportunity of being chosen, therefore helping mitigate bias.

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

> You may end up selecting samples from the same group or geographic region, resulting in biased data and skewed model results.

PACE: Analyze & Construct Stages

● In general, why are descriptive statistics useful?

> They enable swift exploration of data, including key variables, distributions, central tendency, and spread. They also help detect outliers, missing, and incorrect data.

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

> It allowed me to determine the mean, distribution, standard deviation, whether there was missing data and how significant the outliers in the dataset were.

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

> The null hypothesis states that there is no statistically significant difference or effect. The alternative hypothesis states that there is a statistically significant difference or effect, and it rejects the null hypothesis if there is sufficient evidence to do so.

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

> To achieve the business objective, a null hypothesis was formulated. The hypothesis questions whether there is a variation in fare amounts between credit and cash payments. The default assumption is that there is no difference. However, the alternative assumption suggests that there is a difference in fare amounts between credit and cash payments.

● What conclusion can be drawn from the hypothesis test?

> We can conclude that there is a statistically significant difference in fare payments between credit cards and cash, as indicated by the very low p-value.

PACE: Execute Stage

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

> On average, credit card payments by riders exceed those made in cash, possibly due to the convenience of carrying a single card.

- What recommendations do you propose based on your results?

> It has been found that riders tend to spend more money when using credit cards than cash. Therefore, the company should explore ways to encourage riders to use credit cards instead of cash in order to increase revenue. One possible solution could be to introduce a points program or other benefits that are associated with using credit cards, or by investigating methods to make paying with credit far