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

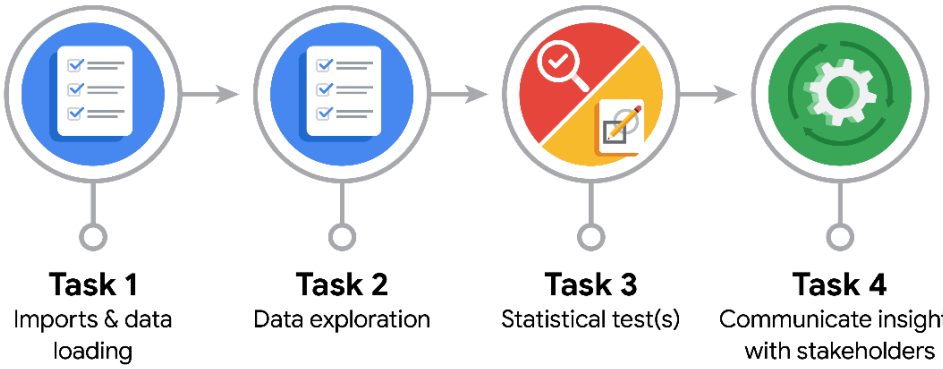
Relevant Interview Questions

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?

To conduct descriptive statistics and a hypothesis test, namely a two-sample hypothesis test in the form of a t-test for the means.

- What is your research question for this project?

Is there a statistically significant difference in the mean amount of drives between iPhone users and Android users?

- What is the importance of random sampling?

Random sampling ensures that the samples are representative, unbiased and fair.

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

An example of sampling bias that could occur if random sampling wasn't used is undercoverage bias. This is when some members of a population are inadequately represented in a sample.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics are useful as they help to better understand the data and to get quick initial statistics of it.

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

It helped to show the central tendency, dispersion and spread of the data.

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

The null hypothesis states that the observation occurs by chance and is not statistically significant, whereas the alternative hypothesis states that the observation doesn't occur by chance and is in fact statistically significant.

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

I began with the null hypothesis, with the meaning of a null hypothesis stating equality and that there is no statistical significance between the observed data. Therefore, I stated for my null hypothesis that there is no statistically significant difference in the mean drives between device users. For my alternative hypothesis, my project is looking at if there is a statistical difference in the means and not if the sample mean is greater or less than the population mean. Therefore, I stated for my alternative hypothesis that there is a statistical significance in the mean drives between device users. This in turn led to the alternative hypothesis meaning inequality and so lead to a two-tailed distribution.

- What conclusion can be drawn from the hypothesis test?

With the P-Value being 0.31 which is greater than our significance level of 5%, we fail to reject the null hypothesis and therefore there is no statistically significant difference in the mean number of drives between drivers who use iPhones and those who use Androids.

PACE: Execute Stage

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

The key business insight that emerged from the hypothesis test was that the drivers who use iPhone devices have on average a similar number of drives as those who use Android devices.

- What recommendations do you propose based on your results?

The recommendations that I propose based on my results are to explore additional behavioural variables to identify more meaningful predictors of user churn. Another is to conduct further hypothesis tests across different user segments. One recommendation is to increase the sample size in order to strengthen statistical power. The final recommendation I propose is to transition to multivariate modelling to better explain churn behaviour.