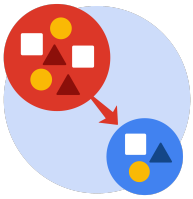
Course Four

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

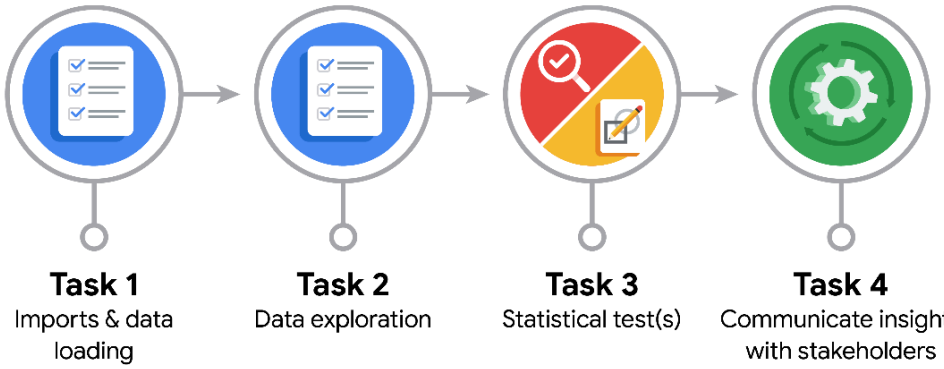
Relevant Interview Questions

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 understand why users churn, and what demographic is the one that is more probable to churn.

- What is your research question for this project?

What's the demographic of users who tend to churn?

- What is the importance of random sampling?

The importance of random sampling is that the samples are not biased, because your data comes from a distribution that is representative of the population.

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

If you only sample people from convenience or from a selected location or social class, the data will be skewed from the sample and may not be representative of the whole population. For example if you only take as sample your friends for a socioeconomic study, the sample might be biased towards people with similar socioeconomic status as yours.

PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

They are really useful because you can easily observe the maximum value, the minimum value, the mean etc..., with that you can comprehend fast the overall distribution of the numerical data, as well you can discover any outlier or wrong data.

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

It gives you a overall glance of how data is distributed, by analyzing min, max, mean, median, mode, standard deviation,

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

In hypothesis testing, the null hypothesis states that the difference between the mean of the population and the mean of the sample is by chance and not statistically difference, while the alternative hypothesis states that the difference is statistically significant and not by chance.

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

You first need to comprehend what question you want to answer using data, after that you can easily state the null and alternative hypothesis, then you need to state the significance level and finally compute the p value.

- What conclusion can be drawn from the hypothesis test?

We can state if the proposition is statistically significant or is by chance.

PACE: Execute Stage

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

From this hypothesis testing we can conclude that there is not a statistically significant difference between android and iphone users by mean drives.

- What recommendations do you propose based on your results?

Explore correlations between other variables, to identify possible correlations between user churn rate.