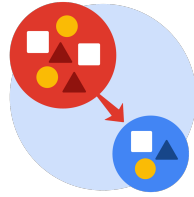


## Course Four

### From Data to Insight: The Power of Statistics



SAHAR HAKAMI

#### 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?

Build churn prediction model, by analyzing the data we can specify the important factors that contribute to making users churn

- What is your research question for this project?

What factors influence the user to churn?

- What is the importance of random sampling?

It can be hard to analyze the data for the whole population, taking a sample that have the characteristics help in drawing conclusions and insight from the data that represent the whole population.

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

Not taking the sample randomly may result in a sample not representing the whole population, some users for example in the population might be chosen to be in the sample more than other.



## PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

There are a lot of numerical variables in the dataset, these variables might have hidden patterns that can be analyzed to draw conclusion on the whole population.

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

They can help you understand the characteristics, patterns, and trends of your data, as well as compare different groups or variables. One of the main benefits of using descriptive statistics is that they can simplify and organize large amounts of data into a few numbers or graphs. This can make it easier to grasp the main features and patterns of your data, as well as identify any outliers or errors. Descriptive statistics can also help you compare different groups or variables within your data. We can detect outliers by computing the distribution of the data points.

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

The null hypothesis states that the result of some variables occur by chance, but the alternative tells us that there is significant statistics to the result, and it is not accruing by chance.

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

The null hypothesis state that there is no affect from an independent variable on the dependent y variable, while the alternative hypothesis tells us that there is a relation between the independent x and the outcome dependent y.

- What conclusion can be drawn from the hypothesis test?

What are the variables that have relationship with the outcome variable?



### PACE: Execute Stage

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

After running t-test on the 2 different groups of users: iPhone and android users.

We got a p-value=0.143 which is greater than 0.05, thus we could not reject our null hypothesis, which states that there is no difference in the mean of drives for the groups.

- What recommendations do you propose based on your results?

Failing to reject the null indicates that our sample did not provide sufficient evidence to conclude that the effect exists. However, at the same time, that lack of evidence doesn't prove that the effect does not exist. Maybe this is a problem with the dataset sampling method and data, more investigation needed to prove this assumption, we can also test other factor that might affect the outcome variable.