## **Summary:**

# What to do when there is no data

Earlier, you learned how you can still do an analysis using proxy data if you have no data. You might have some questions about proxy data, so this reading will give you a few more examples of the types of datasets that can serve as alternate data sources.

## **Proxy data examples**

Sometimes the data to support a business objective isn’t readily available. This is when proxy data is useful. Take a look at the following scenarios and where proxy data comes in for each example:

| **Business scenario** | **How proxy data can be used** |
| --- | --- |
| A new car model was just launched a few days ago and the auto dealership can’t wait until the end of the month for sales data to come in. They want sales projections now. | The analyst proxies the number of clicks to the car specifications on the dealership’s website as an estimate of potential sales at the dealership. |
| A brand new plant-based meat product was only recently stocked in grocery stores and the supplier needs to estimate the demand over the next four years. | The analyst proxies the sales data for a turkey substitute made out of tofu that has been on the market for several years. |
| The Chamber of Commerce wants to know how a tourism campaign is going to impact travel to their city, but the results from the campaign aren’t publicly available yet. | The analyst proxies the historical data for airline bookings to the city one to three months after a similar campaign was run six months earlier. |

# Sample size calculator

In this reading, you will learn the basics of sample size calculators, how to use them, and how to understand the results. A **sample size calculator** tells you how many people you need to interview (or things you need to test) to get results that represent the target population. Let’s review some terms you will come across when using a sample size calculator:

* **Confidence level**: The probability that your sample size accurately reflects the greater population.
* **Margin of error**: The maximum amount that the sample results are expected to differ from those of the actual population.
* **Population**: This is the total number you hope to pull your sample from.
* **Sample**: A part of a population that is representative of the population.
* **Estimated response rate**: If you are running a survey of individuals, this is the percentage of people you expect will complete your survey out of those who received the survey.

**Statistical Power:**

* Statistical power is the probability of getting meaningful results from a test or study.
* It is an important concept for data analysts to understand when conducting hypothesis testing.
* A large sample size is necessary to ensure that the results of a test are not due to random chance.
* A statistical power of at least 0.8 or 80% is generally considered necessary for results to be considered statistically significant.
* In the example provided, a restaurant chain wants to test the effectiveness of a marketing campaign for a new milkshake flavor before launching it nationwide, and must determine the appropriate sample size to ensure statistically significant results.

**More on Sample Size:**

* Sample size is a part of a population that is representative of the entire population, used in businesses to analyze data without having to analyze the entire population.
* Sample size can be calculated using calculators found online, by inputting the confidence level, population size, and margin of error.
* Confidence level is the probability that the sample accurately reflects the greater population, with a higher level of confidence usually desired in industries such as pharmaceuticals.
* Margin of error tells how close the sample size results are to the results that would be obtained if the entire population was used.
* An example is provided of a middle school study on student candy preferences, where the population size is 500, confidence level is 95%, and margin of error is 5%, resulting in a minimum sample size of 218 students.
* Confidence level and margin of error are independent of each other and do not need to add up to 100%.