Selecting the right data

Following are some data-collection considerations to keep in mind for your analysis:

How the data will be collected

Decide if you will collect the data using your own resources or receive (and possibly purchase it) from another party. Data that you collect yourself is called first-party data.

Data sources

If you don't collect the data using your own resources, you might get data from second-party or third-party data providers. **Second-party data** is collected directly by another group and then sold. **Third-party data** is sold by a provider that didn't collect the data themselves. Third-party data might come from a number of different sources.

Solving your business problem

Datasets can show a lot of interesting information. But be sure to choose data that can actually help solve your problem question. For example, if you are analyzing trends over time, make sure you use time series data — in other words, data that includes dates.

How much data to collect

If you are collecting your own data, make reasonable decisions about sample size. A random sample from existing data might be fine for some projects. Other projects might need more strategic data collection to focus on certain criteria. Each project has its own needs.

Time frame

If you are collecting your own data, decide how long you will need to collect it, especially if you are tracking trends over a long period of time. If you need an immediate answer, you might not have time to collect new data. In this case, you would need to use historical data that already exists.

Use the flowchart below if data collection relies heavily on how much time you have:

Data collection considerations

1. Select the right type of data

2. Determine the time frame

Collect new data?

Use existing data?

3. Decide how data will be collected

or

3. Choose data sources

4. Decide how much data to collect

or

4. Decide what data to use