General Assembly

STUDY GUIDE

# THE DATA FRAMEWORK

---

The data analysis framework is a cyclical and iterative process that helps define and guide the workflow of an analysis project. It consists of six phases.

### Phase 1: Frame

Start with a clear, focused question that can be answered with data. These questions are made more specific by adding a focus — which is the specific, reasonable population or area in question — or a metric, which refers to the criteria or variables we're using to measure success or failure.

- Ex: "How should we determine where to open a new Starbucks in U.S. cities?" (focus)
- Ex: "How does GDP per capita impact the quantity and success of Starbucks locations in U.S. cities?" (key metric)

Based on the question you develop, a hypothesis is formed.

- Ex: "Cities with higher GDP per capita will support more Starbucks stores than those with lower GDP per capita."

The hypothesis defines the data we'll need to test it.

- Ex: City-level data, GDP per capita by city, which cities Starbucks stores are in and how many are in each city

### Phase 2: Extract

Find the data you need — both public data and internal data — and prepare it for analysis by cleaning the data.

### Phase 3: Wrangle / Prepare

Data is cleaned by identifying and handling problematic values, such as blanks, misspellings, and formatting inconsistencies.

### Phase 4: Analyze

With your clean data, begin running analyses to reveal relationships, patterns, and trends.

Gain a high-level understanding of your data through the use of summary statistics, graphics, and charts.

### Phase 5: Interpret

Take things a step further by interpreting the data to find relationships, patterns, and trends.

The Data Analytics Framework is cyclical and iterative. You may find yourself going around in circles before you're able to draw a strong conclusion.

Based on what you find in your data, you have three options for how to proceed:

1. Return to Extract: Find data that might reveal other factors that could affect your analysis.
2. Return to Analyze: Gain a better understanding of the data you're using data then attempt alternate methods of analysis.
3. Continue to Communicate!

### Phase 6: Communicate

You started this data analytics process by framing a clear, accurate question. Now, you'll be sharing your insights.

You'll need to convey:

- What was done.
- What was learned.
- What's next.