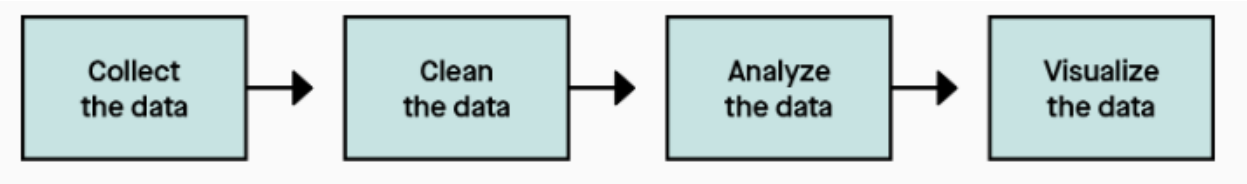
Data Analysis Process

Data analysis is a multistep process. We learned that there are four steps of the data analysis process:



1. Collect

For example, data can be collected through surveys, online website tracking, or transaction tracking. Regardless of which method is used, data that is relevant to your analysis goal is gathered and organized.

2. Clean

Most raw data, which is data that has been collected with no additional processing, isn't clean. It may have errors or formatting issues. This can become a problem if it isn't fixed. If the data type of a field is wrong, then formulas that you use on the data may produce wrong results. If there are blanks or strange symbols, then you may be missing important data from your analysis. Duplicate rows can give you incorrect totals and averages.

Cleaning the data, or data cleansing, is the process of making sure that data types are correct and use consistent formatting, that there are no duplicate records, and that any blank or missing values are correctly filled in with estimates. At this step, your job is to go through the data and look for any inconsistencies. This also gives you an opportunity to get more familiar with your data. You'll learn how to clean data effectively later in this course.

3. Analyze

Once your data is clean, it's finally ready for analysis! Analyzing data is the process of interpreting the data and using it to draw conclusions. Tables are a great way to view data so that you can see patterns or trends that you may have missed from just looking at the data. Statistical tests can highlight important relationships between fields and desired outputs. Dynamic models, where the input scenarios can be changed to see different resulting outputs, can also show how the data fields relate to one another.

4. Visualize

After analyzing data, you may want to share your results and findings with others. If someone isn't as familiar with the data as you are, data visualizations can help you communicate your analysis in a clear and engaging way.

There are a lot of data analysis tools available including: Spreadsheets, Databases, Programming languages, Statistics, and Data Visualization tools.