MODULE 02

DATA LIFE CYCLE: A data lifecycle is the process of managing data from its creation to its disposal, similar to the life cycle of other resources. The data life cycle provides a generic or common framework for how data is managed.

There are six stages to the data life cycle:

1. **Plan:** Decide what kind of data is needed, how it will be managed, and who will be responsible for it.
2. **Capture:** Collect or bring in data from a variety of different sources.
3. **Manage:** Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.
4. **Analyze:** Use the data to solve problems, make decisions, and support business goals.
5. **Archive:** Keep relevant data stored for long-term and future reference.
6. **Destroy:** Remove data from storage and delete any shared copies of the data.

More on the phases of data analysis

Each step in the data analysis process—ask, prepare, process, analyze, share, and act—plays a crucial role in extracting meaningful insights from data. As you navigate through each phase, from asking the right questions to taking informed actions, you harness the true power of data. In this reading, you'll explore how the data analysis process guides this program.

The ask phase

At the start of any successful data analysis, the data analyst:

- Takes the time to fully understand stakeholder expectations
- Defines the problem to be solved
- Decides which questions to answer in order to solve the problem

Qualifying stakeholder expectations means determining who the stakeholders are, what they want, when they want it, why they want it, and how best to communicate with them. Defining the problem means looking at the

current state and identifying the ways in which it’s different from the ideal state. With expectations qualified and the problem defined, you can derive questions that will help achieve these goals.
In an upcoming course, you'll learn how to ask effective questions and define the problem by working with stakeholders. You'll also cover strategies that can help you share what you discover in a way that keeps people interested.

The prepare phase

In the prepare phase, the emphasis is on identifying and locating data you can use to answer your questions. In an upcoming course, you'll learn more about the different types of data and how to identify which kinds of data are most useful for solving a particular problem. You'll also discover why it's so important that data and results are objective and unbiased. In other words, any decisions made from an analysis should always be based on facts and be fair and impartial.

The process phase

In this phase, the aim is to refine the data. Data analysts find and eliminate any errors and inaccuracies that can get in the way of results. This usually means:

- Cleaning data
- Transforming data into a more useful format
- Combining two or more datasets to make information more complete
- Removing outliers (data points that could skew the information)

After data analysts process data, they check the data they prepared to make sure it's complete and correct. This phase is all about getting the details right. Accordingly, the data analyst will refine strategies for verifying and sharing their data cleaning with stakeholders. In an upcoming course, you’ll use spreadsheets and structured query language, or SQL, to clean data.

The analyze phase

With a solid foundation of well-defined questions and clean data, you’ll delve into the analyze phase. This is when you turn the data you’ve gathered, prepared, and processed into actionable information. Data analysts use many powerful tools in their work. In one upcoming course you'll continue using two of them: spreadsheets and SQL. In another upcoming course you’ll explore using the programming language R to work with and analyze data.

The share phase

This phase is exactly what it sounds like: It’s time to share what you’ve learned with your stakeholders! In this part of the program, you'll learn how data analysts interpret results and share them with others to help stakeholders make effective, data-driven decisions. In the share phase, visualization is a data analyst's best friend. So, an upcoming course will highlight why visualization is essential to getting others to understand what your data is telling you. In another upcoming course, you’ll learn how to visualize data with R.

The act phase

The data analysis journey culminates in the act phase, when data insights are put to work. For you, this action involves preparing for your job search and having the chance to complete a case study project. It's a great opportunity for you to bring together everything you've worked on throughout this course. Plus, adding a case study to your portfolio helps you stand out from other candidates!

Key data analyst tools:

Spreadsheets

Data analysts rely on spreadsheets to collect and organize data. Two popular spreadsheet applications you will probably use a lot in your future role as a data analyst are Microsoft Excel and Google Sheets.
Spreadsheets structure data in a meaningful way by letting you

- Collect, store, organize, and sort information
- Identify patterns and piece the data together in a way that works for each specific data project
- Create excellent data visualizations, like graphs and charts.

Databases and query languages

A database is a collection of structured data stored in a computer system. Some popular Structured Query Language (SQL) programs include MySQL, Microsoft SQL Server, and BigQuery.
Query languages

- Allow analysts to isolate specific information from a database(s)

- Make it easier for you to learn and understand the requests made to databases
- Allow analysts to select, create, add, or download data from a database for analysis

Visualization tools

Data analysts use a number of visualization tools, like graphs, maps, tables, charts, and more. Two popular visualization tools are Tableau and Looker.
These tools

- Turn complex numbers into a story that people can understand
- Help stakeholders come up with conclusions that lead to informed decisions and effective business strategies
- Have multiple features

Tableau's simple drag-and-drop feature lets users create interactive graphs in dashboards and worksheets
Looker communicates directly with a database, allowing you to connect your data right to the visual tool you choose
A career as a data analyst also involves using programming languages, like R and Python, which are used a lot for statistical analysis, visualization, and other data analysis.

Choose the right tool for the job:

Spreadsheets	Databases
Accessed through a software application	Database accessed using a query language
Structured data in a row and column format	Structured data using rules and relationships
Organizes information in cells	Organizes information in complex collections
Provides access to a limited amount of data	Provides access to huge amounts of data
Manual data entry	Strict and consistent data entry
Generally one user at a time	Multiple users
Controlled by the user	Controlled by a database management system