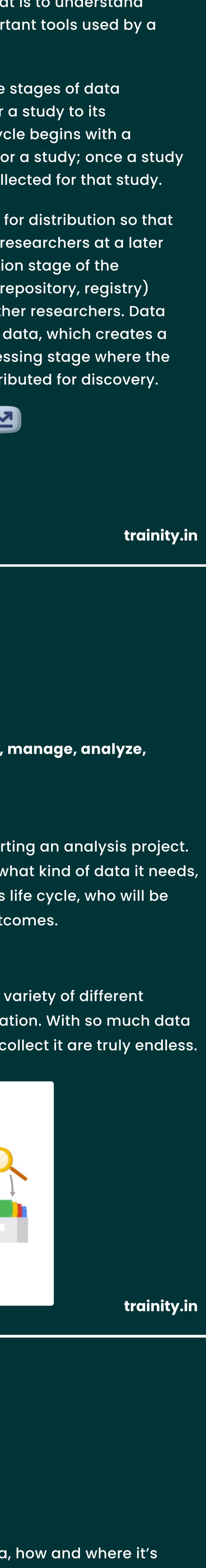


Data Lifecycle Management

A COMPREHENSIVE GUIDE



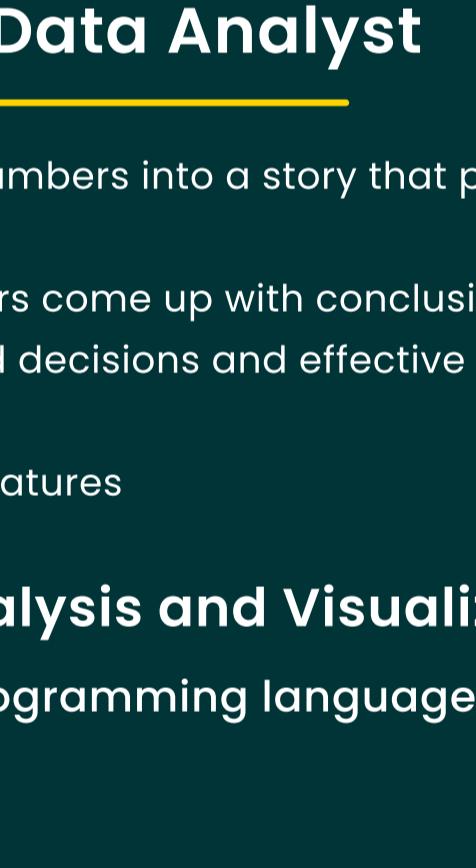
trainity.in

1

- Till now, we covered the overview of data analytics, processes & frameworks and some business cases. We would soon be starting with actual tools of a data analysis and solving case studies in upcoming weeks. But before that, we are left with the final step. That is to understand the Data Lifecycle and the most important tools used by a data analyst on a day to day basis.

The data lifecycle represents all of the stages of data throughout its life from its creation for a study to its distribution and reuse. The data lifecycle begins with a researcher(s) developing a concept for a study; once a study concept is developed, data is then collected for that study.

After data is collected, it is processed for distribution so that it can be archived and used by other researchers at a later date. Once data reaches the distribution stage of the lifecycle, it is stored in a location (i.e. repository, registry) where it can then be discovered by other researchers. Data discovery leads to the repurposing of data, which creates a continual loop back to the data processing stage where the repurposed data is archived and distributed for discovery.



trainity.in

2

Data Life Cycle

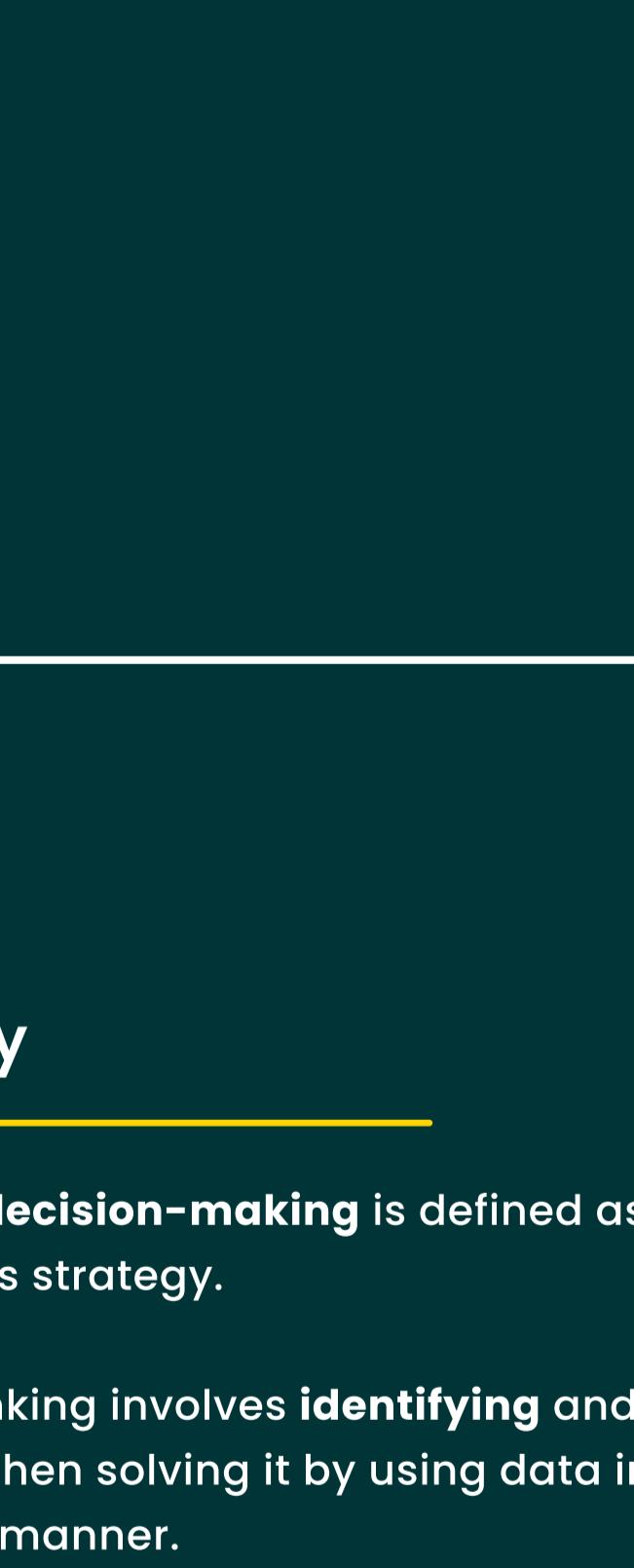
The life cycle of data is **plan, capture, manage, analyze, archive and destroy**.

1. Plan:

This actually happens well before starting an analysis project. During planning, a business decides what kind of data it needs, how it will be managed throughout its life cycle, who will be responsible for it, and the optimal outcomes.

2. Capture:

This is where data is collected from a variety of different sources and brought into the organization. With so much data being created everyday, the ways to collect it are truly endless.



trainity.in

3

Data Life Cycle

3. Manage:

This involves how we care for our data, how and where it's stored, the tools used to keep it safe and secure, and the actions taken to make sure that it's maintained properly.

4. Analyze:

In this phase, the data is used to solve problems, make great decisions, and support business goals.

5. Archive:

Archiving means storing data in a place where it's still available, but may not be used again. During analysis, analysts had a huge amount of data. So it makes sense to archive the data which is no longer relevant to us.

6. Destroy:

This is important for protecting a company's private information, as well as private data about its customers.

Let's understand Data Life Cycle, by watching the [video](#) below.

Click here

trainity.in

4

Toolkit of a Data Analyst

Let's look at the tools you need to be familiar with to become a successful data analyst. In coming modules, we will go in-depth with each tool along with practical assignment & case studies.

• 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. Digital worksheets structure data in a meaningful way by letting you

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

trainity.in

5

Toolkit of a Data Analyst

• 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 Big Query. Query languages;

1. Allow analysts to isolate specific information from a database(s)
2. Make it easier for you to learn and understand the requests made to databases
3. 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.

Tableau's simple drag-and-drop feature lets users create interactive graphs in dashboards and worksheets while Looker communicates directly with a database, allowing you to connect your data right to the visual tool you choose. These tools

trainity.in

6

Toolkit of a Data Analyst

• Statistical Analysis and Visualization:

Involves using programming languages, like R and Python

1. Turn complex numbers into a story that people can understand

2. Help stakeholders come up with conclusions that lead to informed decisions and effective business strategies
3. Have multiple features

• Data-driven decision-making:

is defined as using facts to guide business strategy.

2. Analytical thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

3. The five key aspects to analytical thinking are:

1. Visualization
2. Strategy
3. Problem – Orientation
4. Correlation
5. Big – picture and Detailed – oriented thinking

4. Some of the questions data analysts ask when they're on the hunt for a solution are:

1. What is the root cause of a problem
2. Where are the gaps in our process
3. What did we not consider before

5. Some of the questions data analysts ask when they're on the hunt for a solution are:

1. What is the root cause of a problem
2. Where are the gaps in our process
3. What did we not consider before

trainity.in

7

Summary

• Data analytics is the science of analyzing raw data in order to make conclusions about that information.

• Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information.

• Data analytics is important because it helps businesses optimize their performances.

• Data analytics process or framework is what drives the entire analysis cycle from asking the right business questions, collecting, cleaning, analyzing and visualizing insights powered by data.

• Phases of Data Analysis process includes (**SIX STEP PROCESS**):

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

trainity.in

8

Summary

• Data-driven decision-making is defined as using facts to guide business strategy.

• Analytical thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

• The five key aspects to analytical thinking are:

1. Visualization
2. Strategy
3. Problem – Orientation
4. Correlation
5. Big – picture and Detailed – oriented thinking

• Some of the questions data analysts ask when they're on the hunt for a solution are:

1. What is the root cause of a problem
2. Where are the gaps in our process
3. What did we not consider before

• Some of the questions data analysts ask when they're on the hunt for a solution are:

1. What is the root cause of a problem
2. Where are the gaps in our process
3. What did we not consider before

trainity.in

9

Summary

• Data is basically a collection of facts or information.

• Data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

• Companies hire data analysts to control the waves of data they collect every day, make sense of it, and then draw conclusions or make predictions.

• The data lifecycle represents all of the stages of data throughout its life from its creation for a study to its distribution and reuse.

• Data discovery leads to the repurposing of data, which creates a continual loop back to the data processing stage where the repurposed data is archived and distributed for discovery.

• The life cycle of data is –

1. Plan
2. Capture
3. Manage
4. Analyze
5. Archive
6. Destroy

trainity.in

10

Summary

• The tools that we need for data analyst are –

1. Spreadsheets – Microsoft Excel and Google Sheet
2. Databases and query languages – MySQL, Microsoft SQL server and Big Query
3. Visualization tools – Tableau and Looker
4. Statistical Analysis and visualization – R and Python

trainity.in

trainity.in