



An introduction to data and data analytics

What are data?

What are data?

Data are **collections of distinct values** represented by characters that are stored in or used by a computer.



AB

Letter



01

Digit



&<

Special
character

A **character**, abbreviated as **char**, is a single object used to represent text, numbers, and symbols.

Computers understand **ASCII** codes (numbers between 0 and 255) and translate characters into **binary** (0s and 1s) in order to process them. These forms can also be data.

```
Electrical, 1018445016, no power, -47072, -3771400, 15.07.2022
```

```
69 108 101 99 116 114 105 99 97 108 44 32 49 48 49 56 52 52 53
```

```
01000101 01101100 01100101 01100011 01110100 01110010 01101001  
01100011 01100001 01101100 00101100 00100000
```

Examples of data:

What are data?

Data can be of **types** beyond a single character.



Boolean



Text



Number

Boolean (logical true or false), text (strings), and numbers (float or integer) are commonly stored in **tabular format**. In greater quantities, these types are stored in database tables.



Image



Sound



Video

Image, sound, and video files are **more complex to store and analyse** in tabular format. They can be represented in a table through the storage path of the data file. They can also be stored in **database tables** as BLOBs (binary large objects).

What is a table?

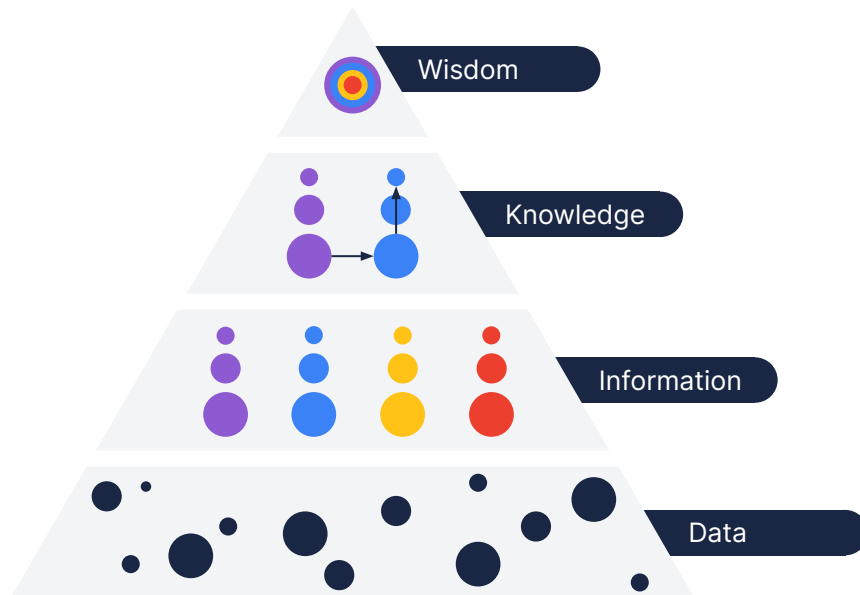
A data **table** is a grid with data arranged in labelled rows and columns.

Column names	Division	Location	Billing Class	Amount	Quantity	Number of contract accounts
Row 1	ELECTRICITY	ACACIA PARK	Commercial customers	R3 710,700	1 000,000 KVA	1
Row 2	ELECTRICITY	AMANDELRUG	Domestic customers	R1 927,420	768 000,000 W	1
Row 3	ELECTRICITY	ATHLONE	Commercial customers	R49 252,220	22 722,594 KWH	4
Row 4	ELECTRICITY	ATHLONE	Departmental	R8 329,520	0,924 KWH	5
	ELECTRICITY	BELLVILLE CBD	Departmental	R174 840,060		2
	ELECTRICITY	BURGUNDY	Not assigned	R83 140,040	44 586,560 KWH	1
	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6

The DIKW pyramid

The DIKW pyramid represents the **relationship** between **data**, **information**, **knowledge**, and **wisdom**.

- **Wisdom** is the ability to make **sound judgements** and **decisions** based on experience, **knowledge**, and understanding.
- **Knowledge** is the **understanding** or **comprehension of the information** we derive from data.
- **Information** is **processed** and **organised** data, often presented within a given context as to make it useful.
- **Data** are **raw**, **unprocessed**, and **unorganised** facts that are seemingly random and useless until they're processed and organised.



What is information?

Information, on the other hand, is **processed and organised data** presented in a meaningful context.



Information uses data as **building blocks to derive meaningful inferences** in accordance with its contextual requirement.



It facilitates **crucial decision-making** and is applicable in many industries such as finance, healthcare, retail, logistics, etc.



It is presented to stakeholders in the form of **spreadsheets, graphs, reports, payslips**, etc.

Data versus information

Data

- Generally **raw, unorganised**, and **unfiltered**.
- Comprise of all **elements or facts required** for analysis.
- **Independent** of information.
- **Meaningless** on its own.
- **Not sufficient for decision-making**.

Information

- **Refined, structured**, and **evolved** from raw data.
- Puts facts or elements into **context**.
- **Dependent** on data.
- **Meaningful**; derived from analysed and interpreted data.
- **Sufficient for decision-making**.

Examples:

A book contains **words** (data) and as one reads the book to completion, one **infers** what the book is about (information).

Every day a cashier **records sales items** (data). At the end of the month, the manager looks at the total number of sales and deduces if they made **a profit or loss** (information).

What are data used for?

Decision-making

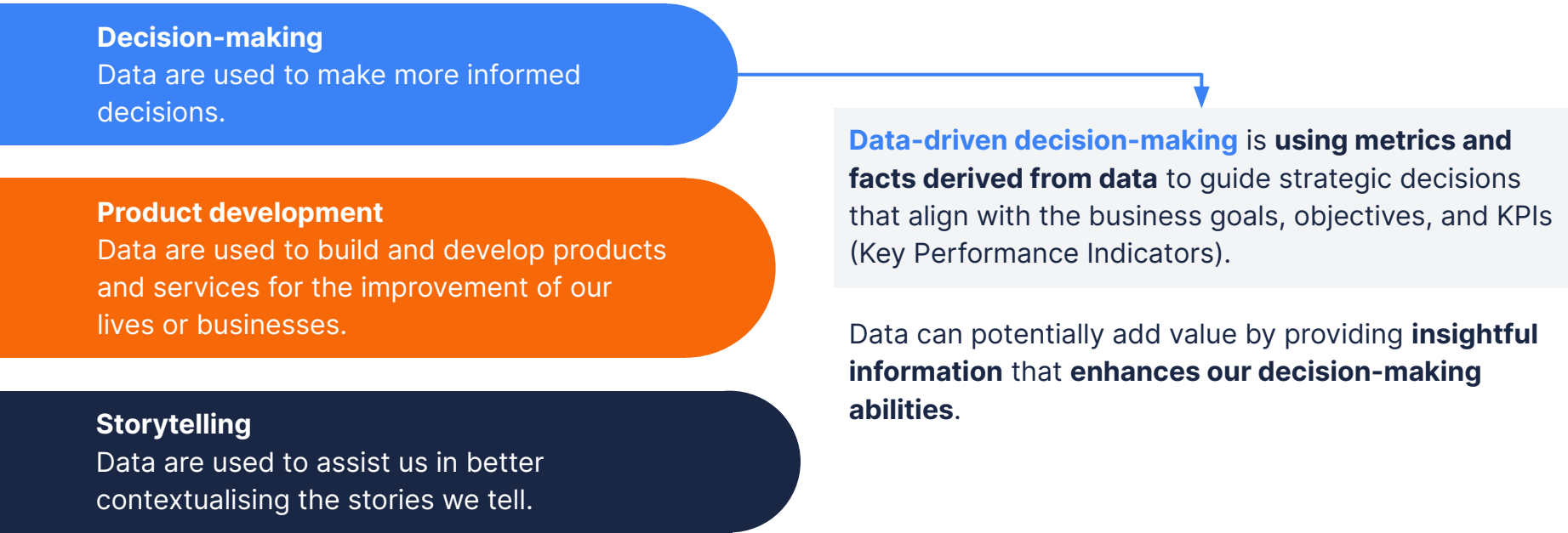
Data are used to make more informed decisions.

Product development

Data are used to build and develop products and services for the improvement of our lives or businesses.

Storytelling

Data are used to assist us in better contextualising the stories we tell.



Data-driven decision-making is using metrics and facts derived from data to guide strategic decisions that align with the business goals, objectives, and KPIs (Key Performance Indicators).

Data can potentially add value by providing **insightful information** that **enhances our decision-making abilities**.

Some examples of **real-world uses** include prediction of future sales, purchasing behaviours, fraud detection products, marketing campaign analysis, and improving customer acquisition and retention.

The data lifecycle

The data lifecycle describes the **stages** data have to go through **before meaningful insight** can be extracted to meet our **objectives**.

01. Data collection

Data are gathered from various sources such as sensors, surveys, and databases.

02. Data storage

Data are stored in a secure and accessible location.

03. Data processing

Data are cleaned and transformed, which may include statistical analysis.

04. Data visualisation

Data and insights are visualised using charts and dashboards so that it is easier to understand.

05. Data dissemination

Data, insights, and visualisations are shared with the relevant stakeholders.

06. Data archiving

Data are stored in a secure, long-term archive when it is no longer needed for current analysis.

07. Data destruction

Data are securely and permanently deleted when it is no longer needed, to ensure regulatory compliance.