

**INTRODUCTION**

Data science is an [interdisciplinary](https://en.wikipedia.org/wiki/Interdisciplinary) academic field that uses [statistics](https://en.wikipedia.org/wiki/Statistics), [scientific computing](https://en.wikipedia.org/wiki/Scientific_computing), [scientific methods](https://en.wikipedia.org/wiki/Scientific_method), processes, [algorithms](https://en.wikipedia.org/wiki/Algorithm) and systems to extract [knowledge](https://en.wikipedia.org/wiki/Knowledge) and insights from potentially noisy, structured, or [unstructured data](https://en.wikipedia.org/wiki/Unstructured_data). Basically data science is all about making meaning out of data.  It uses techniques and theories drawn from many fields within the context of [mathematics](https://en.wikipedia.org/wiki/Mathematics), statistics, [computer science](https://en.wikipedia.org/wiki/Computer_science), [information science](https://en.wikipedia.org/wiki/Information_science), and [domain knowledge](https://en.wikipedia.org/wiki/Domain_knowledge).

Data science and data analysis are both important disciplines in the field of [data management](https://en.wikipedia.org/wiki/Data_management) and analysis, but they differ in several key ways.

While both fields involve working with data, data science is more of an [interdisciplinary field](https://en.wikipedia.org/wiki/Interdisciplinary_field) that involves the application of statistical, computational, and [machine learning](https://en.wikipedia.org/wiki/Machine_learning) methods to extract insights from data and make predictions, while data analysis is more focused on the examination and interpretation of data to identify patterns and trends.

Data analysis and data science are distinct yet interconnected disciplines within the broader field of [data management](https://en.wikipedia.org/wiki/Data_management) and analysis.

Data analysis focuses on extracting insights and drawing conclusions from [structured data](https://en.wikipedia.org/wiki/Structured_data), while data science involves a more comprehensive approach that combines [statistical analysis](https://en.wikipedia.org/wiki/Statistical_analysis), computational methods, and [machine learning](https://en.wikipedia.org/wiki/Machine_learning) to extract insights, build predictive models, and drive data-driven [decision-making](https://en.wikipedia.org/wiki/Decision-making). Data analysis is the collection, transformation, andorganization of data to draw conclusions, make predictions, and drive informed decision-making. Any time you observe and evaluate something in the world, you’re collecting and analyzing data. Your analysis helps you find easier ways of doing things, identify patterns to save you time, and discover surprising new perspectives that can completely change the way you experience things.

Both fields play vital roles in leveraging the power of data to understand patterns, make informed decisions, and solve complex problems across various domains which includes but not limited to education, businesses, financial and health sectors.

**DATA ANALYTICS LIFE CYCLE**

In the data analysis process, there are six phases to go through to get the desired result. These are **ASK, PREPARE, PROCESS, ANALYZE, SHARE,** AND **ACT**.

**ASK**

First up, the analysts’ needs to define what the project would look like and what would qualify as a successful result. So, to determine these things, they askeffective questions and collaborate with leaders and managers who are interested in the outcome. Before you start collecting data, you need to first understand what you want to do with it. Take some time to think about a specific business problem you want to address or consider a hypothesis that could be solved with data. From there, you’ll create a set of measurable, clear, and concise goals that will help you solve this problem.

The question to be asked must be **SMART**, which means Specific, Measurable, Action-oriented, Relevant and Time bound.

The question should be **Specific** in the sense that it should define the purpose and context of the project.

**Measurable** and **action-oriented** means that the questions should be able to provide a range of significant importance and also provide a possibility for change if any.

**Relevant** means that the answers to the questions should be able to solve the problem at hand.

**Time-bound** means that the questions asked should have a specific period that it covers.

For example, **The Boateng And Sons Enterprise**, in Kasoa would like to understand the behavior of customers in order to know when to stock more goods and what particular goods to stock. .

Starting with a clear objective is an essential step in the data analysis process. By recognizing the business problem that you want to solve and setting well-defined goals, it’ll be way easier to decide on the data you need to collect and analyze.

These were some of the SMART questions concept being asked:

* What do customers buy more at the shop?
* What period do customers buy those goods and why?
* Will prices of these goods affect their purchases?
* Do these customers have other substitutes?

**PREPARE**

The second phase is to prepare the data. Preparing your data means collecting or using the data relevant to the problem you are trying to solve. In this phase, you need to understand the different metrics that you need for analysis. You will have to find out where this data is located, whether it is internal or external. Deciding the security measure is also very important.

Once there is an understanding of the problem, the next thing is to think about how to solve it. It is time to decide what data needs to be collected in order to answer the questions and how to organize it so that it is useful.

The following aspects should be considered:

* What metrics to measure? (Metrics are quantitative measurements). To answer these questions, there might be a need to also answer sub-questions (e.g., is our time-to-market competitive for product X? If not, what process improvements would help?).
* What factors should be considered?
* Where the data is located (files, database, external system, internal system)?
* If the data will be moved, how it will be stored and what are the needed security measures to protect that data.

Questions to ask you in this step:

* What needs to be figured out how to solve this problem?
* What would help to measure the outcome of any change to the problematic area?
* What research is needed?
* Where is the information held?

With reference to the example given in the **ASK** phase, the data analysts also made sure the Product and Marketing Managers understand how their data would be collected,stored, managed, and protected**.** Collecting and using data ethically is one of the responsibilities of data analysts. In order to maintain confidentiality and protect and store the data effectively, these were the steps taken:

* They restricted access to the data to a limited number of analysts.
* They cleaned the data to make sure it was complete, correct, and relevant. Certain data was aggregated and summarized without revealing individual responses.
* They uploaded raw data to an internal data warehouse for an additional layer of security.

The table below shows some of the items that were bought most in December 2023 at **Boateng and Sons Enterprise** and their various substitutes.

| **DATA FROM BOATENG & SONS ENTERPRISE, DECEMBER 2023** | | | |
| --- | --- | --- | --- |
| ***ITEMS*** | ***PRICES(GHS)*** | ***NO. OF ITEMS BOUGHT*** | ***SUBSTITUTES*** |
| MILO | 50 | 1500 | CHOCOLIM, RICHOCO |
| NIDO | 45 | 1350 | COWBELL, MIKSI |
| DIGESTIVE BISCUITS | 25 | 800 | SHE ME MA, PERK |
| SARDINE(TITUS) | 15 | 500 | LELE,GINO |
| MILK(IDEAL) | 10 | 2000 | CARNATION, NUNU |

**PROCESS**

The third phase is to process the data. Data processing is to find various inaccuracies, errors, inconsistencies in the data and get rid of them so that our primary business problem is not affected. You must clean the data so that the data is consistent and will not affect the credibility of the analysis.

When we start using the data, it might be a combination from different sources, or it might not be of the highest quality. A process known as data cleaning is the fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. What we aim to achieve is cleandata. There are plenty of tools, theories, and methods to use. So, during this step one might:

* Using proper tools to find incorrect and incomplete data.
* Removing inconsistencies in data. Sometimes there might be duplicated entries.
* One of the most important aspects to keep in mind - identifying whether your data is biased. Essentially, data that is biased will not be representative of the population or phenomenon of study, our issue we are trying to solve.

**Questions to ask yourself in this step:**

* Is the data source trustable and data quality high?
* What data errors or inaccuracies could occur within a given dataset.
* What is the best possible answer to the problem being solved?
* How to clean the data so the information is more consistent? (e.g. replace values with mean values, et cetera)

**ANALYZE**

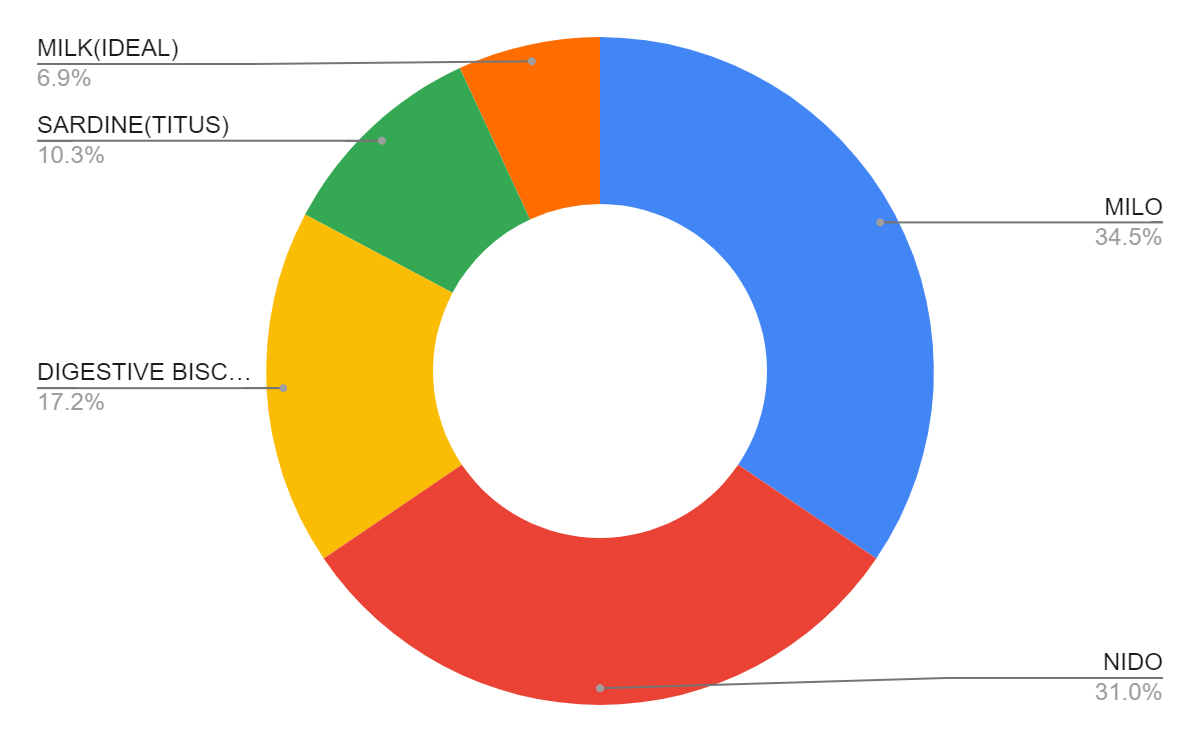
The fourth phase is to analyze the data. The primary goal in this phase is to find the relationships, trends, and patterns that will help you solve your business problem more accurately. In this phase, you will think analytically about the data. You will sort or format the data to understand it deeply. You will try to make sense of the data. You will try to find out what your data is telling you.

Next up is to make some conclusions based on the trustable data. Data Analyses is a skill that takes time to master, but over time the patterns will emerge faster and methods one uses will develop. Main concept is to think analytically about your data, be critical and be creative. There might be a need to sort and format the data to make it easier to process, make a Pivot table, or create awesome graphs! Remember it is a story that must unfold. Further processing might include:

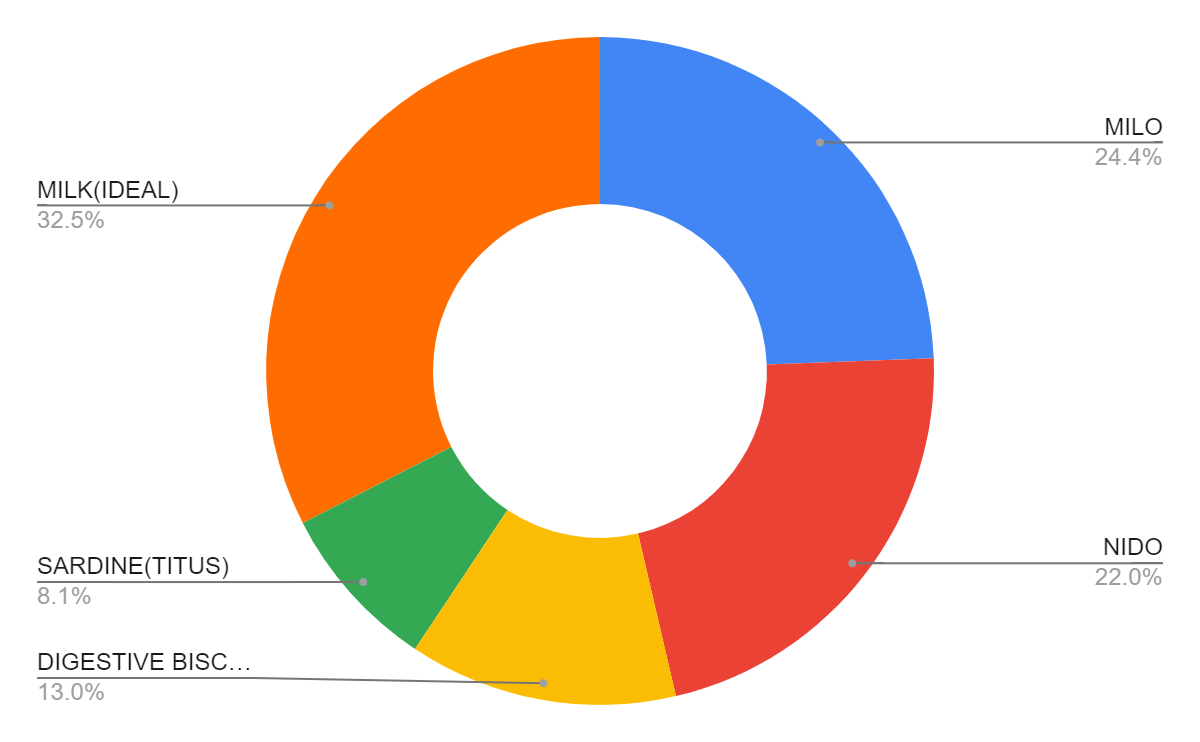
* Performing different calculations get additional metrics.
* Combining additional data attributes from a variety of sources to get a more comprehensive story.
* Create different views for the data. Like tables with your results, filter and pivot them.
* Make it visual if possible! Charts tell more than a thousand words.

**Questions to ask yourself in this step:**

* What story is my data telling me?
* Why can’t it be done?
* Will X (e.g. time, money, manpower or expertise) allow us to solve the issue?
* How will my data help me solve this problem?
* Who needs my company’s product or service?
* What type of person is most likely to use it?



**The first diagram shows the percentage of items based on their values whereas the second displays percentage of items based on their corresponding values per the data provided by Boateng & Sons Enterprise.**

**SHARE**

The fifth phase is to share your data findings. You can do this with the help of visualization because putting information in the image can help people understand the analysis easily. You will think of the various graphs and plots to convey your findings more quickly.

This phase is about getting additional opinions about the findings. This will significantly help to improve the results and ensure that main aspects were taken into account. As there are many ways to share the finding each person has their preference and so does each company. However, many studies reassure that with clear and enticing visuals of the analysis results, the story is better understood. Sharing will certainly help with:

1. **Making better decisions**: The feedback will help to answer the questions that initially were not thought of.
2. **Making more informed decisions:** Feedback will not be merely critical, but also suggestions and additional information on the matter.
3. **Improve the general outcome**: From one angle, the decision will most likely be more informed and better, but also the transparency will grant that there is more support to the findings.

Questions to ask yourself in this step:

* How can I make what I present to the stakeholders engaging and easy to understand?
* What would help me understand this if I were the listener?
* What makes data visualization good?

The analysts for The Boateng and Sons,

* Shared the report with product and marketing managers who helped in getting the data.
* Presented the results to the managers to make sure they had the full picture.
* Asked the managers to personally deliver the results to their teams.

This process gave managers an opportunity to **communicate the results** with the right context. As a result, they could have productive team conversations about next steps to improve customer experience and increase sales.

**ACT**

The sixth phase of the data is to act. In this phase, we will use everything we have learned from our analysis and act upon it. You will provide recommendations to the stakeholder on how to solve the business problem and help them make a good decision.

No analysis conclusion should remain to collect dust on a shelf! Rather some action should be taken. Taking the results and depending on the problem statement, recommendations for further actions can be made. And once the recommendations are ready, the actual decision can be made! Not necessarily is the conductor of the analysis the one to make a decision, it could also mean providing the decision-makers(stakeholders) with recommendations based on the findings so they can make data-driven decisions. But the key here is data-driven decisions.

Questions to ask yourself in this step:

* How can the feedback received during the sharing phase (step 5) be used to meet the stakeholder’s needs and expectations?
* What potential solutions to the outlined problem could there be?
* Is this problem worth solving? (Yes, that is also a potential outcome)

The last stage of the process for the team of analysts was to work with leaders within their company and decide how best to implement changes and takeactions based on the findings.

These were their recommendations:

* Standardize the hiring and evaluation process for employees based on the most efficient and transparent practices.
* Conduct the same survey annually and compare results with those from the previous year.

A year later, the same survey was distributed to employees. Analysts anticipated that a comparison between the two sets of results would indicate that the action plan worked. Turns out, the changes improved the retention rate for new employees and the actions taken by leaders were successful!

One of the many things that makes data analytics so exciting is that the problems are always different, the solutions need creativity, and the impact on others can be great — even life-changing or life-saving.

1. **Ask**: Business Challenge/Objective/Question
2. **Prepare**: Data generation, collection, storage, and data management
3. **Process**: Data cleaning/data integrity
4. **Analyze**: Data exploration, visualization, and analysis
5. **Share**: Communicating and interpreting results
6. **Act**:  Putting your insights to work to solve the problem