DATA ANALYTICS REPORT

What is data analytics?

Data analytics is a broad term that encompasses many diverse types of data analysis. Any type of information can be subjected to data analytics techniques to get insight that can be used to improve things. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

For example, manufacturing companies often record the runtime, downtime, and work queue for various machines and then analyze the data to better plan the workloads so the machines operate closer to peak capacity.

Data analytics can do much more than point out deep problems in production. Gaming companies use data analytics to set reward schedules for players that keep the majority of players active in the game. Content companies use many of the same data analytics to keep you clicking, watching, or re-organizing content to get another view.

About data analytics subset

The subset of data analytics is data analysis .

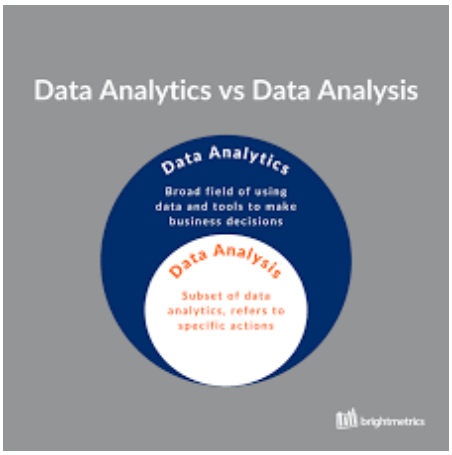
Consider data analysis one slice of the data analytics pie. Data analysis consists of cleaning, transforming, modeling, and questioning data to find useful information. (It's generally agreed that other slices are other activities, from collection to storage to visualization.)

The act of data analysis is usually limited to a single, already prepared dataset. You'll inspect, arrange, and question the data. Today, in the 2020s, a software or "machine" usually does a first round of analysis, often directly in one of your databases or tools. But this is augmented by a human who investigates and interrogates the data with more context.

When you're done analyzing a dataset, you'll turn to other data analytics activities to:

- Give others access to the data
- Present the data (ideally with data visualization or storytelling)
- Suggest actions to take based on the data

A vital point of data analysis is that the analysis already captures data, meaning data from the past.


Data Analytics vs Data Analysis
Data Analytics
Broad field of using data and tools to make business decisions
Data Analysis
Subset of data analytics, refers to specific actions
brightmetrics

Types of data analytics

The four types of data analysis are:

- Descriptive Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

Below, we will introduce each type and give examples of how they are utilized in business.

Descriptive Analytics

The first type of data analysis is descriptive analysis. It is at the foundation of all data insight. It is the simplest and most common use of data in business today. Descriptive analysis answers the “what happened” by summarizing past data, usually in the form of dashboards.

The biggest use of descriptive analysis in business is to track Key Performance Indicators (KPIs). KPIs describe how a business is performing based on chosen benchmarks.

Business applications of descriptive analysis include:

- KPI dashboards
- Monthly revenue reports
- Sales leads overview

Diagnostic Analytics

After asking the main question of “what happened”, the next step is to dive deeper and ask why did it happen? This is where diagnostic analysis comes in.

Diagnostic analysis takes the insights found from descriptive analytics and drills down to find the causes of those outcomes. Organizations make use of this type of analytics as it creates more connections between data and identifies patterns of behavior.

A critical aspect of diagnostic analysis is creating detailed information. When new problems arise, it is possible you have already collected certain data pertaining to the issue. By already having the data at your disposal, it ends having to repeat work and makes all problems interconnected.

Business applications of diagnostic analysis include:

- A freight company investigating the cause of slow shipments in a certain region
- A SaaS company drilling down to determine which marketing activities increased trials

Predictive Analytics

Predictive analysis attempts to answer the question “what is likely to happen”. This type of analytics utilizes previous data to make predictions about future outcomes.

This type of analysis is another step up from the descriptive and diagnostic analyses. Predictive analysis uses the data we have summarized to make logical predictions of the outcomes of events. This analysis relies on statistical modeling, which requires added technology and manpower to forecast. It is also important to understand that forecasting is only an estimate; the accuracy of predictions relies on quality and detailed data.

While descriptive and diagnostic analysis are common practices in business, predictive analysis is where many organizations begin show signs of difficulty. Some companies do not have the manpower to implement predictive analysis in every place they desire. Others are not yet willing to invest in analysis teams across every department or not prepared to educate current teams.

Business applications of predictive analysis include:

- Risk Assessment
- Sales Forecasting
- Using customer segmentation to determine which leads have the best chance of converting
- Predictive analytics in customer success teams

Prescriptive Analytics

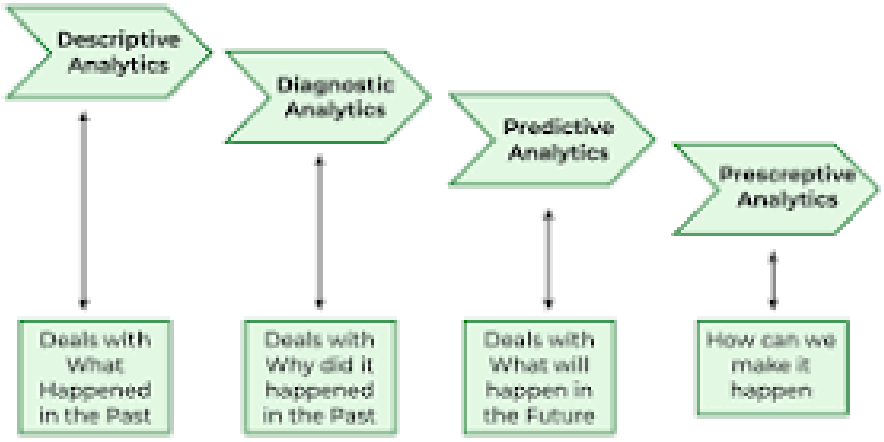
The final type of data analysis is the most sought after, but few organizations are truly equipped to perform it. Prescriptive Analytics is the frontier of data analysis, combining take in a current problem or decision.

Prescriptive analysis utilizes state of the art technology and data practices. It is a huge organizational commitment and companies must be sure that they are ready and willing to put forth the effort and resources.

Artificial Intelligence (AI) is a perfect example of prescriptive analytics. AI systems consume a large amount of data to continuously learn and use this information to make informed decisions. Well-designed AI systems are capable of communicating these decisions and even putting those decisions into action. Business processes can be performed and optimized daily without a human doing anything with artificial intelligence.

Currently, most of the big data-driven companies (Apple, Facebook, Netflix, etc.) are utilizing prescriptive analytics and AI to improve decision making.

For other organizations, the jump to predictive and prescriptive analytics can be insurmountable. As technology continues to improve and more professionals are educated in data, we will see more companies entering the data-driven realm.



Applications of Data Analytics

- ***Retail*** Understanding customer needs and buying habits to predict trends, launch new products, and boost sales.
- ***Healthcare*** Analysing patient data for lifesaving diagnosis and treatment. Data analytics also helps in developing new drugs.
- ***Manufacturing*** Data analytics helps in solving complex supply chain issues, labour constraints, and breakdown of equipment.
- ***Banking*** Pointing out probable loan defaulters and detecting frauds
- ***Logistics*** Developing new business models and optimizing routes.

Tools of Data Analytics

1. Microsoft Excel

Excel at a glance:

- **Type of tool:** Spreadsheet software.
- **Availability**: Commercial.
- **Mostly used for:** Data wrangling and reporting.
- **Pros:** Widely-used, with lots of useful functions and plug-ins.
- **Cons:** Cost, calculation errors, poor at handling big data.

Excel: the world's best-known spreadsheet software. What's more, it features calculations and graphing functions that are ideal for data analysis. Whatever your specialism, and no matter what other software you might need, Excel is a staple in the field. Its invaluable built-in features include pivot tables (for sorting or totaling data) and form creation tools.

It has limitations though. For instance, it runs very slowly with big datasets and tends to approximate large numbers, leading to inaccuracies. Nevertheless, it's an important and powerful data analysis tool, and with many plug-ins available, you can easily bypass Excel's shortcomings. Get

2. Python

Python at a glance:

- **Type of tool:** Programming language.

- **Availability:** Open-source, with thousands of free libraries.
- **Used for:** Everything from data scraping to analysis and reporting.
- **Pros:** Easy to learn, highly versatile, widely-used.
- **Cons:** Memory intensive—doesn't execute as fast as some other languages.

A programming language with a wide range of uses, Python is a must-have for any data analyst. Unlike more complex languages, it focuses on readability, and its general popularity in the tech field means many programmers are already familiar with it. Python is also extremely versatile; it has a huge range of resource libraries suited to a variety of different data analytics tasks. For example, the **NumPy** and **pandas** libraries are great for streamlining highly computational tasks, as well as supporting general data manipulation.

Libraries like **Beautiful Soup** and **Scrapy** are used to scrape data from the web, while **Matplotlib** is excellent for data visualization and reporting. Python's main drawback is its speed—it is memory intensive and slower than many languages. In overall though, if you're building software from scratch, Python's benefits far outweigh its drawbacks.

3. R

R at a glance:

- **Type of tool:** Programming language.
- **Availability:** Open-source.
- **Mostly used for:** Statistical analysis and data mining.

- **Pros:** Platform independent, highly compatible, lots of packages.
- **Cons:** Slower, less secure, and more complex to learn than Python.

R, like Python, is a popular open-source programming language. It is commonly used to create statistical/data analysis software. R's syntax is more complex than Python and the learning curve is steeper. However, it was built specifically to deal with heavy statistical computing tasks and is very popular for data visualization. A bit like Python, R also has a network of freely available code, called **CRAN** (the Comprehensive R Archive Network), which offers 10,000+ packages.

It integrates well with other languages and systems (including **big data** software) and can call on code from languages like C, C++, and FORTRAN. On the downside, it has poor memory management, and while there is a good community of users to call on for help, R has no dedicated support team. But there is an excellent R-specific integrated development environment (IDE) called **RStudio**, which is always a bonus!

4. Jupyter Notebook

Jupyter Notebook at a glance:

- **Type of tool:** Interactive authoring software.
- **Availability:** Open-source.
- **Mostly used for:** Sharing code, creating tutorials, presenting work.
- **Pros:** Great for showcasing, language-independent.
- **Cons:** Not self-contained, nor great for collaboration.

Jupyter Notebook is an open-source web application that allows you to create interactive documents. These combine live code, equations, visualizations, and narrative text. Imagine something a bit like a Microsoft word document, only far more interactive, and designed specifically for data analytics! As a data analytics tool, it's great for showcasing work: Jupyter Notebook runs in the browser and supports over 40 languages, including Python and R. It also integrates with big data analysis tools, like Apache Spark (see below) and offers various outputs from HTML to images, videos, and more.

But as with every tool, it has its limitations. Jupyter Notebook documents have poor version control, and tracking changes is not intuitive. This means it's not the best place for development and analytics work (you should use a dedicated IDE for these) and it isn't well suited to collaboration. Since it isn't self-contained, this also means you have to provide any extra assets (e.g. libraries or runtime systems) to anybody you're sharing the document with. But for presentation and tutorial purposes, it remains an invaluable data science and data analytics tool.

5. Apache Spark

Apache Spark at a glance:

- **Type of tool:** Data processing framework.
- **Availability:** Open-source.
- **Mostly used for:** big data processing, machine learning.
- **Pros:** Fast, dynamic, easy to use.
- **Cons:** No file management system, rigid user interface.

Apache Spark is a software framework that allows data analysts and data scientists to quickly process vast data sets. It was first developed in 2012 before being donated to the non-profit Apache Software Foundation. Designed to analyze unstructured big data, Spark distributes computationally heavy analytics tasks across many computers.

While other similar frameworks exist (for example, **Apache Hadoop**) Spark is exceptionally fast. By using RAM rather than local memory, it is around 100x faster than Hadoop.

It even has a library of machine learning algorithms, **MLlib**, including classification, regression, and clustering algorithms, to name a few. On the downside, consuming so much memory means Spark is computationally expensive. It also lacks a file management system, so it usually needs integration with other software, i.e. Hadoop.

6. SAS

SAS at a glance:

- **Type of tool:** Statistical software suite.
- **Availability:** Commercial.
- **Mostly used for:** Business intelligence, multivariate, and predictive analysis.
- **Pros:** Easily accessible, business-focused, good user support.
- **Cons:** High cost, poor graphical representation.

SAS (which stands for Statistical Analysis System) is a popular commercial suite of business intelligence and data analysis tools. It was developed by the

SAS Institute in the 1960s and has evolved ever since. Its main use today is for profiling customers, reporting, data mining, and predictive modeling. Created for an enterprise market, the software is generally more robust, versatile, and easier for large organizations to use. This is because they tend to have varying levels of in-house programming expertise.

But as a commercial product, SAS comes with a hefty price tag. Nevertheless, with cost comes benefits; it regularly has new modules added, based on customer demand. Although it has fewer of these than say, Python libraries, they are highly focused. For instance, it offers modules for specific uses such as anti-money laundering and analytics for the Internet of Things.

7. Microsoft Power BI

Power BI at a glance:

- **Type of tool:** Business analytics suite.

- **Availability:** Commercial software (with a free version available).
- **Mostly used for:** Everything from data visualization to predictive analytics.
- **Pros:** Great data connectivity, regular updates, good visualizations.
- **Cons:** Clunky user interface, rigid formulas, data limits (in the free version).

At less than a decade old, Power BI is a relative newcomer to the market of data analytics tools. It began life as an Excel plug-in but was redeveloped in the early 2010s as a standalone suite of business data analysis tools. Power BI allows users to create interactive visual reports and dashboard, with a minimal learning curve. Its main selling point is its great data connectivity—it operates seamlessly with Excel (as you'd expect, being a Microsoft product) but also text files, SQL server, and cloud sources, like Google and Facebook analytics.

It also offers strong data visualization but has room for improvement in other areas. For example, it has quite a bulky user interface, rigid formulas, and the proprietary language (Data Analytics Expressions, or 'DAX') is not that user-friendly. It does offer several subscriptions though, including a free one. This is great if you want to get to grips with the tool, although the free version does have drawbacks—the main limitation being the low data limit (around 2GB).

8. Tableau

Tableau at a glance:

- **Type of tool:** Data visualization tool.
- **Availability:** Commercial.

- **Mostly used for:** Creating data dashboards and worksheets.
- **Pros:** Great visualizations, speed, interactivity, mobile support.
- **Cons:** Poor version control, no data pre-processing.

If you're looking to create interactive visualizations and dashboards without extensive coding expertise, Tableau is one of the best commercial data analysis tools available. The suite handles large amounts of data better than many other BI tools, and it is very simple to use. It has a visual drag and drop interface (another definite advantage over many other data analysis tools). However, because it has no scripting layer, there's a limit to what Tableau can do. For instance, it's not great for pre-processing data or building more complex calculations.

While it does contain functions for manipulating data, these aren't great. As a rule, you'll need to carry out scripting functions using Python or R before importing your data into Tableau. But its visualization is pretty top-notch, making it very popular despite its drawbacks. Furthermore, it's mobile-ready. As a data analyst, mobility might not be your priority, but it's nice to have if you want to dabble on the move!

9. KNIME

KNIME at a glance:

- **Type of tool:** Data integration platform.
- **Availability:** Open-source.
- **Mostly used for:** Data mining and machine learning.

- **Pros:** Open-source platform that is great for visually-driven programming.
- **Cons:** Lacks scalability, and technical expertise is needed for some functions.

Last on our list is KNIME (Konstanz Information Miner), an open-source, cloud-based, data integration platform. It was developed in 2004 by software engineers at Konstanz University in Germany. Although first created for the pharmaceutical industry, KNIME's strength in accruing data from numerous sources into a single system has driven its application in other areas. These include customer analysis, business intelligence, and machine learning.

Its main draw (besides being free) is its usability. A drag-and-drop graphical user interface (GUI) makes it ideal for visual programming. This means users don't need a lot of technical expertise to create data workflows. While it claims to support the full range of data analytics tasks, in reality, its strength lies in data mining. Though it offers in-depth statistical analysis too, users will benefit from some knowledge of Python and R. Being open-source, KNIME is very flexible and customizable to an organization's needs—without heavy costs. This makes it popular with smaller businesses, who have limited budgets.

Now that we've checked out all of the data analysis tools, let's see how to choose the right one for your business needs.

How to choose a data analysis tool

Alright, so you've got your data ready to go, and you're looking for the perfect tool to analyze it with. How do you find the one that's right for your organization?

First, consider that there's no one singular data analytics tool that will address all the data analytics issues you may have. When looking at this list, you may look at one tool for most of your needs, but require the use of a secondary tool for smaller processes.

Second, consider the business needs of your organization and figure out exactly who will need to make use of the data analysis tools. Will they be used primarily by fellow data analysts or scientists, non-technical users who require an interactive and intuitive interface—or both? Many tools on this list will cater to both types of users.

Third, consider the tool's data modeling capabilities. Does the tool have these capabilities, or will you need to use SQL or another tool to perform data modeling prior to analysis?

Fourth—and finally! —consider the practical aspect of price and licensing. Some of the options are totally free or have some free-to-use features (but will require licensing for the full product). Some data analysis tools will be offered on a subscription or licensing basis. In this case, you may need to consider the number of users required or—if you're looking on solely a project-to-project basis—the potential length of the subscription.