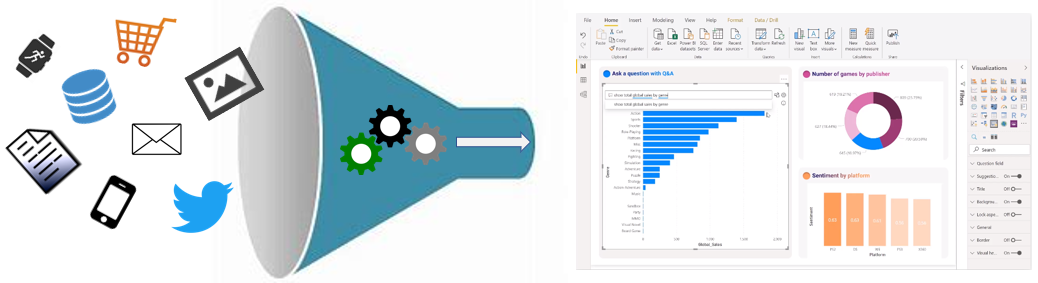
**Introduction**

Completed100 XP

* 5 minutes

As a data analyst, you are on a journey. Think about all the data that is being generated each day and that is available in an organization, from transactional data in a traditional database, telemetry data from services that you use, to signals that you get from different areas like social media.

[](https://learn.microsoft.com/en-us/training/modules/data-analytics-microsoft/media/abundance-data-ss.png#lightbox)

For example, today's retail businesses collect and store massive amounts of data that track the items you browsed and purchased, the pages you've visited on their site, the aisles you purchase products from, your spending habits, and much more.

With data and information as the most strategic asset of a business, the underlying challenge that organizations have today is understanding and using their data to positively affect change within the business. Businesses continue to struggle to use their data in a meaningful and productive way, which impacts their ability to act.

A retail business should be able to use their vast amounts of data and information in such a way that impacts the business, including:

* Tracking inventory
* Identifying purchase habits
* Detecting user trends and patterns
* Recommending purchases
* Determining price optimizations
* Identifying and stopping fraud

Additionally, you might be looking for daily/monthly sale patterns. Common data segments that you might want to examine include day-over-day, week-over-week, and month-over-month so that you can compare how sales have been to where they were in the same week last year, for example.

The key to unlocking this data is being able to tell a story with it. In today's highly competitive and fast-paced business world, crafting reports that tell that story is what helps business leaders take action on the data. Business decision makers depend on an accurate story to drive better business decisions. The faster a business can make precise decisions, the more competitive they will be and the better advantage they will have. Without the story, it is difficult to understand what the data is trying to tell you.

However, having data alone is not enough. You need to be able to act on the data to affect change within the business. That action could involve reallocating resources within the business to accommodate a need, or it could be identifying a failing campaign and knowing when to change course. These situations are where telling a story with your data is important.

The underlying challenge that businesses face today is understanding and using their data in such a way that impacts their business and ultimately their bottom line. You need to be able to look at the data and facilitate trusted business decisions. Then, you need the ability to look at metrics and clearly understand the meaning behind those metrics.

This requirement might seem daunting, but it's a task that you can accomplish. Your first step is to partner with data experts within your organization, such as data engineers and data scientists, to help get the data that you need to tell that story. Ask these experts to participate in that data journey with you.

Your journey of telling a story with data also ties into building that data culture within your organization. While telling the story is important, *where* that story is told is also crucial, ensuring that the story is told to the right people. Also, make sure that people can discover the story, that they know where to find it, and that it is part of the regular interactions.

Data analysis exists to help overcome these challenges and pain points, ultimately assisting businesses in finding insights and uncovering hidden value in troves of data through storytelling. As you read on, you will learn how to use and apply analytical skills to go beyond a single report and help impact and influence your organization by telling stories with data and driving that data culture.

**Next unit: Overview of data analysis**

# Overview of data analysis

Completed100 XP

* 6 minutes

Before data can be used to tell a story, it must be run through a process that makes it usable in the story. Data analysis is the process of identifying, cleaning, transforming, and modeling data to discover meaningful and useful information. The data is then crafted into a story through reports for analysis to support the critical decision-making process.

[](https://learn.microsoft.com/en-us/training/modules/data-analytics-microsoft/media/reporting-trusted-data-ss.png#lightbox)

As the world becomes more data-driven, storytelling through data analysis is becoming a vital component and aspect of large and small businesses. It is the reason that organizations continue to hire data analysts.

Data-driven businesses make decisions based on the story that their data tells, and in today's data-driven world, data is not being used to its full potential, a challenge that most businesses face. Data analysis is, and should be, a critical aspect of all organizations to help determine the impact to their business, including evaluating customer sentiment, performing market and product research, and identifying trends or other data insights.

While the process of data analysis focuses on the tasks of cleaning, modeling, and visualizing data, the concept of data analysis and its importance to business should not be understated. To analyze data, core components of analytics are divided into the following categories:

* Descriptive
* Diagnostic
* Predictive
* Prescriptive
* Cognitive

## Descriptive analytics

Descriptive analytics help answer questions about what has happened based on historical data. Descriptive analytics techniques summarize large datasets to describe outcomes to stakeholders.

By developing key performance indicators (KPIs), these strategies can help track the success or failure of key objectives. Metrics such as return on investment (ROI) are used in many industries, and specialized metrics are developed to track performance in specific industries.

An example of descriptive analytics is generating reports to provide a view of an organization's sales and financial data.

## Diagnostic analytics

Diagnostic analytics help answer questions about why events happened. Diagnostic analytics techniques supplement basic descriptive analytics, and they use the findings from descriptive analytics to discover the cause of these events. Then, performance indicators are further investigated to discover why these events improved or became worse. Generally, this process occurs in three steps:

1. Identify anomalies in the data. These anomalies might be unexpected changes in a metric or a particular market.
2. Collect data that's related to these anomalies.
3. Use statistical techniques to discover relationships and trends that explain these anomalies.

## Predictive analytics

Predictive analytics help answer questions about what will happen in the future. Predictive analytics techniques use historical data to identify trends and determine if they're likely to recur. Predictive analytical tools provide valuable insight into what might happen in the future. Techniques include a variety of statistical and machine learning techniques such as neural networks, decision trees, and regression.

## Prescriptive analytics

Prescriptive analytics help answer questions about which actions should be taken to achieve a goal or target. By using insights from prescriptive analytics, organizations can make data-driven decisions. This technique allows businesses to make informed decisions in the face of uncertainty. Prescriptive analytics techniques rely on machine learning as one of the strategies to find patterns in large datasets. By analyzing past decisions and events, organizations can estimate the likelihood of different outcomes.

## Cognitive analytics

Cognitive analytics attempt to draw inferences from existing data and patterns, derive conclusions based on existing knowledge bases, and then add these findings back into the knowledge base for future inferences, a self-learning feedback loop. Cognitive analytics help you learn what might happen if circumstances change and determine how you might handle these situations.

Inferences aren't structured queries based on a rules database; rather, they're unstructured hypotheses that are gathered from several sources and expressed with varying degrees of confidence. Effective cognitive analytics depend on machine learning algorithms, and will use several natural language processing concepts to make sense of previously untapped data sources, such as call center conversation logs and product reviews.

### Example

By enabling reporting and data visualizations, a retail business uses descriptive analytics to look at patterns of purchases from previous years to determine what products might be popular next year. The company might also look at supporting data to understand why a particular product was popular and if that trend is continuing, which will help them determine whether to continue stocking that product.

A business might determine that a certain product was popular over a specific timeframe. Then, they can use this analysis to determine whether certain marketing efforts or online social activities contributed to the sales increase.

An underlying facet of data analysis is that a business needs to trust its data. As a practice, the data analysis process will capture data from trusted sources and shape it into something that is consumable, meaningful, and easily understood to help with the decision-making process. Data analysis enables businesses to fully understand their data through data-driven processes and decisions, allowing them to be confident in their decisions.

As the amount of data grows, so does the need for data analysts. A data analyst knows how to organize information and distill it into something relevant and comprehensible. A data analyst knows how to gather the right data and what to do with it, in other words, making sense of the data in your data overload.

## Next unit: Roles in data

**Roles in data**

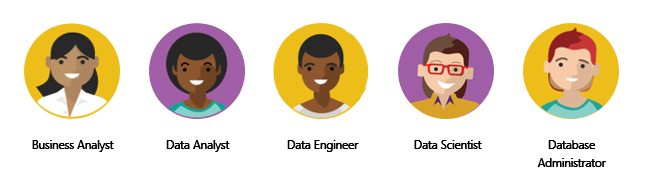
Completed100 XP

* 8 minutes

Telling a story with the data is a journey that usually doesn't start with you. The data must come from somewhere. Getting that data into a place that is usable by you takes effort that is likely out of your scope, especially in consideration of the enterprise.

Today's applications and projects can be large and intricate, often involving the use of skills and knowledge from numerous individuals. Each person brings a unique talent and expertise, sharing in the effort of working together and coordinating tasks and responsibilities to see a project through from concept to production.

In the recent past, roles such as business analysts and business intelligence developers were the standard for data processing and understanding. However, excessive expansion of the size and different types of data has caused these roles to evolve into more specialized sets of skills that modernize and streamline the processes of data engineering and analysis.

[](https://learn.microsoft.com/en-us/training/modules/data-analytics-microsoft/media/roles-data-ss.png#lightbox)

The following sections highlight these different roles in data and the specific responsibility in the overall spectrum of data discovery and understanding:

* Business analyst
* Data analyst
* Data engineer
* Data scientist
* Database administrator

**Business analyst**

While some similarities exist between a data analyst and business analyst, the key differentiator between the two roles is what they do with data. A business analyst is closer to the business and is a specialist in interpreting the data that comes from the visualization. Often, the roles of data analyst and business analyst could be the responsibility of a single person.

**Data analyst**

A data analyst enables businesses to maximize the value of their data assets through visualization and reporting tools such as Microsoft Power BI. Data analysts are responsible for profiling, cleaning, and transforming data. Their responsibilities also include designing and building scalable and effective data models, and enabling and implementing the advanced analytics capabilities into reports for analysis. A data analyst works with the pertinent stakeholders to identify appropriate and necessary data and reporting requirements, and then they are tasked with turning raw data into relevant and meaningful insights.

A data analyst is also responsible for the management of Power BI assets, including reports, dashboards, workspaces, and the underlying datasets that are used in the reports. They are tasked with implementing and configuring proper security procedures, in conjunction with stakeholder requirements, to ensure the safekeeping of all Power BI assets and their data.

Data analysts work with data engineers to determine and locate appropriate data sources that meet stakeholder requirements. Additionally, data analysts work with the data engineer and database administrator to ensure that the analyst has proper access to the needed data sources. The data analyst also works with the data engineer to identify new processes or improve existing processes for collecting data for analysis.

**Data engineer**

Data engineers provision and set up data platform technologies that are on-premises and in the cloud. They manage and secure the flow of structured and unstructured data from multiple sources. The data platforms that they use can include relational databases, nonrelational databases, data streams, and file stores. Data engineers also ensure that data services securely and seamlessly integrate across data platforms.

Primary responsibilities of data engineers include the use of on-premises and cloud data services and tools to ingest, egress, and transform data from multiple sources. Data engineers collaborate with business stakeholders to identify and meet data requirements. They design and implement solutions.

While some alignment might exist in the tasks and responsibilities of a data engineer and a database administrator, a data engineer's scope of work goes well beyond looking after a database and the server where it's hosted and likely doesn't include the overall operational data management.

A data engineer adds tremendous value to business intelligence and data science projects. When the data engineer brings data together, often described as data wrangling, projects move faster because data scientists can focus on their own areas of work.

As a data analyst, you would work closely with a data engineer in making sure that you can access the variety of structured and unstructured data sources because they will support you in optimizing data models, which are typically served from a modern data warehouse or data lake.

Both database administrators and business intelligence professionals can transition to a data engineer role; they need to learn the tools and technology that are used to process large amounts of data.

**Data scientist**

Data scientists perform advanced analytics to extract value from data. Their work can vary from descriptive analytics to predictive analytics. Descriptive analytics evaluate data through a process known as exploratory data analysis (EDA). Predictive analytics are used in machine learning to apply modeling techniques that can detect anomalies or patterns. These analytics are important parts of forecast models.

Descriptive and predictive analytics are only partial aspects of data scientists' work. Some data scientists might work in the realm of deep learning, performing iterative experiments to solve a complex data problem by using customized algorithms.

Anecdotal evidence suggests that most of the work in a data science project is spent on data wrangling and feature engineering. Data scientists can speed up the experimentation process when data engineers use their skills to successfully wrangle data.

On the surface, it might seem that a data scientist and data analyst are far apart in the work that they do, but this conjecture is untrue. A data scientist looks at data to determine the questions that need answers and will often devise a hypothesis or an experiment and then turn to the data analyst to assist with the data visualization and reporting.

**Database administrator**

A database administrator implements and manages the operational aspects of cloud-native and hybrid data platform solutions that are built on Microsoft Azure data services and Microsoft SQL Server. A database administrator is responsible for the overall availability and consistent performance and optimizations of the database solutions. They work with stakeholders to identify and implement the policies, tools, and processes for data backup and recovery plans.

The role of a database administrator is different from the role of a data engineer. A database administrator monitors and manages the overall health of a database and the hardware that it resides on, whereas a data engineer is involved in the process of data wrangling, in other words, ingesting, transforming, validating, and cleaning data to meet business needs and requirements.

The database administrator is also responsible for managing the overall security of the data, granting and restricting user access and privileges to the data as determined by business needs and requirements.

**Next unit: Tasks of a data analyst**