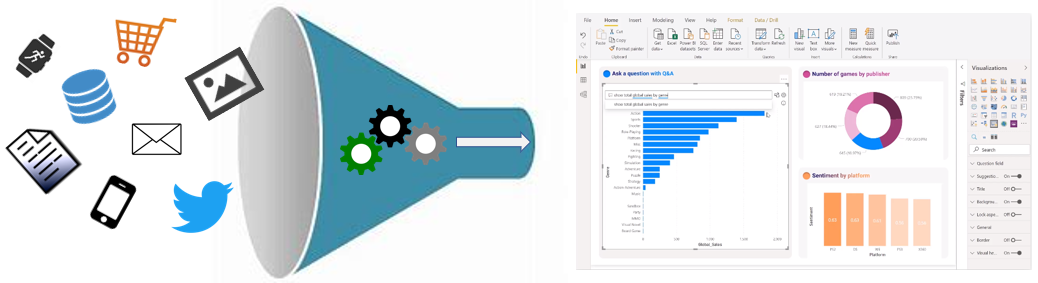
# **Introduction**

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.



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.

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.

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.

# **Overview of data analysis**

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.



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 semantic models 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 semantic models. 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.

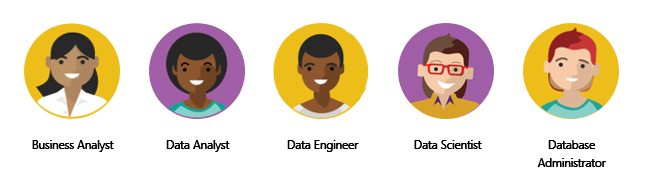
As the amount of data grows, so does the need for data analysts. A data analyst knows how to organize information and distill it into something relevant and comprehensible. A data analyst knows how to gather the right data and what to do with it, in other words, making sense of the data in your data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

# **Roles in data**

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.



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 semantic 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 semantic models 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 semantic 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.

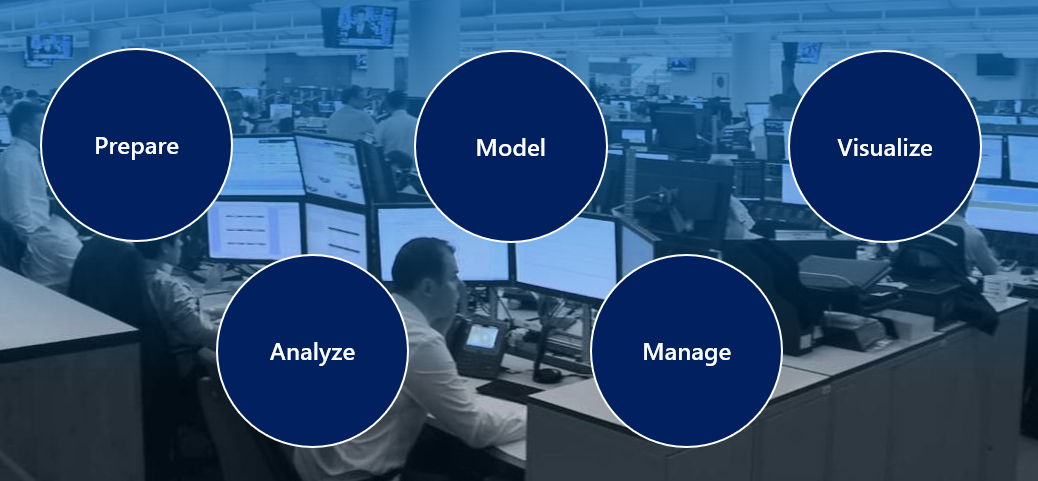
The role of a database administrator is different from the role of a data engineer. A database administrator monitors and manages the overall health of a database and the hardware that it resides on, whereas a data engineer is involved in the process of data wrangling, in other words, ingesting, transforming,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

# **Tasks of a data analyst**

A data analyst is one of several critical roles in an organization, who help uncover and make sense of information to keep the company balanced and operating efficiently. Therefore, it's vital that a data analyst clearly understands their responsibilities and the tasks that are performed on a near-daily basis. Data analysts are essential in helping organizations gain valuable insights into the expanse of data that they have, and they work closely with others in the organization to help reveal valuable information.

The following figure shows the five key areas that you'll engage in during the data analysis process.



## **Prepare**

As a data analyst, you'll likely divide most of your time between the prepare and model tasks. Deficient or incorrect data can have a major impact that results in invalid reports, a loss of trust, and a negative effect on business decisions, which can lead to loss in revenue, a negative business impact, and more.

Before a report can be created, data must be prepared. Data preparation is the process of profiling, cleaning, and transforming your data to get it ready to model and visualize.

Data preparation is the process of taking raw data and turning it into information that is trusted and understandable. It involves, among other things, ensuring the integrity of the data, correcting wrong or inaccurate data, identifying missing data, converting data from one structure to another or from one type to another, or even a task as simple as making data more readable.

Data preparation also involves understanding *how* you're going to get and connect to the data and the performance implications of the decisions. When connecting to data, you need to make decisions to ensure that models and reports meet, and perform to, acknowledged requirements and expectations.

Privacy and security assurances are also important. These assurances can include anonymizing data to avoid oversharing or preventing people from seeing personally identifiable information when it isn't needed. Alternatively, helping to ensure privacy and security can involve removing that data completely if it doesn't fit in with the story that you're trying to shape.

Data preparation can often be a lengthy process. Data analysts follow a series of steps and methods to prepare data for placement into a proper context and state that eliminate poor data quality and allow it to be turned into valuable insights.

## **Model**

When the data is in a proper state, it's ready to be modeled. Data modeling is the process of determining how your tables are related to each other. This process is done by defining and creating relationships between the tables. From that point, you can enhance the model by defining metrics and adding custom calculations to enrich your data.

Creating an effective and proper semantic model is a critical step in helping organizations understand and gain valuable insights into the data. An effective semantic model makes reports more accurate, allows the data to be explored faster and efficiently, decreases time for the report writing process, and simplifies future report maintenance.

The model is another critical component that has a direct effect on the performance of your report and overall data analysis. A poorly designed model can have a drastically negative impact on the general accuracy and performance of your report. Conversely, a well-designed model with well-prepared data will ensure a properly efficient and trusted report. This notion is more prevalent when you are working with data at scale.

From a Power BI perspective, if your report is performing slowly, or your refreshes are taking a long time, you will likely need to revisit the data preparation and modeling tasks to optimize your report.

The process of preparing data and modeling data is an iterative process. Data preparation is the first task in data analysis. Understanding and preparing your data before you model it will make the modeling step much easier.

## **Visualize**

The visualization task is where you get to bring your data to life. The ultimate goal of the visualize task is to solve business problems. A well-designed report should tell a compelling story about that data, which will enable business decision makers to quickly gain needed insights. By using appropriate visualizations and interactions, you can provide an effective report that guides the reader through the content quickly and efficiently, therefore allowing the reader to follow a narrative into the data.

The reports that are created during the visualization task help businesses and decision makers understand what that data means so that accurate and vital decisions can be made. Reports drive the overall actions, decisions, and behaviors of an organization that is trusting and relying on the information that is discovered in the data.

The business might communicate that they need all data points on a given report to help them make decisions. As a data analyst, you should take the time to fully understand the problem that the business is trying to solve. Determine whether all their data points are necessary because too much data can make detecting key points difficult. Having a small and concise data story can help find insights quickly.

With the built-in AI capabilities in Power BI, data analysts can build powerful reports, without writing any code, that enable users to get insights and answers and find actionable objectives. The AI capabilities in Power BI, such as the built-in AI visuals, enable the discovering of data by asking questions, using the Quick Insights feature, or creating machine learning models directly within Power BI.

An important aspect of visualizing data is designing and creating reports for accessibility. As you build reports, it is important to think about people who will be accessing and reading the reports. Reports should be designed with accessibility in mind from the outset so that no special modifications are needed in the future.

Many components of your report will help with storytelling. From a color scheme that is complementary and accessible, to fonts and sizing, to picking the right visuals for what is being displayed, they all come together to tell that story.

## **Analyze**

The analyze task is the important step of understanding and interpreting the information that is displayed on the report. In your role as a data analyst, you should understand the analytical capabilities of Power BI and use those capabilities to find insights, identify patterns and trends, predict outcomes, and then communicate those insights in a way that everyone can understand.

Advanced analytics enables businesses and organizations to ultimately drive better decisions throughout the business and create actionable insights and meaningful results. With advanced analytics, organizations can drill into the data to predict future patterns and trends, identify activities and behaviors, and enable businesses to ask the appropriate questions about their data.

Previously, analyzing data was a difficult and intricate process that was typically performed by data engineers or data scientists. Today, Power BI makes data analysis accessible, which simplifies the data analysis process. Users can quickly gain insights into their data by using visuals and metrics directly from their desktop and then publish those insights to dashboards so that others can find needed information.

This feature is another area where AI integrations within Power BI can take your analysis to the next level. Integrations with Azure Machine Learning, cognitive services, and built-in AI visuals will help to enrich your data and analysis.

## **Manage**

Power BI consists of many components, including reports, dashboards, workspaces, semantic models, and more. As a data analyst, you are responsible for the management of these Power BI assets, overseeing the sharing and distribution of items, such as reports and dashboards, and ensuring the security of Power BI assets.

Apps can be a valuable distribution method for your content and allow easier management for large audiences. This feature also allows you to have custom navigation experiences and link to other assets within your organization to complement your reports.

The management of your content helps to foster collaboration between teams and individuals. Sharing and discovery of your content is important for the right people to get the answers that they need. It is also important to help ensure that items are secure. You want to make sure that the right people have access and that you are not leaking data past the correct stakeholders.

Proper management can also help reduce data silos within your organization. Data duplication can make managing and introducing data latency difficult when resources are overused. Power BI helps reduce data silos with the use of shared semantic models, and it allows you to reuse data that you have prepared and modeled. For key business data, endorsing a semantic model as certified can help to ensure trust in that data.

The management of Power BI assets helps reduce the duplication of efforts and helps ensure security of the data.