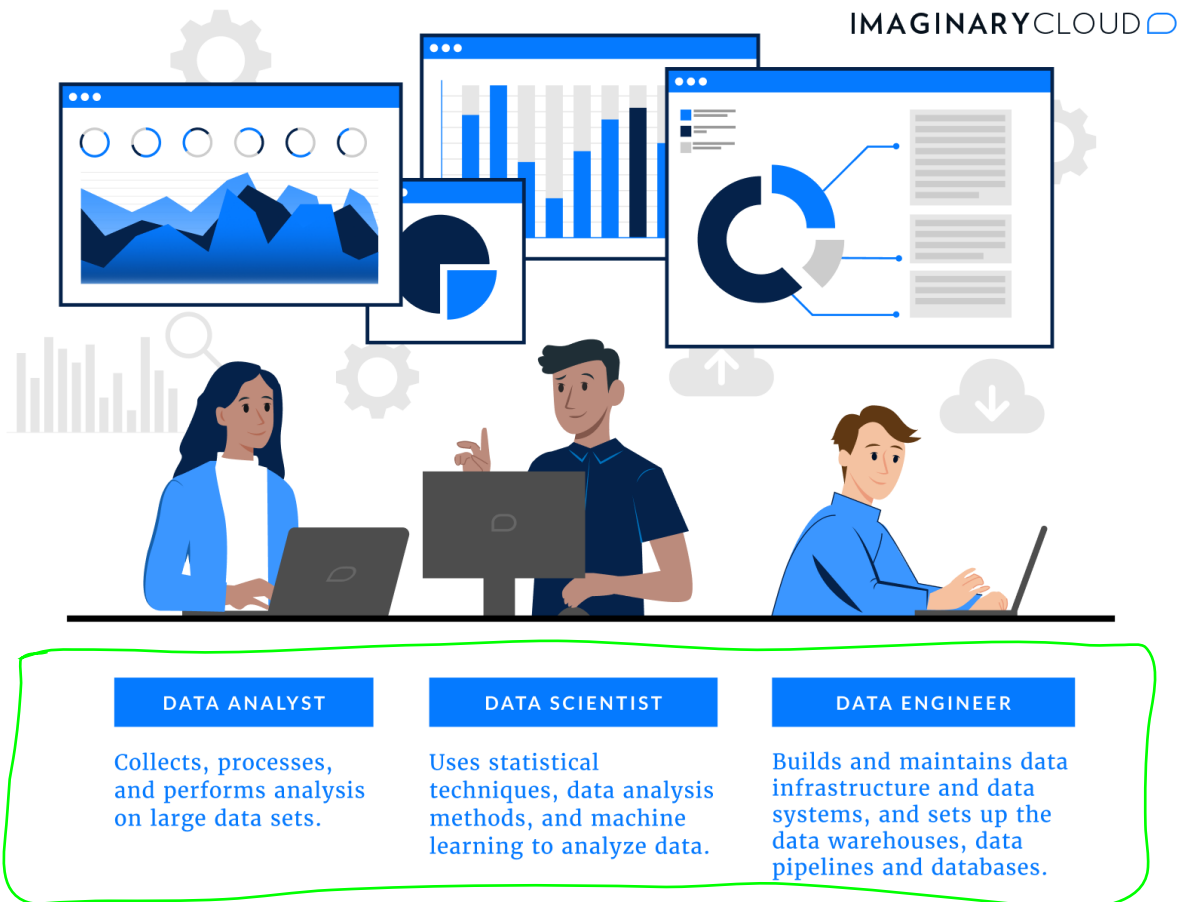


Data Analyst vs Data Scientist vs Data Engineer: What is the difference?



We live in an **information-driven world, in the era of Big Data, and organizations strongly rely on data for decision-making.** There are several roles in the industry today that deal with data, but a lot of people have misconceptions about them, especially when it comes to understanding the exact roles of a *data analyst vs data scientist vs data engineer*.

These job roles come to cause a little bit of confusion for both data job seekers and data hiring managers alike. If we were to put their roles on a spectrum, we could put data scientists right in-between data analysts and data engineers because it takes aspects of each of these two job roles and also builds on them with additional responsibilities and expectations. But let's start by clarifying why **data science is becoming increasingly important for businesses** and how each of the data roles can help leverage the companies' results.

How is Data Science applied in businesses?

In today's business environment, **data is a necessity** and, if well tamed, it can quickly become a **competitive advantage**. More and more companies are hiring data professionals in order to **maximize business revenue, forecast sales, and reduce costs**.

Web and mobile apps, the Internet of Things (IoT), and the advancement of AI technology have implemented

big data solutions so simple that even small and medium-sized companies can benefit from them. Businesses may use big data analytics to make better decisions and improve operational efficiency in several ways. **So what are the main applications and benefits of data science in businesses?**

- **Customer Segmentation**

We can use data science to perform customer segmentation or clustering by dividing your customers into groups or clusters with common features. When it comes to marketing, this can be crucial since it allows you to create focused, individualized marketing ad campaigns that can help increase sales and boost conversion rates. Data professionals will use machine learning techniques such as **K-means** to cluster data points together.

- **Sales Forecasting**

Companies can leverage data science to develop predictive models to forecast sales in the future. Predictive models attempt to forecast future sales based on historical data. What's more, data science also forecasts future daily sales based on factors like promotions and seasonal influences like Black Friday and Christmas sales.

- **Customer's loyalty prediction**

Data Science can help predict customer churn by identifying customers who are more likely to stop buying your products. This works by collecting data about your customers and assigning churn risk scores. Once data professionals understand the patterns, the company can launch a targeted retention strategy for customers who are more likely to leave and predict those who are likely to stay.

- **Recommendation system**

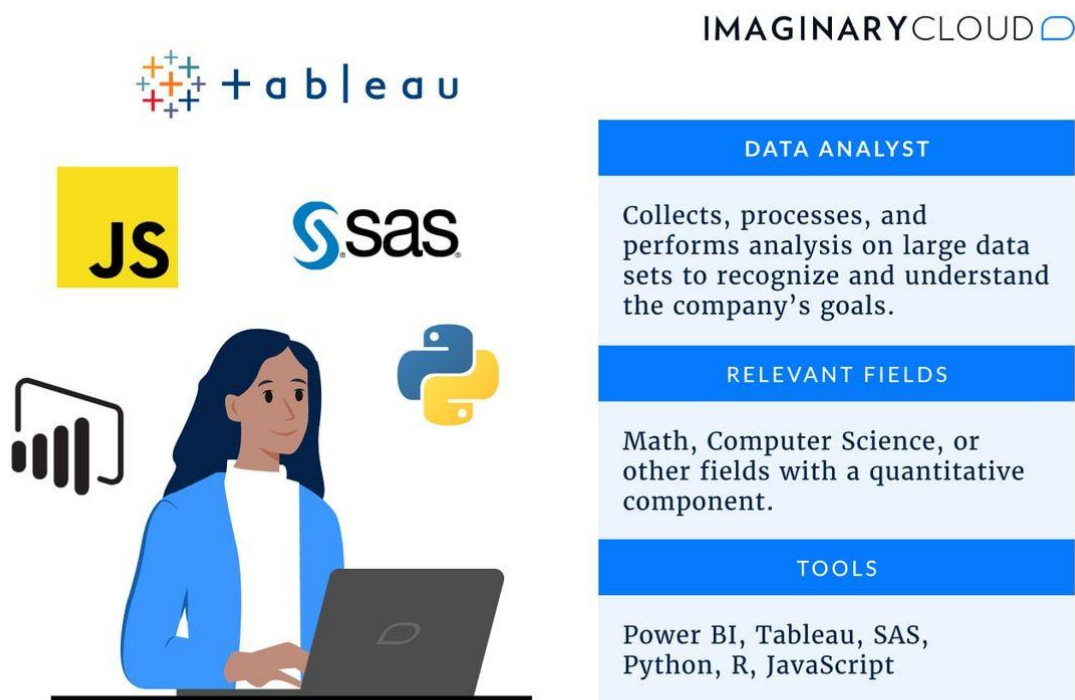
If it makes sense for your business, you may develop a recommendation system, like Netflix, Spotify, or even Amazon are doing. Implementing a recommendation system can predict the probability of a customer purchasing a product and then suggest other products as a cross-selling strategy. Doing so can improve sales and engage your customers.

- **Sentiment analysis**

We can use natural language processing and build a predictive model to predict if customers are happy or not. We can, for example, break sentiments (positive and negative) into smaller sub sentiments

such as “Happy”, “Love”, “Surprise”, “Sad”, “Fear”, “Angry” as per the needs or business requirement. This will help businesses conclude what they should be focusing on more or which areas or products need improvement.

What is a Data Analyst?



What is a Data Analyst?

A *Data Analyst* is responsible for **collecting, processing, and performing analysis** on large data sets. They deal with data wrangling, data modeling, and reporting, bringing in technical expertise to ensure the **quality and accuracy of the data**; after this, they process, design, and present their findings in a way that's meaningful to help the end consumer, businesses, or organizations make better decisions.

After a few years of experience, a **Data Analyst** can move into a **Data Scientist** and a **Data Engineer**, as we will see below.

Responsibilities of a Data Analyst

The first responsibility of a Data Analyst is **to recognize and understand the company's goals**. This, in turn, helps streamline the whole analysis process. They are required to **assess the available resources, comprehend the business problem, and gather the right set of data**. This step is done by **collaborating with different members such as Data Scientists, Business Analysts, and Programmers**. Other main responsibilities include:

- **Gather data** from various databases and warehouses through querying;

- Write complex SQL queries and scripts to gather and extract information;
- Filter and clean data to get the required information;
- Data mining where data is extracted from different sources and then organized in order to mine new information from it;
- Identify and analyze trends in complex data sets using statistical tools;
- Create summary reports through data visualization tools for the leadership teams to make timely decisions.

Skillset of a Data Analyst

The backgrounds of data analysts tend to vary a lot.

Traditionally, a data analyst would be someone with a bachelor's or master's degree in math or computer science. However, the modern data analyst can also have a background in Natural Sciences, something business-related, or any other field with a quantitative component.

The education required to become a Data Analyst is not very strict, and it's more based on one's ability to work with and understand data.

This being said, key skills of Data Analysts should include:

- Some amount of programming skills to independently search the data for answers;
- Understanding of relevant fields like Engineering in Computer Science, Information Technology or Electrical Engineering;
- Familiar with Microsoft Excel, programming languages such as Python, R, and JavaScript to help one write programs to solve complex problems;
- Good knowledge of Statistical Analytics tools such as SAS miner and SSAS;
- Write SQL queries and procedures;
- Perform A/B testings;
- Strong understanding of statistics and machine learning algorithms – these include concepts such as hypothesis testing, probability distributions, and various classification and clustering techniques;
- Create appealing reports with the help of charts and graphs using data visualization tools such as Power BI and Tableau;
- Good presentation skills to convey the right ideas to the clients and stakeholders in a clearer way.

Does data analysis require coding?

The simple answer is no. Some may do it, but a data analyst is not required to code. He doesn't need to be an expert or know any programming language deeply, though understanding SQL and Python is a competitive advantage. Data Analysts mostly use R or Tableau to create high-quality interactive maps, charts, and other visualizations.

What is a Data Scientist?



IMAGINARYCLOUD 

DATA SCIENTIST

Uses different statistical techniques, data analysis methods, and machine learning in order to understand and analyze data that will help draw business conclusions.

RELEVANT FIELDS

Machine Learning, Deep Learning, Mathematics, Statistics

TOOLS

C++, Java, Perl, Spark, Hive, Pig, Hadoop, Python

What is a Data Scientist? A **Data Scientist** is a professional who uses different **statistical techniques, data analysis methods, and machine learning** to understand and

analyze data that will help draw business conclusions.

We can classify data science professionals as **research-focused, business-focused, or development-focused.**

- **Research-focused** data scientists, also known as **Machine Learning Researchers or Research Scientists**, are on a mission to transform the space they're working in, which usually translates into the development or implementation of new machine learning techniques. They work in complex problem spaces such as those in **machine vision or natural language processing**, and also problem spaces that have huge amounts of data - like social media, for example. They will usually **use scripting languages like Python, deep learning tools, and frameworks like TensorFlow.**
- **Business-focused** data science professionals implement **established scientific methods to help businesses make decisions powered by data.** This means **1) understanding the business problem** and **2) knowing how to use data to solve it.** Again, they usually **use scripting languages, like Python, combined with machine learning, statistics libraries, and SQL** to identify appropriate predictive modeling, statistical testing, or analytical

approaches to push the data through and solve a problem.

- **Development-focused data scientists** are the ones who **scale data science processes** or build the data science-related components of applications. Whether it's putting machine learning models into production or building the infrastructure for working with big data, they can be called the **enablers of leveraging data at scale**. They're often called **Machine Learning Engineers, Data Engineers, or Machine Learning Developers**.

They also proactively **fetch information** from sources galore and analyze it to understand better how the business is performing, **building AI tools that automate certain processes** within the company.

Simply put, a **Data Scientist derives meaning out of messy and unstructured data** turning it easier to read and understand.

Responsibilities of a Data Scientist

Data scientists are responsible for cleaning, processing, and manipulating data using several data analytics

tools. Besides those mentioned above, other key responsibilities include:

- Perform ad-hoc data mining;
- Collect large sets of structured and unstructured data from sources galore;
- Interpret the data using statistical methods, designing and evaluating advanced statistical models to work on big data;
- Regularly build predictive models and machine learning algorithms to work on past volumes of data;
- Use visualization packages and tools to create reports and dashboards for relevant stakeholders;
- Work side-by-side with Data Analysts and Data Engineers to formulate the analysis results.

Skillset of a Data Scientist

Data scientists have many of the same skills as Data Analysts. Still, they are often a little more IT-heavy, meaning that they're able to create the databases themselves and pull together a lot of disparate information that might exist in different sources. Because of these expectations, programming skill expectations for Data Scientists are much higher. They're required to have good experience in

programming languages like **Python, C++, or Java**, as well as proficiency in **SQL**.

Other important skills of a Data Scientist include:

- In-depth knowledge of **machine learning and deep learning**;
- Being familiar with [Apache Spark](#), [Apache Hive](#), and [Apache Pig](#) is desirable, along with the knowledge of [Hadoop](#);
- **Data visualization and business intelligence skills for creating reports and dashboards**;
- Communicate and present information and ideas clearly.

Can a data analyst become a data scientist?

The short answer is **yes**. One of the more effective ways to become a data scientist is to start as a data analyst, as both job roles are relatively similar.

Many people ask **which is better, a data analyst or data scientist?** But it's important to clarify that a **data science role is not better than a data analyst role**. It's simply a position that leverages a slightly different set of skills.

Simply put, a data science role is more suited for those who enjoy coding more. Many analysts do code, but they leverage other tools like Tableau and Power BI, as we saw above.

Here are a few tips to making the transition from Data Analyst to Data Scientist:

- **Show that you can code:** as we've already mentioned, data analysts code slightly less than data scientists; when you're transitioning, it's important to showcase your programming ability in Python or R. The best way to do this is through a portfolio on GitHub, for example.
- **Highlight your strengths:** data analysts often have high levels of business understanding and logic, so make sure that you show the value you've been creating at your current company or in your projects.
- **Do Data Science work in your current projects:** look for opportunities to practice data science in your current projects; although you're an analyst, you still have access to data, and it never hurt to go a

little above and beyond to experiment with some more advanced algorithms.

- **Aim at upskilling:** continued education in the form of certificates or university programs is a great way to learn and condense data science concepts; this route may not be for everyone, but it could be useful for some.
- **Join communities to network:** leveraging your existing network is an excellent way to get started; you'll be surprised at how many opportunities you might find when you're talking or just around like-minded people. Online communities, for example, are an incredible place to develop your skills further, too.

What is a Data Engineer?



DATA ENGINEER
Focused on building and maintaining data infrastructure and data systems, and setting up the data warehouses, data pipelines and databases that the Data Analysts and Data Scientists use to access and work with the data.
RELEVANT FIELDS
Data Engineering, Computer Science, IT, Software Development
TOOLS
Hadoop, MapReduce, Pig, Hive, SQL and NoSQL

What is a Data Engineer? A **Data Engineer job description** falls into the category of a software engineer that is focused on building and maintaining data infrastructure and data systems. Data Engineers are the ones setting up the data warehouses, data pipelines, and databases that the Data Analysts and Data Scientists use to access and work with the data.

A **Data Engineer** is also perhaps the most well-defined role of the three, and you can probably see the most consistency with this one. Let's take a better look at the responsibilities and skills of a data engineer.

Responsibilities of a Data Engineer

- Build and maintain ETL (Extract, Transform, and Load) pipelines and data infrastructure;
- Cloud computing;
- Big data and distributed computing frameworks;
- Create and integrate APIs - take a look at the top 6 API testing tools here;
- Machine learning deployment and integration;
- Develop, construct, test and maintain the architecture of large-scale processing systems and databases to make sure that business needs are fulfilled;
- Provide and implement ways to improve the data's reliability, efficiency, and quality.

Skillset of a Data Engineer

Since Data Engineers are architects and caretakers of the data, their role mainly concentrates on database systems. What skills are required for a data engineer? They include:

- Experience in Hadoop, MapReduce, Pig, Hive, and Data Streaming;
- In-depth knowledge of database systems with knowledge in SQL and NoSQL;

- Background in **Software Development, Computer Science, Applied Math or Statistics**;
- Strong computer science skills.

Data Engineer vs Data Scientist

The **key difference between a Data Engineer and a Data Scientist** is education and skills. Let's think of data analytics like a timeline. **Data engineers work at the very beginning of it on the back-end, whereas data scientists tend to take over where data engineers leave off, finding meaning and insights from it for the organization.**

As already seen, **a data scientist is generally good at mathematics and statistics.** He will usually be **proficient in programming and have a penchant for machine learning and artificial intelligence modeling.** A thorough understanding of the domain he is operating in is also an important skill to have, in order **to gather business intelligence that can help the business achieve success.** Lastly, a **Data Scientist is also good at visually and verbally communicating insights** from data with team leaders and business stakeholders.

On the other hand, the **Data Engineer is a programmer** proficient in Python, Java, and Scala and adept at **handling distributed systems to analyze big**

amounts of data. His primary responsibility, as already explained, is creating free-flowing data pipelines using big data technologies for real-time or static data analytics.

All in all, these two roles use similar skill sets, so it's safe to say that both data engineers and data scientists work with big data. Nonetheless, the data scientist is typically a better analyst than a programmer while the data engineer is a better programmer than an analyst. The two roles are complementary, not interchangeable, and they work best together when they're made to perform tasks that match their strengths.