Data Scientist vs Data Engineer vs Data Analyst – What really differentiates them?

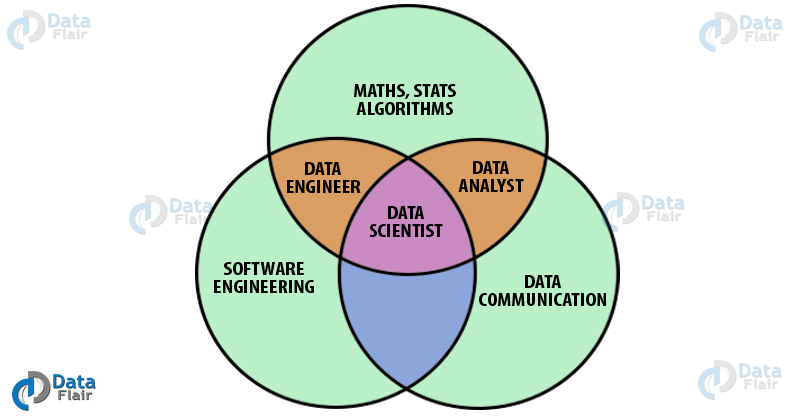
BY [DATAFLAIR TEAM](https://data-flair.training/blogs/author/dfteam2/) · MAY 23, 2019

Have you ever wondered what differentiates data scientist from a data analyst and a data engineer? What is the differentiating factor that helps them to analyze the data from a different point of view? The answer is their core **TASK!**

The task of a Data Scientist is to unearth future insights from raw data. Data engineer focuses on development and maintenance of data pipelines. Data analyst mainly take actions that affect the company’s scope.

**Still confused right? Don’t worry this is just a brief.**In this article, I am providing you a detailed comparison, Data Scientist vs Data Engineer vs Data Analyst. First, you will learn what is a Data Scientist, Data Engineer, and Data Analyst and then you will find the comparison and salary of the three.

I assure you that by the end of the article, you will finalize the best trending Data job for you. So, without wasting more time let’s start.

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/sites/2/2019/05/data-scientist-vs-data-engineer-vs-data-analyst.jpg)

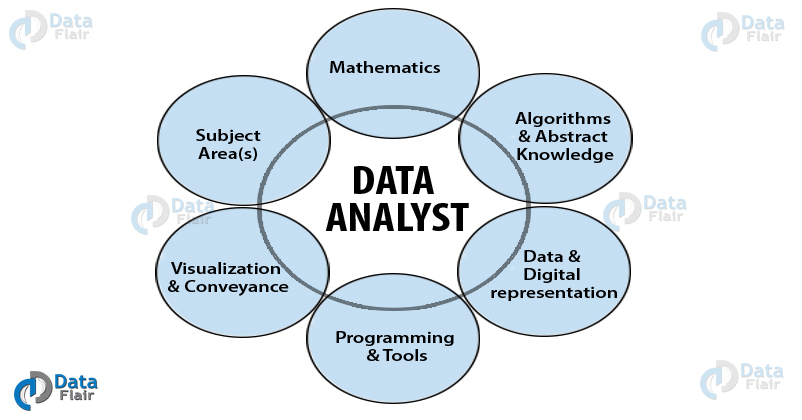
What is Data Analyst?

*The process of the extraction of information from a given pool of data is called data analytics. A data analyst is a person who engages in this form of analysis.* A data analyst extracts the information through several methodologies like data cleaning, data conversion, and data modeling. There are several industries where data analytics is used, such as – technology, medicine, social science, business etc. Industries are able to analyze trends in the market, requirements of their clients and overview their performances with data analysis. This allows them to make careful data-driven decisions.

The two most important techniques used in data analytics are descriptive or summary statistics and inferential statistics. A Data Analyst is also well versed with several visualization techniques and tools. It is utmost necessary for the data analyst to have presentation skills. This allows them to communicate the results with the team and help them to reach proper solutions.

***You must check the latest***[***guide on Maths and Statistics***](https://data-flair.training/blogs/math-and-statistics-for-data-science/)***by experts***

Data Analytics allows the industries to process fast queries to produce actionable results that are needed in a short duration of time. This restricts data analytics to a more short term growth of the industry where quick action is required. Two of the popular and common tools used by the data analysts are **SQL** and Microsoft Excel.

[](https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/sites/2/2019/05/What-is-data-analyst.jpg)

What is Data Engineer?

*A Data Engineer is a person who specializes in preparing data for analytical usage*. Data Engineering also involves the development of platforms and architectures for data processing. In other words, a data engineer develops the foundation for various data operations. A Data Engineer is responsible for designing the format for data scientists and analysts to work on.

Data Engineers have to work with both structured and unstructured data. Therefore, they [***need expertise in SQL***](https://data-flair.training/blogs/sql-tutorial/) and NoSQL databases both. Data Engineers allow data scientists to carry out their data operations. Data Engineers have to deal with Big Data where they engage in numerous operations like data cleaning, management, transformation, data deduplication etc.

A Data Engineer is more experienced with core programming concepts and algorithms. The**role of a data engineer** also follows closely to that of a software engineer. This is because a data engineer is assigned to develop platforms and architecture that utilize guidelines of software development. **For example**, developing a cloud infrastructure to facilitate real-time analysis of data requires various development principles. Therefore, building an interface API is one of the job responsibilities of a data engineer.

***A top skill that gets you hired is Big Data.***[***Start learning Big Data with industry experts***](https://data-flair.training/big-data-hadoop/)***.***

Furthermore, a data engineer has a good knowledge of engineering and testing tools. It is up to a data engineer to handle the entire pipelined architecture to handle log errors, agile testing, building fault-tolerant pipelines, administering databases and ensuring a stable pipeline.

Tools used by Data Engineers

Some of the tools that are used by **Data Engineers** are –

* **Hadoop**

Apache Hadoop is an open-source Big Data Platform which is the bread and butter for all the data engineers. It comprises of Hadoop Distributed Framework or HDFS which is designed to run on commodity hardware. A Data Engineer must be well versed with Hadoop as it is the standard Big Data platform for many industries.

* **Apache Spark**

Spark is a fast processing, analytical big data platform provided by Apache. It was developed as an improvement over Hadoop which could only handle batch data. However, Spark provides support for both batch data as well as streaming data.

***It is the right time to***[***start your Hadoop and Spark learning***](https://data-flair.training/hadoop-spark-developer-course/)

* **Kubernetes**

Kubernetes was developed by Google for cluster orchestration, scaling and automating the application deployment. It is a recent technology that has revolutionized the world of cloud computing.

* **Java**

Java is the most popular programming language that is used for developing enterprise software solutions. A Data Engineer must know this programming language in order to develop pipelines and data infrastructure.

* **Yarn**

Yarn is a part of the Hadoop Core project. It allows several data-processing engines to handle data on a single platform. It is an efficient tool to increase the efficiency of the Hadoop compute cluster.

What is Data Scientist?

Data Science is the most trending job in the technology sector. It has quickly emerged to be crowned as the “Sexiest Job of the 21st century”. Almost everyone talks about Data Science and companies are having a sudden requirement for a greater number of data scientists. While Data Science is still in its infantile stage, it has grown to occupy almost all the sectors of industry. Every company is looking for data scientists to increase their performance and optimize their production.

***Explore the best tips to***[***get your first Data Science Job***](https://data-flair.training/blogs/get-your-first-job-in-data-science/)

There is a massive explosion in data. This explosion is contributed by the advancements in computational technologies like High-Performance Computing. This has given industries a massive opportunity to unearth meaningful information from the data.

Companies extract data to analyze and gain insights about various trends and practices. In order to do so, they employ specialized data scientists who possess knowledge of statistical tools and programming skills. Moreover, a data scientist possesses[***knowledge of machine learning algorithms***](https://data-flair.training/blogs/machine-learning-algorithms/). These algorithms are responsible for predicting future events. Therefore, data science can be thought of as an ocean that includes all the data operations like data extraction, data processing, data analysis and data prediction to gain necessary insights.

However, Data Science is not a singular field. It is a quantitative field that shares its background with math, statistics and computer programming. With the help of data science, industries are qualified to make careful data-driven decisions. Data is everywhere, and as a result, there are a plethora of data science positions. However, due to a high learning curve, there is a shortage in supply for data scientists. This has resulted in a massive income bubble that provides the data scientists with lucrative salaries.

Data Analyst Vs Data Engineer Vs Data Scientist – Definition

* A **data analyst** is responsible for*taking actionable that affect the current scope of the company*. A **data engineer** is responsible for*developing a platform that data analysts and data scientists work on*. And, a **data scientist** is responsible for *unearthing future insights from existing data* and helping companies to make data-driven decisions.
* A**data analyst** does not directly participate in the decision-making process, rather, he helps indirectly through*providing static insights about company performance*. A**data engineer** is not responsible for decision making. And, a**data scientist** participates in the active *decision-making process* that affects the course of the company.
* A **data analyst** uses *static modeling techniques* that summarize the data through descriptive analysis. On the other hand, a **data engineer** is responsible for the *development and maintenance of data pipelines*. A **data scientist** uses *dynamic techniques* like [***Machine Learning to gain insights about the future***](https://data-flair.training/blogs/machine-learning-tutorial/).
* Knowledge of machine learning is not important for **data analysts**. However, this is mandatory for **data scientists**. A **data engineer** need not require the knowledge of machine learning but he is required to have the knowledge of core computing concepts like programming and algorithms to build robust data systems.
* A **data analyst** only has to deal with*structured data*. However, both **data scientists and data engineers** deal with*unstructured data as well*.
* A**data analyst** and **data scientist** are both required to be proficient in data visualization. However, this is not required in the case of a **data engineer**.
* Both **data scientists and analysts** need not have *knowledge of application development* and working of the APIs. However, this is the most essential requirement for a **data engineer**.

Data Analyst Vs Data Engineer Vs Data Scientist – Responsibilities

Following are the main responsibilities of a **Data Analyst** –

* Analyzing the data through descriptive statistics.
* Using database query languages to retrieve and manipulate information.
* Perform data filtering, cleaning and early stage transformation.
* Communicating results with the team using data visualization.
* Work with the management team to understand business requirements.

A **Data Engineer** is supposed to have the following responsibilities –

* Development, construction, and maintenance of data architectures.
* Conducting testing on large scale data platforms.
* Handling error logs and building robust data pipelines.
* Ability to handle raw and unstructured data.
* Provide recommendations for data improvement, quality, and efficiency of data.
* Ensure and support the data architecture utilized by data scientists and analysts.
* Development of data processes for data modeling, mining, and data production.

A **Data Scientist** is required to perform responsibilities –

* Performing data preprocessing that involves data transformation as well as data cleaning.
* Using various machine learning tools to forecast and classify patterns in the data.
* Increasing the performance and accuracy of machine learning algorithms through fine-tuning and further performance optimization.
* Understanding the requirements of the company and formulating questions that need to be addressed.
* Using robust storytelling tools to communicate results with the team members.

Data Analyst Vs Data Engineer Vs Data Scientist – Skills

In order to become a **Data Analyst**, you must possess the following skills –

* Should possess the strong mathematical aptitude
* Should be well versed with Excel, Oracle, and [SQL](https://en.wikipedia.org/wiki/SQL).
* Possession of problem-solving attitude.
* Proficient in the communication of results to the team.
* Should have a strong suite of analytical skills.

Following are the key skills required to become a data engineer –

* Knowledge of programming tools like Python and Java.
* Solid Understanding of Operating Systems.
* Ability to develop scalable ETL packages.
* Should be well versed in SQL as well as NoSQL technologies like Cassandra and MongoDB.
* He should possess knowledge of data warehouse and big data technologies like Hadoop, Hive, Pig, and Spark.
* Should possess creative and out of the box thinking.

For becoming a **Data Scientist**, you must have the following key skills –

* Should be proficient with Math and Statistics.
* Should be able to handle structured & unstructured information.
* In-depth knowledge of tools like R, Python and SAS.
* Well versed in various machine learning algorithms.
* Have knowledge of SQL and NoSQL.
* Must be familiar with Big Data tools.

Data Analyst Vs Data Engineer Vs Data Scientist – Salary Differences

* On average, a**Data Analyst** earns an annual salary of**$67,377**
* A**Data Engineer** earns**$116,591 per annum**
* And a**Data Scientist**, on average, makes **$117,345 in a year**

***Update your skills and get***[***top Data Science jobs***](https://data-flair.training/blogs/top-data-science-jobs/)

Summary

So, this is all about Data Scientist vs Data Engineer vs Data Analyst. We went through the various roles and responsibilities of these fields. Hope now you understand which is the best role for you. I love Data Scientist job and recommend you the same as it is the most sexiest job of the 21st century. So, what are you waiting for? Start working on yourself and get a good job.

**All the best!**

Share your thoughts on the article through comments. Your feedback is appreciable.