

What's the difference between a data analyst, scientist and engineer?

 dataquest.io/blog/data-analyst-data-scientist-data-engineer

June 15, 2017

Data is increasingly shaping the systems that we interact with every day. Whether you're using Siri, searching Google, or browsing your Facebook feed, you're consuming the results of data analysis. Given its transformational ability, it's no wonder that so many data-related roles have been created in the past few years.

The responsibilities of these roles range from predicting the future, to finding patterns in the world around you, to building systems that manipulate millions of records. In this post, we'll talk about the various data-related roles, how they fit together, and help you figure out which role is the right fit.

What is a data analyst?

Data Analysts deliver value to their companies by taking data, using it to answer questions, and communicating the results to help make business decisions. Common tasks done by data analysts include data cleaning, performing analysis and creating data visualizations.

Depending on the industry, the data analyst could go by a different title (e.g. Business Analyst, Business Intelligence Analyst, Operations Analyst, Database Analyst). Regardless of title, the data analyst is a generalist who can fit into many roles and teams to help others make better data-driven decisions.

The data analyst in depth

The data analyst has the potential to turn a traditional business into a data-driven one. While often data analyst positions are 'entry level' jobs in the wider field of data, not all analysts are junior level. As effective communicators with a mastery over technical tools, data analysts are critical for companies that have segregated technical and business teams.

Their core responsibility is to help others track progress and optimize their focus. How can a marketer use analytics data to help launch their next campaign? How can a sales representative better identify which demographics to target? How can a CEO better understand the underlying reasons behind recent company growth? These are all questions that the data analyst provides the answer to by performing analysis and presenting the results. They undertake the complex job of working with data to deliver value to their organization.

An effective data analyst will take the guesswork out of business decisions and help the entire organization thrive. The data analyst must be an effective bridge between different teams by analyzing new data, combining different reports, and translating the outcomes. In turn, this is what allows the organization to maintain an accurate pulse check on its growth.

The nature of the skills required will depend on the company's specific needs, but these are some common tasks:

- Cleaning and organizing raw data.
- Using descriptive statistics to get a big-picture view of their data.
- Analyzing interesting trends found in the data.
- Creating visualizations and dashboards to help the company interpret and make decisions with the data.
- Presenting the results of a technical analysis to business clients or internal teams.

The data analyst brings significant value to both the technical and non-technical sides of an organization. Whether running exploratory analyses or explaining executive dashboards, the analyst fosters greater connection between teams.

What is a data scientist?

A data scientist is a specialist that applies their expertise in statistics and building machine learning models to make predictions and answer key business questions.

A data scientist still needs to be able to clean, analyze, and visualize data, just like a data analyst. However, a data scientist will have more depth and expertise in these skills, and will also be able to train and optimize machine learning models.

The data scientist in depth

The data scientist is an individual that can provide immense value by tackling more open-ended questions and leveraging their knowledge of advanced statistics and algorithms. If the analyst focuses on understanding data from the past and present perspectives, then the scientist focuses on producing reliable predictions for the future.

The data scientist will uncover hidden insights by leveraging both supervised (e.g. classification, regression) and unsupervised learning (e.g. clustering, neural networks, anomaly detection) methods toward their machine learning models. They are essentially training mathematical models that will allow them to better identify patterns and derive accurate predictions.

The following are examples of work performed by data scientists:

- Evaluating statistical models to determine the validity of analyses.
- Using machine learning to build better predictive algorithms.
- Testing and continuously improving the accuracy of machine learning models.
- Building data visualizations to summarize the conclusion of an advanced analysis.

Data scientists bring an entirely new approach and perspective to understanding data. While an analyst may be able to describe trends and translate those results into business terms, the scientist will raise new questions and be able to build models to make predictions based on new data.

What is a data engineer?

Data engineers build and optimize the systems that allow data scientists and analysts to perform their work. Every company depends on its data to be accurate and accessible to individuals who need to work with it. The data engineer ensures that any data is properly received, transformed, stored, and made accessible to other users.

The data engineer in depth

The data engineer establishes the foundation that the data analysts and scientists build upon. Data engineers are responsible for constructing data pipelines and often have to use complex tools and techniques to handle data at scale. Unlike the previous two career paths, data engineering leans a lot more towards a software development skillset.

At larger organizations, data engineers can have different focuses such as leveraging data tools, maintaining databases, and creating and managing data pipelines. Whatever the focus may be, a good data engineer allows a data scientist or analyst to focus on solving analytical problems, rather than having to move data from source to source.

The data engineer's mindset is often more focused on building and optimization. The following are examples of tasks that a data engineer might be working on:

- Building APIs for data consumption.
- Integrating external or new datasets into existing data pipelines.
- Applying feature transformations for machine learning models on new data.
- Continuously monitoring and testing the system to ensure optimized performance.

Your Data-Driven Career Path

Now that we've explored these three data-driven careers, the question remains - where do you fit in? The key is to understand that these are three fundamentally different ways to work with data.

The data engineer is working on the "back-end," continuously improving data pipelines to ensure that the data the organisation relies upon is accurate and available. They will leverage all sorts of different tools to ensure the data is processed correctly and that the data is available to the user when they need it. A good data engineer saves a lot of time and effort for the rest of the organization.

The data analyst may then extract a new dataset using the custom API that the engineer built and begin identifying interesting trends in that data, as well as running analyses on these anomalies. The analyst will summarize and present their results in a clear way that allows their non-technical teams to better understand where they are and how they're doing.

Finally, the data scientist will likely build upon the analyst's initial findings and research into even more possibilities to derive insights from. Whether by training machine learning models or by running advanced statistical analyses, the data scientist is going to provide a brand new perspective into what may be possible for the near future.

Regardless of your specific path, curiosity is a natural pre-requisite of all three of these careers. The ability to use data to ask better questions and run more precise experiments is the entire purpose of a data-driven career. Furthermore, the data science field is constantly evolving and thus, there is a great need to continuously learn more.

So, to all the current and future data analysts, scientists, and engineers out there — good luck and keep learning!