

# Data Science Career Paths: Different Roles in the Industry

---

Roger Huang

Data Science

## Data Science Career Paths: Introduction

**We've just come out with the first bootcamp with a data science job guarantee to help you break into a data science career. As part of that exercise, we dove deep into the different roles within data science.**

Around the world, organizations are creating more data every day, yet most are struggling to benefit from it. According to McKinsey, the US alone will face a shortage of 150,000+ data analysts and an additional 1.5 million data-savvy managers.

But really, what is a data-savvy manager? What does the term even mean? Are all data scientists made equally?

Data science teams are presented with a host of problems. They might be called upon to analyze whether Tweets sent to a company are positive or negative, or they might have to trace where sales are coming from. Different organizations will have different data problems—each problem comes with its own complexities. Solving different data science problems can **require different skill sets**.

Data science teams come together to solve some of the hardest data problems an organization might face. Each individual will have a different part of the skill set required to complete a data science project from end to end. The roles within data science are really a set of complementary roles that each have a specific vocabulary. There are data scientists—but there are also data engineers, and data analysts!

We realize that this can be confusing for a newcomer to the field. We want to demystify the different roles within data science so you can understand the nuances within the field — here goes:



PRESENTS

# THE DIFFERENT DATA SCIENCE ROLES



## UNFILLED JOBS



**150,000**  
DATA ANALYSTS



**1.5 MILLION**  
DATA-SAVVY MANAGERS



## WHO ARE THEY?



### DATA SCIENTISTS

A.K.A. STATISTICIANS,  
DATA MANAGERS

**\$115,000** USD  
AVERAGE SALARY

A data scientist will be able to take data science projects from end-to-end. They can help store large amounts of data, create predictive modeling, and



### DATA ENGINEERS

A.K.A. DATA ARCHITECTS,  
DATABASE ADMINISTRATORS

**\$100,000** USD  
AVERAGE SALARY

These are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and



### DATA ANALYSTS

A.K.A. BUSINESS ANALYSTS

**\$65,000** USD  
AVERAGE SALARY

When somebody helps people from across the company understand specific queries with charts, they are filling the data analyst role.

predictive modeling processes and tell about the findings.

### SKILLS

Mathematics,  
Programming  
Communication



### TOOLS

SQL, Python, R



cleaning up data sets, and implementing requests that come from data scientists.

### SKILLS

Programming  
Mathematics  
Big Data



### TOOLS

Hadoop, NoSQL, Python



analyst role.

### SKILLS

Statistics  
Communication  
Business Knowledge



### TOOLS

Excel, Tableau, SQL



## TYPICAL MAJORS



Mathematics, Economics, Computer Science, Physics, Statistics, Engineering and Business

## OPEN JOB POSITIONS

Data from Indeed.com

~215,000 

## TOP HIRING LOCATIONS

in the United States

in the United States



**New York**      **San Francisco**      **Seattle**

**THINGS YOU'LL CATCH DATA SCIENTISTS SAYING**

*"My classifier gave me 93% accuracy on the first try! [Pause] Something must be wrong with the data ..."*

**TEAMWORK?**  
**HOW DO THE DIFFERENT ROLES CONNECT?**



The Venn diagram consists of three overlapping circles. The top circle is teal and labeled 'Math, Stats, Algorithms'. The bottom-left circle is teal and labeled 'Software engineering'. The bottom-right circle is teal and labeled 'Data communication'. The intersection of the top and bottom-left circles is labeled 'DATA ENGINEER'. The intersection of the top and bottom-right circles is labeled 'DATA ANALYST'. The intersection of all three circles is labeled 'DATA SCIENTIST'.

 **Springboard**

© 2016 Springboard.com

There are **data scientists** who fine-tune the statistical and mathematical models that are applied onto data. When somebody is applying their theoretical knowledge of statistics and algorithms to find the best way to solve a data science problem, they are filling the role of data scientist. When somebody builds a model to predict the number of credit card defaults in the next month, they are wearing the data scientist hat.

A data scientist will be able to take a business problem and translate it to a data question, create predictive models to answer the question and storytell about the findings.

**Statisticians** that focus on implementing statistical approaches to data, and **data managers** who focus on running data science teams tend to fall in the data scientist role.

Data scientists **are the bridge** between the programming and implementation of data science, the theory of data science, and the business implications of data.

**Skills You'll Need:** Knowledge of algorithms, statistics, mathematics, and broad knowledge of programming languages such as R and Python. Broad knowledge of how to structure a data problem, from framing the right questions to ask, to communicating the results effectively.

**Salaries:** Data scientists need to have a broad set of skills that covers the **theory, implementation and communication** of data science. They also tend to be the highest compensated group with an average salary above \$115,000 USD.

**Sample Job Posting:** This data scientist posting at Apple is looking for scientists who are both passionate about creating data driven systems and which have experience in statistical programming. You can truly see the versatility of the data scientist role in this description! The data scientist in question will play an important role in providing fast searches for Spotlight on Safari.

**Typical Majors:** Mathematics, economics, computer science, physics

**Open Job Positions on Indeed.com:** ~22,000 (18% over \$115,000 salary estimate)

**Industries that are Hiring Data Engineers:** Software, medicine, audio companies

**Top Hiring Locations in the United States:** New York City, San Francisco, Seattle

**Things You'll Catch Them Saying:** "My classifier gave me 93% accuracy on the first try! [Pause] Something must be wrong with the data ..."

## Data Engineers

There are **data engineers**, who rely mostly on their software engineering experience to handle large amounts of data at scale. These are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists. They typically know a broad variety of programming languages, from Python to Java. When somebody takes the predictive model from the data scientist and implements it in code, they are typically playing the role of a data engineer.

**Data architects** that focus on structuring the technology that manages data models and **database administrators** who focus on managing data storage solutions tend to be part of the category of data engineers.



**Skills You'll Need:** A deep knowledge of data storage and warehousing solutions (SQL and NoSQL – based flavors), and programming frameworks such as Hadoop and Spark that can help you source data and process it.

**Salaries:** Data engineers often focus on the implementation of data science by making sure code is clean, and technical systems are well-suited to the amount of data passing back and forth for analysis. They tend to be middle of the pack when it comes to compensation, with an average salary around \$100,000 USD.

**Sample Job Posting**Shopify is a Canadian startup that allows you to open an e-commerce store without having to build anything in code. Their posting for a data engineer requires you to have extensive software development experience along with extensive database experience. They are looking for people who are proficient in Python and Scala. They need “passionate software and operations engineers who are excited about data.”

**Typical Majors:** Computer science, engineering.

**Open Job Positions on Indeed.com:** ~98,000 (17% over \$115,000 salary estimate)

**Industries that are Hiring Data Engineers:** Software, aerospace, information technology

**Top Hiring Locations in the United States:** San Francisco, New York City, Seattle

**Things You'll Catch Them Saying:** “My data pipeline would be perfect if it wasn't for the people using it.”

## Data Analysts

Lastly, there are **data analysts** who look through the data and provide reports and visualizations to explain what insights the data is hiding. When somebody helps people from across the company understand specific queries with charts, they are filling the data analyst role.

**Business analysts** are a subset of data analysts that are more concerned with the business implications of the data and the actions that should result. Should the company invest more in project X or project Y? Business analysts will leverage the work of data science teams to communicate an answer.

**Skills You'll Need:** Data analysts will need a solid grasp of data manipulation (using programs like Excel) and data communication.

**Salary:** Data analysts tend to be the least compensated among the data science roles, with an average salary of around \$65k USD. This is largely because data analysis is more of an entry-level role that calls upon less of the skillset needed in data science.

**Sample Job Posting**Stripe helps process payments across the web for some of the largest web platforms in the world. Their data analyst position calls for somebody who is excited to apply their analytical skills to understand user behavior—and who will work closely with business and product teams to answer important data questions.

**Typical Majors:** Business, economics, statistics

**Open Job Positions on Indeed.com:** ~95,000

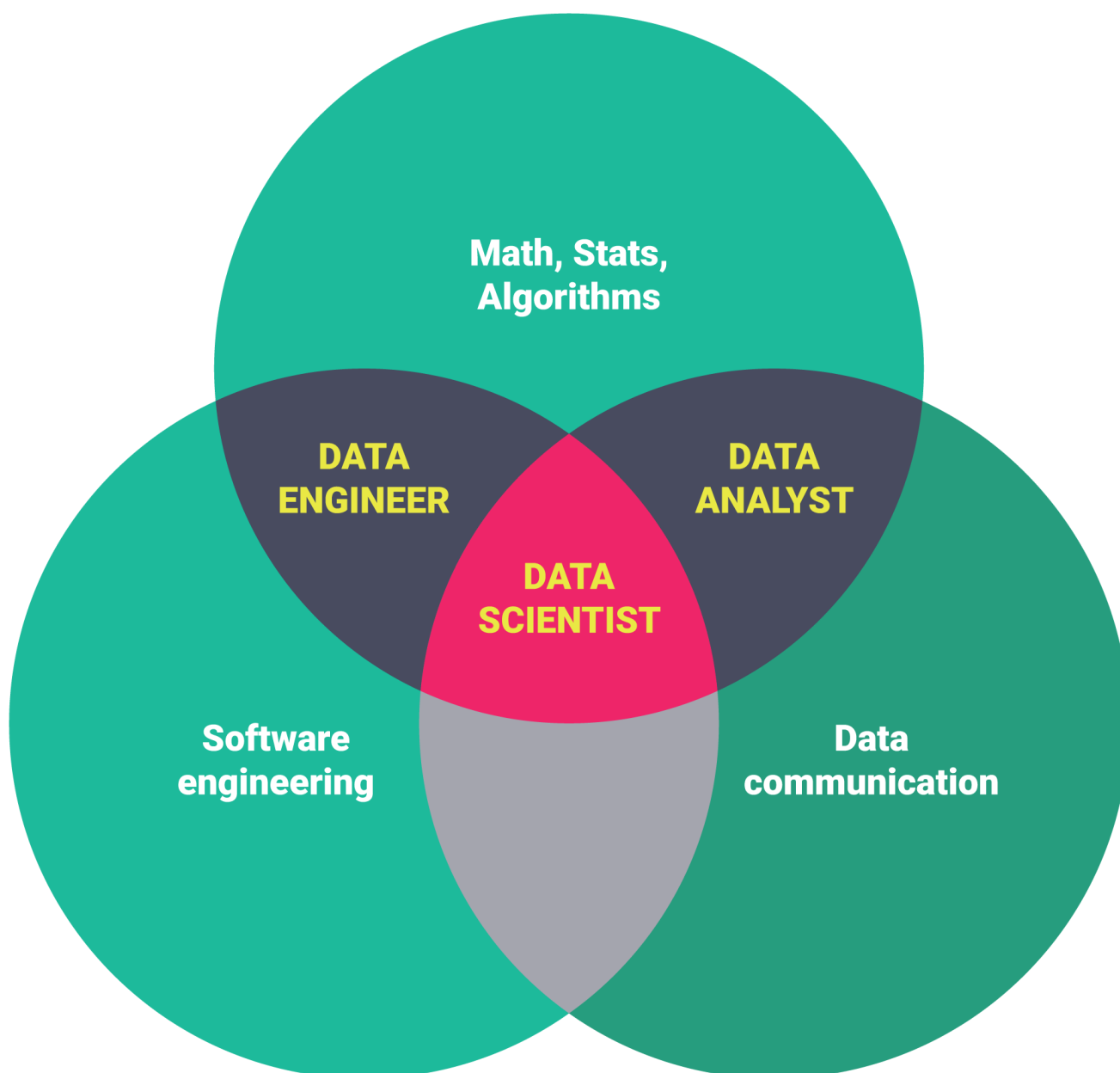
**Industries that are Hiring Data Analysts:** Consulting, healthcare, banking

***Top Hiring Locations in the United States:*** New York City, Washington DC, Chicago

***Things You'll Catch Them Saying:*** "Microsoft Excel is so slow today!"

## Takeaway

You can roughly say that data engineers **rely more on engineering skills**, data scientists rely more on **their training in mathematics and statistics**, and business analysts rely more heavily on **their communication skills and their domain expertise**. You can be sure that people who occupy these roles will have varying amounts of skills outside of their specialities.



## Skillsets

Broadly speaking, there are three distinct skillsets that must be reconciled in data science.

1. **Algorithms:** You understand the theory of data science, the statistics, modelling rules and mathematics that are at the heart of any data problem. You understand how experiments are designed and measured. You understand the **algorithms and theory behind data science**.
2. **Engineering:** You understand the engineering required to source, process and store data. You should be aware of programming languages and distributed computing schemes that will help you deal with massive amounts of data at scale. You should understand **the programming that applies your theories to massive datasets. (Engineering)**
3. **Communication:** You understand how to **communicate your solutions, and how to relate those solutions to business problems**.

## How The Data Science Roles Look in Practice

Let's go through a sample project. A data science team might be assigned to use deep learning to classify images like Yelp's team did.

Millions of photos are uploaded on Yelp every single day, but it can be hard to get images you want for each restaurant. Sometimes, the photos uploaded are all of the same category—maybe they're all photos of the food, or the outside of the restaurant. A holistic evaluation of a restaurant requires images of many different kinds.

You can use machine learning to automatically categorize which images fall into what category. Computers can, with the help of a training set, tell you whether or not an image is out of the outside of the restaurant or of food.

**Data scientists** would create the model that would help machines create those distinctions. They would be able to think through the types of data they need, from manually tagged photos to keywords in image captions.

**Data engineers** would engineer systems to source all of the image data and store it, as well as implement some of the algorithms determined by data scientists at scale.

**Data analysts** would query and present the business implications of the change. Did it please users? How much more traffic did Yelp generate due to the recent change? These are questions data analysts would ask—they would then communicate the insights they found.

## Sample Data Science Profiles

### Data Scientist





DJ Patil, the Chief Data Scientist of the United States, is the perfect prototype of the Data Scientist. He brings a deep understanding of mathematics from his Ph.D. in applied mathematics. He has created multiple data products, and collaborated with people in various data science roles. He's headed up strategy and led teams to build out entire new extensions of LinkedIn's data, from the creation of "People You May Know", to Talent Match, a function that automatically sources the best candidate for any job posted on LinkedIn.

### **Data Engineer**



Doug Cutting, the creator of Hadoop and a member of Apache's Board of Directors is somebody who has dedicated his time to creating technical solutions to store and process data at scale. Hadoop is widely used to

distribute data across several hardware servers so that huge data sets can become manageable. Doug Cutting is the prototypical example of a data engineer and he is now the chief architect at Cloudera, one of the largest data engineering organizations in the world.

## Data Analyst



The humble author of this piece, while nowhere near as talented as the two individuals referenced above, did serve a brief stint as a data analyst for a pharmaceutical company. He analyzed the sales pipeline with Excel and shared the data in Powerpoint with his supervisor to determine budget choices. He was being a data analyst in that capacity.

## Conclusion

**Data science** is a new and exciting field that requires individuals who fit into specific data science roles to come together and solve cutting-edge problems. We hope we've demystified exactly how those roles work together—we have a whole lot more in our careers guide