

# Data Science: The Big Picture

by Matthew Renze

Start Course

Bookmark

Add to Channel

Download Course

Table of contents

Description

**Transcript**

Exercise files

Discussion

Learnin

## Course Overview

### Course Overview

Hi, I'm Matthew Renze with Pluralsight, and welcome to Data Science: The Big Picture. Data science is the practice of transforming raw data into actionable insight. In our data-driven economy this set of skills is in extremely high demand in commanding significant increases in salary as it's revolutionizing the world around us. However, most developer and IT professionals have not yet learned this valuable set of skills. In this course, we'll answer the following three questions, what is data science, why is it important, and where is this all going. To answer these questions we'll look at a series of trends that are driving the data science revolution, including data analytics, the internet of things, big data, and machine learning. Finally, we'll see how all of these trends are converging to create fully-autonomous, intelligent systems, which will likely revolutionize our economy and our world as we know it. By the end of this course you'll understand data science and the tools, technology, and trends driving the data science revolution. As an introductory course there are no prerequisites for this course; we'll be explaining everything you need to know along the way, so please join us today at Pluralsight and learn how to become part of the data science revolution with Data Science: The Big Picture.

# Introduction

## Introduction

Hello, and welcome to Data Science: The Big Picture. I'm Matthew Renze with Pluralsight, and in this course we'll learn about Data Science from a high-level perspective. Whether you realize it or not, there's a flood of data coming. In fact, the flood is already here and it's growing exponentially each year. In the next few years we're going to see two entirely different outcomes for individuals, businesses, and governments based on whether they learn to use these data to their advantage or not. Either these people, businesses, and governments are going to sink in the coming sea of data, or they're going to learn how to swim. So, I'm currently doing what I can in my professional career to learn how to swim in this new data-driven economy, and I encourage you to do the same. We see the same sentiment expressed by experts across various industries. In fact, I see articles like these popping up on a daily basis. There really seems to be something big happening in the world of data and data science. The reality is that there's a rapidly-growing demand for IT professionals with the ability to transform data into actionable insight. However, there are nowhere near enough people available with the proper skills and tools to meet this demand. This means that people with the right set of skills and the right tools are commanding significantly higher salaries in the IT industry right now. And those of us with data science skills are going to be leading the software industry through the next industrial revolution, an information revolution. So you might be thinking to yourself, you know, I'm not a data scientist, a data analyst, or a data engineer, so how does data science apply to me? The reality is, that as a software developer, a data professional, or an IT professional you work with data every day, and you can learn to apply data science practices to your data to identify patterns, to make better decisions, and to predict outcomes. In fact, if your job doesn't involve working with data it's likely that it may not exist in a few years. As we'll soon see, data science is automating the work around us. So I'm very happy that you decided to learn more about this topic and I'm excited to teach you a bit of what I know about data science.

## Overview

As an overview of this course, first we'll learn about data science, what it is and why it's important. Next, we'll learn about data analytics, the application of data science in the business world. Then we'll learn about the internet of things and how connecting devices and sensors to the cloud provides us with a tremendous stream of data to be analyzed. Next we'll learn about

big data and how to effectively store and process massive amounts of data. Then we'll learn about machine learning, software able to detect patterns, make decisions, and predict outcomes using data without explicitly being programmed to do so. Finally, we'll close the loop to show how all of these data science trends are converging to create fully-autonomous systems that behave rationally in their environment and learn without any human intervention. So let's get started. The purpose of this course is to address the following three questions, what is data science, why is it important, and where is this all going? In addition, by the end of this course I'm hoping that you'll be as excited as I am about the future of data science and what it means for those of us in the software industry. The intended audience for this course are software developers, data professionals, and IT professionals. This content is applicable to a wide audience including novices just starting out in the software industry and wanting to understand how data science will affect your career, all the way up to managers and executives wanting to understand how data science will impact your business. Essentially, this course is for anyone in the IT industry who wants to understand what data science is, why it's so important to our industry, and what direction these trends are moving. Since this is an introductory course there are no required prerequisites; however, I do want to point out that this course is intended to be a very high-level introduction to data science. We will not be seeing any code demos or performing hands-on data science work in this course. If you're looking for a more in-depth, hands-on course, please watch my Pluralsight course on Data Science with R. This course covers how to perform basic data science practices using R, which is one of the most popular programming languages for data science. However, I would still recommend that you start with this course first and then move onto that course next for more in-depth, hands-on knowledge. You can find this course at the following URL. So now, let's learn what data science is and why it's so important.

## Data Science

Now let's learn about data science, what it is, and why it's important. The first question we need to answer is, what is data science? Data science is an interdisciplinary field composed of computer science, math and statistics, and domain expertise that seeks to derive insight from data. Data science is the intersection of these three respective disciplines. The goal of data science is to transform data into knowledge, knowledge that can be used to make rational decisions. So what then is a data scientist? A data scientist is someone that performs data science for a living. However, a data scientist is more than just a scientist, more than just a data analyst, and more than just a developer. They possess all three sets of skills. Individuals with all three sets of skills and proper credentials are currently very rare. Despite how rare actual data

scientists are they are rapidly growing in demand. In addition, there doesn't appear to be any end in sight for the growth of this demand. As a result, in the very near future this specific set of skills will be in high demand, whether you're a data scientist or applying data science practices to your current job role. The rarity of data scientists combined with their high demand leads to the much higher salaries for data scientists and IT professionals with similar skills. The next question we need to address is, what set of skills are necessary for data science? In general, the skills commonly associated with data science are programming computers using programming languages like SQL, Python, and R, working with data, that is collecting, cleaning, and transforming data, creating and interpreting descriptive statistics that is numerically analyzing data, creating and interpreting data visualizations, that is visually analyzing data, creating statistical models and using them for statistical inference, hypothesis testing and prediction, handling big data, that is data sets that are too large to work with via conventional computing architecture, automating decision making and prediction using machine-learning algorithms, and deploying data science solutions into production to communicate results to a wider audience. So what tools are used to perform data science? Regarding the tools used in data science, here are the tools listed by respondents of O'Reilly's data science salary survey in 2015. On the X-axis we have the name of the tool, which includes programming languages, data platforms, or analytics tool, and on the Y-axis we have the percent of respondents reporting that they use the corresponding tool. As we can see, the most popular tool is Structured Query Language, or SQL. SQL, also pronounced S-Q-L, is a programming language used for querying tables of data in a relational database. SQL is a very important language in data science because of how much time data scientists spend exploring and cleaning up data. Next we have Excel. Many data scientists have a general dislike for Excel and prefer more powerful tools for working with data. However, the reality is, is that Excel is very easy to use and quite powerful. As a result, much of the business world today is still run on Excel spreadsheets. After this we have both Python and R. These two languages are the work horses of the data-science world. In almost every survey I've seen these two languages are usually neck and neck in terms of their popularity. However, both languages have their strengths and weaknesses. As a result I think it's important to be comfortable using both of these languages. Now on this, we have several other programming languages, data platforms, and analytics tools. It's important that you become familiar with these tools so that you know what they are, what problems they solve, and their strengths and weaknesses relative to comparable tools. Finally, we need to address the question, how is data science performed? The general data science process works like this. First, we find a question that we want to answer. This can be a hypothesis we want to test, a decision we want to make, or something we want to attempt to predict. Second, we collect data for our analysis. Sometimes this means designing an

experiment to create new data, other times the data already exist and we just need to find them. Third, we prepare the data for analysis, a process often referred to as data munging or data wrangling. We need to clean and transform these data to get them into a form suitable for analysis. Fourth, we create a model for our data. In the most generic sense, this can be a numerical model, a visual model, a statistical model, or a machine learning model. We use this model to provide evidence for or against our hypothesis, to help us make a decision, or to predict an outcome. Fifth, we evaluate the model. We need to determine if our model answers our question, helps us make a decision, or creates an accurate prediction. In addition, we need to make sure that our model is appropriate given our data and the context. Finally, if everything looks good, we deploy our model. This could mean communicating the results of our analysis to others, making a decision and acting upon our decision, or deploying an application into production. We then repeat this process for each question we would like to answer using feedback from our previous results to help guide our process. Data science is typically an iterative process. We typically go through the complete cycle multiple times learning and improving with each iteration. In addition, this process is often non-sequential; we often have to bounce back and forth between steps as we discover problems and learn better ways of solving these problems. In addition, there are times when we don't need to complete the process. Often, we learn that what we're doing isn't working or doesn't make sense given our data or context, so we terminate the process and shift our focus to the next most important question in our to-do list instead. There are some well-established practices for the data science process available, like the CRISP-DM process, which stands for Cross Industry Standard Process for Data Mining. These established processes are useful to help you get started with your data-science process. So now, let's learn why data science has become so important in recent years.

## Importance of Data Science

So why has data science become so important all of a sudden? To answer this question we need to look at a series of emerging trends that are driving the importance of data science. First, we'll learn about data analytics, the trend of applying data science's practices and tools in the business world. Second, we'll learn about the internet of things, the trend of connecting devices and sensors via the cloud, which is generating massive streams of data to be analyzed. Third, we'll learn about big data, a trend of creating tools and systems able to store and process these enormous data sets at scale. Fourth, we'll learn about machine learning, a trend in artificial intelligence of teaching machines to solve problems without explicitly being programmed to do so. Machines able to make decisions and predictions all by identifying statistical patterns in these

massive data sets. Finally, we'll close this loop to show how all four of these trends are converging to create fully-autonomous, intelligent systems, machines capable of acting rationally within their environment, and learning how to optimize their performance over time without any human intervention. For each of our four main topics we'll learn about trends within each topic that have been occurring over the past few decades. We'll also learn where these trends currently are at today, and we'll try to predict where these trends appear to be going in the next decade or so. Ultimately, the importance of data science all boils down to two things, it's being driven by economics and it's being made possible by technology. The cost of collecting, storing, processing, and analyzing data is continuously decreasing, and the value that these data can provide when the data science process is applied continues to increase. As a result, data science has now become a cost-effective strategy for answering questions, making decisions, and predicting outcomes in a wide variety of scenarios in our world. Given this trend, it's unlikely that the demand for data science will decrease any time in the near future. So now, let's wrap things up on this module so that we can get started learning about the various trends driving the importance of data science.

## Summary

In this module, first we learned what data science is. We learned that the goal of data science is to transform data into knowledge, knowledge that can be used to make rational decisions. Then we learned about the skills that are important for data science, the tools that are used, and the data science process. Finally, we provided a basic argument for why data science has become so important in recent years. The remainder of this course will examine this more in depth by looking at each of the four key data science trends and discuss where these trends appear to be going in the future. In the next module we'll learn about data analytics, the application of data science practices to the world of business.

# Data Analytics

## Introduction

Hello again, and welcome to our next module on Data Science: The Big Picture. I'm Matthew Renze with Pluralsight, and in this module we'll learn about data analytics, the application of data science practices to the world of business. As an overview of this module, first we'll learn about

data analysis, the predecessor to data analytics and the core of data science. Next we'll learn about data analytics, the application of data science practices to the world of business. Then we'll learn about data-driven decision making, the future trend of data science in the business world. Finally, we'll see a few examples of industries that have been significantly disrupted by the application of data science to their professions, so let's get started.

## Data Analysis

First, let's take a step back into the past, a few decades ago, to discuss data analysis, the predecessor to data analytics. Just a few decades ago, data analysis was largely the domain of two kinds of people, scientists doing research and fields like physics, chemistry, and economics, and statisticians, which were a bit of rarity in the business world, and when you found them were typically found in industries like insurance and finance. Back then, testing a hypothesis with data wasn't easy. We had to design an experiment, run the experiment, collect data, and then analyze the data to provide evidence for or against our hypothesis. This took significant time, money, and effort. To analyze the data we had complex proprietary software. These tools were expensive, took a long time to learn, and required a deep understanding of statistics in order to use them properly. To communicate our findings we typically used numerical reports, clunky data visualization, and scientific articles. Communicating insight to a wider audience was costly, often ineffective, and relatively slow. Ultimately, in the past, collecting, analyzing, and communicating data was difficult, expensive, and slow. So now, let's look to the present to learn about data analytics, the adoption of data science practices in the world of business.

## Data Analytics

Now let's learn about data analytics, the application of data science practices to the world of business. Today, the marketing departments of the world have renamed data analysis to the much flashier term data analytics. However, despite the flashy new title, data analytics is still good old-fashioned data analysis at the core, but it's improved in several ways. Most importantly, it's spread from the domain of science to the world of business as well. This means that it's now widely acceptable to use data analysis in a scientific method to solve common business problems. In addition, data are now inexpensive and readily available. We can run an A/B test on a website with a simple flip of a feature toggle. This means that we can collect data easily and with significantly less cost. In fact, oftentimes, we don't even need to run an experiment to collect data, the data already exist, we just need to find them and analyze them. We have much better tools

for performing data analysis as well. We now have easy-to-use spreadsheet software like Microsoft Excel for performing numerical analysis. We can also create interactive charts and graphs using software like Tableau, which makes creating data visualizations relatively quick and easy. And we have open-source programming languages like R and Python that make more rigorous data analysis inexpensive and relatively easy as well. We also have a variety of ways to communicate our findings to others. For example, we can convey key performance indicators of our business to executives, management, and employees using dashboards. In addition, we can convey complex information about our business to a wider audience using infographics that allow users to rapidly consume and digest data. Ultimately, today collecting, analyzing, and communicating data is easy, inexpensive, and fast. So now, let's look to the future to see the direction of trends driving data science in the world of business.

## Data-driven Decision Making

So what will the future of data analytics look like? Current trends indicate that we're moving towards a world of data-driven decision making, an age when all non-trivial business decisions will be made using data rather than guesswork or gut instinct. In the future we won't have to spend time collecting data like we did in the past, we'll have almost all the data we need right at our fingertips. This trend is being referred to as the democratization of data. In the business world, this means that employees will have access to significantly more of their company's data. Data isn't just for managers and executives anymore. This means that anyone in the company can run an experiment to test a hypothesis and use data to drive their own decision making. In addition, we see this happening in the public sector with open data initiatives. Governments are opening up their previously private datasets and scientific research datasets are being made public as well. We're also seeing this trend emerge in our personal lives. We have devices like fitness trackers and sleep trackers that allow us to optimize our daily lifestyles. In crowd-sourced initiatives involving personal health data that help people with medication, side effects, and managing symptoms. In addition, we have a new generation of tools that will allow almost anyone to perform their own complex data analyses. We'll have self-service business intelligence tools that allow almost any employee to analyze their organization's data. We'll also have natural-language query applications that allow you to speak a question and receive an answer based on your organization's data, similar to how Amazon's Alexa or Apple's Siri work, however, you'll be asking questions about your own business using your company's own data. To communicate information we'll have new tools that allow us to tell better stories with our data. For example, interactive and immersive data visualization applications using augmented reality and virtual



reality. Imagine putting on a HoloLens or an Oculus Rift and being able to immerse yourself within your company's data. This is leading to the data-driven enterprise, companies that value information as an asset and use data to drive their decision making. Initial findings from MIT research show that companies that are implementing data-driven decision-making practices had 4% higher productivity and 6% higher profits than traditional businesses. In the highly-competitive landscape of the future, companies that are not data-driven may eventually cease to exist. As a result, it's important to ask yourself if your organization is on the road to becoming a data-driven enterprise, and if not, what can you do to help it make better decisions using data? However, empowering people to make better decisions is not the end goal of data science, rather, as we'll soon see, it's just the beginning. So now, let's take a look at a few examples of industries that have been disrupted by the introduction of data science into their professions.

## Examples

Before we wrap up this module let's take a look at three examples of industries that have been completely disrupted by data science. I think it's important to see some real-world examples of how data science is changing the world of business in order to anticipate the impact that data science may likely have your industry. First, the financial industry. Just a couple decades ago the trading floors of Wall Street and other exchange markets were filled with people yelling and screaming to buy or sell stocks, bonds, and other securities. At the intersection of every trade were people using a combination of information, guesswork, and gut instinct to make trading decisions. However, trading room floors are mostly empty these days. Stock trading shifted from human interactions on the trading floor to quantitative analysts sitting behind computer screens in backroom offices. Quantitative analysts, or quants as they're simply known, use data science to scientifically investigate investment hypotheses and build models to predict investment outcomes. In recent years, however, even quants have replaced themselves with automated intelligent trading algorithms. Today, lightning-fast, high-frequency trading algorithms now account for the majority of all trades in the modern stock market. As a result Wall Street has been forever changed by data science. Second, we have baseball. Prior to 2002 picking a line up a players in the major-league baseball draft largely relied on traditional measures of player performance, like stolen bases, runs batted in, and batting average. In addition, the scouting profession was full of human biases that distorted the true value that each player brought to a baseball team. However, in 2002, the Oakland Athletics baseball team used data science to select players that were undervalued by the market. Through rigorous statistical analysis the Oakland A's discovered that metrics like on-base percentage and slugging percentage were better

indicators of offensive success. Despite their significantly smaller budget for player salaries, the Oakland A's created a team that was competitive against the much-better funded teams. This data-driven strategy contributed to the Oakland A's setting a record 20-consecutive wins and brought them to the playoffs in 2002 and 2003. As a result, the game of baseball and how players are valued has also been forever changed by data science. Finally, the insurance industry. The insurance industry was an early adopter of the use of data and statistics in the business world. These early data science practices were used to calculate the aggregate risk of groups of individuals being ensured. However, the insurance industry is currently going through a radical transformation in our merging data-driven economy. We now have significantly more data and much better data available in order to calculate risk. People have personal fitness monitors that contain valuable health and fitness data, automobiles are being outfit with sensors and internet connections that allow them to collect and transmit valuable telemetry data. In addition as autonomous vehicles with near-perfect driving begin hitting the road, the dynamics of auto insurance are certain to change as well. As a result, new companies that are learning how to leverage modern data science practices have the potential to completely disrupt the established insurance industry giants who are scrambling to modernize their systems, employee skills sets, and company culture. We are already seeing some pretty dramatic changes emerging in the insurance industry, and the landscape of this industry will likely continue to change for years to come.

## Summary

In this module, first, we learned about data analysis, the predecessor to data analytics in the core of data science. We learned that collecting, analyzing, and communicating data was difficult, expensive, and relatively slow. Next, we learned about data analytics, the application of data science practices to the world of business. We saw how collecting, analyzing, and communicating data has become relatively easy, inexpensive, and fast. Then, we learned about data-driven decision making, the future trend of data analytics were all nontrivial business decisions are made using data. We also saw how this is leading to the emergence of the data-driven enterprise, companies that value information as an asset and use data to drive their decision making. Finally, we saw a few examples of industries that have been disrupted by data science, and how this radically transformed each of them in ways most people could not have predicted. In the next module we learn about the internet of things and how connecting devices and sensors to the cloud is generating tremendous streams of data to be analyzed.

# Internet of Things

## Introduction

Hello again, and welcome to our next module on Data Science: The Big Picture. I'm Matthew Renze with Pluralsight, and in this module we'll learn about the internet of things, the connection of devices and sensors to the internet creating a tremendous amount of data to be analyzed. As an overview of this module, first we'll learn about the internet of the past, we'll learn about the technological and economic constraints that govern the internet and the limited amounts of data that the internet was generating. Next, we'll learn about the internet of things, the interconnection of devices and sensors via the cloud that's now generating tremendous amounts of data. Then we'll look to the future to learn about the internet of everything. We'll see what types of data may be available to be analyzed in the very near future. Finally, we'll look at a few examples of devices that have recently been connected to the internet to see how the data they generate can create completely new value. So let's get started.

## The Internet

A few decades ago we had the internet, just the plain old internet. The internet was originally designed for people. It was designed to allow humans to communicate, collaborate, and exchange information. The cost to connect to the internet was relatively high, the speed of our internet connections was slow, bandwidth was low, and where it connected to the internet via physical wires. In addition, the web was largely used for consuming information. Producing new information for the internet was rather difficult and required quite a bit of technical knowledge to create even a simple web page. Most of the data that was generated by the internet was collected via basic web forms. These forms allowed people to enter basic information like name, address, telephone number, and email address. The internet really wasn't generating all that much data. Ultimately, in the past, the internet was expensive, slow, and not generating all that much data. So now, let's look to the present to see how the internet of things has changed all of this.

## The Internet of Things

Today, the internet isn't just for people anymore, it's an internet of things as well. The internet of things is the emerging trend of connecting all devices, which are things, via the internet. It's also a trend of connecting sensors that monitor physical objects, which are also things as well. The

internet of things is being driven by economics. The cost to connect a device to the internet is falling, connection speeds are increasing, bandwidth is increasing as well, and wireless connections are now abundant and cheap. This means that it is now cost-effective to connect devices and sensors to the internet to allow them to collect and transmit their data to the cloud. These data, when analyzed, are able to create additional value that justifies the cost of internet enabling these devices. For example, connecting your fitness tracker to the internet allows you to track your physical exercise, sleep patterns, heartrate, and other vital information. Connecting your automobile to the internet provides you with real-time navigation, can help you locate a missing car, and can even lower your auto insurance rates for practicing safe driving habits. In connecting your house to the internet allows you to monitor your house while you're away, manage it remotely, and can even help you lower your energy bill by optimizing energy usage over time. As a result of the economics driving the internet of things we've seen an explosion in the growth of IoT devices in the past decade. In fact, it's predicted that the total number of internet-enabled devices will reach 50 billion devices by 2020, that's over 6 internet-connected devices for every human on the planet. In addition to data being generated by new IoT devices and sensors, today's internet is also deriving additional value from data exhaust. Data exhaust is the byproduct of people's online activities. It's data that's created as a result of someone visiting a website, buying a product, or searching for something using a search engine. For example, every time someone searches for cold or flu symptoms data is generated about the search terms they use and where they were located when they performed the search. Search engine companies can use this data exhaust to determine where cold or flu outbreaks might occur based on the number on people searching for cold or flu symptoms in a specific geographic region. As a result, we can now derive additional value from the internet of the past by analyzing data exhaust to find secondary value streams. With the internet of things and data exhaust the data that are being generated by the internet are no longer limited to just simple web-based forms. Instead, data sources are now devices, sensors, activities, audio, video, text, and more. Swarms of people, devices, and sensors now generate a continuous stream of data into the cloud. Ultimately, today, the internet of things is cheap, fast, and generating tons of data, data that can be analyzed using data science practices to extract new value and insight. So now, let's look to the future to see what the internet of things may eventually look like.

## The Internet of Everything

So what will the internet likely become in the future? Some experts predict that in the future the internet will become the internet of everything. Economics will continue to drive the internet to

become the internet of everything. The cost to connect people, devices, and sensors will continue to drop, the speed will continue to increase, the bandwidth will get wider, and everything becomes wireless. As a result, as long as the value derived from a device's data is greater than the cost to connect the device to the internet, there will be a business case to internet-enable the device. So most, if not all devices will likely become part of the internet of things. In fact, in the future, an internet connection will likely be as common to devices as electricity is today. At this point, all human artifacts will likely be connected to the internet. This means that data sources are essentially anything of important in our world. All people, devices, buildings, cities, etc., essentially become a source of data to be analyzed. We're currently in the midst of building a peripheral nervous system for our planet. We'll connect everything in our world using the internet. However, like our own nervous system these sensors will need to be connected to a central nervous system to store and process these data. Essentially, the internet of things needs a brain, which leads us to our next topic, big data. But first, let's take a look at a few existing devices that are being completely transformed by the internet of things.

## Examples

Before we wrap up this module, let's take a look at three examples of new IoT technologies. I think it's important to see a few real-world examples of how the internet of things is taking traditional items that we've used for decades and is creating new value with the data collected from internet-enabled devices and sensors. In fact, some of these examples might surprise you. First, lighting. The lightbulb has remained relatively unchanged since it was first invented in 1879. We've seen the technology transition from incandescent, to florescent, to now LED lighting, but the basic design and control principles remain the same. However, in recent years, lightbulbs have become an internet of things device. Through the use of power over Ethernet technologies, both power and data can now be sent to lighting fixtures via a single Ethernet cable. This provides extremely fine-grained sensing and control of every lighting asset in a building, which allows new ways to maximize energy efficiency, it allows occupants to control light levels in their own spaces, and allows for new ways to communicate information through lighting. Next, we have the toilet. Once again, flushable toilets have remained relatively unchanged since they were first invented over a century ago. Water comes in and waste products go out, a relatively simple, yet effective technology. However, even toilets are now becoming part of the internet of things. Smart, internet-connected toilets can track how frequently they are flushed, whether they are clogged or not, and can even play your favorite music. In the near future they will likely be equipped with sensors that can detect key health indicators, pregnancy, bacterial infections, and even prostate

cancer. Finally, we have the cow. While most people wouldn't think of cows as an IoT technology, in recent years, cows have begun to be outfitted with mobile IoT sensors. These sensors allow cattle farmers to track their cow's movement, behavior, fertility, and even lactation. In addition, farmers are now using aerial drones equipped with cameras and sensors to monitor their livestock and fields as well. Just imagine all of the things in our world that will become sources of data to be analyzed in the very near future.

## Summary

In this module, first we learned about the internet of the past. We saw that yesterday's internet was expensive, slow, and wasn't generating all that much data. Next, we learned about the internet of things, the interconnection of devices and sensors via the cloud that's generating tremendous amounts of data to be analyzed. Then, we learned about the internet of everything, a possible future when all human artifacts will be connected to the internet, so that their data can provide us with additional value. Finally, we saw three examples of how the internet of things is creating new value using data collected from traditional devices and things in our world. In the next module we'll learn about big data, technologies able to store and process data sets beyond the limitations of traditional computing architecture.

# Big Data

## Introduction

Hello again, and welcome back to Data Science: The Big Picture. I'm Matthew Renze with Pluralsight, and in this module we'll learn about big data, data sets beyond the capabilities of conventional computing architecture and the technologies we use to extract insight from these data sets. As an overview of this module, first, we'll learn about the data of the past. Data sets that were relatively small, slowly processed, and had little diversity. Next we'll learn about big data, data sets beyond the limitations of conventional computing architecture and the technologies we used to extract insight from these data sets. Then we'll look to the future to see how the big data trend may likely return to being just data again. Finally, we'll see a few example of novel insights that have been extracted from large data sets using big data technologies. So let's get started.

# Data

In the past, big data was just called data. Our data sets were very small compared to modern standards. For small data sets we stored and processed them using a single desktop computer. All of the data could fit either in memory, or at least fit on the hard drive and be processed record by record. For larger data sets we stored them on a bigger, more powerful computer called a server. These machines had faster processors, more memory, and larger hard drives. For the largest data sets we stored on an even bigger and more powerful server with an array of hard drives that could fit all of the data on a single machine. In the past, processing our data and our need to respond to data was pretty slow. We generally processed data in daily batches, we just needed to see our sales figures once a day in order to make timely business decisions. In addition, our data was quite homogeneous as well, There really wasn't much diversity. Most data sets were tabular data sets, that is, all the data was stored in the columns and rows of a table, and we stored these tables of data in simple flat files or more complex relational databases. Ultimately, in the past, data sets were small, slow, and had little diversity compared to today's data sets.

## Big Data

Today we have a trend the marketing folks have called big data. So, beyond the marketing hype, what is big data? The term big data is often used to refer to two things. First, big data is often used to refer to the rapid increase in the amount of data being generated by our world. The amount of digital information our world creates and stores is doubling every two years. This means that in the past two years we have created more data than the entire history of the human race, and there doesn't appear to be any end to this trend in sight. In addition, the term big data is more specifically used to describe data sets that are of a volume, velocity, or variety that's beyond the capabilities of conventional computing architecture. Big data is generally said to exist when one or more of these conditions are met. By volume, we mean the size of the data. Big data means data sets too large to process in memory on a conventional computer, or too large to fit the entire data set on a conventional hard drive. As a result, we now have distributed computing technologies like Spark and Hadoop, which spread out data over multiple computers. With distributed computing, both storage and processing of data are spread out amongst several computers. So, when we query a data set, our query is broken up into small pieces, and all of the computers work together to each perform their own piece of the total work on the part of the data set that they're responsible for. Distributed computing also allows us to create data lakes. Rather than storing all of our data in a more expensive relational database or traditional data warehouse, we can now store less- frequently used datasets in their native form in a much less

expensive data lake. Data Lakes create a central repository for all the company's raw data. From here, the data can either be queried directly using distributed queries or subsets of the overall data can be loaded into a data warehouse for self-service ad-hoc queries. By velocity, we mean the speed of extracting insight from the data. Processing data over night in daily batch operations isn't quick enough to remain competitive anymore. So, we have real-time analytics technologies that allow us to make decisions in real time. For example, to detect credit card fraud, we can't just wait overnight for a fraud detection algorithm to analyze the data and slowly make a determination. We need to be able to detect fraudulent activities after the card has been inserted in the credit card reader, but before the transaction completes. In addition, we have streaming analytics tools that allow us to detect patterns in fast moving streams of incoming data. For example, we need streaming analytics to be able to analyze the 10-GB per second streams of data coming from 5, 000 sensors in a modern jet engine to determine if it's about to fail in flight or not. By variety, we mean the various types of data we need to process. We're not just dealing with tabular data in a single database anymore, we're dealing with text, images, audio, video, and more, coming from a wide variety of data sources. To handle the wide variety of data we now have data integration tools that allow us to combine multiple datasets together to create a unified view of our world. In fact, much of the value from analyzing big data sets comes from joining them together to create new insights that were not possible before. These new tools make this type of data integration significantly easier and more cost effective than it was in the past. We also have new tools to extract features of interest from structured, semi-structured, and unstructured data. For example, we can now extract names and properties of object and images, text from audio recordings, keywords, summaries, and sentiment from bodies of text, and analyze videos for content. This means that we now have access to information in these data sets that would have previously required a human to analyze. Ultimately, today's data sets are bigger, faster, and more diverse than the datasets of the past. Big data technologies allow us to handle the volume, velocity, and variety entailed by these modern data sets. So now, let's look to the future to see where these big data technologies appear to be moving.

## Just Data Again

In the future, it appears that we may be going from the era of big data to just data again. Now, I'm not suggesting that data sets will be getting smaller again, in fact, quite the opposite appears to be true. Rather, that computer architecture and software design is fundamentally changing to support big data by default. As a result, it's likely that the marketing buzzword big data will just fade into obscurity. There's a few reasons why this appears to be the case. First, the cost of



storage is going to continue to drop. As a result, storing data will become so low that it'll essentially be free. In fact, many cloud service providers are just giving away free storage in order to entice you to store your data with them so that they can charge you for processing in additional services. In addition, distributed computing technology is changing the way we store our data. We're nearing a point in time when storage becomes completely elastic, there'll be no need to even think about how much disk space you need in order to store all of your data, you'll just store your data in the cloud and the cloud will automatically continue to allocate more space to meet all of your storage needs. In addition, the computer as we know it is going through a transformation. Distributed computing technology is changing the way we process data. Algorithms are becoming highly parallelized and processing power is almost completely elastic as well. There'll be no need to buy a powerful computer in order to handle a data processing task, the cloud will automatically distribute your work load and scale processing power as necessary. We'll stop programming the computer and instead begin programming the cloud. In essence, the data center will become the computer. Finally, our ability to handle a wide variety of data will continue to improve. We're seeing the emergence of automated data integration tools that can eliminate much of the work of integrating various data sets. We're also seeing the emergence of new automated-feature extraction tools that can automatically extract even more complex features from unstructured data sets. This will provide us with new data that we've never before had cost-effective access to. We're building tools to automate the extraction of insight from big data. However, we're beginning to get ahead of ourselves again, because to accomplish this we need machine learning. But first, let's take a look at a few examples of insights that have been derived from analyzing big data.

## Examples

Now let's take a look at three examples of insights that have been extracted from big data. Some of these insights might seem obvious to you, however, others might be a bit of a surprise. First, beer and diapers. Most people would probably assume that beer and disposable diapers wouldn't have much in common. However, in 1992 Osco Drug performed what was at the time a big-data analysis of over 1.2 million sales transactions in 25 stores. Their analysis was attempting to determine which products were commonly purchased together. Of the product couplings identified one of the most surprising insights was the fact that both beer and diapers were commonly sold together. As a result, it's been rumored that some retail stores use this insight to locate beer closer to the disposable diapers to increase the likelihood of impulse purchases. Whether these rumors are true or not, the beer and diapers legend has now become part of big

data folklore. Second, back in 2012 the New York Times wrote an article about how companies learn your shopping habits. One of the more interesting stories in the article was about how statisticians working for the major US retailer Target discovered a set of 25 products that were likely to predict a specific major life event for the person who purchased them. Among the products were mineral supplements, like magnesium, calcium, and zinc, extra-large bags of cotton balls, and larger than usual amounts of unscented lotion. Can you guess which major life event these purchases predicted? Well, if you said pregnancy, you'd be correct. Those 25 product purchases could be used to assign each customer with a pregnancy prediction score. In fact, the data could even be used to estimate the customer's due date within a small window so product coupons could be timed at very specific stages of the customer's pregnancy. It really makes you wonder what other major life events our shopping histories allow companies to predict even before our friends and family know. Finally, big data analysis has brought us another interesting insight involving the products people purchase right before a hurricane makes landfall. As hurricane Francis approached the southeastern part of the United States in 2004 Walmart decided to query its 465 TB of customer sales data to see what products sell most frequently before a hurricane. While one might expect to see sales of flashlights, bottled water, and other essential survival items at the top of the list, one top-selling product was a complete surprise. Interestingly, sales of strawberry Pop-Tarts increased sevenfold in the days leading up to a hurricane. As a result, Walmart sent trucks filled with these breakfast pastries to the stores in the path of the hurricane. The additional Pop-Tarts sold extremely well, and as a result, Walmart now regularly stocks its shelves with extra Pop-Tarts prior to a hurricane.

## Summary

In this module, first, we learned about the data of the past. Data sets that were relatively small, slowly processed, and had little diversity. Next, we learned about big data, datasets of a volume, velocity, or variety beyond the limitations of conventional computing architecture and the technologies that allow us to work effectively with these datasets. Then we learned about the future of the big data trend and how fundamental changes in the way we create and use computers and software may likely return things to being just data again. Finally, we saw a few examples of novel insights that have been extracted from large data sets using big-data technologies. In the next module, we'll learn about machine learning, software able to learn how to solve problems without being explicitly programmed to do so.

# Machine Learning

## Introduction

Hello again, and welcome to our next module on Data Science, The Big Picture. I'm Matthew Renze with Pluralsight, and in this module, we'll learn about machine learning, software able to learn how to solve problems without being explicitly programmed to do so. As an overview of this module, first, we'll learn about the artificial intelligence of the past, the first generation of intelligence software that was able to make rational decisions, but required being explicitly programmed to do so. Next, we'll learn about machine learning, the next generation of artificial intelligence, software able to solve tasks without being explicitly programmed to do so by detecting statistical patterns and data. Then we'll learn about deep learning, the future of artificial intelligence and machine learning which involves deeply stacked layers of machine learning able to solve more complex problems without explicit programming. Finally, we'll see three examples of tasks where deep learning algorithms now outperform their human counterparts. So let's get started.

## Artificial Intelligence

In the past, we had artificial intelligence, however, it wasn't all that intelligent. We had some machines that were capable of making rational decisions, however, they had to be explicitly programmed to make these decisions and they could only operate successfully in very constrained environments. So the bulk of decisions were made by humans. We still used machines to help us collect, store, and analyze data, but ultimately, it was human beings making the decisions. There was a lot of hype about what artificial intelligence would be able to do. Experts predicted that machines would soon replace human labor, but it never happened. By the end of the 90s, machines couldn't even solve basic general-purpose tasks that even a toddler could solve. The inflated hype about the potential of AI and its subsequent disillusionment when it never happened led to what is now referred to as the AI winter. It was a period of time between the late 70s into the early 2000s where funding for research and startups involving AI had almost entirely dried up. However, by the mid-2000s, the AI winter has ended and things are warming up again with the emergence of modern machine learning.

## Machine Learning

Today, most artificial intelligence research is focused on machine learning. So what is machine learning. Machine learning is a subfield of artificial intelligence based on statistics. It involves machines learning how to complete tasks without being explicitly programmed to do so. Essentially, with machine learning we use existing data to learn a function that can make a prediction given new data. For example, imagine that we want to create a function to determine whether a photo contains an image of a cat or not. First, we would need to create a data set that contains images with cats and images without cats. We would then have humans label each photo indicating whether it contains a cat or not. Next, we would apply a machine learning algorithm to that dataset of images with and without cats. The algorithm would learn a function that predicts whether an image contains a cat or not. Finally, if we've done everything correctly, we should be able to provide the function with a new image and it will tell us whether it contains a cat or not. Well this is a vastly oversimplified explanation of machine learning, it captures the essence of what we're attempting to accomplish. Some examples of tasks that machine learning algorithms can perform are classification where we make a decision or a prediction involving two or more categories or outcomes, for example, deciding whether to accept or reject a loan based on data from a customer's financial history. Regression, where we attempt to predict a numeric outcome based on one or more input variables. For example, how much will a house sell for based on the features of the house compared to the sale price of similar houses. Clustering, where we group similar objects together based on similarities and their data, for example, grouping customers into marketing segments based on their income, age, gender, number of children, etc. An anomaly detection where we find observations in the data that are different from the normal data, for example, detecting an unusual spike in the number of negative comments about a new product that's just been released. But it doesn't just stop with tabular data. Machines can perform classification, regression, clustering, and anomaly detection on a variety of data sources including images, text, audio, and video. With machine learning, computers can now be taught or teach themselves how to complete all of these tasks and more. In fact, machine learning algorithms are now better than humans at many tasks, some of which might surprise you. For example, modern machine learning algorithms now outperform humans in most games like Chess, Go, 8-bit video games, and even Jeopardy, handwritten character recognition, predicting a person's age from a photo, lip reading, and much more. In the next generation of machine learning, we'll be able to complete even more complex tasks, which leads us to deep learning, the future of artificial intelligence and machine learning.

## Deep Learning

So what does the future of machine learning look like? In the past few years, we've seen a series of major breakthroughs in machine learning. This has led to a new type of machine learning called deep learning. Deep learning as we'll see is the future of artificial intelligence. Deep learning is a form of machine learning that stacks multiple layers of machine learning models one on top of another to form a hierarchy. For example, if we want to teach a deep neural network how to detect human faces, first in the input layer, we would feed a set of labeled images of human faces into the network in order to teach it what different faces look like. The first hidden layer of the neural network would learn to detect geometric primitives, for example, horizontal lines, vertical lines, diagonal lines, and more. The second hidden layer would learn to detect more complex facial features, for example, an eye, a nose, or a mouth. The third hidden layer would learn to detect the general pattern for entire faces. And the output layer would learn to detect the most abstract representations of a person, for example, the name of the person being recognized. Deep neural networks are significantly more powerful and accurate in their pattern detection, decision making, and prediction accuracy than the previous generation of machine learning algorithms. In fact, these deep learning algorithms are rapidly becoming better than any human on the planet at a variety of complex mental and physical tasks. As a result, these new deep learning algorithms will likely replace human decision makers in a variety of tasks in the next few decades. These recent breakthroughs in deep learning have led to speculation that we may be on the verge of a huge breakthrough in computer science, statistics, and neural science. Some experts believe that we may be on the verge of discovering a universal learning algorithm, a general purpose artificial intelligence algorithm that can evolve to learn any function that a human can learn given sufficient data and training time. We call this artificial general intelligence, and if or when this happens, it will be the single most significant event in human history. Now I should point out that I'm currently cautiously optimistic about this. From what I've seen and the experts I've spoken to, artificial general intelligence seems plausible. However, we saw what happened the last time when there was too much hype in the artificial intelligence industry. Whether this happens or not, machines are about to get much smarter and change our world in very fundamental ways, which leads us to our next topic, closing the loop. But first, let's take a look at three examples of deep learning algorithms that now outperform their human counterparts.

## Examples

Now let's see a few real-world examples of modern machine learning applications to see how they're rapidly becoming better than humans at certain tasks. First, speech recognition. You may have noticed that in the past few years, your smart phone has gotten quite a bit better at voice

recognition. The recent leap in improvement is a result of deep learning algorithms being applied to speech recognition. These algorithms are rapidly approaching human level accuracy and will likely beat humans in the next few years or already have depending upon when you're watching this video. However, computers are about to become much easier to interact with in general because several deep learning techniques can be combined to make them much better predicting your intentions. For example, in addition to speech recognition, computers are now able to perform lip reading better than humans, which when combined with speech recognition improves their ability to accurately recognize commands, especially in noisy environments like automobiles. This can be combined with natural language processing which understands sentence structure, grammar, and context to allow more human-like conversations with computers. And by applying sentiment detection to the words you're speaking combined with emotion detection in your face as you're speaking them, the machine can even understand the emotional state you're in and respond most appropriately. As a result, interacting with computers in the near future will likely become much easier and more natural. Next, image generation. We've already seen how we can use machine learning algorithms to perform classification and regression on images, for example, attempting to predict a person's gender, age, and emotional state. However, what you may not know is that machines are now getting exceptionally good at generating images as well. In fact, the image on the right was created by a deep learning algorithm that attempts to predict what I'll look like 20 years in the future. It was taught how to do this by seeing a large number of images of people aging over the years. As a result, it learned how to predict the changes in facial features that occur as we age so that it can apply them to new images that it's never seen before. The results are better than most humans could produce with image editing software and probably better than most police sketch artists. In fact, the application is so good that it was able to fool the age detection app that I used earlier into thinking that I'm exactly 20 years older in the simulated picture. Finally, medicine. One of the areas where deep learning is predicted to make a considerable impact is the world of medicine. In the past, diagnosis of medical conditions and recommendations for treatments have been the task of human doctors only. However, in recent years, deep learning algorithms have beaten even the top doctors in the world at a variety of medical tasks. Machines can now outperform human doctors in several areas of diagnostics, for example, predicting heart failure as much as nine months before human doctors using traditional means, and detecting traumatic brain injuries, cancer, fractures, fetal distress, and other medical conditions found using various medical imaging techniques. In addition, these machines now outperform human doctors in several areas of treatment as well, for example, tailoring cancer treatment plans based upon individual genetics to maximize the likelihood of success in minimized side effects and prescription recommendations to avoid potential side effects and

other adverse reactions from using combinations of drugs. While human doctors will likely be around for quite some time, some experts predict that in the near future doctors will likely have a consultation with a machine-based medical expert prior to diagnosing or administering treatment for medical conditions.

## Summary

In this module, first, we learned about the artificial intelligence of the past, the first generation of intelligence software that was able to make rational decisions, but required being explicitly programmed to do so. Next, we learned about machine learning, the next generation of artificial intelligence, software able to solve tasks without being explicitly programmed to do so by detecting statistical patterns in data. Then we learned about deep learning, the future of artificial intelligence and machine learning. Deep learning involves deeply stacked layers of machine learning able to solve more complex problems without explicit programming. Finally, we saw three examples of tasks where deep learning algorithms now outperform their human counterparts. In the next module, we'll close the loop to see how all of these data science trends are converging to create fully autonomous intelligent systems.

# Closing the Loop

## Introduction

Hello again, and welcome to our final module on Data Science, The Big Picture. I'm Matthew Renze with Pluralsight, and in this module, we'll learn about the future of data science and we'll wrap things up for this course. As an overview of this module, first, we'll close the loop to learn how all of these data science trends are converging to create fully autonomous intelligent systems. Next, we'll learn about how these autonomous systems will impact our economy and our society. Then we'll learn where to go for more information. And finally, we'll wrap things up for this module and for the course as a whole. So let's get started.

## Closing the Loop

So far in this course, we've discussed data science and the four trends that are driving data science forward. However, now we're going to take things a step further in order to answer the

question, what will the future of data science look like. Essentially, where is this all going? I think the best way to answer this question is to close the loop. First, we have data analytics, the application of data science practices to the world of business. This is driving the widespread adoption of data science as an acceptable method of making rational decisions in the business world. Second, we have the Internet of Things, a peripheral nervous system of devices and sensors generating massive streams of real time data about the world around us. Third, we have big data, a central nervous system of elastic and scalable distributed computing technology able to store and process massive quantities of data in real time. Fourth, we have machine learning, software with the ability to detect patterns, make decisions, and predict outcomes without being explicitly programmed to do so. These four trends are all feeding into one another in a figurative sense. Each of these trends is driving demand for the other trends and each of these trends is generating new ideas and technology, which helps improve neighboring trends. However, these four trends are also feeding into one another in a much more literal sense. First, we have the devices and sensors in the Internet of Things sending massive streams of data to big data infrastructure. Next, the data stored in process by big data infrastructure are fed into machine learning algorithms to identify patterns, make decisions, and predict outcomes. Finally, we close this loop by allowing the machine learning applications to send commands back down to the devices in the Internet of Things. This allows the devices to affect the world around them, then they sense the results of their actions to see whether they produce positive or negative outcomes, feed the data back through their model to learn from the results of their actions, and optimize their behavior over time. Closing the loop completely eliminates the human in the middle and leads to fully autonomous intelligent systems capable of making rational decisions and able to learn how to optimize their behavior over time. Over the next few decades, we'll see the emergence of many of these fully autonomous systems which are currently being referred to using names like smart systems, cloud robotics, and cyber-physical systems. No matter what marketing term ends up sticking, these intelligent systems will likely be disruptive technologies in many industries and aspects of our society. For example, we'll have self-driving cars, which will likely replace all human drivers in cars, semis, boats, and airplanes in the next few decades. We'll have smart buildings that optimize their own energy efficiency and anticipate the needs of their occupants. We'll have general purpose robots that will be able to be taught to perform manual labor rather than being explicitly programmed for each new task. And we'll eventually have smart cities that will optimize municipal resources like electricity, water, waste, and traffic without any human intervention. We'll embed these intelligent systems into all aspects of our environment. Essentially, computers become ubiquitous and invisible as computer-based intelligence will just be woven into the fabric of our society. To see how this will affect our society as a whole, let's



take a step in a different direction now to learn about how these trends will likely affect our economy.

## Economic Impact

Now let's see how the convergence of these data science trends is likely to impact our economy and our society going forward. It's important that you recognize that data science in these fully autonomous intelligent systems are likely going to have a significant impact on our economy and capitalism as we know it. First, it appears this trend is leading to a more non-linear economy. Those with smart machines will have even more power and those without will have progressively less power. In addition, data will likely become one of the most valuable resources in our information economy, so those that have the most data and who are able to extract valuable insight from their data will will tremendous power in our information economy. This will have an enormous impact on our labor economy as well. Modern fully autonomous intelligent systems will likely augment or replace entire sectors of our labor economy in the next few decades. Many jobs involving physical labor and knowledge work are now within reach of modern automation technologies. In fact, some experts are currently attempting to predict which jobs will be most likely to be eliminated based on their repetitiveness and complexity. On the X axis, we have the repetitiveness of a task from high to low. On the Y axis, we have the complexity of a task from low to high. If you were to draw a straight line diagonally through the chart, you would start with the most routine and simple task in the lower left-hand corner and end up with the most non-routine and complex tasks in the upper right-hand corner. Given this matrix of repetitiveness and complexity, we can attempt to predict what type of technology is necessary to automate a variety of occupational tasks. For example, we can see which retail jobs will likely be eliminated in the coming years as machine learning continues to be applied to retail sales. Even the medical industry isn't immune to the coming wave of automation. While these medical tasks are generally more complex and less repetitive, many of these tasks are rapidly becoming within the reach of automation technologies using machine learning and deep learning algorithms. As a result, we'll likely see a rapid increase in the number of machines performing manual labor and knowledge work that was once a job of humans. The rate of industrial robots has already been growing steadily for the past few decades. However, the application of modern data science practices like big data and machine learning will likely only amplify this trend going forward. We're beginning to transition from an economy where most of the work of value is done by humans to one where most of the work of value will be done by machines. As a result, it's important to ask yourself

which side of this new economy will your job be on, the side that's leading the new economy or the side that's being eliminated.

## Where to Go Next

So what does this all mean for us? Where do we go from here? As IT professionals, our work is going to be critical to the future of our world. We'll be the ones building all of the infrastructure to make this new economy possible, so we want to be the ones driving this data science revolution, not the ones taking a backseat for this wild ride. As a result, we need to educate ourselves in order to leverage these new data science skills. So where can we go to learn these new skills? Fortunately for us, there are a number of great educational resources available today. I encourage you to find the courses that make the most sense for you based on your own aspirations and skill level. However, if you're unsure about where to go next, I have a few recommendations for additional Pluralsight courses. First, if you're interested in more hands-on training for data science practices, I recommend starting with my course on Data Science with R. This course will introduce you to a wide variety of data science practices with hands-on informative demos. You'll learn how to perform data analysis, data visualization, handle big data, and create machine learning models all in a single course. Next, if you're interested in learning specifically about data analytics, I recommend Data Analytics: Hands On by Ben Sullins. This course will introduce to a wide array of data analytics tools and practices. If you're interested in learning more about big data, a good next step would be Big Data: The Big Picture by Andrew Brust. You'll learn more about big data and the tools and technologies used in the big data space. If you're interested in learning more about machine learning, I recommend Understanding Machine Learning by David Chappell. This course will provide you with a gentle introduction to the topic of machine learning. Finally, I have several other courses on data science topics at Pluralsight, and I'm always creating new courses. So be sure to check out my author page to see what other data science courses are available. In addition, my website contains a ton of information on data science practices, as well. I have articles, videos, presentations, open source sample projects, and links to additional resources. So be sure to check out all of these excellent sources of information to learn more. Finally, before we wrap things up for this course, feedback is very important to me. I use your feedback to improve each and every one of my courses, so please be sure to take a moment to rate this course, ask questions in the discussion board if there is something about this course that you didn't understand or would like me to clarify, leave comments to let me know what you liked about this course or if you think there is something I can improve upon for future courses, and feel free to

send me a tweet on Twitter if you'd like to provide me with feedback in public. You can find me on Twitter @matthewrenze.

## Course Summary

In this course, first, we learned about data science, what it is and why it's important. Next, we learned about data analytics, the application of data science practices to the world of business. Then we learned about the Internet of Things, the trend of connecting devices and sensors to the cloud, which is generating tremendous amounts of data. Next, we learned about big data, datasets of a volume, velocity, or variety beyond the limitations of conventional computing architecture. Then we learned about machine learning, software we're able to learn how to solve problems without being explicitly programmed to do so. Finally, we closed the loop to see how all of these data science trends are converging to create fully autonomous intelligent systems that will likely revolutionize our economy and the world around us. Thank you for joining me for this course on data science. I hope that you've learned some valuable information that you'll be able to put to great use, and I hope to see you again in another Pluralsight course in the future.

### Course author



Matthew Renze

Matthew is a data science consultant, author, and international public speaker. He has over 17 years of professional experience working with tech startups to Fortune 500 companies. He is a...

### Course info

Level	Beginner
-------	----------

Rating	★★★★☆ (588)
--------	-------------

My rating	★★★★★
-----------	-------

Duration	1h 9m
----------	-------

Released	15 Sep 2017
----------	-------------

Share course

