# SkillsBuild Learning

## What is data science?

---

# It's all about information

## Uncovering the story

**5** Minutes

### Let's explore data science

Our world is rich with **data.** Data scientists make sense of data and use it to solve problems.

Before we dig in, try this warm-up question. If you don't know the answer, just make your best guess. Here goes...

**Do you think that "unstructured data" is:**

1. Not yet organized

2. Impossible to analyze

3. Difficult to fit into a table

4. Useless

Keep your best guess in mind. You'll find the answer later in this module!

To get started on data science, here's a short video that explores the field from different perspectives.

**WHAT IS DATA SCIENCE? INTRODUCTION BY IBM**

What is Data Science? - Introduction by IBM #1

▶

⬀ **Video transcript (assets/VQNZDDPNDRKX1D2P/Vid%20Transcript%20WIDS.pdf)**

To explore the rest of the course, click **I've checked it out!**

---

# How do you handle data?

## Organizing facts

**5** Minutes

### You can organize data in two different ways

Data is raw information. It might be facts, statistics, opinions—any kind of content that's recorded in some format. Numbers? Sure. Facts? Sure. Voices, photos, names, dance moves? Also data.

So when data scientists handle raw information to uncover its story, they start by organizing it into one of two forms: **structured data** or **unstructured data**. The difference between these forms changes how we work with them.

**Structured data**

Structured data is information that can be laid out in rows and columns. You might already have worked with structured data using a spreadsheet like Microsoft Excel. For complex information, data scientists use more powerful tools like SQL, Apache, or R, which can sort through vast

amounts of data stored in many connected tables. Can you organize information within the data into groups based on specific characteristics? Those groups are structured data.

Here's a sample of structured data from your local hardware store.

| Customer name | Last name | Phone number | Last order |
|---|---|---|---|
| 00001 | Ajay | (555) 678-9012 | 03/12/19 |
| 00002 | Thompson | (555) 345-6789 | 08/14/18 |
| 00003 | Smith | (555) 432-1098 | 08/01/19 |
| 00004 | Kim | (555) 665-5443 | 11/16/18 |
| 00005 | Gonzalez | (555) 912-9945 | 12/24/17 |
| 00006 | Wangzi | (555) 212-3767 | 06/30/19 |
| 00007 | Colbert | (555) 866-0922 | 05/21/19 |
| 00008 | Jarrah | (555) 778-1845 | 05/22/18 |
| 00009 | Magnusson | (555) 395-7677 | 01/02/19 |
| 00010 | Cooper | (555) 550-5515 | 10/30/18 |

This table organizes customer information from the hardware store by characteristics such as customer number or name. Each row shows information related to a particular customer, while each column shows one customer characteristic that spans a group of customers.

As you can see, structured data tends to be well-organized, making it easier for data scientists to discover its treasure using common data analysis tools. Spreadsheets are based on tables like this, so they handle structured data very well.

### Unstructured data

Then there's unstructured data, which is a fancy way of saying "everything else." We use this term when there's no built-in organization (or structure) to the data. Unstructured data can be a collection of audio files, or social media posts, or essay texts, or even song lyrics.

Here are two examples to help you see the difference:

1. Your Department of Motor Vehicles takes photographs of everyone who gets a driver's license. A collection of those images is **unstructured data**. (But the table of peoples' names, addresses, and licence numbers that indexes those photos is structured data.)

2. A downloadable library might offer text from thousands of different books. The catalog listing names, authors and dates of those books is **structured data**. (But the text of those books is unstructured data.)

Unstructured data can be harder to work with than structured data, but it's still useful! Suppose a video game company is getting a lot of email bug reports about a new release. The text of those emails is unstructured data. By examining those texts (and perhaps by converting some of their

contents to structured data), a data scientist can figure out patterns and identify the problem so the company can fix it!

# How do you become a data scientist?
## Building a career

**5** Minutes

## What's the path into this field?

In an earlier topic, you saw young professionals talk about data science. Did you notice that no one got into this field the same way? That's because data science is an emerging study that hasn't been defined clearly until recently. (And, in some areas, the definition is still changing.)

It helps to think of data science as **creating knowledge from data**, no matter what technique is used to analyze the data. It's like a treasure hunt. You start with raw text, or numbers, or graphics, or any other kind of data, and you explore that data to discover valuable patterns and insights.

No special certification or skillset qualifies you as a data scientist. You just start analyzing data in any of its many forms. If you use scientific techniques to derive information from data, you're a data scientist, no matter how you entered the field! But there are areas of study that can help you prepare to hunt for data treasure, and we'll look at them later in this course.

In the next video, you'll learn how different paths can lead you to this career destination.

**THE MANY PATHS TO DATA SCIENCE BY IBM**

### What is Data Science? - The Many Paths to Data Science by IBM #2



⬦ Video transcript (assets/PZDQWWYDDNYQ2ADQ/Vid%20Transcript%20Many%20Paths.pdf)

# Will you like the work?

## Hunting treasure

**10** Minutes

### It's a search for understanding

The following video provides a strong point of view about the qualities that a data scientist should have, including curiosity, passion, and the ability to tell a story.

**ADVICE FOR A NEW DATA SCIENTIST BY IBM**

## What is Data Science? - Advice for New Data Scientists by IBM #3

▶

☐ Video transcript (assets/KXPRNNZKGXQQ294G/Vid%20Transcript%20New%20DS.pdf)

# Could you be a data scientist?

Now that you know more about data science and what the work is like, is this a career you'd like to explore?

Find out by answering a few questions in the following text box.

1. What are you passionate about? Write down a field of study, a hobby, or some other topic that you're interested in learning more about. Remember, from soccer to astrophysics to music to hurricane relief, there's data in everything.

2. What data might relate to your interests? Write down examples of what you think is out there.

3. Is that data mostly structured, so you could list it in neat rows and columns? Or is it mostly unstructured, like text or multimedia files?

4. Try asking a question (or two, or three!) that relates to your area of interest and the data that might describe it.

*Enter your response in the text box. (Writing an answer is a good way to process your thoughts. These answers are saved to your computer for your use only.)*

**Save Text**

# Data science as a career

## Infographic

**5** Minutes

## Careers in data science

Data science is a rapidly-growing field with plenty of opportunities for curious minds. Now you might be wondering what industries focus most on data science, what technical skills might be in demand, or what the interesting jobs in the field are. The answer to these questions, plus a few more facts, are included in the following infographic.

### CHECK THIS OUT!

🗗 Infographic: Interested in data science? (PDF)
(https://ibm.box.com/s/7j198hq7kfsxylwrb3hwkzrgavc5yney)