What is data science?

1. What is data science?

00:00 - 00:12

Welcome! My name is Lis. In this course, we'll unpack the what, why, and how of data science. By the end of this course, you'll have a better grasp of how data is being used around you and how you can use data.

2. Let's ask Google!

00:12 - 00:17

If we Google "What is data science?", we'll see a huge amount of confusing information.

3. Making data work for you

00:17 - 00:38

But data science is actually simple. It's a set of methodologies for taking in thousands of forms of data that are available to us today, and using them to draw meaningful conclusions. Data is being collected all around us. Every like, click, email, credit card swipe, or tweet is a new piece of data that can be used to better describe the present or predict the future.

4. What can data do?

00:38 - 01:39

So what can data do? Data can describe our current state, like our energy consumption. This can be accomplished with dashboards or alerts, simplifying time-intensive reporting processes. It can help detect anomalous events, such as fraudulent purchases. If we have data on what has happened previously, we can increase efficiency by automatically detecting a new event that is unexpected or abnormal. Data can also diagnose the causes of observed events and behaviors, for instance your activity on Spotify or Netflix. Rather than determining correlations between small numbers of events, data science techniques help us understand complex systems with many possible causes. Finally, data can predict future events, such as forecasting population size. We can use techniques to take various causes into account and predict potential outcomes. Further, we can evaluate the probability of our prediction mathematically to clarify our level of uncertainty.

5. Why now?

01:39 - 01:56

So now we know what data science is. The next question is why is it so popular? The answer is pretty obvious: we're collecting more data than ever before. Suppose that you visit a car dealership and fill out some information.

6. Why now?

01:56 - 02:04

All of that data is automatically entered into a computer, and combined with the data from hundreds of dealerships into one big database.

7. Why now?

02:04 - 02:40

Once we have that data, it's easy to use the email address that you provided when you bought that car to tie your car purchase data with your data from social media or web browsing. Suddenly, we have a very complete picture about everyone who purchased a car in the last year: their ages, their likes and dislikes, their families, their friends. This additional data can be used to predict what price you can pay for your car, and what other purchases you're likely to make, or how best to sell you insurance for that new car. Data is everywhere, and it's incredibly valuable information for businesses, organizations, and governments.

8. The data science workflow

02:40 - 03:01

So, how do we start to use data? In data science, we generally have four steps to any project. First, we collect data from many sources, such as surveys, web traffic results, geo-tagged social media posts, and financial transactions. Once collected, we store that data in a safe and accessible way.

9. The data science workflow

03:01 - 03:14

At this point, data is in its raw form, so the next step is to prepare data. This includes "cleaning data", for instance finding missing or duplicate values, and converting data into a more organized format.

10. The data science workflow

03:14 - 03:25

Then, we explore and visualize the cleaned data. This could involve building dashboards to track how the data changes over time or performing comparisons between two sets of data.

11. The data science workflow

03:25 - 03:38

Finally, we run experiments and predictions on the data. For example, this could involve building a system that forecasts temperature changes or performing a test to find which web page acquires more customers.

12. Let's practice!

03:38 - 03:50

Now, you know why data science is important and the first four steps in the data science workflow. Let's explore these themes in some practical exercises.