- We ended our last video talking about

In particular, some programming and data analysis skills.

But we'll also remind you about these soft skills

from time to time.

But what data tools should you pick?

According to a recent article on KDnuggets,

based on skills and jobs data from indeed.com,

Python is a clear leader in many data science categories.

Although learning any of the programming languages

shown here, including R Java, C, Scala, and Julia

is a good idea.

There are specific reasons why we pick Python,

and why employers are looking for these skills.

Instead of explaining why Python is a good language

for data science,

let's focus on why data scientists love Python.

In addition to being an easy-to-learn and readable language,

Python is an open language with a vibrant community.

Thanks to the efforts of this community,

it offers an ever-growing set of data management,

analytical processing, and visualization libraries,

some of which we will review in this course.

Such libraries make Python applicable to every step

of the data science process.

Lastly, but very importantly,

the Jupyter Notebooks make Python-based analysis

more producible and repeatable,

as well as provides built-in training

and communication support to help with team communication.

Throughout the rest of this course, we will learn about some

of the most powerful Python libraries

and apply them to case studies ranging

from simple soccer data analysis

to astrophysics and satellite image analysis.

We will start by learning about Jupyter Notebooks,

followed by NumPy and Pandas to ingest

and analyze data efficiently.

We will add the visualization libraries,

including Matplotlib,

and continue with applying machine learning libraries

in Scikit-Learn to create models.

We will add libraries like BeautifulSoup

to easily read an XML and HTML-type data,

and go over some of the examples of working with databases.

We hope you will enjoy this technical programming

and learning journey as much as we did.

the necessary technical skills for a data scientist.

But what is a data scientist overall?

And why does one need Python for data science?

Let's analyze these further in this video.

By the end of this video,

you will be able to list some of the traits

of modern data scientists,

explain why Python is a good programming language

for data science,

and recite four major Python modules

that are useful for data analysis.

You have probably seen diagrams like this one

that describes data science.

Data science happens at the intersection

of computer science, mathematics,

and business or scientific expertise.

If you zoom deeper into this diagram

and open up the sets of expertise,

we would see a variation of this figure.

Even at this level,

all of these boxes require deeper knowledge

and skills in areas like domain expertise,

data engineering, statistics, and computing.

And even deeper analysis of these skills based

on data science job listings

would lead you to skills like machine learning,

statistical modeling, relational algebra,

business passion, problem solving,

and data visualization.

That's a lot of skills to have for a single person.

Given such a wide range of skills

across multiple definitions

of data scientists seems impossible.

Some folks have even begun to ask if data scientists

are like unicorns, meaning they don't exist.

I want to point out that there are data science expert,

who has expertise in more than one of these skills for sure,

but they're relatively rare

and still would probably need help from an expert

on some of these areas.

So in reality,

data scientists are teams of people who act like one.

This is why we say data science is team sport,

referring to the breadth of information

and skills it takes to make it happen.

However, there are still common traits

to our data scientists.

For example, data scientists are passionate about the story

and meaning behind data,

they understand the problem they are trying to solve,

and aim to find the right analytical methods

to solve this problem.

And they all have an interest in engineering solutions

to solve problems.

They also have curiosity about each other's work,

and have communications skills to interact

within the team and present their ideas

and results to others.

These mostly soft skills are very important

for success in any data science team.

In this course,

we'll focus on providing you with technical skills.