

# Best Approach to learn and Practice Data Science

## What is the best way to learn and practice Data Science?

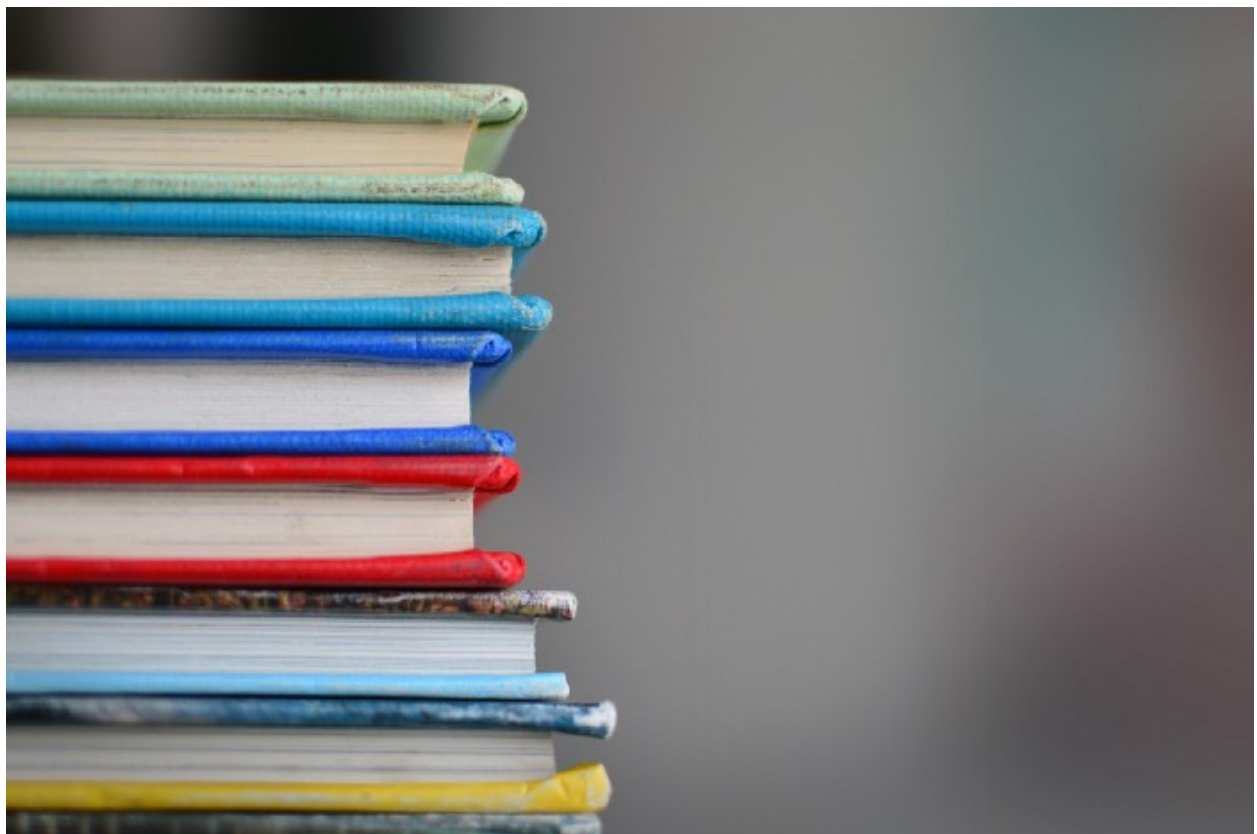
My advice would be to start with practical projects and then slowly progress with theory. [Kaggle notebooks](#) are a great way to learn the practical part.

Ask questions in [Reddit communities](#) or in [Cross Validated community](#).

When you become satisfied with your knowledge of tools and practices, I would suggest you construct the dataset for some problem by yourself (e.g., you can scrape the data) and apply ML algorithms to it. The hardest thing in ML is dataset construction. You might even build a company out of it.

Kaggle is a great way to learn the practical part.

## What are good resources to learn Machine Learning?



I personally like the [Machine Learning](#) Coursera course by Andrew Ng. The course starts easy and then gradually gets harder as it goes. The good thing about it is that it focuses on the fundamentals of Machine Learning.

I suggest that you listen to at least the first few lectures. Don't worry if you don't understand everything because you can always revisit it later. I would also advise that you don't focus just on a single course. We all learn differently, and that's ok.

We all learn differently, and that's OK.

### **I hardly have any technical background. What do you think would be the best approach to learn?**

Don't study alone! Find and join online communities that can help you learn and grow. I wrote about Data Science communities in:

- [Data Science while Working From Home](#)

You can start practicing Machine Learning in Excel. Try to implement a [Linear Regression in Excel](#). It is a great first challenge, and it will get you motivated. Start practicing Machine Learning in Excel.

### **Should I learn Python or R?**



*Photo by [mana5280](#) on [Unsplash](#).*

Let's address the elephant in the room. If you're just starting I would suggest learning Python. The main reasons are:

- rich ecosystem for Data Science, Backend... you name it, Python has it.
- the language is still gaining momentum in popularity.

With Python, you can do the analysis, develop the model from scratch, and then run it in production. While I am sure that models in R also run in production, I haven't heard about one (let me know in the comments if your experience is different).

Don't get me wrong, if you know R, that's totally fine. Data Science teams are usually using both languages as some prefer R and others Python.

In the end, it doesn't really matter as some models have to be reimplemented in a compiled language (Java, Go) to make faster predictions in production.

Python enables you to do the analysis, develop the model from scratch and run it in production.

## **Should I learn SQL?**

This is a great question. The answer is YES — with capital letters.

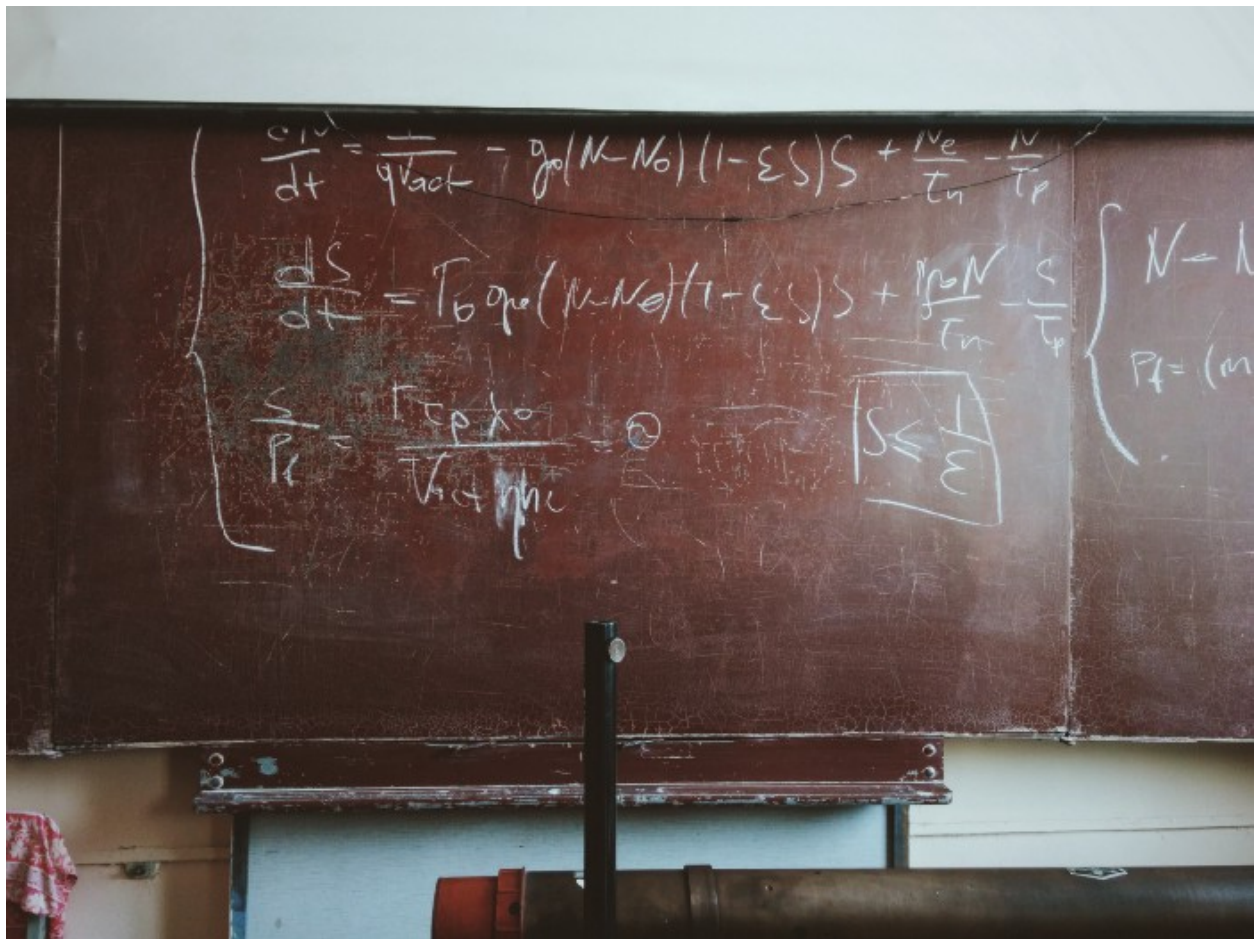
Whether you'll be using the SQL databases or not, you should know the main concepts from relational databases like **joins**, **group by**, **window functions**, **lag**, **lead**, etc. These concepts are essential even when working with pandas, R, or some other tool.

In case you're interested, I also wrote a few articles about SQL:

- [5 mistakes when writing SQL queries](#)
- [Don't Repeat These 5 Mistakes with SQL](#)

The answer is YES — with capital letters.

## Should I take more math classes?



*Photo by [Roman Mager](#) on [Unsplash](#).*

The more math you know, the better for you in the long term. Knowing math will enable you to understand what's happening behind the scenes of a black box Machine Learning model. It's also easier to transfer knowledge from theory to practice.

With math you'll understand what's happening behind the scenes of a black box model.

## **Are math and statistics important in Data Science?**

Math becomes crucial when you need to improve the model. You need math to understand the difference between different types of models, distributions, etc.

Senior Machine Learning Engineers can tell the main properties of a model just by looking at the optimization function.

Math becomes crucial when you try to improve the model.

## **Which classes should I take to be better prepared for the Data Science role?**

My advice would be to think ahead. Each field needs a Data Scientist or will need it in the future. Ask yourself, in which company would you like to get an internship after you finish the study? It will be easier to get a Bioinformatics internship if you already listened to a few related classes.

Think ahead.

## **Do I need a Ph.D. to work in Data Science?**

You don't need a Ph.D. to work in Data Science — meaning doing analysis of real-world data, and applying the Machine Learning models.

If your goal is doing research and developing new Machine Learning algorithms (e.g., working in Deep Mind), then you should pursue a Ph.D.

You don't need a Ph.D. to work in Data Science, but...