A long-term Data Science roadmap which WON'T help you become an expert in only several months

Some thoughts on becoming a data scientist. It isn't easy or fast and requires a lot of efforts, but if you are interested in data science, it is worth it.

Andrew... Follow
Dec 2, 2018 · 7 min...

From time to time I am asked: how does one become a data scientist? What courses are necessary? How long will it take? How did you become a DS? I have answered this question several times, so it seems to me that writing a post could be a good idea to help the aspiring data scientists.

About me

I got a masters degree at MSU Faculty of Economics (Russia, Moscow) and worked for ~4 years as an analyst/consultant in ERP-system implementation sphere. It involved talking with clients, discussing their needs and formalizing them, writing documentation, explaining tasks to programmers, testing the results, organizing projects and many other things.

But it was a stressful job, with lots of problems. What is more important, I did not really like it. Most of the things were not inspiring, though I liked working with data. So, in the Spring-Summer 2016 I have started looking for something else. I got a Green Belt in Lean Six Sigma, but there were no opportunities nearby. One day I have found out about BigData. After a

couple weeks of googling and reading numerous articles I realized that this could be my dream career.

I left my job and 8 months later got my first position as a data scientist in a bank. Since then I have worked in a couple of companies but my passion for data science is still strong. I have completed several courses in ML and DL, made several projects (such as a chat-bot or a digit recognizer app), took part in many ML competitions and activities, got three silver medals on Kaggle and so on. Thus, I have some experience with studying data science and working as data scientist. Of course I have a lot of things to learn and a lot of skills to acquire still.

Disclaimer

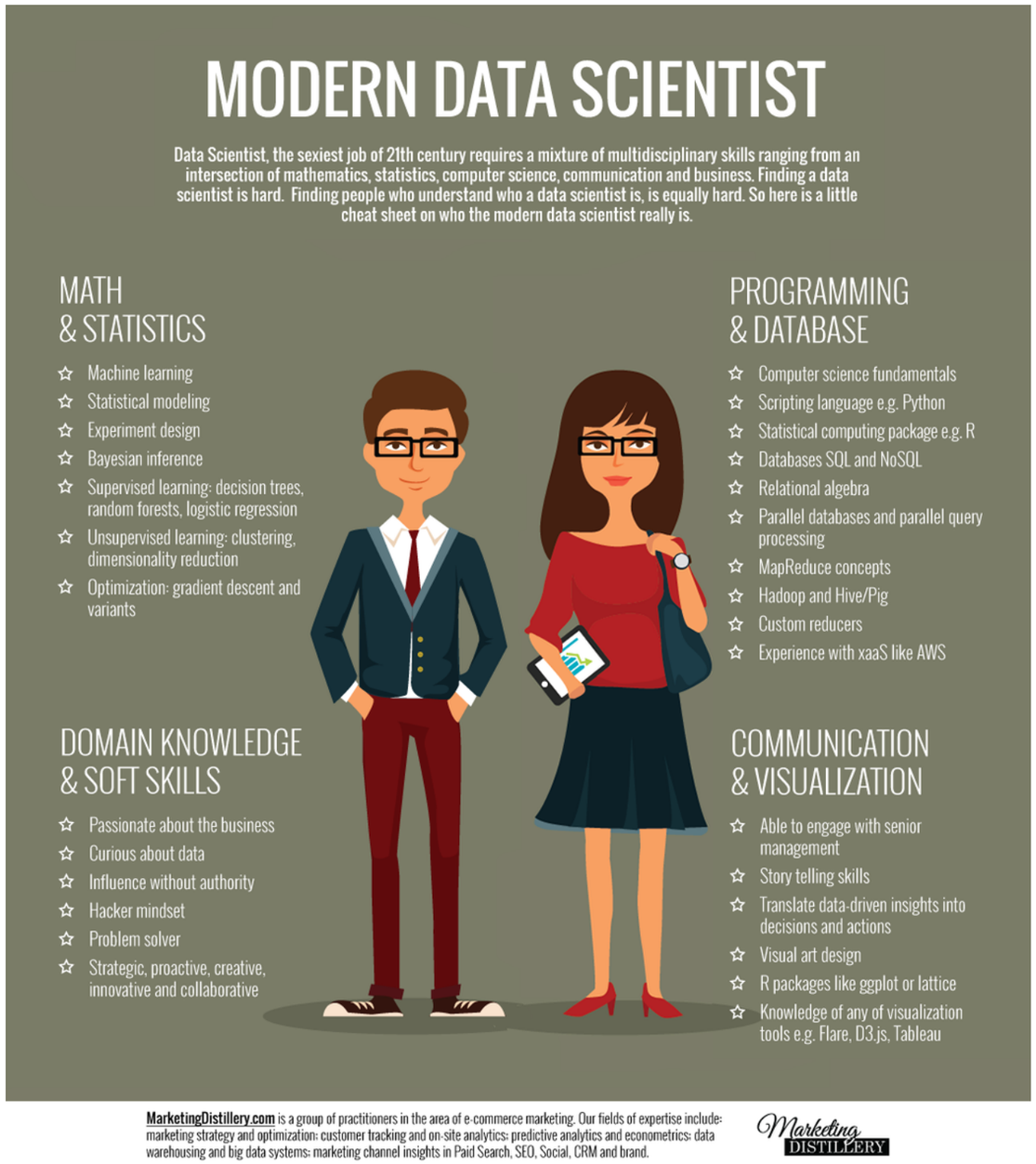
This article contains my opinions. Some people may disagree with them and I want to point out that I do not want to offend anyone. I think that anyone wishing to become a data scientist must invest a lot of time and effort in it or they will fail. Courses or MOOCs claiming that you can become an expert in ML/DL/DS in several weeks or months are not entirely truthful. You can get some knowledge and skills within weeks/months. But without extended practice (which is not a part of most courses) you will not prevail.

You do need internal motivation, but, more importantly, you need discipline, so that you will continue working after the motivation went away.

Let me repeat again — you need to do things by yourself. If you ask the most basic questions without even trying to use Google/StackOverflow or thinking for a couple of minutes, you will never be able to catch up with the professionals.

In most of the courses which I took, only around 10–20% of people completed them. Most of those who dropped out did not have dedication or patience.

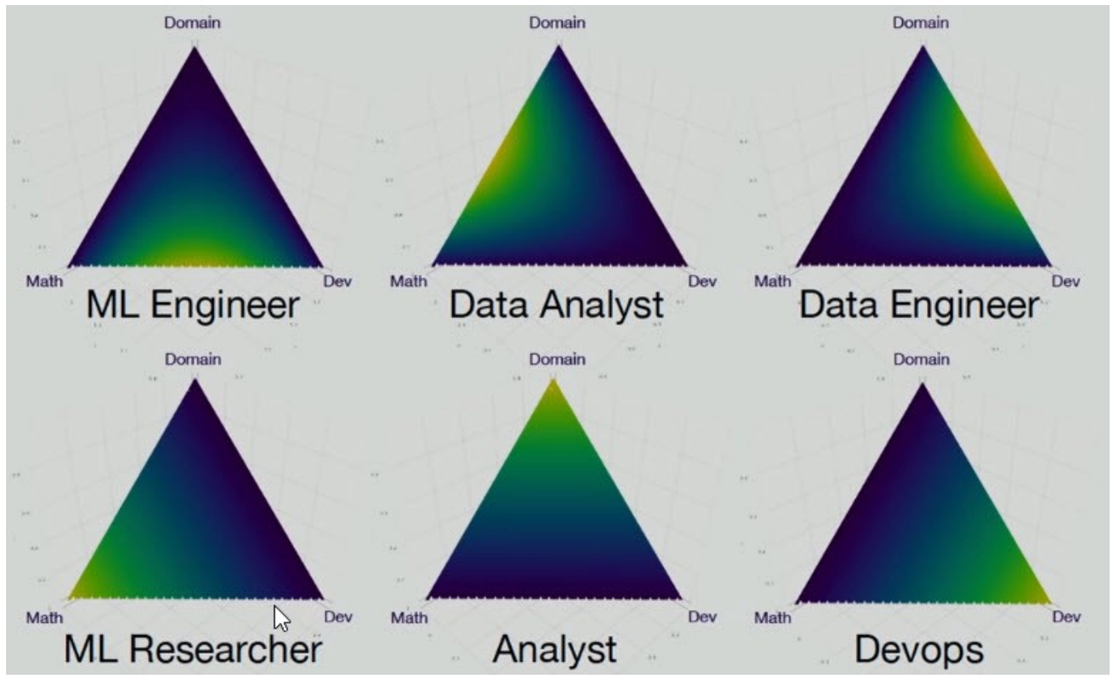
Who is a Data Scientist?



There are many pictures showing data scientist's core skills. For the purposes of this post any of them is good, so let us look at this one. It shows that you need Math & Stats, Programming & Devops, Domain knowledge and Soft skills.

That's a lot! How is it possible to know all of this? Well, it really takes a lot of time. But here are good news: it is not necessary to know everything.

There was an interesting talk on 21 October 2018 at Yandex. It was said that there are many types of specialists, who have different combinations of aforementioned skills.



Data Scientists are supposed to be in the middle, but in fact they can be in any part of triangle, having different levels in any of the three spheres.

In this article I will talk about data scientists as they are usually assumed — those who can talk with customers, perform analysis, build models and

deliver them.

Switching careers? This means you already have something!

Some people say that switchings career is quite difficult. While it is true, have a career to switch from, means you already know something. Maybe you have experience with programming & devops, maybe you have worked in a math/stats heavy sphere or you honed your soft skills everyday. At a bare minimum you have an expertise in your domain. So always try to use your strong sides.

Roadmap from Reddit

In fact there will be two roadmaps :)

The first one is from Reddit, from this topic:

> *First, read fucking Hastie, Tibshirani, and whoever. Chapters 1–4 and 7–8. If you don't understand it, keep reading it until you do.*
>
> *You can read the rest of the book if you want. You probably should, but I'll assume you know all of it.*
>
> *Take Andrew Ng's Coursera. Do all the exercises in python and R. Make sure you get the same answers with all of them.*
>
> *Now forget all of that and read the deep learning book. Put tensorflow and pytorch on a Linux box and run examples until you get it. Do stuff with CNNs and RNNs and just feed forward NNs.*
>
> *Once you do all of that, go on arXiv and read the most recent useful papers. The literature changes every few months, so keep up.*

> *There. Now you can probably be hired most places. If you need resume filler, so some Kaggle competitions. If you have debugging questions, use StackOverflow. If you have math questions, read more. If you have life questions, I have no idea.*

And from one of the comments:

> *Still not enough. Come up with a novel problem where there's no training data and figure out how to collect some. Learn to write a scraper, then do some labeling and feature extraction. Install everything on EC2 and automate it. Write code to continuously retrain and redeploy your models in production as new data becomes available.*

While being short, harsh and very difficult, this guide is quite great and it will get you to a hireable level.

Of course there are many other ways to data science, so I will offer mine. It is not perfect, but it is based on my experience.

My Roadmap

There is one skill which will get you very far. If you do not have it yet, I urge you to develop it. This skill is... formulating thoughts, searching for information, finding it and understanding it. Seriously! Some people cannot formulate thoughts, some are unable to find solutions to the most basic questions, some do not know how to properly create google queries. This is a basic and necessary skill, and you must perfect it!

1. Choose a programming language and study it. Usually it would be Python or R. I highly recommend choosing Python. I won't list the reasons, because there are a lot of arguments about R/Python out there already, I personally think Python is more versatile and useful.

Spend 2–4 weeks on learning the language, so that you can do basic things. Get a general understanding of the libraries used, such as pandas/matplotlib or tydiverse/ggplot2.

2. Go through the ML course by Andrew NG. It is old, but it gives a great foundation. It could be useful to complete the tasks in Python/R, but it is not necessary.

3. Now take one more good course in ML (or a couple of them). For R users I recommend Analytics Edge, for Python users — mlcourse.ai. If you know Russian language, this course on Coursera is also great. In my opinion mlcourse.ai is the best among these three. Why? It provides good theory and some tough assignments, which could already be enough. However, it also teaches people to take part in Kaggle competitions and make standalone projects. This makes it great for practice.

4. Study SQL. In most companies data is kept in relational databases so that you will need to be able to get it. Make yourself comfortable with using select, group by, CTE, joins and other things.

5. Try to work with raw data to get the experience of working with dirty datasets.

6. While the previous point may not be necessary, this one is mandatory: complete at least 1 or 2 complete projects. Perform a detailed analysis and modelling of some dataset, or create an app, for example. The main thing to learn is how to create an idea, plan its implementation, get data, work with it and bring the project to completion.

7. Go to Kaggle, study kernels and take part in competitions.

8. Join a good community. I have joined ods.ai — a community of 15k+ active Russian data scientists (by the way, this community is open to data scientista from any countries) and it helped me a lot.

Studying Deep Learning is a completely different topic.

This is only the beginning. Following this roadmap (or doing something similar) will help you start your journey to becoming a data scientist. The rest is up to you!

Data Science Roadmaps