

The Lifestyle of a Data Scientist

Author, Ajit Singh

A Data Science Foundation Blog

April 2019

www.datascience.foundation

Copyright 2016 - 2017 Data Science Foundation

Why has being a Data Scientist become the hottest profession?

Intuitively, it seems like, with the advantage and insight that Data Science provides, Data Scientists should have been around for much longer, helping us make decisions, isn't it? This question can be answered in one word. **Data.** Data is the bread and butter of a Data Scientist. A Data Scientist sleeps, breathes and eats data. The data that is being produced today is quite mind boggling. We have produced more data in the last two years alone than we did since the dawn of time! And this data only increases every day.

To just let the ever increasing data get produced and stay stagnant did not seem useful. Data was mounds of useful information, a tool, a Pandora's Box, full of untapped potential, waiting to be explored. This is around the time Data Science was born. Data Science is so new a field which not even actual Data Scientists can tell you what exactly the job entails!

Data Analytics has been performed on data for years. It has been around for over 20 years. So, why not just use basic Analytics tools? Why introduce an entirely new profession called Data Science? Because, with the evolution and advancement of Machine Learning, a branch of Artificial Intelligence, the kind of insights that were being mined from data were revolutionary! Statistical and Machine Learning models as simple as Linear Regression were providing immensely helpful insights that were vastly improving business decisions.

For example...

A company may find a pattern with a section of consumers behaving differently. After digging deeper and further analysis, they discover that this subsection of consumers has a similar trait. Now, they can work on ideas to modify the consumer's behaviour or understand what can be done differently to cater this audience.

Let's take an airline company. The company might want to know things affecting costs, things affecting revenue, and things which will help them attract and retain customers. Fuel is a major cost for any airline, so you might do some data analysis to project the future expenditure and buy more fuel when the price is low. Regarding growing the revenue, consider that customers are usually more sensitive to price when buying tickets for personal travel compared to business, as the fare comes out of their own pockets. Here, you might explore the opportunity to attract new customers by giving low-cost fares to popular vacation destinations and offering higher-cost fares with offers that suit the business travelers going.

A website might want to know about a metric such as, how long do people spend on their site and then find certain features that are correlated with it. Then emphasize those features to boost metrics. They might also ask different questions to incorporate few changes. Like, if they can build the findings into the product features, so that if a person is looking for a review of a restaurant, they can be prompted to review it from the homepage itself. Both how they decide which restaurant to show you, because you may have looked at several places and user interface elements have a huge impact on whether you write a review.

Data Science Foundation

A Data Scientist's Day

Data Science being a new profession, even Data Scientists wouldn't be able to tell you what their job entails. If you asked Data Scientists at Facebook, Yahoo, GE, etc. what exactly they do throughout the day, their answers could probably differ a little.

However, we have tried to show you how a possible Data Scientist's day could look like

The Mornings

A Data Scientist would start off his day with a steaming cup of coffee. Nothing works better than to get those grey cells activated than a rush of caffeine! Creativity is a huge part of a

Data Scientist's job. It isn't just about crunching numbers. A Data Scientist is often compared to an artist. He weaves a story out of the insights provided in such a way that, even a layman should be able to understand i©2016 UpX Academ

What's the Breakfast Menu?

At breakfast, the popular choice is eggs and bread with orange juice and fruit. Yes, it is pretty routine. A Data Scientist is human after all!

Newspapers

A Data Scientist would then go on to read the newspaper to stay informed about the world's goings on. His job involves staying on top of everything occurring in the world, since Data Science can be applied to practically every industry, be in banking, sales, finance, education, etc. from, to obtain insights. After this, he'd probably move on to read some mainstream

Data Science magazines and journals. Keeping himself updated on the new research happening in the field every day (And mind you, that is a LOT! A new technology or tool emerges practically every day.) is vital. It is a field so volatile, (in the positive sense) that keeping up is crucial. Are you an aspiring Data Scientist? Then these journals are for you!

- IEEE Transactions on Knowledge and Data Engineering
- Journal of Data Science
- EPJ Data Science Journal
- Information Visualization
- Journal of Machine Learning Research
- Predictive Modelling News
- Transactions on Machine Learning and Data Mining

At work

Data Science Foundation

When a Data Scientist gets to work, the first thing he would be is to communicate with members of the team about which stage the product is in the right now. Based on this, he'd put on his thinking hat and search his brain for WHAT questions should be asked at that particular phase. Because remember, not asking questions is better than asking the wrong question! The questions should maximize the advantage that Data Science would offer at that point to the highest possible extent.

Now, that the right questions have been asked, it's time for him to get his hands dirty and dive into the first step of predictive modelling - Cleaning and organizing the data. Research has shown that this is the least preferred part for a Data Scientist. Compared to his usual work, it is rather dull. Don't get me wrong, though; it is one of the most important parts of data analytics. For example, while performing Linear Regression, if two data points are outliers and they are not —managed|| or —cleaned||, they could jeopardize the entire model!

The next step a Data Scientist would take is a crucial one. He now has to find the right model. This depends on a variety of things. The —kind|| of data, what questions are being asked, and what predictions are being sought. Not all models work for all data. Some models are fit for specific data.

How do they work?

Once the appropriate model is selected, a Data Scientist applies the model. And then, (hopefully), the model gives us the appropriate results. One of the most important jobs that a Data Scientist performs, and one of the most crucial parts of a Data Scientist's day, is understanding and interpreting the results that are obtained after running the model. The results are often a garbled bunch of numbers with random letters in between. Probably only a Data Scientist will truly be able to understand what these results mean, and interpret them to mean something that a company can use to improve operations.

Once provided with the results, insights can be given by a variety of people. Yes, a Data Scientist can usually provide the best insights, but often, a useful insight is caught by someone's eye on the management team, since they deal with the business decisions after all. However, interpreting the results of a model as it is probably beyond the scope of the management.

This is where the next part of a Data Scientist's day comes in. It is the job of a Data Scientist to weave the results into a story (See why they are called artists?), probably through visualization tools that can be understood by anyone that sets their eyes on it (At least, the management!).

This process isn't standalone. It works in a constant feedback loop. Once these insights have been applied, it is the job of a Data Scientist to ask, has used the insights provided the expected results? Are they not as satisfactory? What other parameters can and should be considered while revamping the model? In this way, constant optimization is exercised to make the system as efficient as possible.

Back home

Now that a Data Scientist has performed his official duties, he returns home, (maybe hits the gym?), and since he is so passionate about the budding field, reads up on it from many informative books. As they

Data Science Foundation

say, when you love your work, work stops feeling like work!

Some of the books that are commonplace in a Data Scientist's library are:

- Data Science for Business by Foster Provost and Tom Fawcett
- The Elements of Statistical Learning by Trevor Hastie, Jerome H. Friedman, and Robert Tibshirani
- Python for Data Analysis by Wes McKinney
- Data Science in the Cloud by Stephen F. Elston
- Statistical Inference for Data Science by Brian Caffo
- An Introduction to Statistical Learning by Robert Tibshirani
- Machine Learning for Hackers by Drew Conway & John Myles White
- Agile data science by Russell Journey
- Natural Language Processing with Python by Steven Bird et al.

A Data Scientist then probably just kicks back with a glass of wine and some healthy dinner

(No, this isn't a prerequisite to becoming a Data Scientist!), and relaxes. Quite a neat little day, isn't it? The best part is there is no monotony in the job. Every day brings new challenges, every day a new creative thinking hat needs to be put on, and insights need to be mined. The future of the company probably depends on it. But, no pressure!

At the movies!

A Data Scientist is someone that deals with Mathematics and logic throughout his day. His mind is fine-tuned and conditioned to movies that involve these areas. Apparently, Data Scientists have no concept of —Don't take your work home with you!! Some of the movies that are on every Data Scientist's must watch list are -

- A Beautiful Mind - A true life story where a schizophrenic Maths professor called John Nash is the protagonist. Represents how powerful an impact Mathematics has on his life. He goes on to win the Nobel Prize!
- The Imitation Game - Alan Turing, the protagonist, builds the first computer to crack the German Enigma Machine to eventually execute a series of smart decisions and win WWII. Interestingly enough, it is said that this was around the time discussions about Neural Networks and using data originated among these brilliant British mathematicians!
- Moneyball - Based on a true story, Moneyball deals with using Analytics to find the most optimal sports team by assessing each player's value.

Conclusion

It might seem like all work and no play, but most Data Scientists have quite a relaxed life. If you're thinking about getting into the field, read the mentioned magazines and start on the books, get familiar with various Machine Learning algorithms, and most importantly, get your hands dirty by downloading any of the various data sets from the plethora of sources online, and working on them. You're on the path to becoming a successful Data Scientist!

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: contact@datascience.foundation
Telephone: 0161 926 3641
Atlantic Business Centre
Atlantic Street
Altrincham
WA14 5NQ
web: www.datascience.foundation