

# A Road Map for Data Science



Jared [Follow](#)

May 5, 2019 · 5 mi...



## What is Data Science?

Data science at its most basic level is defined as using data to obtain insights and information that provide some level of value. Data science is evolving fast and has a wide range of possibilities surrounding it and so to limit it by that basic definition is kind of elementary. An extension of the that definition would be that data science is a complex combination of skills such as programming, data visualization, command line tools, databases, statistics, machine learning and more... in order to analyze

data and obtain insights, information, and value from vast amounts of data.

For those of you interested in more specifics of Data Science and what it is you can learn more from this book [here](#)...

## **Python Programming**

The very first thing you should learn is some basic python programming. Learn the Syntax, Variables and Data types, Lists and for Loops, Conditional Statements, Dictionaries and Frequency Tables, Functions, and Object Oriented Python to get started.

For python programming this is the only resource you will ever need...

## **Data Analysis and Visualization**

Now we want to learn data analysis and visualization. First you will want to start off by learning pandas and numpy for cleaning and exploring your data. Then you will want to learn matplotlib for exploratory data visualization and storytelling with your data.

Here is a good book that will get you started with hands-on data analysis.

## **Command Line Tools**

Next you will want to learn how to navigate the file directory, create and delete directories, how to edit and manage files and their permissions, how to work with programs from the command line, and how to create virtual environments. You'll also want to learn about git and GitHub for version control.

I find the best way to get into the command line is to use it on a day to day basis... here is a free crash course on using the command line.

## Databases

You'll want to learn SQL for querying data as well as PostgreSQL for advanced database management. You should also know how to work with APIs and web scraping for creating your own datasets. Also try learning spark and map-reduce.

For SQL here is a great hands-on resource that will have you up and running with SQL in no time.

## Statistics

Next you'll want to learn statistics fundamentals which includes sampling, frequency distributions, the mean, weighted mean, the median, the mode, measures of variability, Z-scores, probability, probability distributions, significance testing, and chi squared tests.

Introduction to Statistical Learning and Elements of Statistical Learning will give you a statistics foundation that will make you the go to person for all things statistics...

## Machine Learning

You will want to learn at least 10 basic algorithms for machine learning: linear regression, logistic regression, SVM, random forests, Gradient Boosting, PCA, k-means, collaborative filtering, k-NN, and ARIMA.

You will also need to understand how to evaluate model performance, hyperparameter optimization, cross-validation, linear and nonlinear functions, basic calculus and linear algebra, feature selection and

preparation, gradient descent, binary classifiers, overfitting and underfitting, decision trees, neural networks, and then you should build something with those skills and even try some kaggle competitions. You can also move on to more advanced topics like NLP and AI if interested in those.

Here is a really good book for getting hands-on with machine learning...

## You Should have a SPECIALIZED SKILL

Once you've gotten the basic skills down I recommend getting really good at one thing such as deep learning, AI, statistics, NLP, or something else because it allows you to be the go to person for a specific skill and it looks really good for a job interview if that's what you are trying to do.

## Projects

You should really build some projects as you go. I recommend building things after you've learned basic python and data visualization tools.

Learning by doing is one of the best ways to truly learn the skills you need in data science and it also proves to others that you actually can build something with data.

## Starting a career in DATA SCIENCE:

**You will want to build 2 advanced projects that you can put onto a resume or in a portfolio:**

- One that shows you can do an end to end data science project
- Then the second one should be a project that showcases your specialized skill
- Make sure your projects are presentable, well-documented, easy to understand, and put them on GitHub

- Create a great resume that stands out and communicates the right information tailored to the specific job you are applying for
- Create a solid LinkedIn profile so recruiters can find you and you can also use LinkedIn to apply for jobs

### **Your Portfolio:**

- Your projects should tell an easy to follow story
- Should clearly visualize your results
- Should be well-documented with high-quality, organized code
- Includes a clear write of what you did and why
- Demonstrates you can do the job of a data scientist

### **Your Resume:**

- Should be easy to find relevant information in 6 seconds or less
- Highlights only the best/most important experiences
- Visually stands out against the sea of cookie-cutter applications
- Use the correct formula to frame your projects and experiences in terms of business impact(even if they were personal/academic projects)
- Format: What you did -> How you did it -> Impact it made
- *Bad: built recommender system in python*
- *Good: built recommender system in python using collaborative filtering and matrix factorizations that resulted in a 3% increase in basket size and a \$3M increase in yearly revenue*

- Make sure your resume is easy to read — use [www.readable.io](http://www.readable.io) and aim for a 5th grade reading level
- Make sure you have the proper keywords that using [www.jobscan.co](http://www.jobscan.co)

### **Your LinkedIn:**

- Translate your experiences from your resume to your LinkedIn
- Create a summary that shows your unique skills and personality
- Take a professional profile pic that is friendly and makes you more trustworthy
- Fill out the skills sections with the right skills so that recruiters find you (cut the extras that clutter your profile)
- Begin applying for jobs through LinkedIn
- Send follow up messages — (find 3–5 key decision makers (these will most likely be people in HR for the company you applied for) and send them follow up messages)
- Quickly and simply show your enthusiasm for their company
- Briefly pitch your unique skills and how they'll help the company (just give a preview of what you can do)
- Keep the follow up messages to 5 sentences max. (shorter is better and more likely to be read)

Thanks for reading my article and I hope you gain something from it.

I am working on creating some tutorials, guides, and a complete course on data science to help all those who need it and I plan to release it very soon get in now before it closes here! Available to a Limited number of students!!!

