The Ultimate Data Scientist Cheat Sheet for Data Scientists

Johnnyvastola
Jul 10 · 5 min read ★

Aspiring or professional data scientist, you know all too well the vast number of SQL commands, algorithms, Excel functions and Googling that go into your job. While the ones you use every day are committed to memory, there are probably times when you need to jump into a project that requires new tools, or a different application of your programming language of choice.

Enter: the ultimate compilation of data science cheat sheets. This is your go-to resource, allowing you to see the most common formulas and algorithms all in one place rather than sifting through pages of Google results for the answers you need.

These links will help speed up your programming and computing so you can move on to solving bigger problems. Read on to find the most useful reference guides for Python, SQL, machine learning and even linear algebra. When you need to copy/paste a line of Python or find the rules for multiplying matrices, we got you.

Machine Learning

Machine Learning is transforming our world, and data scientists are driving that change. Our virtual assistants, Instagram algorithms and Google results all rely on machine learning. But there's a lot of programming that goes into creating the machine learning models that consumers end up interacting with every day. It starts with enormous datasets and lots of clever programming.

For data scientists who focus on machine learning, or analysts who are studying up to enter this growing field, these quick reference tools will be indispensable.

- Machine Learning Glossary

- Machine Learning Formulas

- Machine Learning Algorithms — Python and R

- TensorFlow

- Deep Learning

- Supervised Learning

- Unsupervised Learning

Artificial Intelligence

With artificial intelligence, data scientists are achieving feats we previously thought were impossible. Just a few years ago, the idea of computers making decisions, learning from data and accurately predicting the future would have sounded more sci-fi than science. Times have changed: companies in healthcare, education, finance, and other industries now rely on artificial intelligence to improve their products and better understand their customers.

The capabilities of AI are evolving fast, but the foundation will stay the same. We collected some of the most common terms and algorithms used in the world of AI. Get a quick reminder of how the different neural networks compare, or study up on the most common AI models.

- The 4 Types of AI

- AI Glossary

- Natural Language Processing

- Neural Network Illustrations

- The Most Popular AI Models

- Top AI Researchers and Organizations

SQL

Data scientists across the world use SQL to organize data in tables or work with multiple datasets. SQL is typically used to extract the data needed for a given analysis, while Python and its many specialized libraries are then brought in to do the heavy lifting (more on that to follow).

Here are all the basic SQL commands and functions you'll use as a data scientist. For fans of Microsoft Azure, we included a log query cheat sheet to use with your Azure SQL databases. Rounding out the list are some cheat sheets for MySQL, the specialized system for web databases, and PostgreSQL, which is sometimes employed with other languages like C++ and Java.

- SQL

- SQL Commands

- SQL for Azure Monitor

- MySQL

- PostgreSQL

Python

Python is one of the easiest programming languages to pick up, but it's also one of the most useful for data analysis. It speeds up the process of cleaning and manipulating large datasets, leaving Excel functions in the dust.

Because Python is so essential to the big data field, there are now a whole slew of complementary tools built specifically for data analysis. Libraries like NumPy, SciPy and Pandas accelerate the processing power of Python, making them invaluable for data scientists and a shoo-in for our list. Sci-kit Learn is specific to machine learning, while Bokey and Matplotlib are useful for visualizing the data. The essential functions for each are in the cheat sheets below.

- Python

- Python — Importing Data

- Python — Data Cleansing

- NumPy — Basics

- NumPy — Advanced

- SciPy

- Pandas — Basics

- Pandas — Advanced

- Bokeh

- Matplotlib

- Sci-kit Learn

R

This programming language was designed with one goal in mind: to clean, analyze, and visually represent data. When scientists need to plot out data pulled from thousands or even millions of users, R allows them to do it with just a few powerful lines of code, saving hours of tedious spreadsheet labor. While this is one of the lesser-known languages in the broader world of computer programming, it's a favorite among data scientists for obvious reasons.

Use the following cheat sheets to quickly generate the proper R code for your next data analysis challenge. These are also great starting points for those who are still picking up the basics of R, and the many associated tools like RStudio and tidy evaluation.

- R

- R — Basics

- R Commands

- R Markdown

- RStudio

- Tidy Evaluation with rlang

- Caret Package

- Data Visualizations in R

Math

To state the obvious: data science is a deeply complex field that requires some pretty advanced math. Depending on your area of specialty, you may need to employ calculus, linear algebra, and statistics on a regular basis. In order to make real progress in the field, data scientists need to have a complete understanding of the concepts at work, and how they apply in different scenarios.

You probably have a strong foundation in math if you made the decision to pursue data science in the first place. These tools are not designed to teach calculus or algebra from the ground up, or impart any true understanding of these concepts for a beginner. They're resources for data science students and professionals, to help you quickly find a specific equation or check your work.

Even for experienced data scientists, some of these formulas can get a little fuzzy — especially if you don't use them every day. These are your at-a-glance reference tools with all the equations and definitions data scientists might encounter.

- Probabilities — Basics

- Probabilities — Advanced

- Data Mining

- Calculus

- Linear Algebra — Basics

- Linear Algebra — Deep Learning

Data Science Resources

If you're early in your career or still studying to become a data scientist, you may occasionally have to refresh your memory on key definitions and Excel functions. You'll save time with online guides that offer up direct answers, short definitions, and paste-able formulas.

The following resources will serve as helpful tools to help you get ready for your first data science interview, prepare for a data analysis test, or do a V lookup in Excel.

- Data Science Glossary

- Big Data Overview

- Excel Formulas and Functions

- Basic Programming Guide

- Hadoop

Grow Your Career in Data

Wherever you are on your data science path, it's crucial to keep up with new developments in this fast-changing tech field. Every element of your job is subject to grow and evolve over the years. The programming languages, software, and methodologies that shape data analysis are constantly improving and becoming more robust. It's one of the things that makes this career path so exciting.

So keep on learning and growing professionally. Look for local meetups in your city. If you're interested in delving deeper into one specific area of data science, now is the perfect time to try virtual conferences or webinars on big data, deep learning or AI.

Data Science   Cheatsheet   Machine Learning   Artificial Intelligence

About   Help   Legal

Get the Medium app