DS 4A EMPOWERMENT

correlation.one

DATA SCIENCE FOR ALL RESOURCES

Provided here is a list of curated resources that can help you learn the importance of data science and jumpstart your development as a data-literate individual.

As part of our mission to make data science for all, we have worked hard to curate the best materials across the internet to help you on your journey. We have also ensured that all of these resources are completely free. They provide a strong foundation for you to learn analytics, and will also position you well to continue your learning either in a future DS4A Empowerment cohort or elsewhere.

CONTENT

HOW TO USE THESE RESOURCES

Going through all the resources provided below would take hundreds of hours. We do not expect that to work for most people. So how should you use these resources?

First, consider what you want to get out of these resources. Do you want foundational data literacy to stand out in your current field? Do you want to be able to code with a focus on data? Do you want a career switch to data science?

Second, consider how much time you're willing to put into this. If you want meaningful results, you need to make an honest effort and actually put time in consistently and over a sustained period of time. With that said, you do not need to sink in hundreds of hours to see noticeable development, as we have curated the most-effective-per-unit-time resources for you.

The table below outlines multiple routes you can take through these resources based on your answers to the above questions (of course, feel free to define your own path). Regardless of what route is best for you, please go through the "Why is AI Important" and "How are Data Science, Data Literacy, and AI Different" sections to develop a better understanding of why data science and data literacy are important.

HOURS	LEARNING OBJECTIVE		
	DATA LITERACY	**CODING + SOME DATA**	**DATA SCIENCE**
20	Read chapter 1 of Data Science: A Gentle Introduction, find your favorite code-free data science resource and become an expert with that tool	Work through the learnpython.org tutorials, read chapters 1 & 2 of Data Science: A Gentle Introduction, complete all the Pandas intermediate level Python tutorials	Work through the learnpython.org tutorials, read chapters 1, 2, & 7 of Data Science: A Gentle Introduction, complete all the Pandas intermediate level Python tutorials
50	Read all of Data Science: A Gentle Introduction, find your favorite code-free data science resource and become an expert with that tool	Work through all of Python for Everybody, read chapters 1 & 2 of Data Science: A Gentle Introduction, complete all the Pandas intermediate level Python tutorials	Work through chapters 1 - 11, 15, & 16 of Python for Everybody, work through chapters 1, 2, & 9 of Openstax's Intro Stats, complete all the Pandas intermediate level Python tutorials
100+	Work through the learnpython.org tutorials, read all of Data Science: A Gentle Introduction, complete all the Pandas intermediate level Python tutorials, learn 2 code-free data science tools (that have different use cases)	Work through all of Python for Everybody, read through all of Data Science: A Gentle Introduction, complete all the Pandas intermediate level Python tutorials, try out several advanced tutorials from realpython	Work through chapters 1 - 11, 15, & 16 of Python for Everybody and work through all of Think Stats, try out several data science/ML advanced tutorials from realpython

If you're coming in with prior stats or Python experience, the best route for you is probably not one of the above. If this is the case for you, we encourage you to pick and choose from the resources listed here that best fill in your missing skills.

WHY IS AI IMPORTANT?

If you work in finance, wouldn't you want to know which loan applicants will likely default? Or if you were in sales, which people are most likely to buy your product?

Many people would approach this by applying rules of thumb or by using their intuition. These approaches are somewhat effective, but also give incorrect conclusions fairly often. Whereas, if you could predict the outcome correctly nearly every time, you could save a lot of time and money.

This is the crux of AI: being able to optimally predict an outcome based on the given information. Instead of using rules of thumb, AI analyzes all the data you have on past loan defaults or successful sales to make the right decision. As long as you have enough data on past dealings, you can use AI to improve your future work.

Regardless of what industry you work in, AI can noticeably improve your dealings in it. For examples of how AI is changing your industry or others, we suggest you read the second chapter in the State of AI 2019.

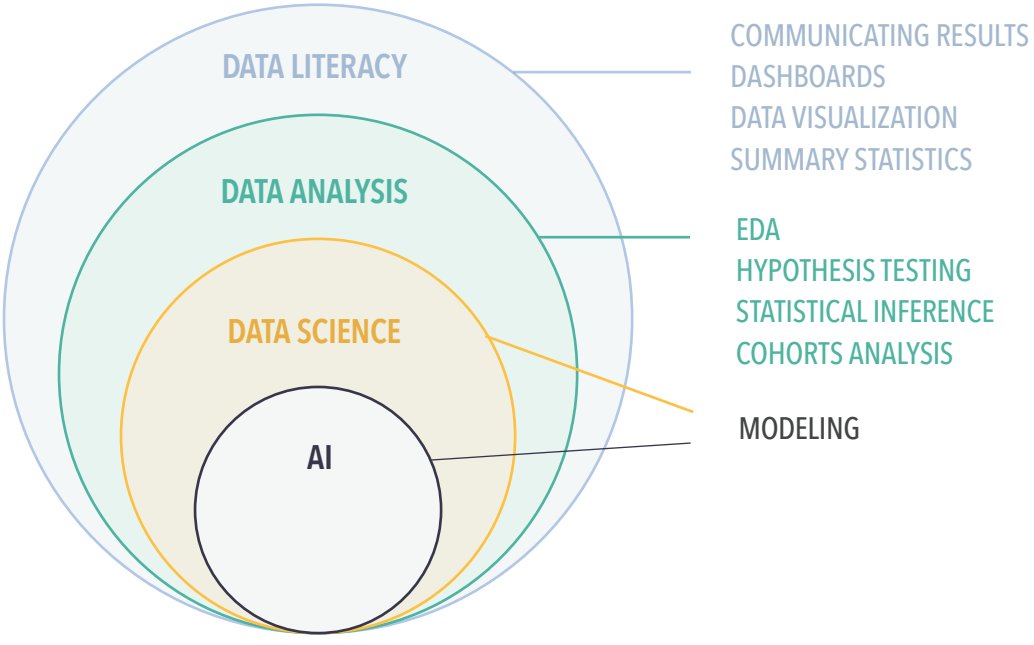
HOW ARE DATA SCIENCE, DATA LITERACY, AND AI DIFFERENT?

AI is just one component of data science. It is an advanced form of data modelling where the "author" of the model does not know exactly how the model works (just as we do not know exactly how our brains function), yet the model still gives useful results.

Typical data science models are explicit - that is, the author knows exactly what the program is doing and why. These models are extremely valuable and require a similar amount of expertise as AI to create. The difference is that AI can tackle some more complex problems (like artificial vision) whereas data science models efficiently tackle more tractable problems (like predicting future sales accounting for seasonality).

In contrast to AI and data science, data literacy has us take a step back from the weeds of manipulating data and instead has us communicating with data. That is, being able to read and interpret visualizations and summary statistics to draw accurate conclusions from data. Moreover, data literacy allows us to effectively communicate those conclusions to anyone, regardless of their background.

The below graphic provides an overview of the relationship between data literacy, data analysis, data science, and AI.



DATA SCIENCE: KEY TOPICS

Hopefully at this point you see the value of AI and the data science that underpins it. Since you cannot create useful AI without a firm foundation in data science, what are the key topics and skills you should know to build this foundation?

IBM created an excellent guide that answers just that. For a clean overview, go through pages 3 and 4. If you want a detailed breakdown of the topics, go through the tables on pages 6-11.

Fundamentally, you need to understand some statistics and some programming to create useful insights from data. For this reason, we suggest that you explore the Python and statistics resources provided below.

If coding sounds too intimidating, we also suggest several resources where you can apply data science techniques without having to code! (see the Code-Free Data Science Resources section).

STATISTICS RESOURCES

Below is a table of excellent resources to learn statistics with. Each resource comes with a time estimate to read through the resource thoroughly + time for dedicated practice to solidify your learning. The read times are based on slow, high absorption reading (your mileage may vary). Do not just skim through these books if you are looking to learn.

RESOURCE	KEY CONSIDERATIONS	WHY USE THIS RESOURCE?	ESTIMATED TIME COMMITMENT
Data Science: A Gentle Introduction	- Very easy to read - Presented from a data science (not statistics perspective) - You will need to find your own practice problems	Great if you want to learn statistics through data science applications. Gives you a better sense of purpose (to be a data scientist) as you work through the book	30h + 20h of practice
Think Stats	- Includes code on Github as examples and homeworks - Well written	If you already know Python, it will teach you both statistics and Python's stats libraries	30h + 20h of practice
Openstax's Introductory Statistics	- Many detailed exercises throughout the chapters - Provided "try it" exercise solutions	You will likely find the text easier to follow than OpenIntro. It will give you a very solid statistics foundation	90h + 0h (practice included throughout)
OpenIntro Statistics	- Complementary videos for most sections	Ideal if you want both lectures and a textbook. It will give you a solid statistics foundation	45h + 15h of extra practice
Statistics for beginners video-course	- Full video table of contents for easy navigation - Slow and detailed explanations	The video is concise and presents a manageable amount of content to learn	7.5h video (watch at x1.25 speed) + 25h of practice

PYTHON RESOURCES

In data science, two programming languages are dominant: Python and R. R has a stronger presence in academia, however we recommend Python as it is the more popular language generally and has a very active community.

STARTING POINT

If you have no familiarity with Python, we suggest you start with the Hello World! tutorial from learnpython.org. This site provides guides that will walk you through how to code in Python right from the beginning up to some basic data science applications. You can expect to learn Python's syntax from these tutorials, but do not expect to develop a deep understanding of programming. We suggest you complete all of the tutorials up to and including Pandas Basics. Working through these tutorials will likely take you 7 to 10 hours to complete.

If you want a deeper understanding of programming, or if you prefer a book to walk you through Python, we suggest Python for Everybody (which includes recorded lectures and graded assignments to assist your learning). Be sure to complete the exercises as you go through the book, otherwise you will not retain your learnings. In total, going through this book and practicing will likely take you around 40 to 50 hours.

WHERE TO CODE

When you start playing around with your own basic Python programs, we suggest you use Code Skulptor. This is an in-browser version of Python that will let you test and run your code without needing to install anything.

Once you move to more advanced programs (such as reading from and writing to files), you will have to install python on your own computer. We suggest you do so by installing Anaconda. Anaconda is a pain-free way to install Python quickly. It also installs an application called Spyder (an IDE), which you can write and run your code in. For your convenience, here are the Anaconda download links for Windows, macOS, and Linux.

ADVANCING YOUR PYTHON

Once you have the fundamentals of Python figured out, if you would like to further develop your Python skills, we suggest the intermediate level tutorials from realpython.com (not all of these are free but our suggested ones are). Below we recommend some of the more useful ones for working with data. Pick and choose the ones that interest you; do not feel as though you need to complete them all!

- Jupyter Notebook Introduction
- The Pandas DataFrame
- Plotting With Pandas
- Tips and Trick With Pandas
- Correlation With Python (Pandas)
- Reading and Writing CSV Files
- Manipulating Excel With Python
- Data Management With Python (SQL)
- Web Scraping in Python
- Working With JSON Data in Python
- Advanced Python Import Techniques
- Object-Oriented Programming
- Natural Language Processing in Python
- Making Web Requests With Python

CODE-FREE DATA SCIENCE RESOURCES

Although having the programming experience lets you make more advanced models, you can use services that handle the coding for you. This can both save time for experienced programmers, and bring the power of data science to those lacking a coding background.

Here are some tools you can use to analyze data with the power of data science:

Google Data Studio -- here are some good resources to learn the tool:
- The ultimate guide to google data studio
- These short video tutorials on how to use the tool

This article lists 8 good code-free data science options with varying use cases

The key to getting value from these code-free options is knowing the right questions to ask (so that you generate impactful insights). This article explains the importance of this step and how to best approach it.

PROVIDE YOUR FEEDBACK

We are trying our hardest to provide you with the best free resources to jumpstart your data science journey. We would love to hear from you about your experience using these resources (which you liked, which you didn't) and if you think we missed some resources that would improve this package.

To provide your feedback and suggestions, or if you would like to talk with us, please contact:
resources@ds4a.com

THANK YOU