

[Support me through my Patreon](#) and [Join the Slack Channel!](#)

DS with TS

Terence's Data Science Repository

If this helped you in any way, I would greatly appreciate it if you supported me [here](#).

I want to accelerate this and turn it into the most comprehensive resource for everyone to have access to.

Table of Content

1. Introduction
2. Mathematics and Statistics
3. Programming Fundamentals
4. Data Collection
5. Data Preparation
6. Data Modelling (Machine Learning Algorithms)
7. Machine Learning Theory
8. Interview Prep Resources
9. Crash Courses & Cheat Sheets
10. Intermediate Data Science Applications
11. Data Science Project Ideas & Walkthroughs
12. Miscellaneous
13. Favorite YouTube Channels

1. Introduction

First, I want to say that I'm not an expert by any means. However, I am sharing my insights and knowledge from my experiences, so feel free to take what you want from this.

My assumption is that as an aspiring Data Scientist, you'll want to **fully understand** the concepts and details of various machine learning algorithms, data science concepts, and so forth.

Therefore, I recommend that you start with the building blocks before you even look at machine learning algorithms or data science applications. If you don't have a basic understanding of calculus & integrals, linear algebra, and statistics, you'll have a hard time understanding the mechanics behind various algorithms. Likewise, if you don't have a basic understanding of Python, you'll have a hard time implementing your knowledge in real life applications.

In this resource, you'll find a repository of resources that I've created and discovered throughout my data science journey, which I believe are the best resources in terms of explaining concepts as simply as possible.

If you find this helpful, I would greatly appreciate it if you can [support me on Patreon](#). Otherwise, feel free to share this with your peers if you find this useful. :)

If you would like to connect, feel free to do so with the links below:

My LinkedIn: <https://www.linkedin.com/in/terenceshin/>

My Github: <https://github.com/terenceshin>

With that said, I wish you the best of luck!

2. Mathematics and Statistics

Like anything else, you have to learn the fundamentals before you get to the fun stuff. TRUST ME, I would've had a much easier time if I started with learning mathematics and statistics before getting into any machine learning algorithms.

Integrals

Integrals are essential when it comes to probability distributions and hypothesis testing. While you don't need to be an expert, it's in your best interest to learn the fundamentals of integrals.

The first two articles are for those who want to get an idea of what integrals is all about or for those who simply need a refresher. If you know absolutely nothing about integrals, I recommend that you complete Khan Academy's course. Lastly, I've provided a link to a number of practice problems to hone your skills.

- [Introduction to Integrals](#) (article)
- [AP Integrals Crash Course](#) (article)
- [Khan Academy: Integral Calculus](#) (course)
- [Practice Questions](#) (*start at unit 6*)

Statistics

If there was one topic that you should focus the majority of your time on, it's statistics. After all, a data scientist is really a modern statistician. I'm going to split this part into two sections: written crash courses and university courses.

If you don't have the time to go watch the university courses or you simply need a refresher, look at the written crash courses. Otherwise, it's in your best interest to go through the university courses.

[Support me through my Patreon](#) and [Join the Slack Channel!](#)

In the very least, you should look at the written crash courses and understand all of the concepts. Likewise, you should go through the beginner courses.

Written Crash Courses

- [Basic Statistics](#)
- [Introduction to Probability](#)
- [Introduction to Random Variables](#) [NEW]
- Introduction to Probability Distributions [coming soon]
- Introduction to Estimation [coming soon]
- [Introduction to Hypothesis Testing](#) [new version soon]
- Introduction to Regression

University Courses

- Beginner Courses
 - [Statistical Concepts and Reasoning](#)
 - [Elementary Statistics](#)
 - [Introduction to Probability Theory](#)
 - [Introduction to Mathematical Statistics](#)
- Intermediate/Advanced Courses
 - [Applied Statistics](#)
 - [Regression Methods](#)
 - [Analysis of Variance and Design of Experiments](#)
 - [Design of Experiments](#)
 - [Analysis of Discrete Data](#)
 - [Applied Multivariate Statistical Analysis](#)
 - [Sampling Theory](#)
 - [Applied Data Mining and Statistical Learning](#)
 - [Applied Time Series Analysis](#)

Linear Algebra

Linear Algebra is especially important if you want to get into deep learning, but even then, it's good to know for other fundamental machine learning concepts, like PCA and eigenvectors.

- [Khan Academy: Linear Algebra](#)

3. Programming Fundamentals

Just as having a fundamental understanding of math and stats is important, having a fundamental understanding of programming will make your life much easier, especially when it comes to implementation. Therefore, I recommend that you take the time to learn basic SQL and Python before diving into machine learning algorithms.

SQL

It's entirely up to you whether you want to learn Python or SQL first, but if you were to ask me, I'd start with SQL. Why? It's easier to learn and it's useful to know if you work for a company that works with data, EVEN if you're not a data scientist. Below are several resources that you can use to learn and practice SQL. If you get through all of this, you should be confident in your SQL abilities!

Learn SQL

- [Learn Basic SQL in 5 Steps](#)
- [Mode's SQL Tutorials](#)
- [6 Resources for Advanced SQL](#)
- [Ten SQL Concepts You Should Know for Data Science Interviews](#)

Practice SQL

- [Leetcode Questions](#)

[Support me through my Patreon](#) and [Join the Slack Channel!](#)

- [Hackerrank Questions](#)
- [Case Studies](#)

Python

I started with Python and I'll probably stick with Python for the rest of my life. It's so far ahead in terms of open source contributions and it's very simple to learn. Feel free to go with R if you want, but I have no opinions or advice in regards to it. Personally, I found that learning Python by 'doing' is much more helpful. That being said, after going through several Python crash courses, I found this one to be the most comprehensive (and it's free!). *Feel free to reach out to me if you know a better resource than this!*

- [Introduction to Python Programming - Georgia Tech](#)

Pandas

Arguably the most important library to know in Python is Pandas, which is specifically meant for data manipulation and analysis. This library will be extremely useful when it comes to the Data Preparation step (#5).

Below are two resources that should ramp you up pretty quickly. The first link is a tutorial on how to use Pandas and the second link provides dozens and dozens of practice problems that you can use to solidify your learnings!

- [Learn Pandas with Kaggle.](#)
- [Practice Pandas with dozens of practice problems!](#)

4. Data Collection

If you've gotten this far, congrats! You're now ready to start your very own machine learning project. The first thing you'll need to do is think of a problem that you would like to solve with data, and gather your data. Below are three ways that you can collect data:

1. Get data from public sources
2. Get data via web scraping

3. Get data via APIs

5. Data Preparation

[An Extensive Step By Step Guide for Data Preparation](#)
[A Comprehensive Guide to Data Exploration](#)

6. Data Modelling

Machine Learning Algorithms

- Linear Regression
 - [Comprehensive Summary](#)
 - [Mathematical Explanation](#)
 - [Video](#)
- Logistic Regression
 - Comprehensive Summary [coming soon]
 - [Video](#)
- K-Nearest Neighbours
 - Comprehensive Summary [coming soon]
 - [Video \(In-depth explanation\)](#)
- Decision Trees [coming soon]
 - [Video](#)
- Naive Bayes
 - [Comprehensive Summary](#)
 - [Video](#)
- Support Vector Machines (SVMs)
 - Comprehensive Summary [coming soon]
 - [Python Implementation](#)
 - [Mathematical Explanation](#)
- Neural Networks
 - [Comprehensive Summary](#)
- Random Forests
 - [Video](#)

[Support me through my Patreon](#) and [Join the Slack Channel!](#)

- AdaBoost
 - [Comprehensive Summary](#)
 - [Video](#)
- Gradient Boost
 - [Video](#)
- XGBoost
 - [Video](#)
- K-Means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)
 - [Video](#)

Fundamental Machine Learning Concepts

- [Bias and Variance Tradeoff](#)
- [Regularization](#)
- [Confusion matrix and relevant metrics](#)
- [AUC and ROC \(video\)](#)
- [Bootstrap Sampling](#)
- [Ensemble Learning, Bagging, and Boosting](#)
- Scaling vs Standardization vs Normalization [coming soon]

Data Science Project Ideas & Walkthroughs

Data Science Projects

- Collaborative Filtering Recommendation System for Movies ([Github](#)) [NEW]
- Predicting Wine Quality w/ Classification Techniques ([Article](#), [Github](#))
- A/B Testing with Chi-Squared Test ([Article](#), [Github](#))
- Predicting Used Car Prices ([Article](#), [Github](#))
- Coronavirus Data Visualizations using Plotly ([Article](#), [Github](#))
- How to Build a Simple Marketing Mix Model ([Article](#), [Github](#))

[Support me through my Patreon](#) and [Join the Slack Channel!](#)

Data Science Project Ideas

- [14 Data Science Projects to do During Your 14 Day Quarantine](#)
- [7 Data Science Project Ideas for Aspiring Data Scientists](#)

Interview Prep Resources

Interview Questions and Answers

- 100+ Data Scientist Interview Practice Problems ([Subscribe to Patreon to see it](#))
- 30+ Data Science Interview Questions from FAANG Companies ([Subscribe to Patreon to see it](#))

Other Resources

- [How to Explain Each Machine Learning Model at an Interview](#)
- [Ten SQL Concepts You Should Know for Data Science Interviews \[NEW\]](#)

Crash Courses and Cheat Sheets

- [All Machine Learning Models Explained in 6 Minutes](#)
- [A Guide for Exploratory Data Analysis \(EDA\)](#)
- [How to Evaluate Your Machine Learning Models](#)
- [A Pandas Glossary for Data Science](#)

Intermediate Data Science Applications

Recommendation Systems

- Video: [How does Netflix recommend movies? Matrix Factorization](#)

[Support me through my Patreon](#) and [Join the Slack Channel!](#)

- Video: [Singular Value Decomposition](#)
- Course: [Machine Learning and AI Foundations - Recommendations](#)
- Project walkthrough: [Building a Movie Recommendation Engine](#)

Survival Analysis

- [Part 1: What is Censoring?](#)
- [Part 2: Survival Function, Hazard, Hazard Ratio](#)
- [Part 3: Kaplan Meier vs Exponential vs Cox Proportional Hazards](#)

Miscellaneous

- [How Artificial Intelligence is Transforming Business in 2020](#)
- [What Happened to the Apple Card?](#)

Favorite YouTube Channels

- [StatQuest](#)
- [sentdex](#)
- [3Blue1Brown](#)
- [Luis Serrano](#)
- [Tech with Tim](#)