

# KDnuggets

[Subscribe to KDnuggets](#)



[Submit a blog](#)  
[Win a Reward!](#)




- [Blog](#)
- [Opinions](#)
- [Tutorials](#)
- [Top stories](#)
- [Courses](#)
- [Datasets](#)
- [Education: Online](#)
- [Certificates](#)
- [Events / Meetings](#)
- [Jobs](#)
- [Software](#)
- [Webinars](#)



[Everything you need to know about data prep. Get \*Data Preparation for Dummies\* ebook](#)

[Submit a blog](#) to KDnuggets -- [Top Blogs Win A Reward](#)

Topics: [AI](#) | [Data Science](#) | [Data Visualization](#) | [Deep Learning](#) | [Machine Learning](#) | [NLP](#) | [Python](#) | [R](#) | [Statistics](#)

[KDnuggets Home](#) » [News](#) » [2021](#) » [May](#) » [Tutorials, Overviews](#) » A Guide On How To Become A Data Scientist (Step By Step Approach) ( [21:n20](#) )

## A Guide On How To Become A Data Scientist (Step By Step Approach)

[<= Previous post](#)

[Next post =>](#)

Like 312

Share 312

Tweet

Share

Share

498

Tags: [Career Advice](#), [Data Science Skills](#), [Data Scientist](#), [Python](#), [R](#), [SQL](#), [Statistics](#)

Becoming a Data Scientists is an exciting path, but you cannot learn data science within one year or six months—instead, it's a lifetime process that you have to follow with proper dedication and hard work. To guide your journey, the skills outlined here are the first you must acquire to become a data scientist.



### SAS Talent Development

---

[comments](#)

By [Aditya Agarwal](#), Graduate Student at Northeastern U.

There are tons of resources and links out there, but often we get confused on which resources to follow. Don't worry, I have got you covered. I have attached the links to several YouTube channels, blogs, courses, and other websites that I found appropriate for a beginner.

You can also use the Data Science Community Websites Like Analytics Vidhya and Kaggle for implementing your learning and getting hands-on experience in Data Science.



Data Science Roadmap

## STEP 1: Choose A Programming Language (Python / R)

The first step while starting the Data Science Journey is to get familiar with a programming language. Between the two, Python is the most preferred coding language and is adopted by most Data Scientists. It is easy to understand, versatile, and supports various in-built libraries such as Numpy, Pandas, Matplotlib, Seaborn, Scipy, and many more.

1. [FreeCodeCamp's Python Tutorial](#) (Recommended)
2. [Kaggle's Python Course](#)
3. [Krish Naik's Python Tutorial](#) (Recommended)
4. [Udemy's Python for Data Science and Machine Learning Bootcamp](#)
5. [Coursera Python Course](#)

*NOTE:* While learning Python, one should know essential Python variables, data types, OOPs concepts, Numpy, Pandas, Matplotlib, and Seaborn.

## STEP 2. Statistics

For becoming a Data Scientist, having knowledge of statistics and probability is as essential as having salt in food. Knowing them will help the data scientists interpret large data sets, get insights from them, and analyze them better.

1. [Krish Naik's Statistics Playlist](#) (Recommended)
2. [Coursera Statistics Course](#)
3. [Khan Academy Statistics And Probability Course](#)
4. [FreeCodeCamp Statistics Course](#) (Recommended)

*NOTE:* Statistics provides the ideas about Mean, Median, Mode, Range, Variance, Standard Deviation, Graphs or Plotting, Populations, and Samples.

## STEP 3: Learn SQL

Structured Query Language (SQL) is used for extracting and communicating with large databases. One should focus on understanding the different types of normalization, writing nested queries, using co-related questions, group-by, performing join operations, etc., on the data and extract in raw format. This data will then further be cleaned either in Microsoft Excel or by using Python libraries.

1. [FreeCodeCamp SQL](#) (Recommended)
2. [Intro To SQL By Kaggle](#) (Recommended)
3. [Advanced SQL By Kaggle](#)
4. [Edureka's SQL Playlist](#)

*NOTE:* In SQL, one should know about creating tables, inserting data, updating data, deleting data, and performing some basic query operations.

## STEP 4. Data Cleaning

When a Data Scientist is given a project, the majority of the time goes into cleaning the data set, removing unwanted values, handling missing values. It can be achieved by using some inbuilt python libraries like Pandas and Numpy. One should also know how to manipulate data using Microsoft Excel.

1. [Blog — Cleaning Data Using Python](#) (Recommended)
2. [Edureka's Microsoft Excel Course](#)
3. [Learning Pandas By Kaggle](#) (Recommended)

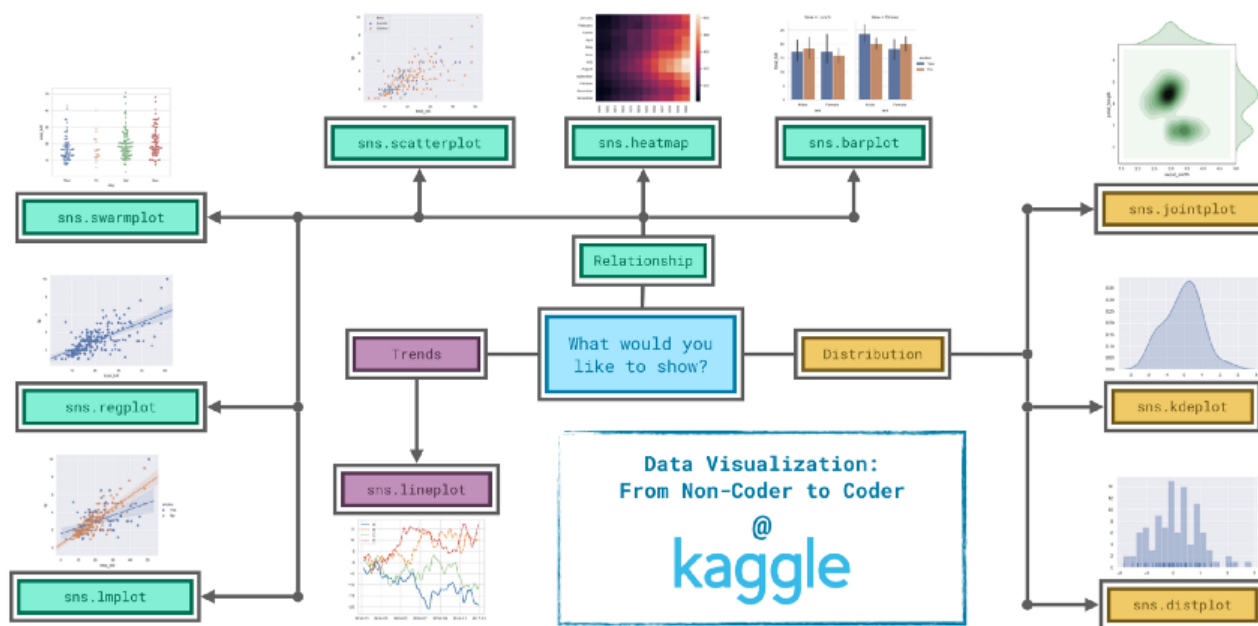
*NOTE:* In Microsoft Excel, you should know basic data filtering or sorting, Functions or Formulas, Vlookup, Pivot table and charts, and Tables, etc.

## STEP 5: Exploratory Data Analysis

Exploratory data analysis is the essential part when talking about data science. The data scientist has many tasks, including finding data patterns, analyzing data, finding the appropriate trends in the data and obtaining valuable insights, etc., from them with the help of various graphical and statistical methods, including:

- A) Data Analysis using Pandas and Numpy
- B) Data Manipulation
- C) Data Visualization

1. [Intro To EDA By Code Heroku's](#) (Recommended)
2. [Blog — Performing EDA on Iris Data Set](#) (Recommended)
3. [Coursera Course On EDA, Statistics, Probability](#)



*Types of plots in the Seaborn Python library.*

## STEP 6: Learn Machine Learning Algorithms

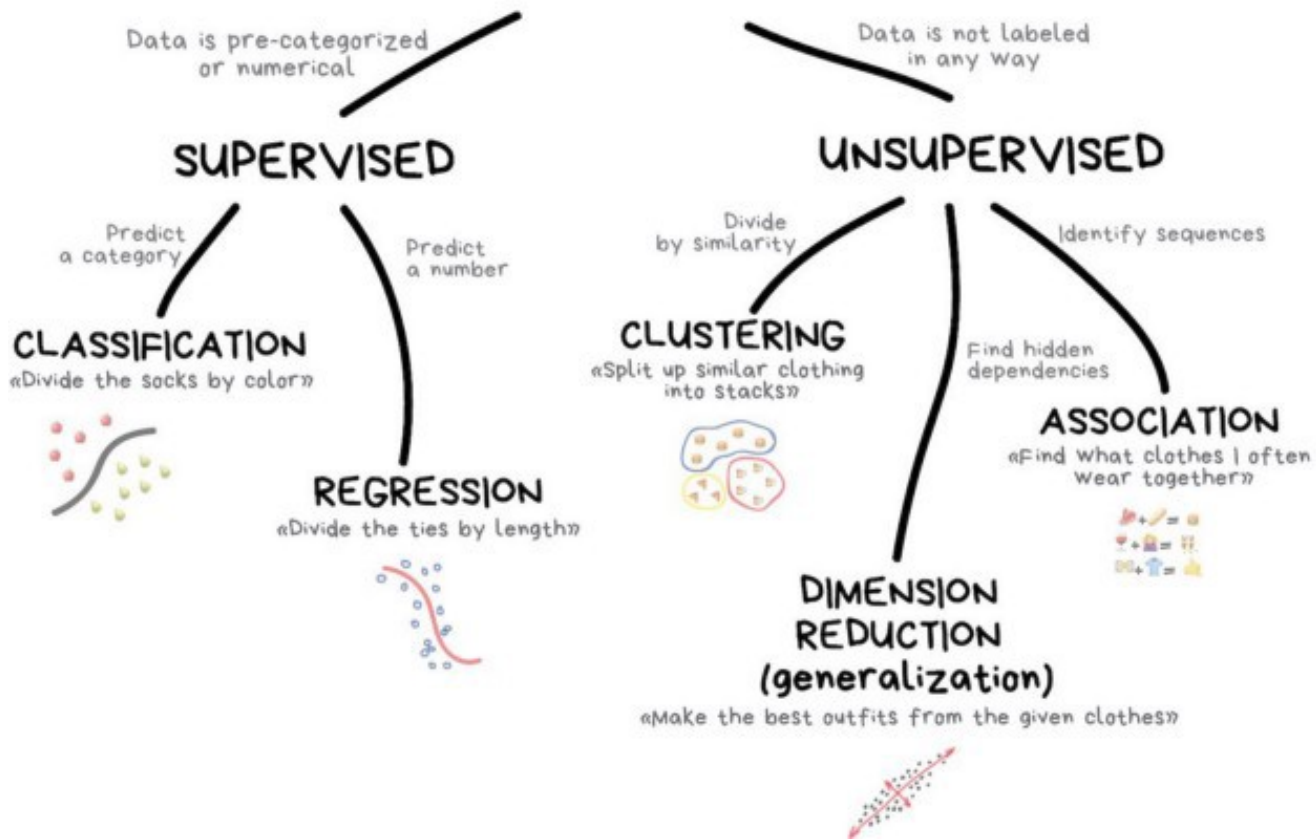
According to Google, “Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.”

It is the most crucial step in a life cycle of a data scientist where one has to build various models using machine learning algorithms and should be able to predict and come with the most optimum solution to solve any problem.

1. [Machine Learning By Andrew NG](#) (Recommended)
2. [Deep Learning By Krish Naik](#)
3. [Intro To ML By Kaggle](#) (Recommended)
4. [Machine Learning By Krish Naik](#) (Recommended)
5. [Coursera Deep Learning Specialization](#)



# CLASSICAL MACHINE LEARNING



*Machine Learning landscape.*

## Step 7: Practice On Analytics Vidhya and Kaggle

After acquiring the basics of Data Science, now it's time to get hands-on experience in its part. There are many online platforms, like Kaggle and Analytics Vidhya, that can provide you with hands-on experience with both beginner and advanced level data sets. They can help you to understand various machine learning algorithms, different analyzing techniques, etc.

You can follow the below approach to know how effectively you can use these platforms.

1. You can start by first downloading the datasets and analyzing the data, and implementing all the different techniques you have learned.
2. Next, you can check on other people's notebooks and understand how they have solved a particular problem or gained insights from the data. (This method will certainly make you more confident and help to improve your knowledge.)
3. After you are confident enough, you can participate in Competitions organized by both Kaggle and Analytics Vidhya. This will not only help you to sharpen your skills in Data Science but also to learn Data Science better.

[Original](#). Reposted with permission.

### Related:

- [7 Must-Haves in your Data Science CV](#)
- [Data Scientist vs Machine Learning Engineer – what are their skills?](#)
- [15 Habits I Learned from Highly Effective Data Scientists](#)

What do you think?

## Important Update

When you log in with Disqus, we process personal data to facilitate your authentication and posting of comments. We also store the comments you post and those comments are immediately viewable and searchable by anyone around the world.

Please access our Privacy Policy to learn what personal data Disqus collects and your choices about how it is used. All users of our service are also subject to our Terms of Service.

Proceed

[<= Previous post](#)  
[Next post =>](#)

## Top Stories Past 30 Days

### Most Popular

1. [Data Scientist, Data Engineer & Other Data Careers, Explained](#)
2. [Vaex: Pandas but 1000x faster](#)
3. [Data Science Books You Should Start Reading in 2021](#)
4. [Data Preparation in SQL, with Cheat Sheet!](#)
5. [Charticulator: Microsoft Research open-sourced a game-changing Data Visualization platform](#)

### Most Shared

1. [Data Scientist, Data Engineer & Other Data Careers, Explained](#)
2. [Essential Linear Algebra for Data Science and Machine Learning](#)
3. [Data Scientist vs Machine Learning Engineer – what are their skills?](#)
4. [DeepMind Wants to Reimagine One of the Most Important Algorithms in Machine Learning](#)
5. [How to Determine if Your Machine Learning Model is Overtrained](#)

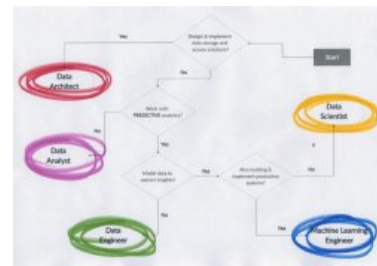
## Latest News

- [State of Mathematical Optimization Report, 2021](#)
- [Essential Math for Data Science: Basis and Change of Basis](#)
- [4 Tips for Dataset Curation for NLP Projects](#)
- [Choosing the Right BI Tool for Your Business](#)
- [AIRSIDE LIVE Is Where Big Data, Data Security and Data ...](#)
- [Great New Resource for Natural Language Processing Rese...](#)

Top Stories  
Last Week

[Most Popular](#)

1. [Data Scientist, Data Engineer & Other Data Careers, Explained](#)
2. [Vaex: Pandas but 1000x faster](#)
3. [A checklist to track your Data Science progress](#)
4. [How to Determine if Your Machine Learning Model is Overtrained](#)
5. [The Most In Demand Skills for Data Engineers in 2021](#)



### Most Shared

1. [Data Scientist, Data Engineer & Other Data Careers, Explained](#)
2. [How to Determine if Your Machine Learning Model is Overtrained](#)
3. [Vaex: Pandas but 1000x faster](#)
4. [Awesome list of datasets in 100+ categories](#)
5. [Binary Classification with Automated Machine Learning](#)

### More Recent Stories

- [Great New Resource for Natural Language Processing Research an...](#)
- [AI Books you should read in 2021](#)
- [Top Data and Analytics Trends](#)
- [Budgeting For Your AI Training Data: Consider These 3 Factors](#)
- [Topic Modeling with Streamlit](#)
- [Top Programming Languages and Their Uses](#)
- [Essential Machine Learning Algorithms: A Beginner's Guide](#)
- [KDnuggets 21:n20, May 26: Data Engineer, Data Scientists &#...](#)
- [The Rise of Vector Data](#)
- [Where Did You Apply Analytics, Data Science, Machine Learning ...](#)
- [These Soft Skills Can Make or Break Your Data Science Career](#)
- [Write and train your own custom machine learning models using ...](#)
- [Top Stories, May 17-23: Data Scientist, Data Engineer & O...](#)
- [How to Deal with Categorical Data for Machine Learning](#)
- [A Guide On How To Become A Data Scientist \(Step By Step Approach\)](#)
- [Data Validation in Machine Learning is Imperative, Not Optional](#)
- [6 Business Trends Benefiting Data Scientists](#)
- [How to pitch to VCs, explained: The Deck We Used to Raise Capi...](#)
- [Building RESTful APIs using Flask](#)
- [DataOps: 5 things that you need to know](#)

[KDnuggets Home](#) » [News](#) » [2021](#) » [May](#) » [Tutorials, Overviews](#) » [A Guide On How To Become A Data Scientist \(Step By Step Approach\)](#) ( [21:n20](#) )

© 2021 KDnuggets. | [About KDnuggets](#) | [Contact](#) | [Privacy policy](#) | [Terms of Service](#)

[Subscribe to KDnuggets News](#)

X