

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

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.



[Capture the magic of data science for kids.](#)

[Get a FREE e-book or 50% off hardcopy. Get tonight's bedtime story](#)

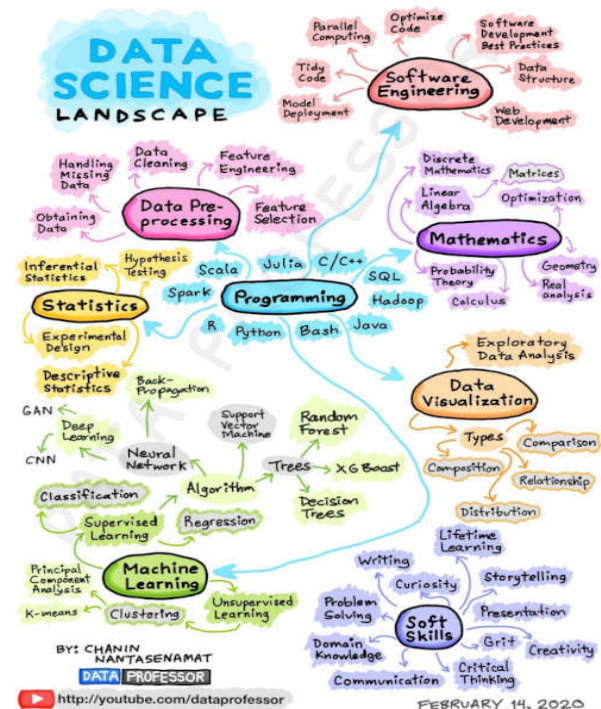
[comments](#)

[By Aditya Agarwal, Graduate Student at North-eastern 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)

