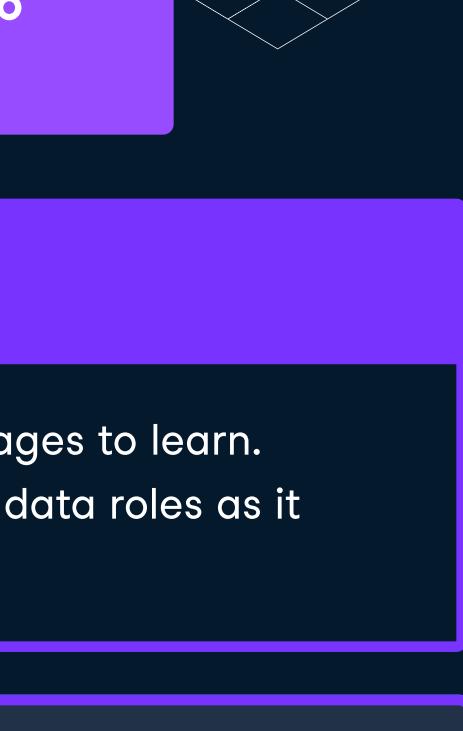




The Data Scientist Learning Path Checklist



Data science is a popular and lucrative career that involves analyzing and managing data, using machine learning and programming skills, and understanding business needs. It requires a variety of skills, including data analysis, business acumen, communication skills, and more. Use this checklist to guide your data science learning journey.

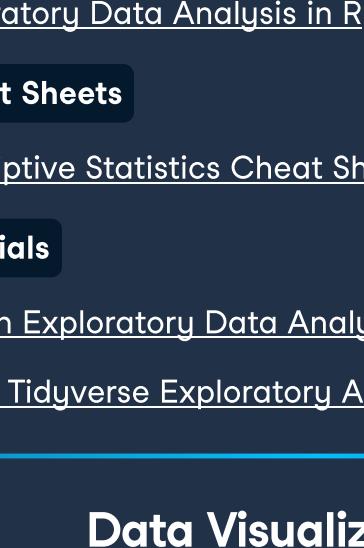


Choose your tool

When getting started with data science, it is important to choose which programming languages to learn. Two popular choices are R and Python. Additionally, learning SQL is important for almost all data roles as it is a standard language for working with databases.



R is a programming language and software environment for statistical computing and graphics. It is widely used by data scientists for statistical analysis, data visualization, and machine learning.



SQL (Structured Query Language) is a programming language used to manage and manipulate data stored in relational databases. It is used to create, modify, and query databases, as well as to control access to the data within them. It is widely used in most data roles today.



Python is a popular programming language for data science due to its useful libraries and easy syntax. It can be used for various data science tasks, such as data cleaning, statistical analysis, and machine learning. Python is the most popular data science programming language.

Skills checklist

Learn on DataCamp

Apply your skills

Exploratory Data Analysis

Descriptive Statistics

- Calculate metrics on measures of location like mean and median, measure of variation like range and standard deviation, and other characteristics of features
- Calculate metrics like correlation to understand the relationships between feature

Courses

[Introduction to Statistics in Python](#)

[Introduction to Statistics in R](#)

[Exploratory Data Analysis in Python](#)

[Exploratory Data Analysis in R](#)

Cheat Sheets

[Descriptive Statistics Cheat Sheet](#)

Tutorials

[Python Exploratory Data Analysis Tutorial](#)

[Video: Tidyverse Exploratory Analysis](#)

Projects

[A Visual History of Nobel Prize Winners](#)

[Optimizing Online Sports Retail Revenue](#)

Workspace Template

[Explore a DataFrame](#)

Live Trainings

[Analyzing Carbon Footprints in SQL](#)

[Exploring World Cup Data in Python](#)

Data Visualization

- Create plots like bar plots, histograms and box plots to visualize single features.
- Create plots like scatter plots, line plots and heat maps to visualize relationships between features.

Courses

[Introduction to Data Visualization with Seaborn](#)

[Introduction to Data Visualization with Plotly in Python](#)

[Introduction to Data Visualization with ggplot2](#)

[Interactive Data Visualization with plotly in R](#)

Cheat Sheets

[Data Visualization Cheat Sheet](#)

[Python Seaborn Cheat Sheet](#)

[Plotly Express Cheat Sheet](#)

[ggplot2 Cheat Sheet](#)

Tutorials

[Python Seaborn Tutorial For Beginners](#)

[Graphics with ggplot2 Tutorial](#)

Projects

[Visualizing COVID-19 in R](#)

[Modeling the Volatility of US Bond Yields in R](#)

[Exploring the Bitcoin Cryptocurrency Market in Python](#)

[Real-time Insights from Social Media Data in Python](#)

Workspace Template

[Visualize Correlation with a Diagonal Correlation Plot in Python](#)

Live Trainings

[Data Visualization in Python for Absolute Beginners](#)

[Visualizing Video Game Sales Data with ggplot2 in R](#)

Data Management

Importing & Reading Data

- Import data from common file formats like CSV and spreadsheets.
- Import data by querying SQL databases.
- Import data via web APIs.

Courses

[Introduction to Importing Data in Python](#)

[Intermediate Importing Data in Python](#)

[Streamlined Data Ingestion with pandas](#)

[Introduction to Importing Data in R](#)

[Intermediate Importing Data in R](#)

[Introduction to SQL](#)

Cheat Sheet

[Importing Data in Python Cheat Sheet](#)

Tutorials

[Pandas Tutorial: Importing Data with read_csv\(\)](#)

[Web Scraping With Python and Beautiful Soup](#)

[How to Import Data Into R: A Tutorial](#)

[Importing Data Into R - Part Two](#)

Projects

[Importing and Cleaning Data](#)

[The Android App Market on Google Play](#)

Workspace Template

[Visualize Historical Stock Data with a Candlestick Chart](#)

Live Trainings

[Analyzing Streaming Service Content in SQL](#)

[Analyzing Students' Mental Health in SQL](#)

Data Wrangling

- Perform common data manipulations such as sorting, subsetting, adding new features, and aggregating.
- Join two datasets together via inner, left and other joins.
- Pivot a rectangular dataset to convert rows to columns or columns to rows.

Courses

[Data Manipulation with pandas](#)

[Joining Data with pandas](#)

[Reshaping Data with pandas](#)

[Data Manipulation with dplyr](#)

[Joining Data with dplyr](#)

[Reshaping Data with tidyr](#)

[Joining Data in SQL](#)

Cheat Sheets

[Pandas Cheat Sheet for Data Science in Python](#)

[Data Manipulation with dplyr in R Cheat Sheet](#)

[SQL Joins Cheat Sheet](#)

[Pandas Cheat Sheet: Data Wrangling in Python](#)

Tutorials

[Joining DataFrames in pandas Tutorial](#)

[Joins in SQL Tutorial](#)

Projects

[What and Where are the World's Oldest Businesses?](#)

[Streamlining Employee Data](#)

Workspace Template

[Merge DataFrames](#)

Live Training

[Analyzing NASA Planetary Exploration Budgets in SQL](#)

[Analyzing Students' Mental Health in SQL](#)

Data Cleaning

- Identify and fix issues with data constraints such as wrong data types, numbers out of range, or duplicate values.
- Identify and fix issues with text and categorical data such as invalid categories or incorrect formatting.
- Identify and fix issues with data uniformity such as incorrect units, incorrect date formats, and inconsistency between features.
- Identify and fix issues with missing data values.

Courses

[Cleaning Data in Python](#)

[Cleaning Data in R](#)

[Cleaning Data in SQL](#)

Infographic

[Data Cleaning Checklist](#)

Tutorials

[Data Cleaning Tutorial](#)

[Cleaning Data in SQL](#)

Projects

[Exploring the Bitcoin Cryptocurrency Market in Python](#)

[Real-time Insights from Social Media Data in Python](#)

Business Acumen

Business Goals

- Make recommendations for analytic approaches based on business goals
- Judge performance of analytic results against KPIs or other relevant business criteria

Courses

[Data-Driven Decision Making for Business](#)

[Analyzing Business Data in SQL](#)

Tutorials

[The Many Business Applications of Machine Learning](#)

[Customer Lifetime Value](#)

Webinar

[Fighting Customer Churn with Data](#)

Projects

[Importing and Cleaning Data](#)

[The Android App Market on Google Play](#)

Workspace Template

[Merge DataFrames](#)

Live Training

[Analyzing Streaming Service Content in SQL](#)

[Analyzing Students' Mental Health in SQL](#)

Organizational Knowledge

- Understand the impact of data science projects on your business.
- Understand which teams or employees need to be involved in a data project, and in what capacity.

Courses

[Data Science for Business](#)

[Machine Learning for Business](#)

Cheat Sheet

[Data Science Cheat Sheet for Business Leaders](#)

Tutorial

[The Impact of Machine Learning Across Verticals and Teams](#)

Projects

[Which Debts Are Worth the Bank's Effort?](#)

Workspace Template

[Feature Engineering for Fraud Detection](#)

[User Retention by Cohort](#)

Live Training

[Analyzing a Marketing Funnel in Spreadsheets](#)

[Visualizing Cost Savings in Tableau](#)

Programming for Data Science

Computational Thinking

- Use common programming constructs like flow control and iteration.
- Understand functions and functional programming to write repeatable code for analysis.

Courses

[Intermediate Python](#)

[Writing Functions in Python](#)

Intermediate R

[Introduction to Writing Functions in R](#)