The Data Scientist Learning Path Checklist

datacamp

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: 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: 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: 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

Skills checklist
- ☐ 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

Learn on DataCamp

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

Apply your skills

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

Skills checklist
- ☐ 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.

Learn on DataCamp

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

Apply your skills

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

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

Learn on DataCamp

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

Apply your skills

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

Skills checklist
- ☐ 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.

Learn on DataCamp

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

Apply your skills

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

Data Cleaning

Skills checklist
- ☐ 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.

Learn on DataCamp

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

Apply your skills

Projects
- Exploring the Bitcoin Cryptocurrency Market in Python
- Real-time Insights from Social Media Data in Python

Business Acumen

Business Goals

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

Learn on DataCamp

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

Apply your skills

Projects
- Comparing Search Interest with Google Trends
- Optimizing Online Sports Retail Revenue

Workspace Template
- Predict CTR and Evaluate ROI
- Calculate Customer Churn Metrics

Organizational Knowledge

Skills checklist
- ☐ 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.

Learn on DataCamp

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

Apply your skills

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

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

Learn on DataCamp

Courses
- Intermediate Python
- Writing Functions in Python
- Intermediate R
- Introduction to Writing Functions in R

Tutorials
- Python Loops Tutorial
- A Loops in R Tutorial - Usage and Alternatives

Apply your skills

Projects
- Functions for Food Price Forecasts
- Writing Functions for Product Analysis

Workspace Template
- Group and Aggregate data with custom functions

Production Coding

Skills checklist
- ☐ Make use of version control like git for managing code
- ☐ Use error handling, assertions, and unit tests to ensure code quality
- ☐ Write documentation to make your code understandable by others
- ☐ Develop packages to make your code reusable

Learn on DataCamp

Courses
- Introduction to Version Control with git
- Software Engineering for Data Scientists in Python
- Developing Python Packages
- Developing R Packages

Cheat Sheet
- Git Cheat Sheet

Tutorials
- Exception and Error Handling in Python
- Unit Testing in Python Tutorial
- What is Git? - The Complete Guide to Git

Apply your skills

Projects
- Functions for Food Price Forecasts
- Writing Functions for Product Analysis

Model Development

Model Design

Skills checklist
- ☐ Choose an appropriate model type (regression, classification, clustering, etc.) based on your dataset and the analysis goals

Learn on DataCamp

Courses
- Supervised Learning with scikit-learn
- Unsupervised Learning in Python
- Supervised Learning in R: Classification
- Supervised Learning in R: Regression
- Unsupervised Learning in R

Cheat Sheets
- Supervised Machine Learning Cheat Sheet
- Unsupervised Machine Learning Cheat Sheet

Tutorial
- 8 Machine Learning Models Explained in 20 Minutes

Apply your skills

Projects
- Predicting Credit Card Approvals
- Predict Taxi Fares with Random Forest
- Classify Song Genres from Audio Data
- Find Movie Similarity from Plot Summaries
- Clustering Heart Disease Patient Data
- ASL Recognition with Deep Learning

Workspace Template
- Disney Movies and Box Office Success

Feature Engineering

Skills checklist
- ☐ Extract problem-relevant information from existing features, like getting the day of week from a datetime variable, or getting an "is working age" indicator from a data of birth.
- ☐ Combine multiple features into new features, for example summing regional sales into total sales, or calculating profit as revenue minus costs.
- ☐ Use external datasets to define new features, for example using a geographic API to get the city from a longitude and latitude, or using a computer vision API to determine if an image contains people.
- ☐ Use imputation to estimate missing values.

Learn on DataCamp

Course
- Feature Engineering for Machine Learning in Python
- Preprocessing for Machine Learning in Python
- Feature Engineering in R

Tutorial
- Machine Learning with Kaggle: Feature Engineering

Apply your skills

Projects
- Customer Analytics: Preparing Data for Modeling
- Predict Taxi Fares with Random Forest
- Classify Song Genres from Audio Data
- Find Movie Similarity from Plot Summaries

Workspace Template
- Encoding Categorical Variables

Live Training
- Sentiment Analysis and Prediction in Python

Model Fitting

Skills checklist
- ☐ Can generate training and testing splits from a dataset, including using cross-validation.
- ☐ Uses hyperparameter tuning to optimize model performance.

Learn on DataCamp

Course
- Hyperparameter Tuning in Python
- Modeling with tidymodels in R
- Hyperparameter Tuning in R

Cheat Sheet
- Scikit-Learn Cheat Sheet: Python Machine Learning

Tutorial
- Hyperparameter Optimization in Machine Learning Models

Apply your skills

Projects
- What Makes a Pokémon Legendary?
- Predict Taxi Fares with Random Forests

Workspace Template
- Machine Learning with Python
- Machine Learning with R

Live Training
- Predicting Hotel Booking Cancellations in Python
- Analyzing a Time Series of the Thames River in Python

Model Validation

Skills checklist
- ☐ Can evaluate supervised learning model performance using metrics like accuracy, precision and recall.
- ☐ Can evaluate unsupervised learning model performance using metrics like homogeneity, completeness, and silhouette coefficient.

Learn on DataCamp

Course
- MLOPs Concepts
- MLOps Deployment and Life Cycling
- Model Validation in Python
- Cluster Analysis in Python
- Cluster Analysis in R

Tutorial
- Python Machine Learning: Scikit-Learn Tutorial

Apply your skills

Projects
- Clustering Bustabit Gambling Behavior
- Degrees That Pay You Back

Workspace Template
- Evaluate your ML Model using the F-score

Live Training
- How to Explain Black-Box Machine Learning Models

Statistical Experimentation

Sampling Methods

Skills checklist
- ☐ Understand statistical distributions like the normal, uniform and Poisson distributions
- ☐ Choose appropriate sampling methods to answer your questions while avoiding bias

Learn on DataCamp

Course
- Foundations of Probability in Python
- Foundations of Probability in R
- Sampling in Python
- Sampling in R

Apply your skills

Projects
- Health Survey Data Analysis of BMI

Hypothesis Testing

Skills checklist
- ☐ Understand null and alternative hypotheses
- ☐ Know when and how to use hypothesis tests like the t-test, Chi-squared test, and Mann-Whitney U test
- ☐ Interpret test statistics and p-values

Learn on DataCamp

Course
- Hypothesis Testing in Python
- Hypothesis Testing in R
- Foundations of Inference in Python
- Foundations of Inference in R

Tutorials
- Hypothesis Testing in Machine Learning
- What is A/B Testing?

Apply your skills

Projects
- Dr. Semmelweis and the Discovery of Handwashing
- Mobile Games A/B Testing with Cookie Cats

Data Communication

Data Storytelling

Skills checklist
- ☐ Create a narrative that describes your motivation, methods, results, and conclusions
- ☐ Ensure your narrative is consistent with the findings of the data
- ☐ Edit your stories to remove extraneous details

Learn on DataCamp

Course
- Communicating Data Insights

Cheat Sheet
- Data Storytelling & Communication Cheat Sheet

Webinars
- Storytelling for More Impactful Data Science
- Effective Data Storytelling: How to Turn Insights into Action

Podcast
- The Data Storytelling Skills Data Teams Need

Apply your skills

Workspace Template
- Tips for Reporting in Workspace

Live Training
- Data Visualization in Python for Absolute Beginners

Understand your Audience

Skills checklist
- ☐ Understand your audience's prior knowledge and interests
- ☐ Tailor your message to resonate with the audience, even if they are non-technical

Learn on DataCamp

Course
- Data Communication Concepts

Tutorials
- Seven Tricks for Better Data Storytelling: Part I
- Seven Tricks for Better Data Storytelling: Part II

Webinars
- Effective Data Storytelling: How to Turn Insights into Action

Apply your skills

Live Training
- Exploring World Cup Data in Python