coursera

< Data Science

Data Science Specialization

Launch Your Career in Data Science. A ten-course introduction to data science, developed and taught by leading professors.

4.5 38,168 ratings

Jeff Leek, PhD +2 more instructors

Offered By

Enroll for Free

Starts Aug 22

Try for Free: Enroll to start your 7-day full access free trial

Financial aid available

467,169 already enrolled

Included with PLUS Unlimited access to 7,000+ courses, Projects, Specializations, and Professional Certificates. **Learn More**

About How It Works Courses Instructors Enrollment Options FAQ

WHAT YOU WILL LEARN

- Use R to clean, analyze, and visualize data.
- Navigate the entire data science pipeline from data acquisition to publication.
- Use GitHub to manage data science projects.
- Perform regression analysis, least squares and inference using regression models.

SKILLS YOU WILL GAIN

Github Machine Learning R Programming Regression Analysis Data Science Rstudio Data Analysis Debugging Data Manipulation Regular Expression (REGEX) Data Cleansing Cluster Analysis

About this Specialization

94,082 recent views

Ask the right questions, manipulate data sets, and create visualizations to communicate results.

This Specialization covers the concepts and tools you'll need throughout the entire data science pipeline, from asking the right kinds of questions to making inferences and publishing results. In the final Capstone Project, you'll apply the skills learned by building a data product using real-world data. At completion, students will have a portfolio demonstrating their mastery of the material.

Shareable Certificate
Earn a Certificate upon completion

100% online courses
Start instantly and learn at your own schedule.

Flexible Schedule
Set and maintain flexible deadlines.

Beginner Level
You should have beginner level experience in Python. Familiarity with regression is recommended

Approximately 11 months to complete
Suggested pace of 7 hours/week

English
Subtitles: English, Arabic, French, Portuguese (European), Italian, Vietnamese, Korean, German, Russian, Spanish, Chinese (Simplified), Portuguese (Brazilian), Japanese

Thanks to learning on Coursera, I'm able to add my courses and certificates to my LinkedIn & resume that make me stand out from my peers.

— Ellen R.

1 2 3

How the Specialization Works

Take Courses

A Coursera Specialization is a series of courses that helps you master a skill. To begin, enroll in the Specialization directly, or review its courses and choose the one you'd like to start with. When you subscribe to a course that is part of a Specialization, you're automatically subscribed to the full Specialization. It's okay to complete just one course — you can pause your learning or end your subscription at any time. Visit your learner dashboard to track your course enrollments and your progress.

Hands-on Project

Every Specialization includes a hands-on project. You'll need to successfully finish the project(s) to complete the Specialization and earn your certificate. If the Specialization includes a separate course for the hands-on project, you'll need to finish each of the other courses before you can start it.

Earn a Certificate

When you finish every course and complete the hands-on project, you'll earn a Certificate that you can share with prospective employers and your professional network.

There are 10 Courses in this Specialization

COURSE 1

The Data Scientist’s Toolbox

4.6 33,163 ratings

In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that

In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, GitHub, R, and RStudio.

COURSE 2

R Programming

4.5 21,693 ratings

In this course you will learn how to program in R and how to use R for effective data analysis. You will learn how to install and configure software necessary for a statistical programming environment and describe generic programming language concepts as they are implemented in a high-level statistical language. The course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, profiling R code, and organizing and commenting R code. Topics in statistical data analysis will provide working examples.

COURSE 3

Getting and Cleaning Data

4.5 7,940 ratings

Before you can work with data you have to get some. This course will cover the basic ways that data can be obtained. The course will cover obtaining data from the web, from APIs, from databases and from colleagues in various formats. It will also cover the basics of data cleaning and how to make data "tidy". Tidy data dramatically speed downstream data analysis tasks. The course will also cover the components of a complete data set including raw data, processing instructions, codebooks, and processed data. The course will cover the basics needed for collecting, cleaning, and sharing data.

COURSE 4

Exploratory Data Analysis

4.7 5,981 ratings

This course covers the essential exploratory techniques for summarizing data. These techniques are typically applied before formal modeling commences and can help inform the development of more complex statistical models. Exploratory techniques are also important for eliminating or sharpening potential hypotheses about the world that can be addressed by the data. We will cover in detail the plotting systems in R as well as some of the basic principles of constructing data graphics. We will also cover some of the common multivariate statistical techniques used to visualize high-dimensional data.

COURSE 5

Reproducible Research

4.6 4,116 ratings

This course focuses on the concepts and tools behind reporting modern data analyses in a reproducible manner. Reproducible research is the idea that data analyses, and more generally, scientific claims, are published with their data and software code so that others may verify the findings and build upon them. The need for reproducibility is increasing dramatically as data analyses become more complex, involving larger datasets and more sophisticated computations. Reproducibility allows for people to focus on the actual content of a data analysis, rather than on superficial details reported in a written summary. In addition, reproducibility makes an analysis more useful to others because the data and code that actually conducted the analysis are available. This course will focus on literate statistical analysis tools which allow one to publish data analyses in a single document that allows others to easily execute the same analysis to obtain the same results.

COURSE 6

Statistical Inference

4.2 4,351 ratings

Statistical inference is the process of drawing conclusions about populations or scientific truths from data. There are many modes of performing inference including statistical modeling, data oriented strategies and explicit use of designs and randomization in analyses. Furthermore, there are broad theories (frequentists, Bayesian, likelihood, design based, ...) and numerous complexities (missing data, observed and unobserved confounding, biases) for performing inference. A practitioner can often be left in a debilitating maze of techniques, philosophies and nuance. This course presents the fundamentals of inference in a practical approach for getting things done. After taking this course, students will understand the broad directions of statistical inference and use this information for making informed choices in analyzing data.

COURSE 7

Regression Models

4.4 3,307 ratings

Linear models, as their name implies, relates an outcome to a set of predictors of interest using linear assumptions. Regression models, a subset of linear models, are the most important statistical analysis tool in a data scientist's toolkit. This course covers regression analysis, least squares and inference using regression models. Special cases of the regression model, ANOVA and ANCOVA will be covered as well. Analysis of residuals and variability will be investigated. The course will cover modern thinking on model selection and novel uses of regression models including scatterplot smoothing.

COURSE 8

Practical Machine Learning

4.5 3,200 ratings

One of the most common tasks performed by data scientists and data analysts are prediction and machine learning. This course will cover the basic components of building and applying prediction functions with an emphasis on practical applications. The course will provide basic grounding in concepts such as training and tests sets, overfitting, and error rates. The course will also introduce a range of model based and algorithmic machine learning methods including regression, classification trees, Naive Bayes, and random forests. The course will cover the complete process of building prediction functions including data collection, feature creation, algorithms, and evaluation.

COURSE 9

Developing Data Products

4.6 2,236 ratings

Show Less

A data product is the production output from a statistical analysis. Data products automate complex analysis tasks or use technology to expand the utility of a data informed model, algorithm or inference. This course covers the basics of creating data products using Shiny, R packages, and interactive graphics. The course will focus on the statistical fundamentals of creating a data product that can be used to tell a story about data to a mass audience.

Instructors

COURSE 10

Data Science Capstone

4.5 1,202 ratings

PhD

Associate Professor, Biostatistics

The project class will allow students to create a usable/public data product that can be used to show your skills to potential employers. Projects will be drawn from real-world problems and will be conducted with industry, government, and academic partners.

Bloomberg School of Public Health

1,487,856 Learners
32 Courses

Roger D. Peng, PhD

Associate Professor, Biostatistics

Bloomberg School of Public Health

1,454,514 Learners
37 Courses

Brian Caffo, PhD

Professor, Biostatistics

Bloomberg School of Public Health

1,478,258 Learners
30 Courses

Offered by

Johns Hopkins University

The mission of The Johns Hopkins University is to educate its students and cultivate their capacity for life-long learning, to foster independent and original research, and to bring the benefits of discovery to the world.

INDUSTRY PARTNERS

Start Today with a 7-Day Free Trial

Financial aid available

- Shareable Specialization and Course Certificates
- Self-Paced Learning Option
- Course Videos & Readings
- Practice Quizzes
- Graded Assignments with Peer Feedback
- Graded Quizzes with Feedback
- Graded Programming Assignments

Enroll for Free

Starts Aug 22

Shareable on LinkedIn

You can share your Course Certificates in the Certifications section of your LinkedIn profile, on printed resumes, CVs, or other documents.

Frequently Asked Questions

- What is the refund policy?
- Can I just enroll in a single course?
- Is financial aid available?
- Can I take the course for free?
- Is this course really 100% online? Do I need to attend any classes in person?
- How long does it take to complete the Specialization?
- How often is each course in the Specialization offered?
- What background knowledge is necessary?
- Do I need to take the courses in a specific order?
- Will I earn university credit for completing the Specialization?
- What will I be able to do upon completing the Specialization?
- Can I sign up for the course without paying or applying for financial aid?

More questions? Visit the **Learner Help Center**.

About
What We Offer
Leadership
Careers
Catalog
Coursera Plus
Professional Certificates
MasterTrack® Certificates
Degrees
For Enterprise
For Government
For Campus
Become a Partner
Coronavirus Response

Learners
Partners
Developers
Beta Testers
Translators
Blog
Tech Blog
Teaching Center

More

Press
Investors
Terms
Privacy
Help
Accessibility
Contact
Articles
Directory
Affiliates
Modern Slavery Statement

Learn Anywhere