

Data Engineer vs Data Scientist

Which one should you choose?

May 17, 2023



Hello!

I'm Amy, Head of Analytics & Data Science Curriculum at DataCamp



Amy Peterson

Head of Analytics & Data Science Curriculum

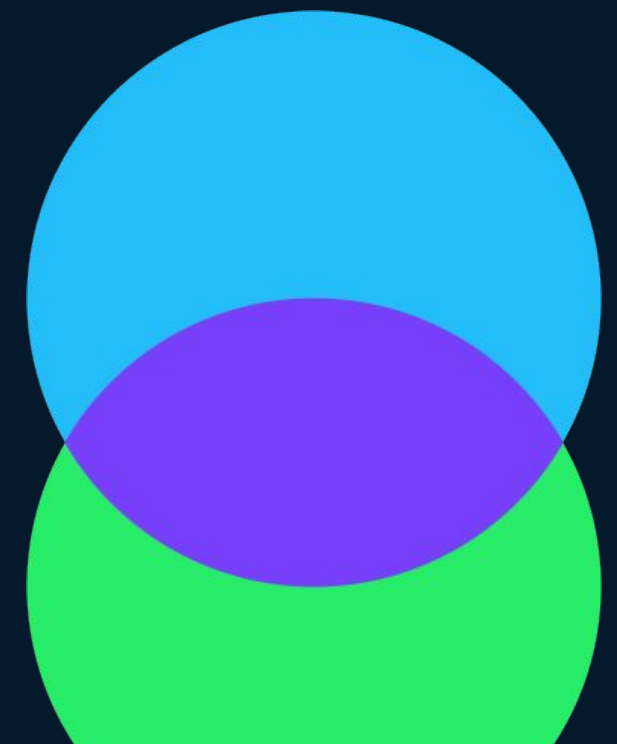
 [amypeterson6](#)



Vicky Kennedy

VP Certification

 [vickykennedy](#)



Agenda



- 1 Understanding the Roles
- 2 Understanding the Careers
- 3 Learning the Skills
- 4 Achieving Certification
- 5 Q&A





1

Understanding the Roles





“Data scientists utilize their analytical, statistical, and programming skills to collect, analyze, and interpret large data sets.”

[Glassdoor](#)



Data Scientists

Competencies and skills



Data Scientists

Technical skills

1 DATA MANAGEMENT

- *Importing, joining, and aggregating data*
- *Data cleaning tasks*
- *Assess data quality and perform validation tasks*

2 EXPLORATORY ANALYSIS

- *Use statistics and visualizations to analyze and identify trends in data sets:*
 - *Calculating metrics*
 - *Creating data visualizations*

3 STATISTICAL EXPERIMENTATION

- *Describe statistical concepts of hypothesis testing*
- *Apply sampling methods to data*
- *Implement methods for performing statistical tests*

4 PROGRAMMING

- *Data Science tasks*
- *Write repeatable production quality code for analysis*
- *Demonstrate best practices in production code - version control, testing, and package development*

5 MODEL DEVELOPMENT

- *Feature engineering, model selection, evaluation*
- *Implement standard modeling approaches for supervised and unsupervised learning problems*
- *Use suitable methods to assess model performance*



Data Scientists

Soft skills

1 BUSINESS ACUMEN

- *Make recommendations for analytic approaches based on business goal*
 - *Provide recommendations for future action*
- *Judge performance of analytic results against business criteria*

2 DATA COMMUNICATION

- *Data storytelling, data visualization, presentation, and writing skills*
- *Deliver a verbal presentation addressing business goals, outcomes, and recommendations*
- *Provide a written explanation of findings and reasoning for selected approaches*





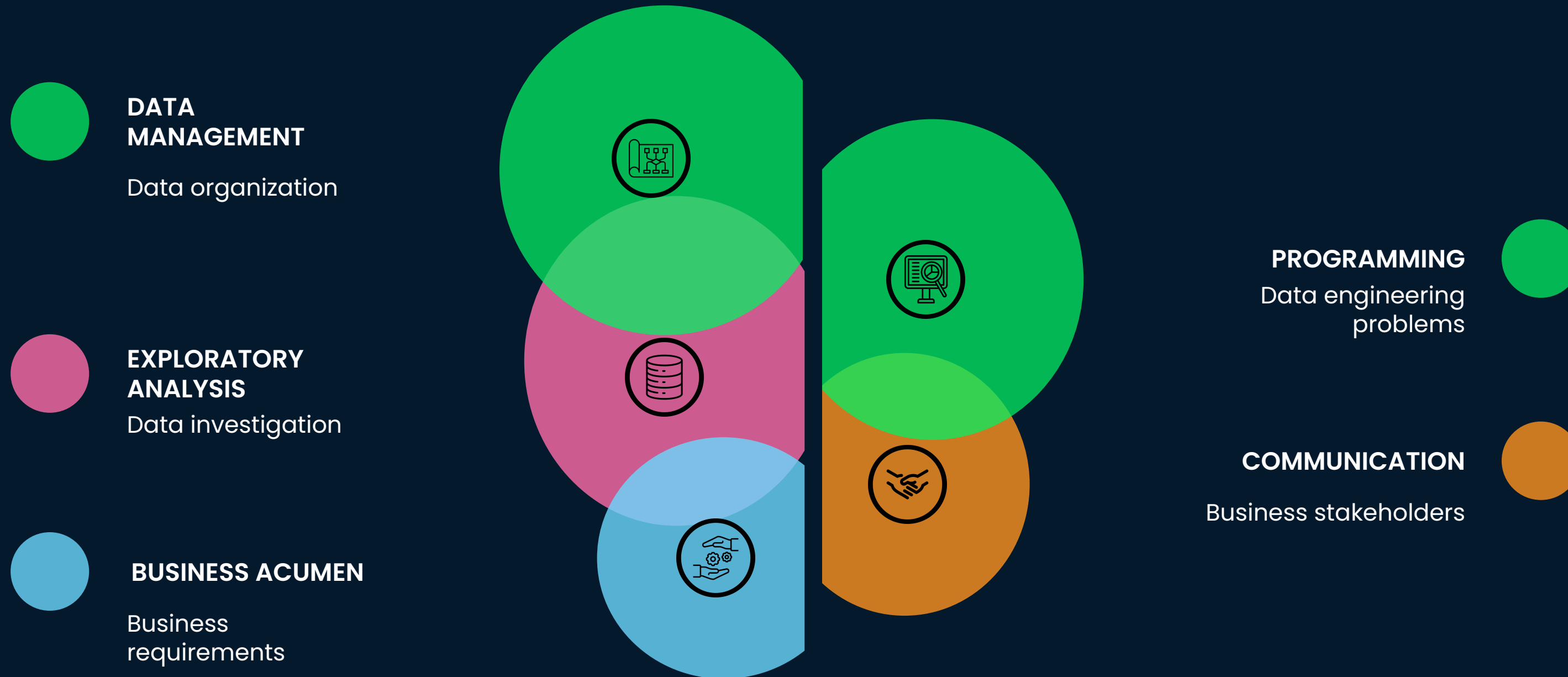
“Data engineers are responsible for laying the foundations for the acquisition, storage, transformation, and management of data in an organization.”

[\[DataCamp\] How to Become a Data Engineer in 2023](#)



Data Engineers

Competencies and skills



Data Engineers

Technical skills

1 DATA MANAGEMENT

- *Importing, joining, and aggregating data*
- *Data cleaning tasks*
- *Assess data quality and perform validation tasks*
- *Interpret a database schema and explain database design concepts*
- *Explain concepts in data storage and architecture*

2 EXPLORATORY ANALYSIS

- *Use data visualization tools to demonstrate characteristics of data and represent relationships between features*

3 PROGRAMMING

- *Data Engineer tasks*
- *Write repeatable production quality code for analysis*
- *Demonstrate best practices in production code – version control, testing, and package development*
- *Demonstrate software engineering principles, including OOP, to write modular code*
- *Demonstrate software development and deployment best practices and workflows*



Data Engineers

Soft skills

1 BUSINESS ACUMEN

- *Make recommendations for data pipeline design based on business requirements and problem statements*
- *Make recommendations for data approaches based on business goal*

2 COMMUNICATION

- *Collaboration with technical and non-technical stakeholders*
- *Deliver a verbal presentation of findings*
- *Provide a written explanation of findings*





2

Understanding the Careers



Data Science Workflow



Data Collection
& Storage



Data
Preparation



Exploration &
Visualization



Experimentation &
Prediction

Data Science Workflow



Data Collection
& Storage

Data
Preparation

Exploration &
Visualization

Experimentation &
Prediction

Data
Engineer

Data
Scientist

Project comparison

Data Engineer

Clean and prepare data for analysis

Example tasks:

- Data ingestion
- Data storage
- Build data pipelines
- Clean raw data using SQL
- Conduct exploratory analysis

Data Scientist

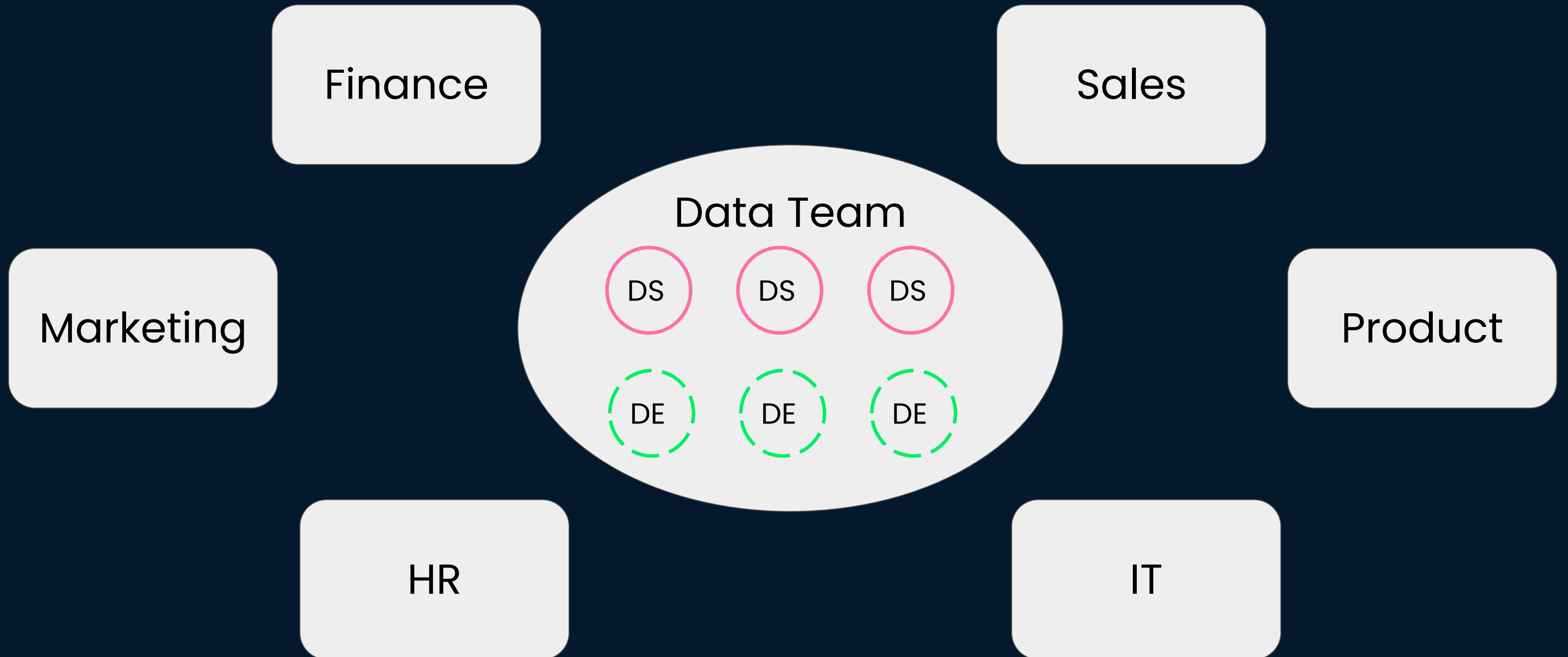
Conduct analysis to answer business problems and develop models to enable future decisions

Example tasks:

- Predict sales
- Recognize patterns
- Identify fraud
- Conduct exploratory analysis
- Predict consumer behavior



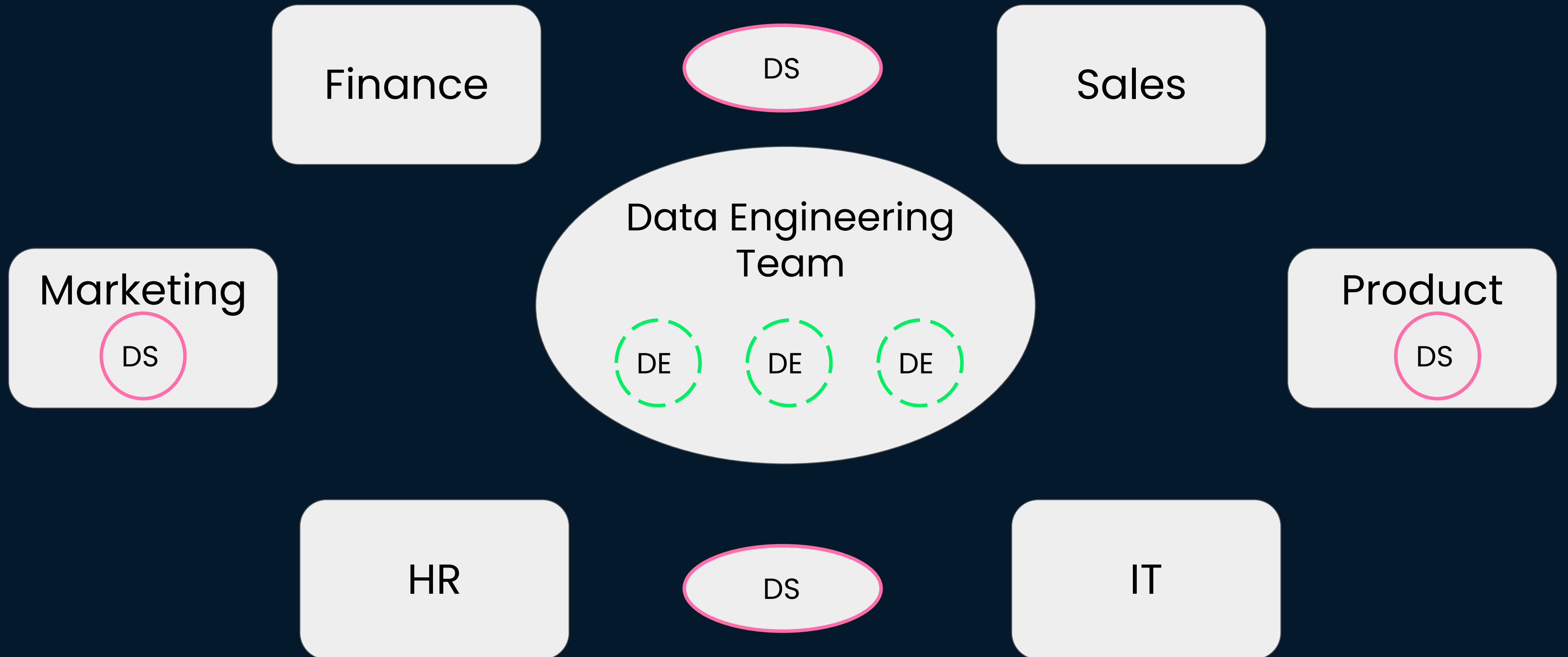
Data Team Structure: Centralized



[Medium] How should our company structure our data team?



Data Team Structure: Embedded



[Medium] How should our company structure our data team?



Choosing a career

Day to day
role

Technical skills

Projects and
tasks

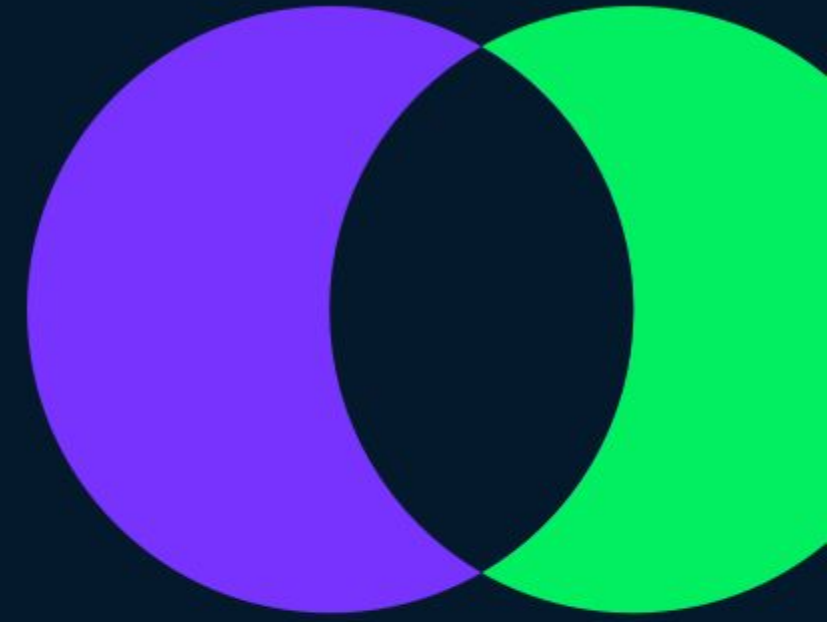
Competencies





3

Learning the Skills



Data Scientist Content

Python

Data Scientist with Python



23 Courses



90 hours

Data Scientist Professional with Python



31 Courses



116 hours

R

Data Scientist with R



22 Courses



88 hours

Data Scientist Professional with R



28 Courses



107 hours

CERTIFICATION AVAILABLE

Data Scientist with Python **ENROLLED**

Learn Python for data science and gain the career-building skills you need to succeed as a data scientist, from data manipulation to machine learning!

In this track, you'll learn how this versatile language allows you to import, clean, manipulate, and visualize data—all integral skills for any aspiring data professional or researcher.

...

[Read More](#)

Python 90 hours 23 Courses 6 Projects 3 Skill Assessments



Data Engineer Content

Data Engineer



17 Courses



57 hours

Content coming soon!

- [Introduction to Snowflake](#)
- [Introduction to dbt](#)
- Introduction to Databricks*
- Introduction to Testing in Python
- Introduction to the Modern Data Architecture
- Introduction to Data Pipelines
- Data Engineering projects

* Conceptual

CAREER TRACK

Data Engineer

In this track, you'll discover everything you need to know to become a data engineer by learning Python, SQL, and Git from scratch. You'll discover how to interact with relational databases to query, input, and modify data and get hands-on experience in importing and cleaning data in Python, optimizing your code for efficiency, and writing tests to validate your code.

Throughout this track, you'll learn some of the essential data engineering tools,...

[Read More](#)

[Switch Track](#)





4

Achieving Certification



DataCamp Certification is an **official recognition** that a person has achieved the required skill level as a data professional.

Why DataCamp Certification?



Tool / Product Agnostic

We objectively certify based on industry competencies, not product knowledge.



Cost effective

Certification access is included in Learn subscriptions.



Personal Grading

Real-world application via case study submissions are personally graded by humans



Trusted & Credible

Offered by the leader in data science education



Value for individuals

- Validate the knowledge and skills they've learned outside of academic settings
- Prove they can apply knowledge to business problems
- Build confidence by knowing they're job ready
- Stand out to employers



Certification Comparison

Data Engineer

Associate (SQL + Python)

Exam DE101 & DE102 + DEA501 Practical

Professional

Roadmap 2024

Domains Assessed

Data Management

Exploratory Analysis

Programming for Data Engineering

Data Scientist

Associate (SQL + R or Python)

Exams DS101 & DS102 + DSA501 Practical

Professional (SQL + R or Python)

Exams DS101 & DS201 + DSP501 Practical

Domains Assessed

Data Management

Exploratory Analysis

Statistical Experimentation

Programming for Data Science

Model Development

Data Communication

Business Acumen (Pro only)



Certification Process

- Registration
 - Begins 30 day countdown
- 1 or 2 Timed Exams (45–60 questions each)
 - Must be passed within 2 attempts
- Practical Exam
 - Must be passed within 2 attempts
 - Hand graded by a human
 - See [example](#) presentation
 - [Practical Hub](#)
- If failed, locked out of that certification for 14 days

The screenshot displays the certification process interface. At the top, a dark banner reads "Become a Certified Data Scientist Associate" with a subtext "To get certified we'll be testing the skills required to succeed as a Data Scientist Associate." A countdown timer shows "29d : 23h : 59m : 53s" and "REMAIN TO COMPLETE CERTIFICATION". Below this, three exam options are listed:

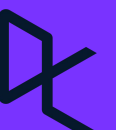
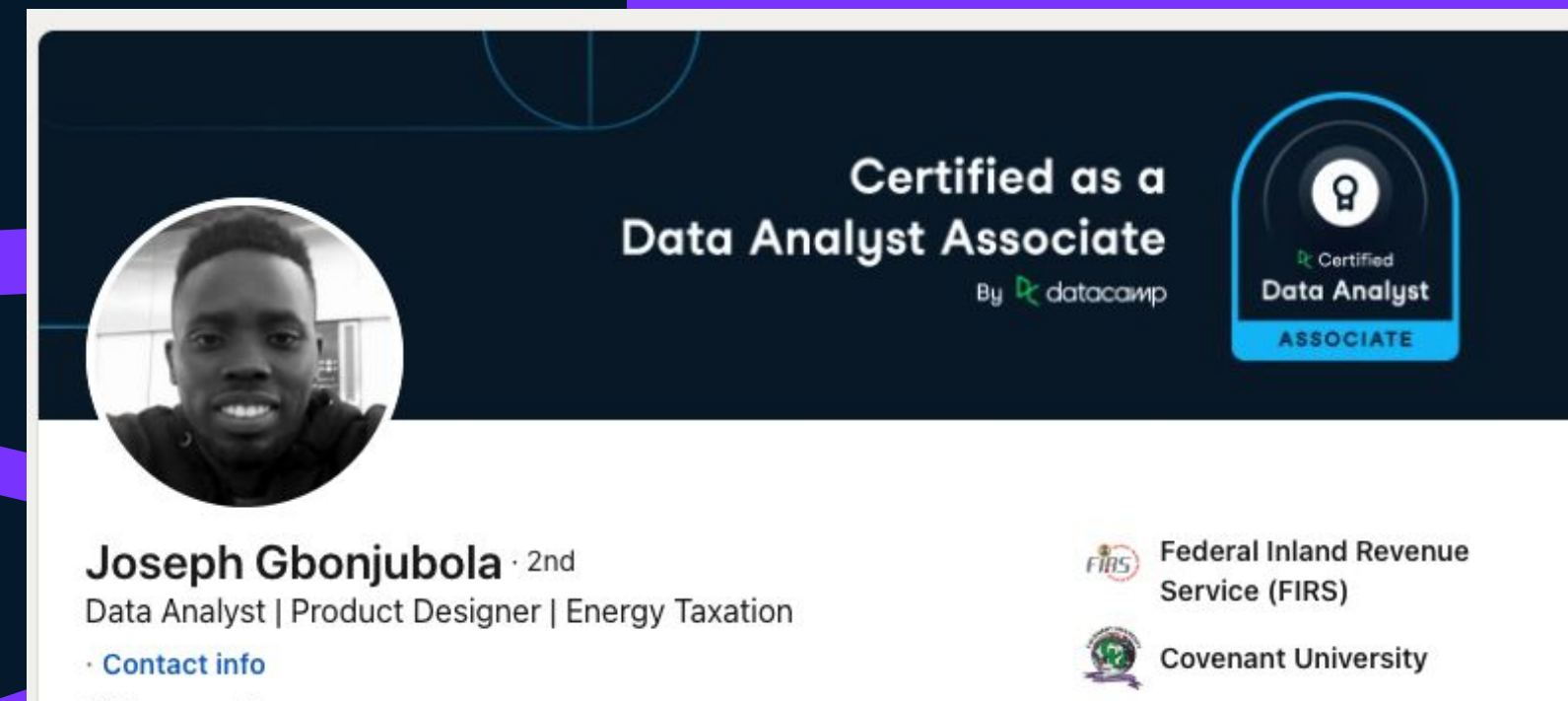
- Timed Exam DS101** (1 HOUR): This exam assesses your knowledge and abilities in statistical experimentation and exploratory analysis using R or Python. It takes approximately 60 minutes to complete and must be completed in one session. A green "Start the Exam" button is visible.
- Timed Exam DS102** (LOCKED): This exam assesses your skills in data management, modeling, and programming for data science using R or Python.
- Practical Exam DS501P** (LOCKED): Analyze a dataset in either R or Python and submit a summary report.

At the bottom right, there is a link to "Withdraw from this certification".



Upon Certification

- Candidates receive verified certificate, badges, and social media banners
- Access to DC Certified Community



BECOME A CERTIFIED DATA SCIENTIST 🔑

Prepare for Certification

DataCamp Certification tests the skills employers are looking for so you can prove you're up to the task. **Prepare for certification in two ways:**

OPTION 1

Complete this career track

- Best for beginners
- Offers a comprehensive curriculum to build all the skills needed by today's data scientists
- Test-drive your skills as you go with interactive exercise and hands-on practice

[Return to Track](#)

OPTION 2

Build a personalized learning plan (Beta)

- Best for those with some prior knowledge. Assess your current skill level across all tested domains and allow us to build a plan just for you
- Achieve your goal faster by focusing only where you need to go

[Enroll in Plan](#)

What you'll get once certified:



Options to prepare

- Option 1: Complete career track
- Option 2: (SQL, Python only) Complete personalized learning plan

And / Or

- Leverage our “readiness quiz”
- Or self-study using the study guide downloads

Not sure you're ready? Take the quiz.

This 3 minute quiz will determine your readiness for certification with a set of rapid questions.

[Take the Quiz →](#)

Data Analyst

[Download Study Guide](#) ↓

Many certified individuals completed courses and tracks on DataCamp to prepare for the certification. We recommend that you complete [Data Analyst with SQL](#) career track for the associate and professional certification and either the [Data Analyst with R](#) or [Data Analyst with Python](#) career tracks for the professional certification.

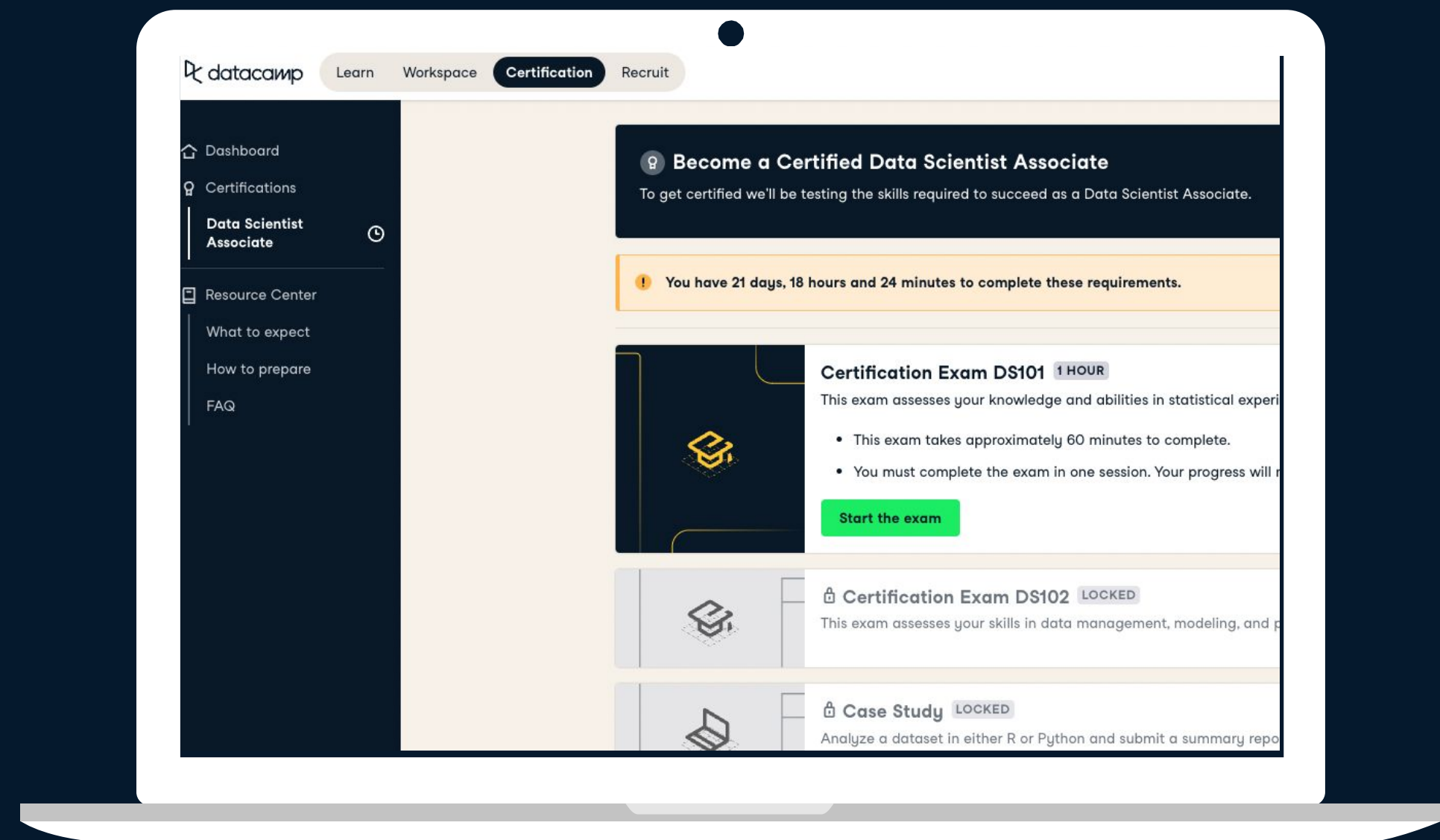
Skill Assessments

If you aren't sure whether your skills are at the level required for certification, you might want to take the following [assessments](#) and see if your score is at or above the associate or professional level.

SKILL ASSESSMENT	ASSOCIATE	PROFESSIONAL
Data Management in SQL (PostgreSQL)	134	134
Data Analysis in SQL (PostgreSQL)	115	115
Importing & Cleaning Data in R/Python	n/a	111
Data Manipulation in R/Python	n/a	111
Statistical Fundamentals in R/Python	n/a	125

Demo time

Let's take a quick tour
of the experience.





5

Q&A





Thank you