REPORT

State of Mathematical Optimization in Data Science 2022

GUROBI OPTIMIZATION

Table of Contents

Overview

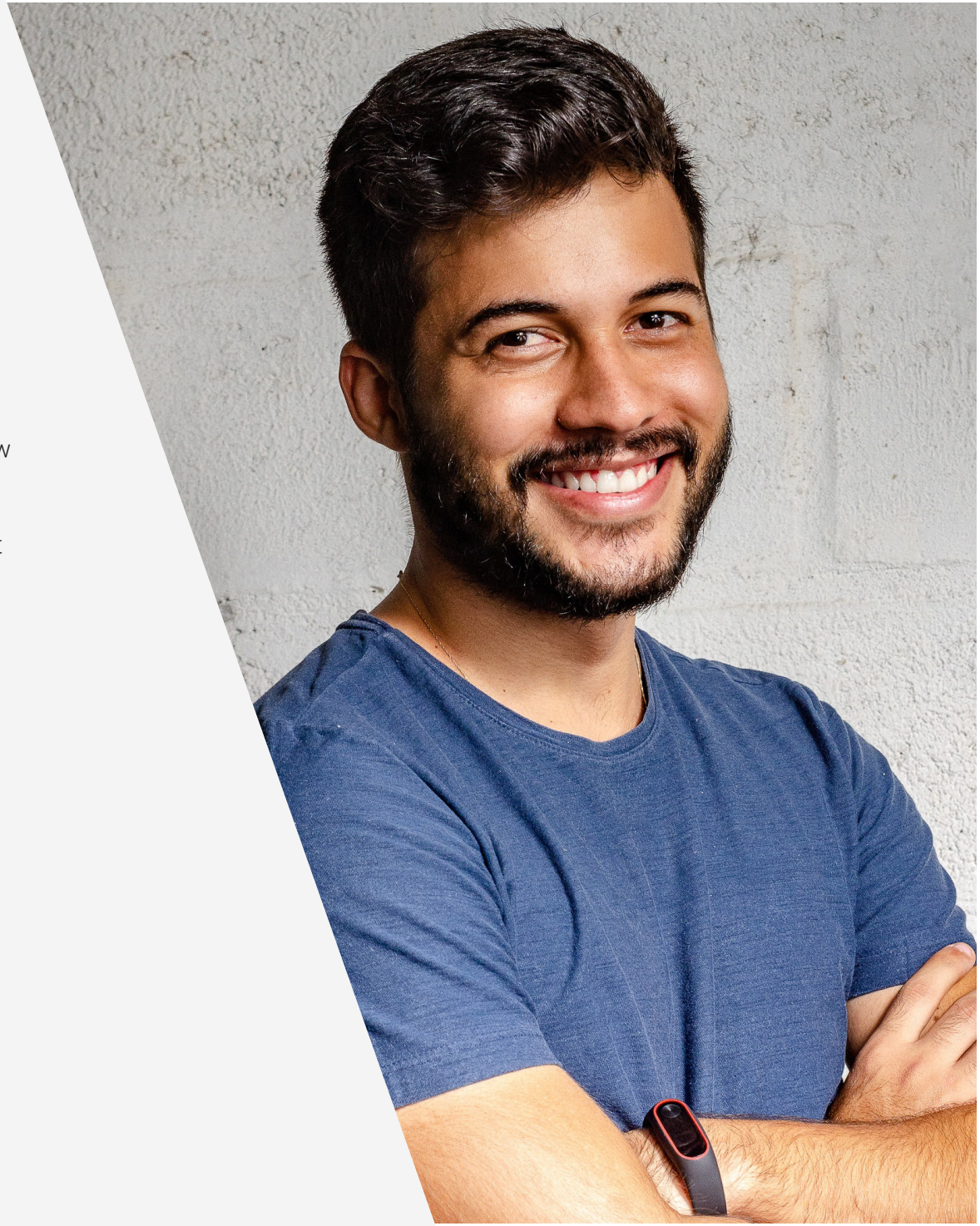
To explore how data scientists are using mathematical optimization today, Gurobi Optimization surveyed 369 Data Science Central subscribers in June 2021. This survey is the second in a series, following Gurobi Optimization's inaugural data science survey in March 2021, which surveyed 722 KD Nuggets subscribers.

Methodology

This report focuses on respondents who are current data practitioners, independent consultants working on data science projects, and data science/analytics leaders. The survey logic does not include those who identify as "academic" or "other."

Key Findings

- Most data scientists have heard of mathematical optimization, but less than half can accurately identify how it's applied.
- Of those who have used mathematical optimization, most have only experienced it through limited-capacity solvers and open-source tools.
- Most data scientists are self-taught professionals and indicate an interest in learning mathematical optimization as well.
- Data scientists see the value in using machine learning and mathematical optimization together.
- By adopting mathematical optimization skills, data scientists expect to be able to solve new types of problems, such as resource allocation, supply chain decision-making, and revenue-maximizing strategies.



Profile

EDUCATION

75% of respondents have advanced degrees, with 55% holding a master's degree and 19.4% holding a PhD.

EXPERIENCE

Most respondents are experienced practitioners, with 47% having worked in a quantitative/technical role for 11+ years. This was followed by a fairly even distribution of experience: 16.9% at 1-3 years, 16.9% at 4-6 years, and 19.4% at 7-10 years.

INDUSTRIES

The most prominent industries represented in the study include Technology (25.3%), Financial Services (15.6%), and Business Services (7.9%).

TEAM FUNCTION

The majority of respondents described their team function as Data Science (25.8%), followed by Analytics (15.4%), IT (10.7%), and Business Intelligence (10.7%).

COMPANY SIZE

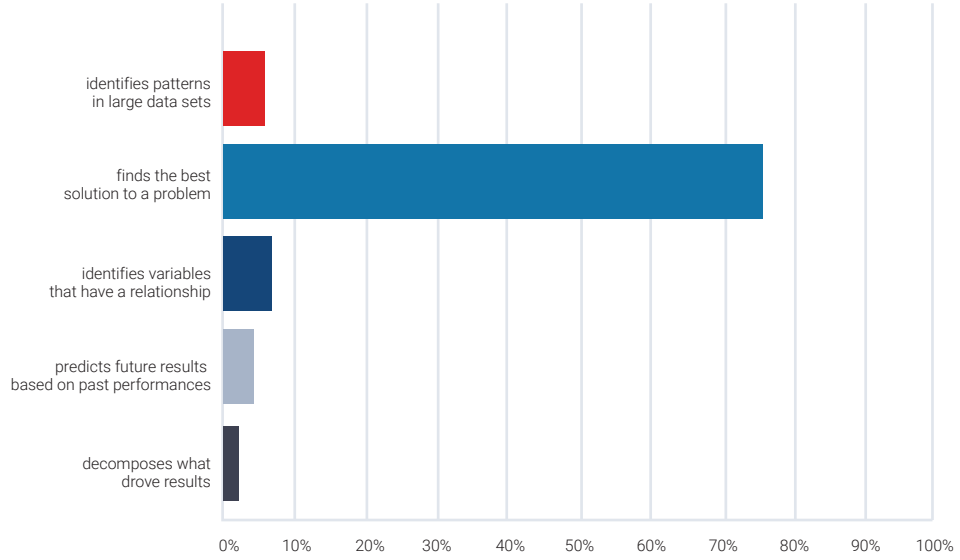
Most respondents work for a small company with less than $10M in revenue (33.9%) or an enterprise with $1B+ in revenue (25.3%).

TEAM MODELS

Most respondents (42%) indicated that their data science teams function as a hybrid model—where data science professionals are embedded with teams and stakeholders that report to a central function and leader.

For those who have either a centralized or decentralized team model (i.e., not a hybrid combination of the two), there was a fairly even split. 26.4% report having one or more centralized groups (i.e., centralized), and 27.6% report having data scientists, analytics, and engineers embedded with the business and operating as part of that organization (i.e., decentralized).

Please check the box that describes the power of mathematical optimization based on your understanding.



ANSWER CHOICES	RESPONSES	
identifies patterns in large data sets	6.74%	12
finds the best solution to a problem	75.84%	135
identifies variables that have a relationship	7.30%	13
predicts future results based on past performances	6.74%	6
decomposes what drove results	3.37%	6
Total		**178**

Mathematical optimization in data science

Awareness

60.5% of respondents report being aware of mathematical optimization, and 75% of this group can accurately identify the ideal uses for mathematical optimization. This reveals that less than half of respondents truly understand mathematical optimization.

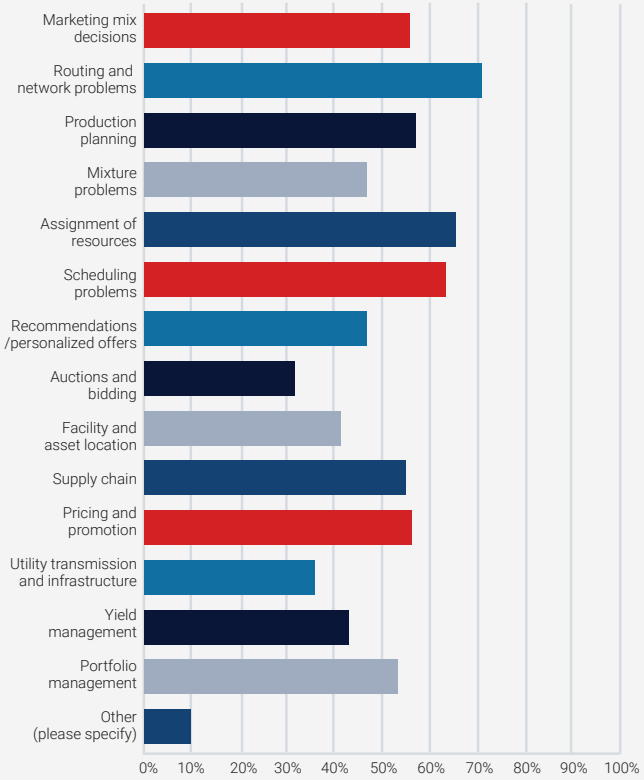
In this case, the only correct answer is “finds the best solution to a problem.” The other options are representative of machine learning, data analytics, and statistics.

This indicates that there’s an opportunity for data scientists to discover mathematical optimization and its unique capabilities.

Application

When presented with a list of mathematical optimization applications, respondents most often selected Routing and Network Design, Assignment of Resources, Scheduling Problems, Portfolio Management, Supply Chain, and Pricing—even though all of the answers are correct. This indicates that respondents do not understand the full range of problems mathematical optimization can solve.

Which of the following could you apply mathematical optimization to? (Mark all that apply)



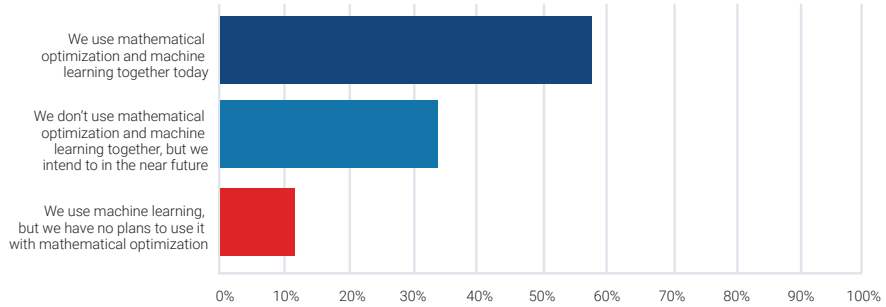
ANSWER CHOICES		RESPONSES	
Marketing mix decisions		55.81%	96
Routing and network problems		70.93%	122
Production planning		56.98%	98
Mixture problems		47.09%	81
Assignments of resources		65.70%	113
Scheduling problems		63.37%	109
Recommendations/personalized offers/next best alternative		47.09%	81
Auctions and bidding		31.98%	55
Facility and asset location		41.28%	71
Supply chain		55.23%	95
Pricing and promotion		56.40%	97
Utility transmission and infrastructure		35.47%	61
Yield management		43.02%	74
Portfolio management		53.49%	92
Other (please specify)	Responses	9.88%	17
Total Respondents: 172			

Combining machine learning and mathematical optimization

Of those who report being aware of mathematical optimization, 57.0% currently use mathematical optimization and machine leaning together, and an additional 32.3% plan to in the near future. Nearly all respondents (94%) agreed that mathematical optimization and machine learning can be used in a complementary manner.

These findings indicate that respondents recognize that machine learning and mathematical optimization are complementary technologies.

Please indicate how you believe mathematical optimization and machine learning relate for your company's operations



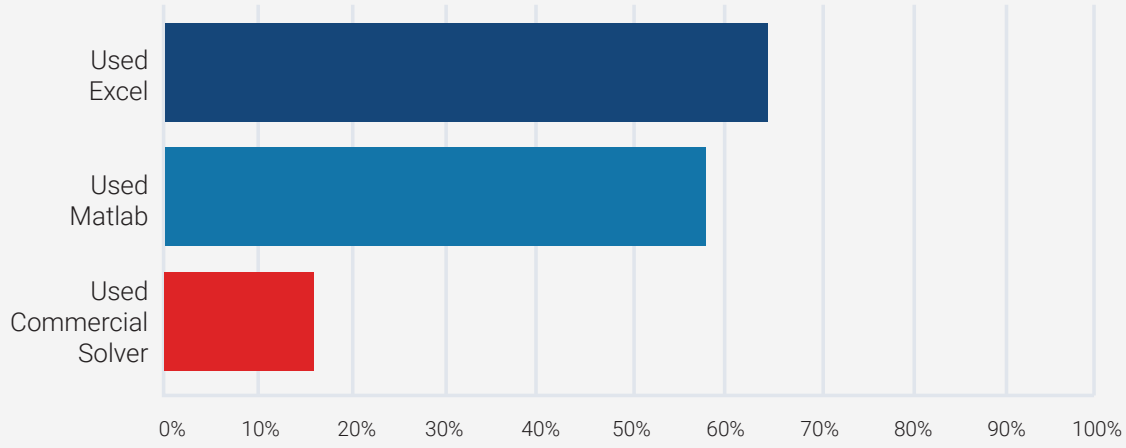
ANSWER CHOICES	RESPONSES	
We use mathematical optimization and machine learning together today.	57.14%	76
We don't use mathematical optimization and machine learning together today, but we intend to in the near future.	32.33%	43
We use machine learning, but we have no plans to use it with mathematical optimization.	10.53%	14
Total		**133**

Experience with solvers

A majority of respondents (63.9%) have used Excel to conduct mathematical optimization calculations, followed by 58.5% who have used MATLAB. Fewer than 20% reported using a commercial solver.

This low percentage could be a result of data scientists having a poor experience with open-source solvers and being reluctant to try a commercial solver.

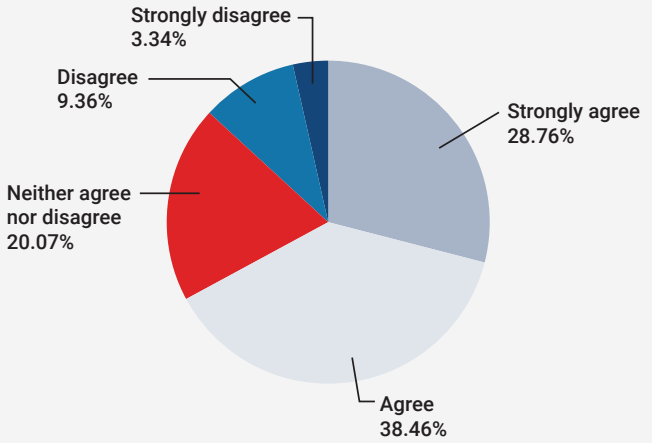
Regardless, it indicates a significant opportunity for data scientists to discover the power that a commercial solver can provide—especially for projects that involve 200+ decision variables, what-if analyses, and real-time decision-making.



Open source vs. commercial

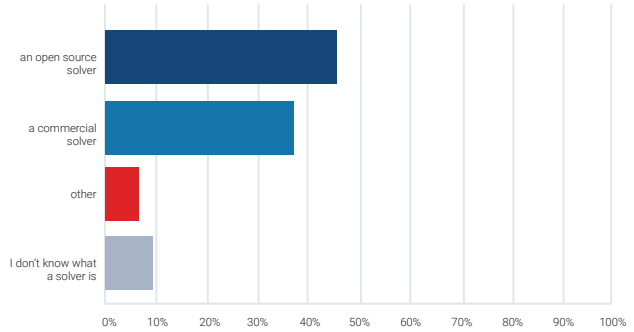
Most respondents agreed (38.5%) or strongly agreed (28.8%) that their company embraces open source and uses it widely. So it's no surprise that 45.8% of those who use mathematical optimization have used an open-source solution, and 37.4% have used a commercial solver.

This is a shift, however, from the 2020 survey, where nearly 60% had used open-source and 25% had used commercial. This could indicate a growing interest in commercial solvers.



67% agree that their company embraces open source and uses it widely

Have you used:



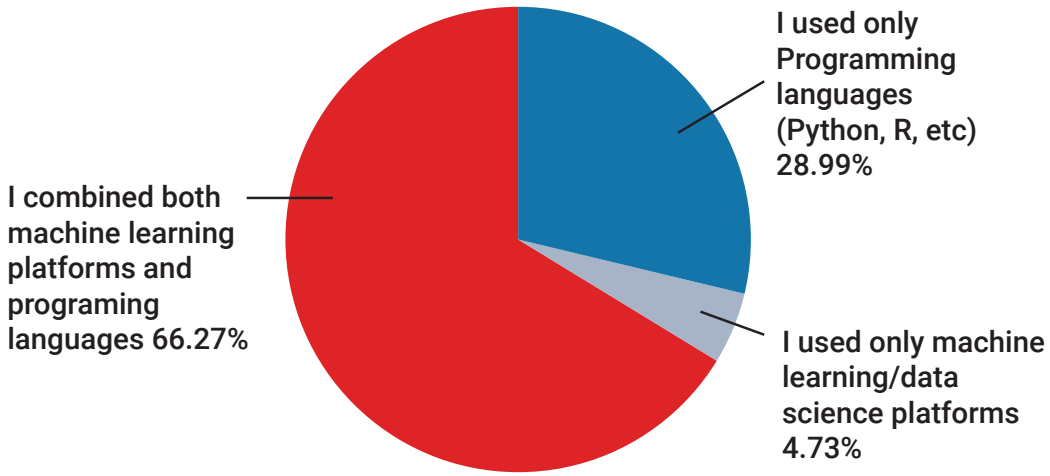
ANSWER CHOICES	RESPONSES	
an open source solver	45.80%	60
a commercial solver	37.40%	49
other	6.87%	9
I don't know what a solver is	9.92%	13
Total		**131**

Programming vs. platforms

Respondents indicated that programming languages are still essential to their jobs, with under 5% using a machine learning or data science platform alone. The majority use either a combination of machine learning platforms and programming languages (66.3%) or programming languages alone (28.9%).

This indicates that data scientists—especially those who are mathematically minded—are well-positioned to learn mathematical optimization, since it requires programming skills (Python, R, etc.).

Please select the workflow/methods that best describe your data science work over the past 6 months:



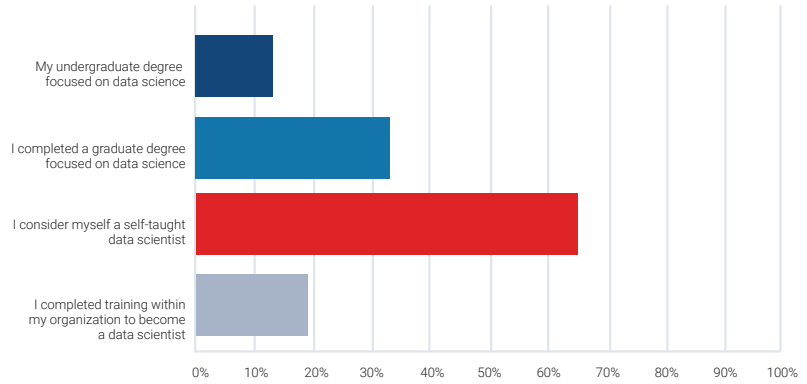
Opportunities for data scientists

A culture of learning

Most respondents (64%) consider themselves to be self-taught data scientists—having learned through massive open online courses (MOOCs), certificate programs, or self-guided education.

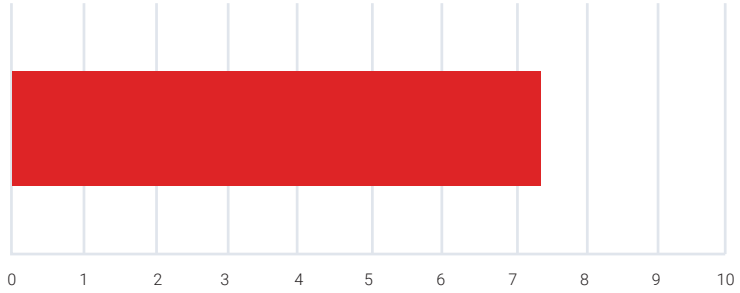
This indicates that data scientists, when given access to training, are fully capable of picking up mathematical optimization skills and applying them to new problem types.

What was your educational path that led to a career in data science? (Please select all that apply)



ANSWER CHOICES	RESPONSES	
My undergraduate degree focused on data science	12.58%	39
I completed a graduate degree focused on data science	31.94%	99
I consider myself a self-taught data scientist (MOOC (massive open online courses) /certificate programs/self-guided education	64.19%	199
I completed training within my organization to become a data scientist	19.68%	61
Total		**131**

If an industry player created a broad and varied study training agenda for math optimization (videos, hands-on practice, case studies, webinars, etc.) and proficiency required 30 hours of study, on a scale of 1-10 how likely would you be to complete the course? (1 = Very Unlikely; 10 = Very Likely)



A growing interest in learning optimization

77.2% of respondents reported they would be interested in learning more about mathematical optimization. And when asked how likely they would be to complete 30 hours of mathematical optimization training (on a scale of 1-10), the average response was 7 (likely).

Conclusion

Although most data scientists have heard of—or had some experience with—mathematical optimization, the majority have only scratched the surface. By limiting their experience to open-source solvers and limited-capacity solvers, they are missing out on the power and speed that a commercial solver can deliver.

Data scientists who want to add mathematical optimization to their analytics toolbox can start here. Once they've picked up some basic skills, they can start applying those skills to their organization. A quick-win could be to tackle a resource allocation challenge—such as allocating budgets, equipment, supplies, people, etc.

The data scientist who learns mathematical optimization can then move on toward even-more-valuable initiatives, such as:

- Improving supply chains and networks
- Making long-term, capital-intensive decisions
- Improving revenue growth and profitability

To take the next step toward mathematical optimization, watch the recent webinar, "Adding Optimization to Your Data Science Analytics Toolbox." You'll discover the why's and how's of mathematical optimization, as well as real-world examples of machine learning and optimization in action.

For more information

www.gurobi.com

GUROBI
OPTIMIZATION