2019 ANACONDA STATE OF DATA SCIENCE

In Spring 2019, Anaconda Inc. launched its second annual survey of the Anaconda community to understand what's on the minds, hearts, and screens of the open source data science world.

Because the Anaconda community is so broad, their feedback tells an interesting story about the ongoing evolution of data science:

1. **Multiple areas of business now recognize the data science imperative and have started laying the groundwork for systemic organizational transformation.**

2. **Data science literacy is growing across business disciplines and is becoming critical for nearly all enterprise job titles.**

3. **The demands of data science require the use of a wide variety of tools, and the open source ecosystem has delivered on this requirement more effectively than proprietary tools.**

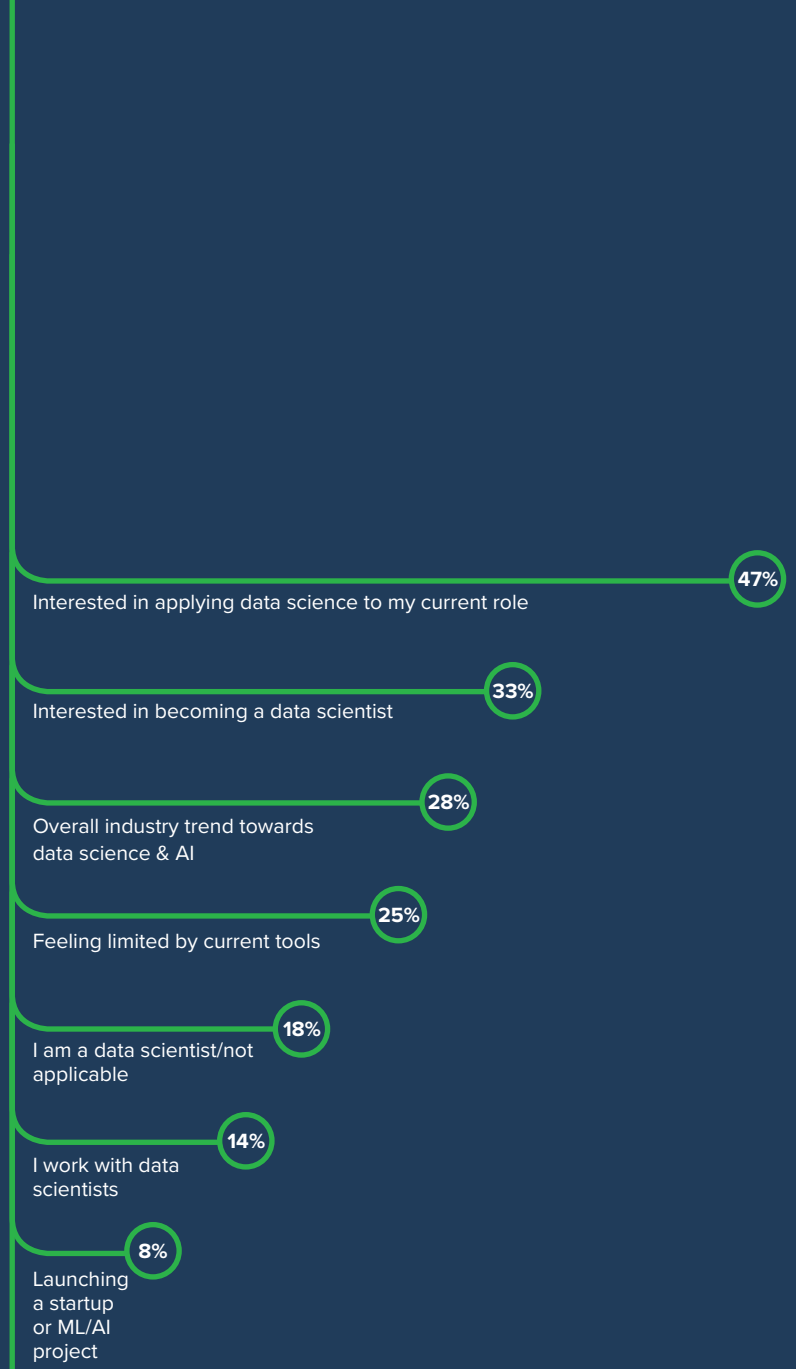
4. **The Anaconda community tells us that Python 3 adoption has passed the tipping point. All organizations should initiate a migration plan, if they haven't already.**

In the following pages, we'll take a closer look at the survey results and what they reveal about the ***state of data science.***

1. THE GROUNDWORK FOR SYSTEMIC ORGANIZATIONAL TRANSFORMATION IS ALREADY BEING LAID.

A business revolution is underway, with 47% of respondents learning data science to apply it in their current role. These professionals are learning data science in order to become what Anaconda CTO Peter Wang calls “data science literate” — they’re learning to think quantitatively about business challenges and developing fluency with the tools and concepts of predictive and prescriptive modeling.

Some have called these users “Citizen Data Scientists,” but we believe that terminology needlessly siloes people. It creates divisions between technical data scientists and business users, who need to progress together to create a truly data-driven organization. The truth is that anybody can become data science literate, regardless of their experience level or professional background. Python and other open source tools, which are accessible for all users, have democratized data science and put it within reach of a wide range of professionals.



WHY ARE YOU LEARNING ABOUT DATA SCIENCE?

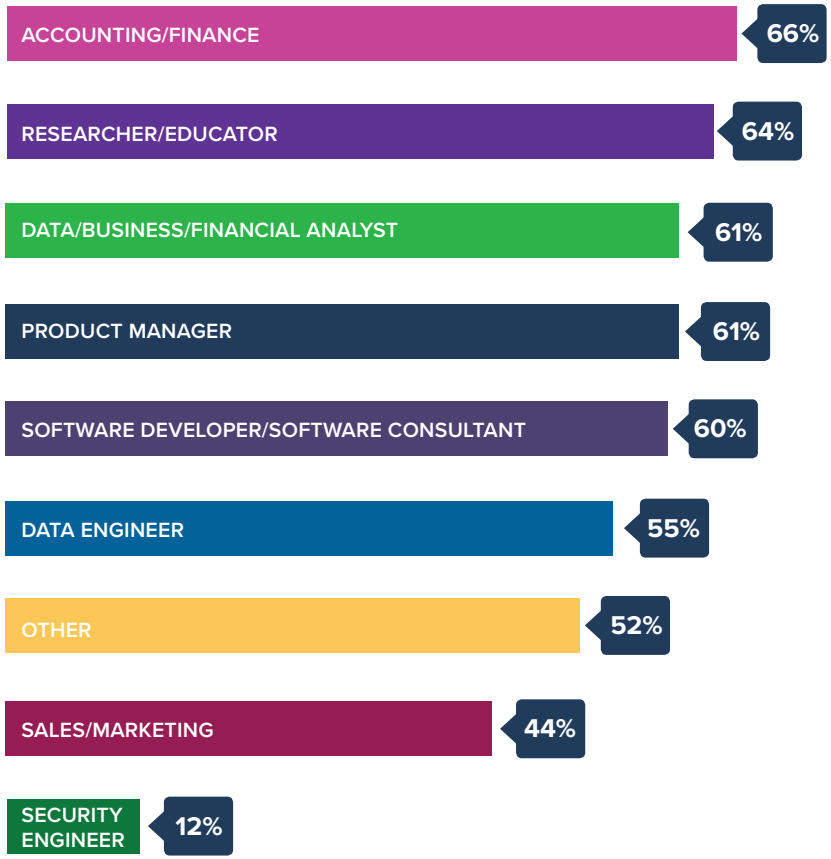
PERCENTAGE OF RESPONDENTS

A follow-up question investigated the types of jobs held by these data science learners. Once again, their responses pointed to a revolution. Interest in data science spans a very broad range of job functions, and, as these professionals attain literacy, they will accelerate the paradigm shift to truly data-driven businesses.

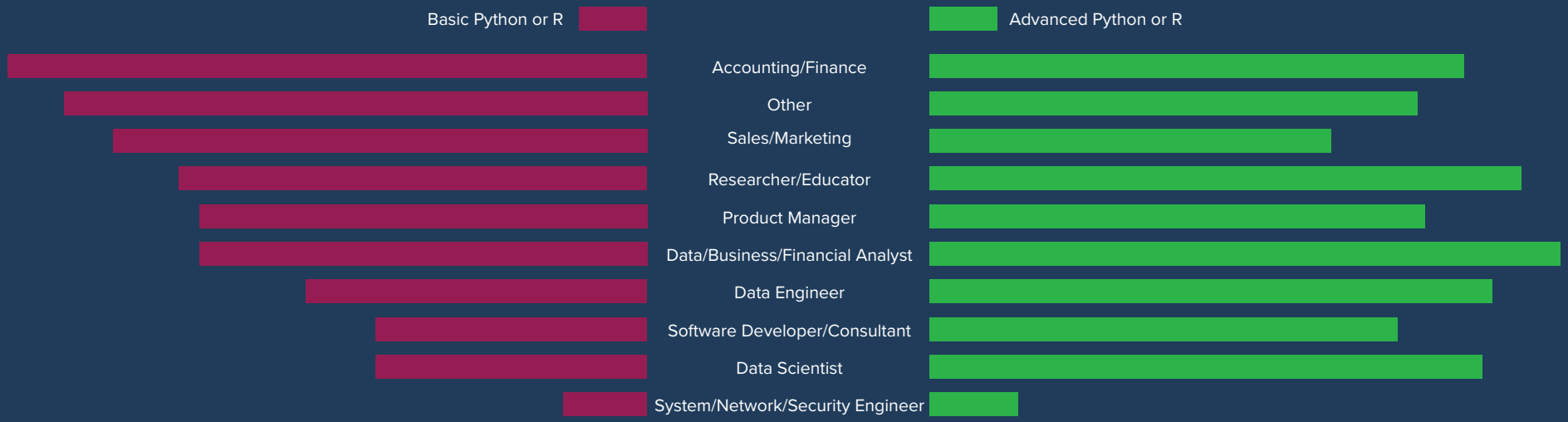
The more areas of the business that embrace data science, the more silos will fall, and the potential of data science will be realized more quickly. Everything's connected in a data-driven organization. We are entering a period of seismic change and things may get a little messy. The good news: those that tackle the challenges of a data-science-enabled transformation head-on will also realize the benefits fastest.

WHO IS LEARNING DATA SCIENCE TO APPLY IT TO THEIR CURRENT ROLE?

RESPONDENTS BY JOB FUNCTION



2. DATA SCIENCE LITERACY WILL BE CRITICAL FOR NEARLY ALL ENTERPRISE JOB TITLES.



WHAT SKILLS ARE YOU CURRENTLY STUDYING OR PLANNING TO LEARN IN THE NEXT 6-12 MONTHS?

Numerous enterprises are beginning to differentiate their products, services, and approaches to the market by leveraging data science. This means that the ability to think quantitatively, understand advanced statistical approaches, and use computer code to express ideas is becoming important for a broader range of roles in business than ever before. Accordingly, we see that professionals across a wide variety of disciplines are upgrading their skills.

Anaconda asked survey respondents “What skills are you currently studying or planning to learn over the next six to 12 months?” Python/R was the clear leader. We drilled down further to find out the skill levels of different job functions. It makes sense that most respondents in accounting, finance, sales, and marketing are still learning basic Python or R — these disciplines are adding new skills to complement tools like Microsoft Excel and Tableau. In other disciplines, we see that more respondents have already invested the time to learn basic Python/R and are largely moving on to more advanced techniques. As their knowledge and skills grow in the next six months, data science literacy will expand across multiple job functions.

As data science transforms the enterprise to become truly data-driven, professionals of all stripes must be able to keep up. The ability to leverage data science will no longer be a niche skillset held by a small group of specialists. Data science literacy will be required to do great work in any aspect of the business — whether it’s accounting, product design, business analysis, research, software development, or data science. The revolution is coming, and businesses must be ready.

3. THE OPEN SOURCE ECOSYSTEM DELIVERS THE FREEDOM AND INNOVATION DATA SCIENCE NEEDS.

When we asked survey participants about their tool usage, we expected to see strong results for Jupyter Notebooks, JupyterLab, and RStudio — they're all designed specifically for interactive computing. That said, among respondents who frequently use at least one of those options, ***85% indicated that they use more than one tool.***

THE PERCENTAGE OF PEOPLE USING THE FOLLOWING NOTEBOOKS AND IDES IN THEIR DEVELOPMENT ENVIRONMENT

79% Jupyter Notebooks

54% JupyterLab

35% RStudio

30% PyCharm

25% VS Code

20% VIM/Emacs

16% Colab

8% Zeppelin

One of our optional survey questions sought to gauge affinity for various data science tools and technologies. In a survey of the Anaconda community, it's unsurprising that we'd see a lot of positive feedback for the platform. What's more interesting is that RStudio and Microsoft Excel also had strong positive responses among the 2109 participants that answered this question.

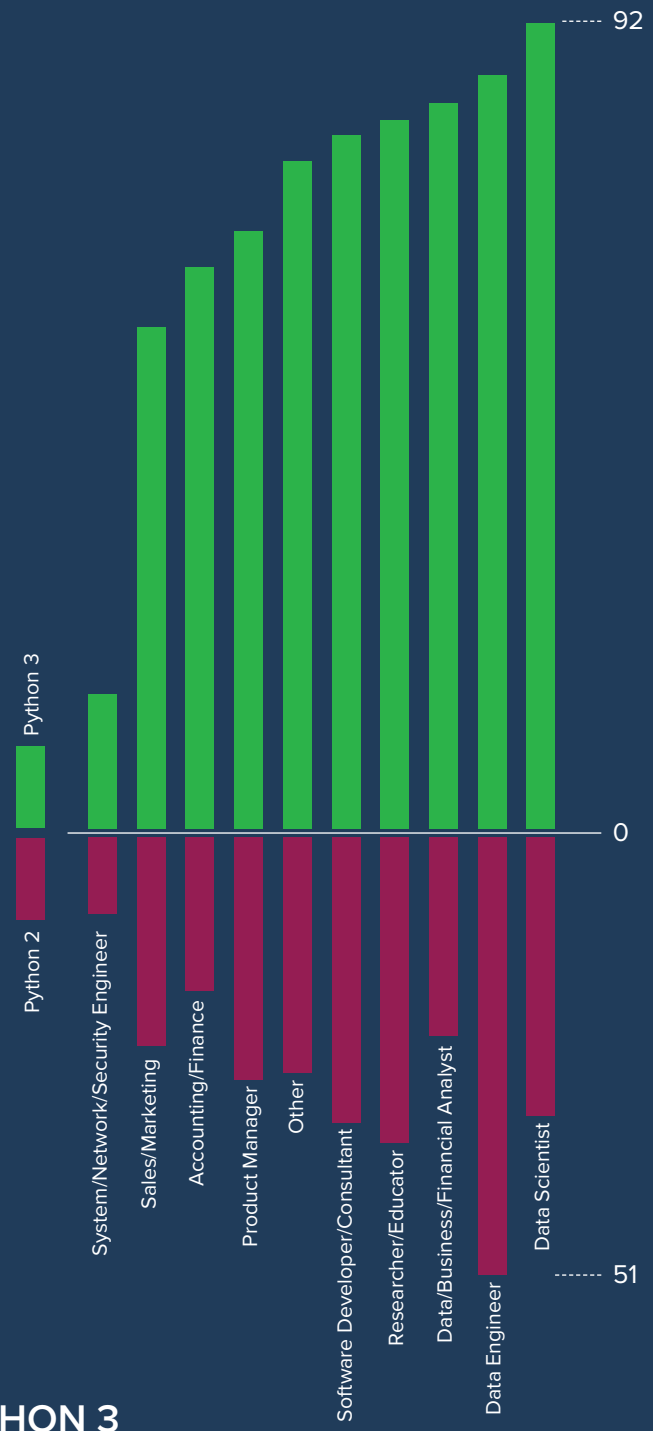
Because data science is dynamic and exploratory at its core, data scientists need freedom to use the right tool(s) for the job in different circumstances, and to experiment with new packages and languages as desired. At the same time, data science demands innovative tools to address newly discovered challenges and opportunities. Only the open source ecosystem delivers that freedom and innovation at scale.

PERCENTAGE OF PEOPLE WHO ANSWERED THAT THEY "LIKE" AND "STRONGLY LIKE" THE FOLLOWING TOOLS/TECH.

- 92% Anaconda
- 67% RStudio
- 67% Excel
- 52% MATLAB
- 47% Tableau
- 31% Mathematica
- 28% Hadoop

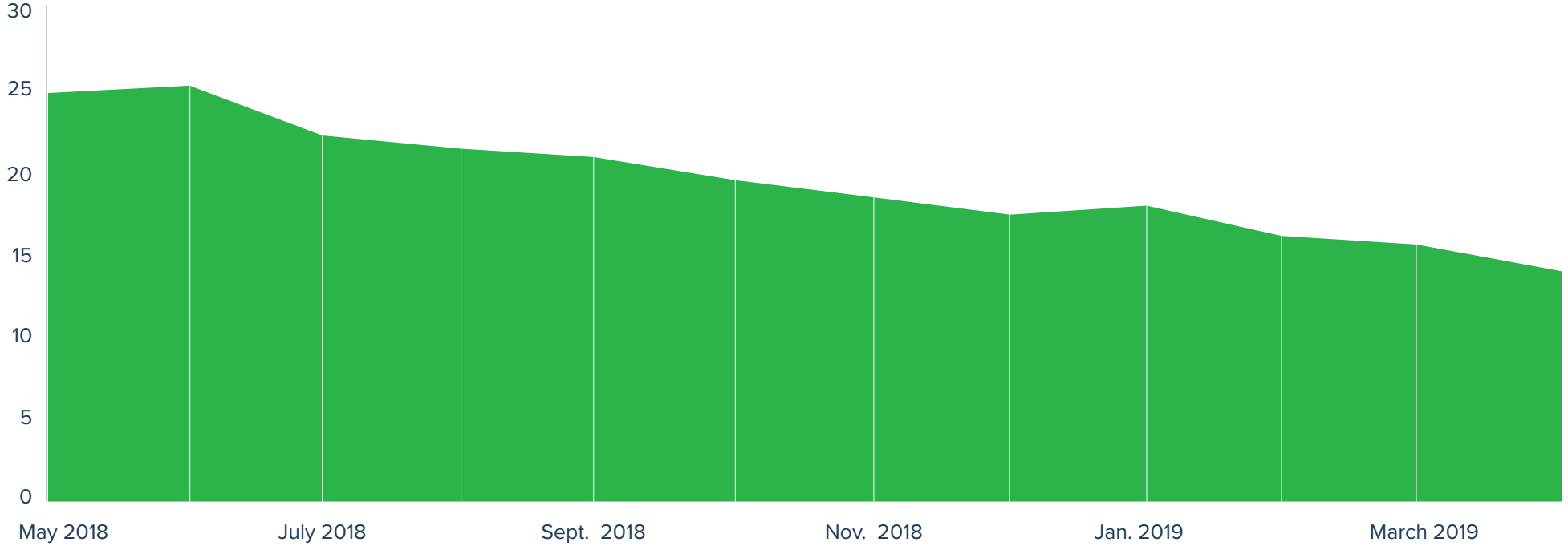
4. THE TIME TO MIGRATE TO PYTHON 3 IS NOW.

This is admittedly a more specific finding than our previous, but with end of life for Python 2 fast approaching, the migration question is on everyone's minds. Our survey showed Python 3 adoption percentages growing across multiple fields, alongside a steady decrease in the percentage of Anaconda users who are downloading Python 2 packages.



PYTHON 2 VS. PYTHON 3 USAGE BROKEN DOWN BY JOB FUNCTION.

PERCENTAGE OF USERS WHO ARE DOWNLOADING PYTHON 2 PACKAGES



Based on this data, we believe ***now is a critical time for organizations to initiate and then accelerate their migration strategies***. Those who don't risk getting left behind.

Note: Our blog End of Life (EOL) for Python 2.7 is coming. Are you ready? provides a good starting point for Python 3 considerations.

OUR CLOSING THOUGHTS

The 2019 State of Data Science report reflects the rapid evolution and adoption of AI and augmented analytics throughout the enterprise. (Augmented analytics is the automation of data insights using machine learning.) AI business value is expected to reach 3.9 trillion in 2022.* Our data science community survey demonstrates how all professions have begun to see the utility of data science, and many are beginning to apply it to their current roles within the organization.

Our survey also reflects the importance of open source tools and collaboration. While Jupyter Notebooks is the leading data science tool, many of these users also use RStudio, VS Code, and PyCharm, among other tools. For data science to be successful at the enterprise level, it will be necessary for business professionals, analysts, developers, and data scientists to collaborate using multiple tools within a secure environment.

* Forbes: Roundup Of Machine Learning Forecasts And Market Estimates For 2019

METHODOLOGY

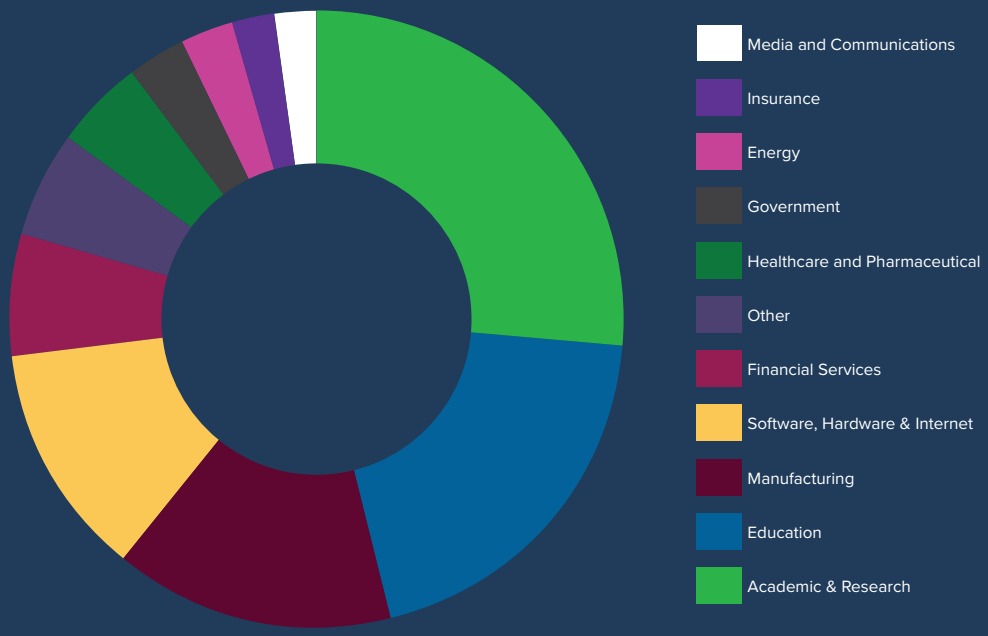
THANK YOU FOR YOUR SUPPORT.

Everyone at Anaconda is committed to furthering the development of the open source data science community by providing an open platform, innovative tools, and useful thought leadership. The Anaconda State of Data Science report is just one more way we hope to empower your success.

SURVEY METHODOLOGY AND RESPONDENTS

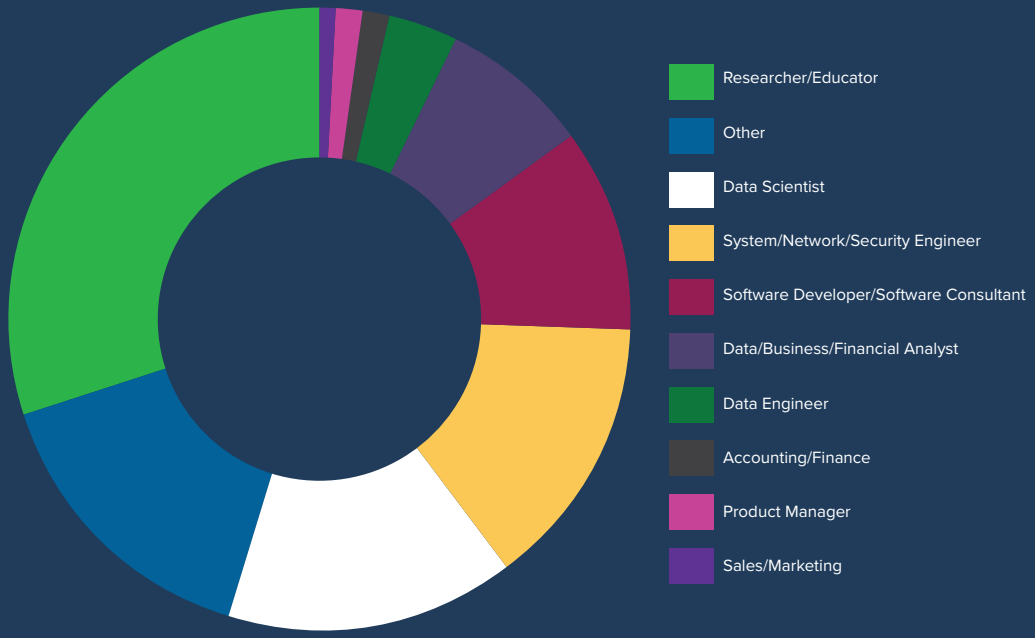
We conducted an online survey of the Anaconda community between March 5 and April 30, 2019, ultimately receiving 4,970 responses. The majority of respondents were in academic research and education, followed by those in manufacturing and software/hardware/Internet-based careers.

RESPONDENTS BY INDUSTRY



Among respondents, a strong majority identified as researchers or educators. Data Scientists made up roughly 16 percent, followed closely by System/Network/Security Engineers. Interestingly, the second-largest group of respondents did not identify with any of the provided job function choices. This throws a spotlight on the breadth of the Anaconda community.

RESPONDENTS BY JOB TITLE



ADDITIONAL INSIGHTS

Roughly 30% said they work at enterprise-level organizations, followed by companies with 101-500 employees, and self-employed individuals.

A strong majority of respondents identified as individual contributors and/or professionals who are not in management roles.

The majority of respondents have degrees in something other than computer science, with a nearly equal split between those with bachelor and graduate degrees.

ANACONDA, INC.

With more than 15 million users, Anaconda is the world's most popular data science platform and the foundation of modern machine learning. Our flagship product, Anaconda Enterprise, delivers data science and machine learning at speed and scale, unleashing the full potential of our customers' data science and machine learning initiatives. Visit anaconda.com/enterprise to learn more.