



Kaggle's State of Data Science and Machine Learning 2019

Enterprise Executive Summary



Kaggle’s State of Machine Learning and Data Science 2019 survey is the most comprehensive dataset available on the state of machine learning and data science today. This executive summary is limited to responses from individuals who selected their current job title as “data scientist.”



Overview

Kaggle’s third annual survey of its community shows the worldwide reach of data science. **Based on responses from 19,717 Kaggle members, this report focuses on the 21% that are currently employed as data scientists.** Overall, we see a relatively young, highly educated community working at companies of all sizes that are still figuring out the best way to adopt machine learning technologies.

The Kaggle community includes learners and practitioners of all levels. This analysis focuses on the professional data scientists within the community—their education, employment, and the tools used to perform their work. You’ll see certain regions—most notably the United States and India—represented at the extremes of the data.

Data scientists have adopted cloud computing in their work, though not as a replacement for local developer environments. Nevertheless, many have significant budgets for cloud tools, with the United States spending beyond others. Google Cloud Platform usage grew compared to the 2018 survey, with overall usage second to AWS. Among cloud machine learning tools, use of Google Cloud AutoML nearly doubled since last year.

While many data scientists have advanced degrees, most continue to learn new data science skills. Blogs, Kaggle forums, Coursera, and YouTube are among the common methods of ongoing education. With many companies still new to machine learning, it’s clear there is still a need for both instruction and practical application of techniques.

Report Methodology

The content of this report focuses on respondents who are currently employed and chose their current job title as “data scientist”. There are many other job titles that support data science and machine learning workflows and you can find their responses in the [complete 2019 survey dataset on Kaggle](#).

Many survey questions were multiple choice with the ability for respondents to select all options that applied to them. For that reason, you may see visualizations where the total percentage is more than 100%. All monetary amounts captured in the report are in USD.

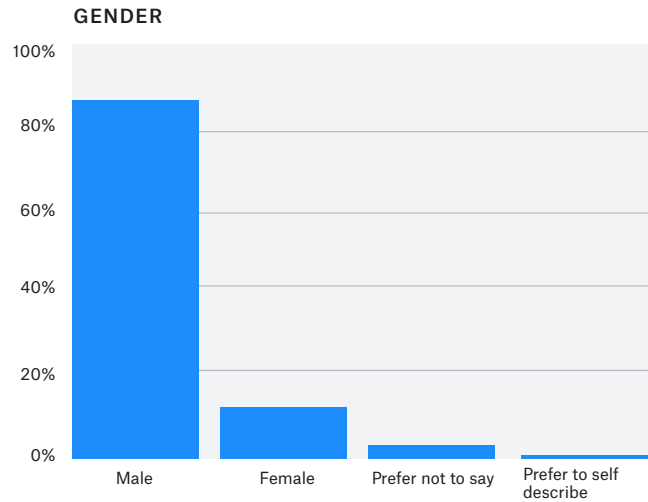
Key Results

- Data science is mostly male, an imbalance that has remained unchanged from previous years.
- Over half of data scientists are less than thirty years old.
- Unsurprisingly, data scientists are highly educated, with well over half obtaining advanced degrees.
- More than half of respondents have fewer than five years of coding experience and even less experience with machine learning.
- Salaries for data scientists in the United States far exceed other countries.
- Most data scientists work at either small or very large organizations.
- More than half of companies are new to machine learning.
- Local development environments are the most common way data scientists perform their work.
- Nearly one in four professional data scientists have still not adopted cloud computing.
- TensorFlow and Keras continue to be the dominant deep learning framework.
- The United States spends far more on machine learning and cloud computing products than the rest of the world.
- Simple methods, such as linear regressions and decision trees, dominate despite the power of more complex techniques.
- Usage of Google Cloud AutoML nearly doubled compared to last year.

Data Scientist Profile

Gender

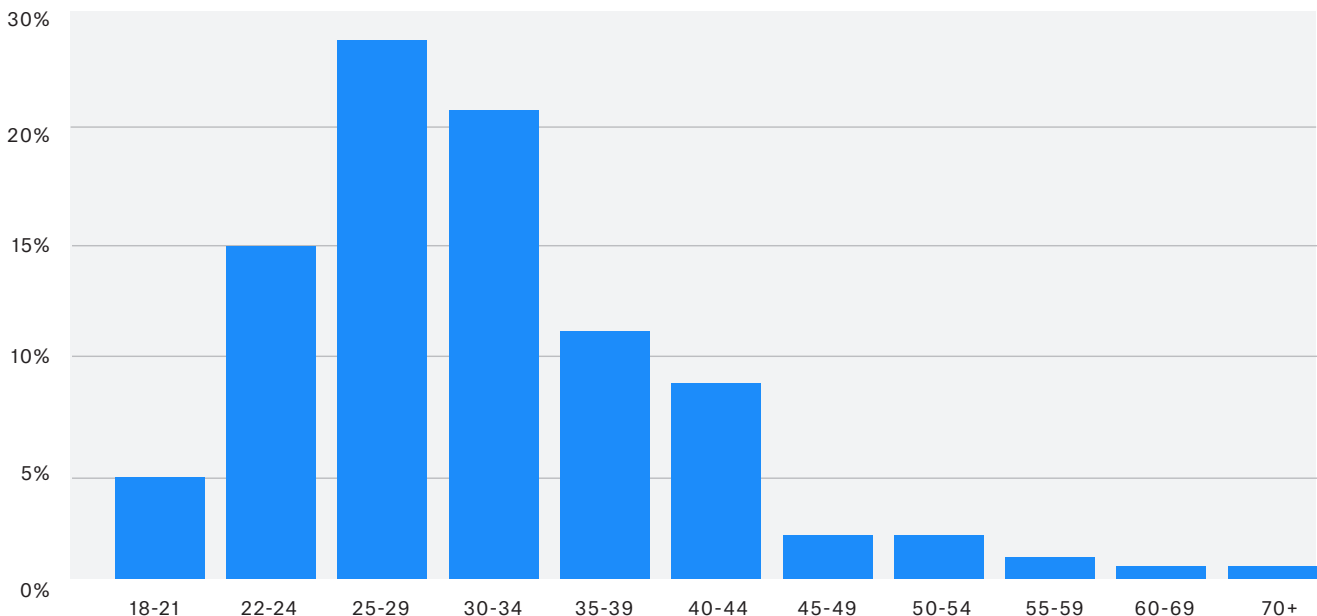
There's still a significant gender gap for data scientists, with **84% of users identifying as males**. The United States has a slightly smaller gender gap at 79%, while Japan has a slightly higher one at 90%. The results are relatively consistent regardless of region and has not changed since results of earlier Kaggle surveys.



Age

Millennials dominate data science, with 25-29 year olds being the most common age bracket. **In India, the numbers skew even younger, where 41% are 19-24.** However, adults of all ages are exploring data science, with 18% of all respondents 40 or older.

AGE



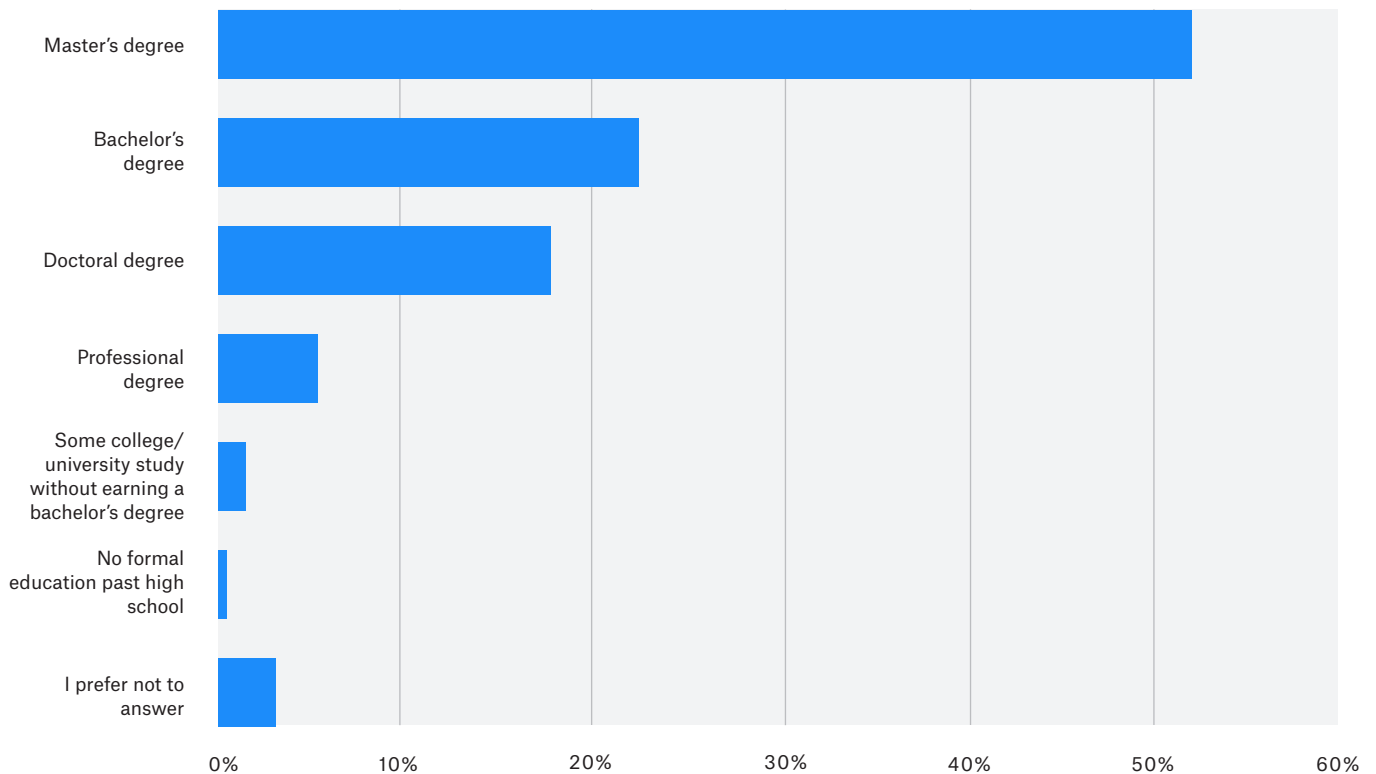
Education

Higher Education

The data scientist community is highly educated. Looking at only employed data scientists, **over 70% of respondents have a degree above a bachelor's degree**, with a majority (~52%) having a master's degree. While 19% of the total respondents have PhDs, it varied greatly by country. Germany had the highest percentage of respondents who held doctoral degrees with 38%, while India had the lowest percentage with under 5%.

More than 99.5% of data scientists pursued some education after high school.

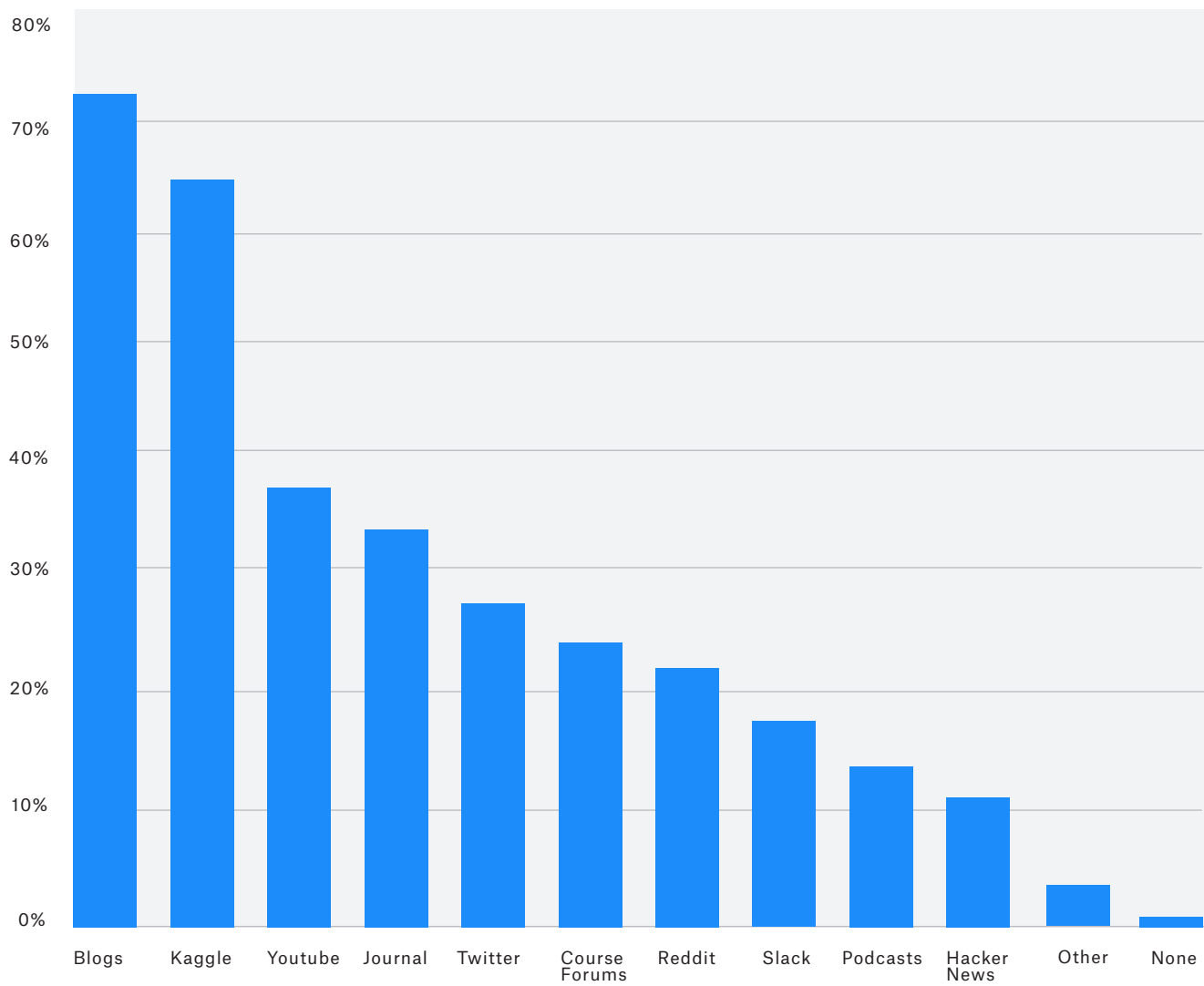
EDUCATION



Ongoing Learning

Over 70% of data scientists said they learned through reading blogs. Using the Kaggle forums is also popular among Kaggle users, with over 65% using them. There were many other responses, **but one thing is certain: the vast majority of data scientists are still learning**; only ~2% of respondents said they don't use any media to improve their data science skills.

MEDIA CONSUMPTION

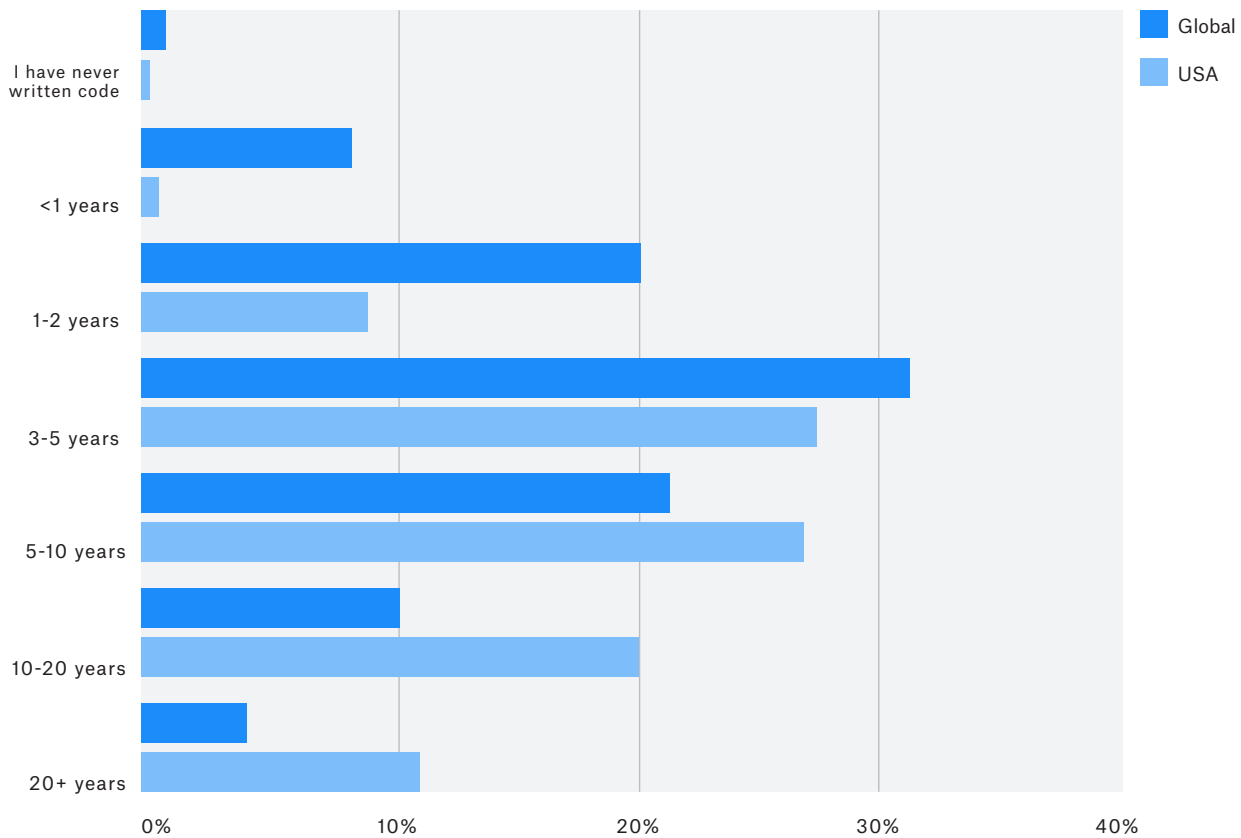


Data Science & Machine Learning Experience

The global data scientist community consists of an equal amount of new learners and seasoned veterans.

The most common (33%) range is 3-5 years experience. Roughly one-third have less than three years experience and another third have more than five years experience.

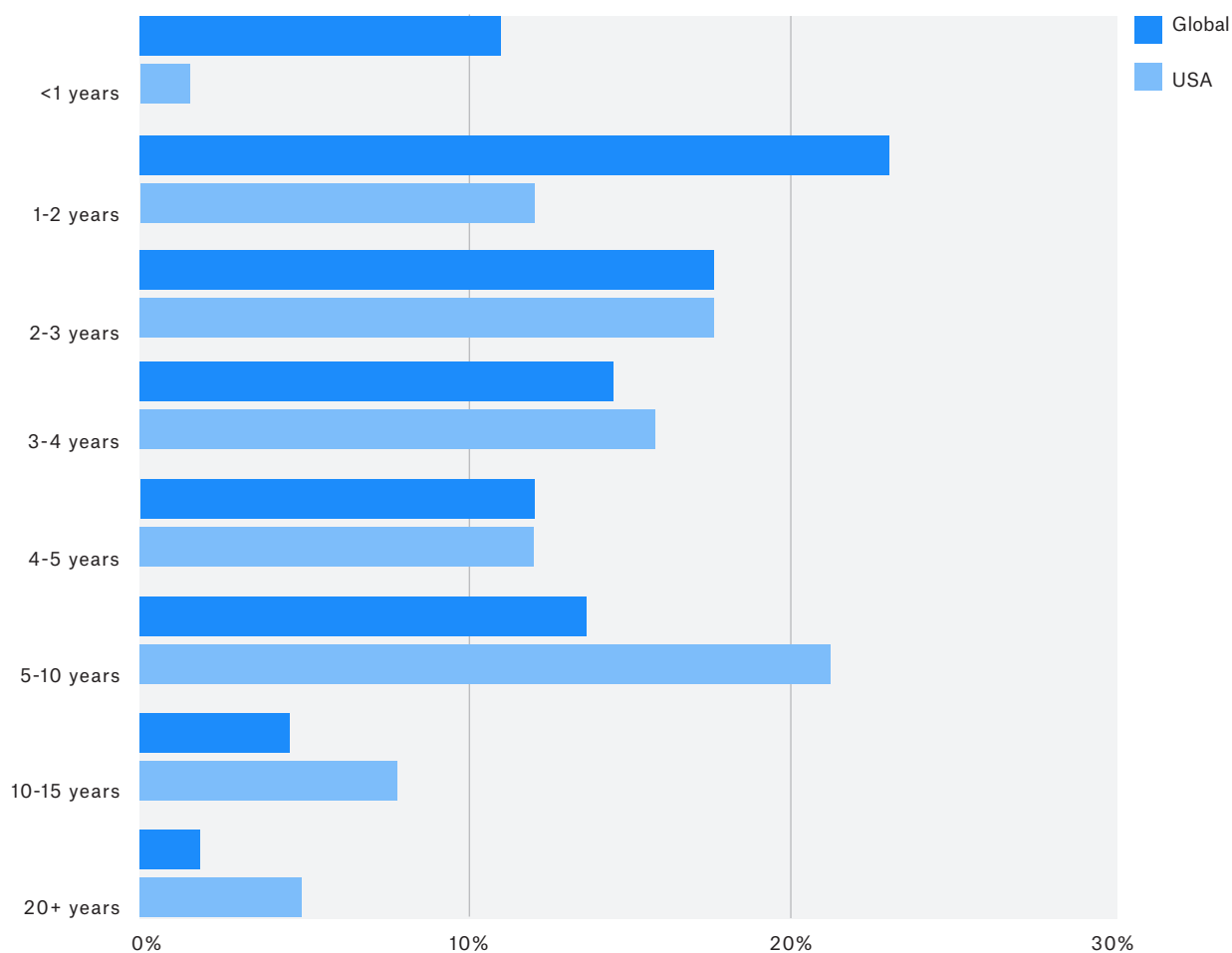
TIME SPENT LEARNING CODE



Machine learning is less normally distributed. While most have more than a year of experience, **35% are still in their first two years of using machine learning**. About 6% have more than 10 years of machine learning experience.

Compared to the international figures, the US has more data scientists in their first two years (17%) and more with 10+ years of experience (14%).

TIME SPENT LEARNING MACHINE LEARNING



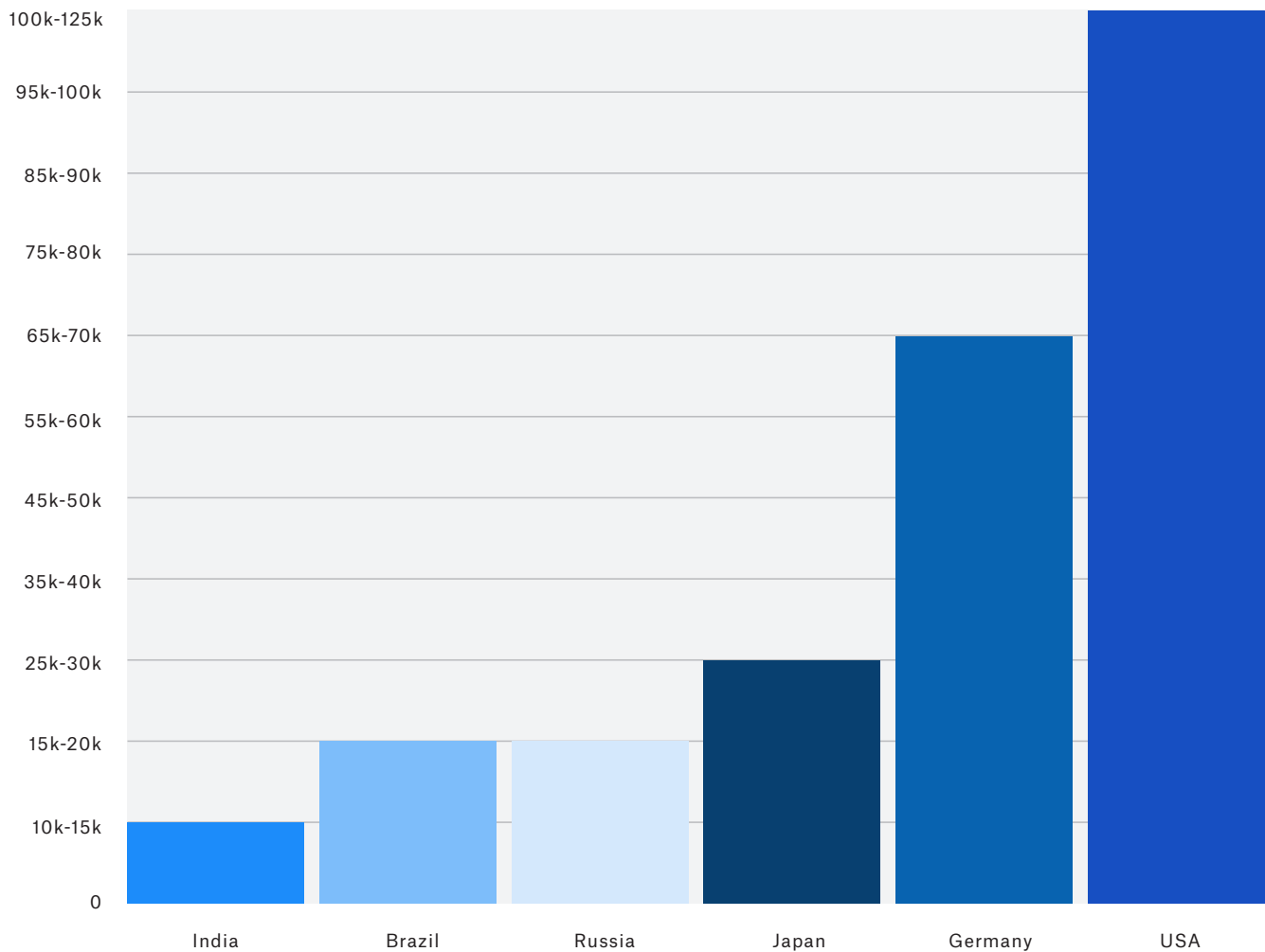
Employment

Pay

We asked data scientists about their salary, employer type, and how they spend their time. Results varied considerably by country, especially when it came to pay.

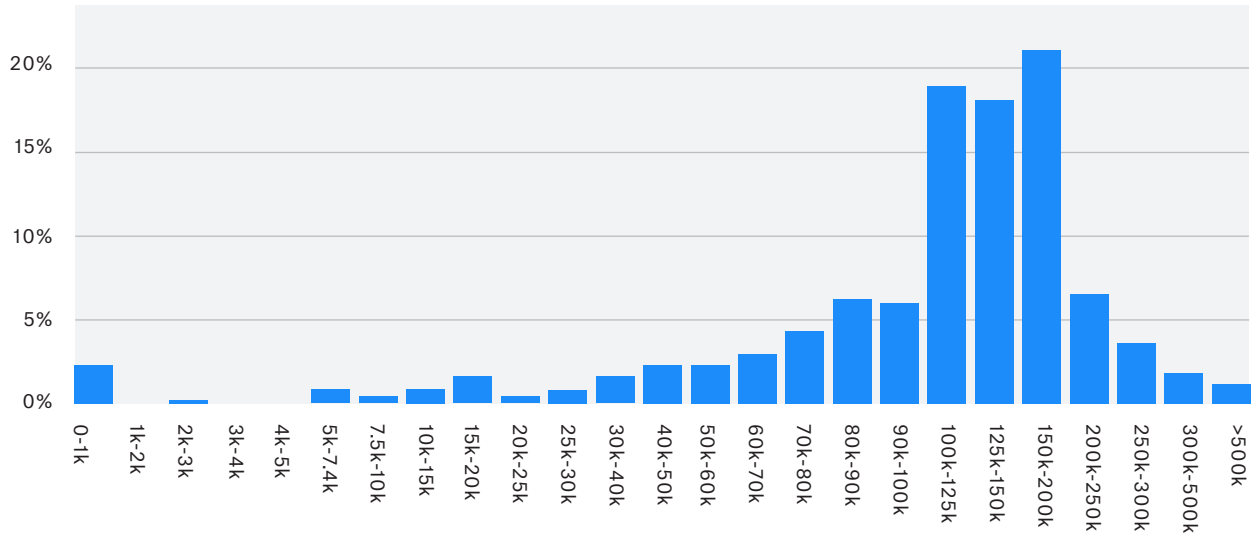
United States data scientists average higher wages than others surveyed, followed by Germany and Japan. India, on the other hand, sees lower salaries, with nearly 20% of Indian respondents earning less than \$1,000 annually.

GLOBAL SALARIES (\$USD)



Those employed as data scientists in the United States fall within ranges near the top of the scale used in our survey. The majority make between \$100,000 to \$200,000.

SALARY

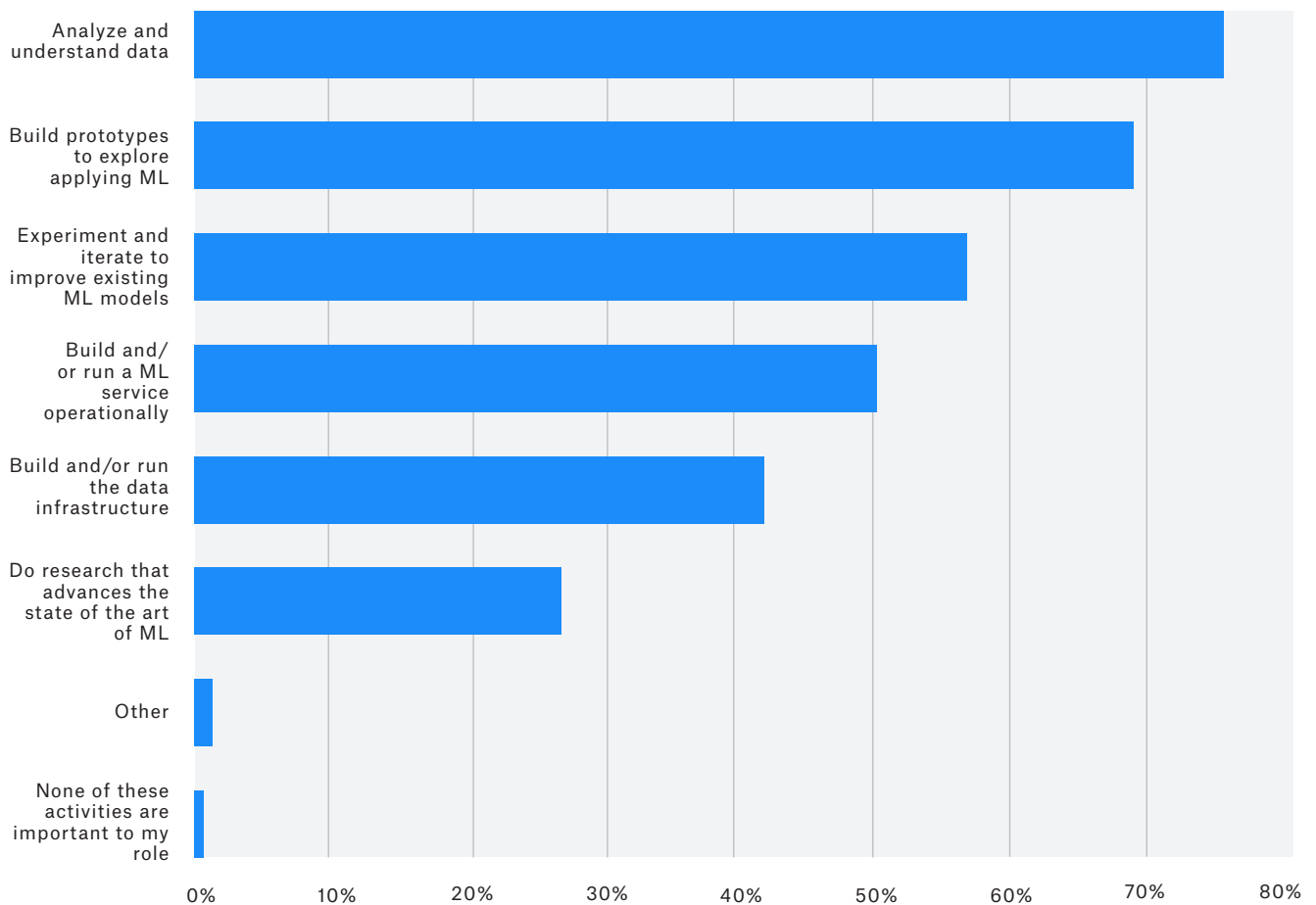


Time Spent

What do users say is the most common duty of being a data scientist? More than complex machine learning, **over 75% suggested understanding and analyzing the data is a common activity.** Perhaps this explains how Kagglers are able to create so many great EDA kernels in the first hour of every new competition!

Machine learning does play a healthy role in the work of a data scientist. **Prototyping and experimenting with machine learning were mentioned by more than half of respondents.**

HOW DATA SCIENTISTS SPEND THEIR TIME



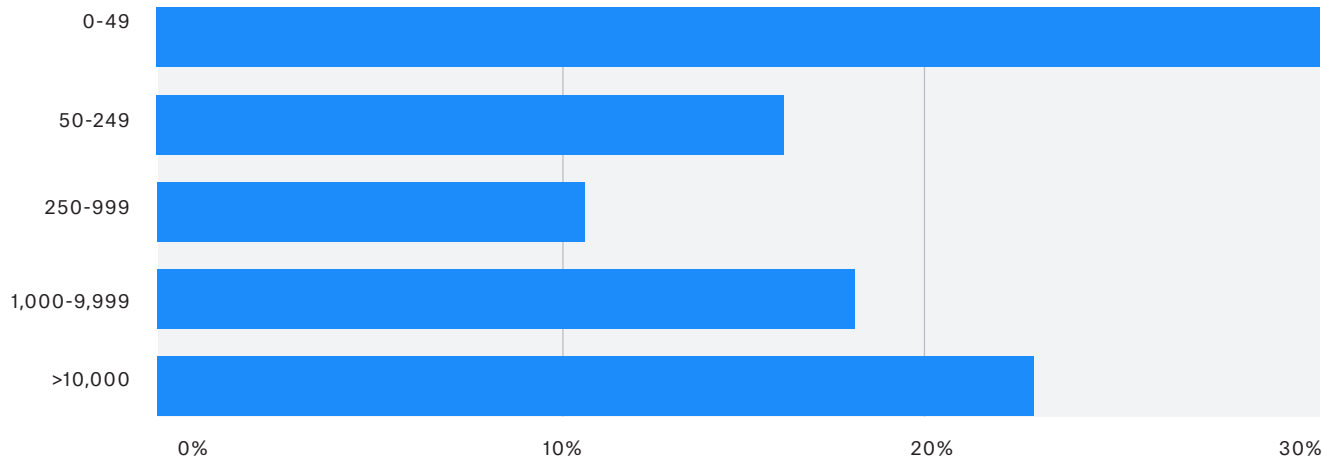
Companies Employing Data Science

We asked data scientists more about the organizations where they worked: number of employees, team sizes, and how the companies have adopted machine learning practices.

Data scientists tend to congregate at both ends of the company size spectrum.

The most common responses were from representatives of companies with less than 50 employees. Next came the much larger companies with more than 10,000 employees.

COMPANY SIZE (# OF EMPLOYEES)

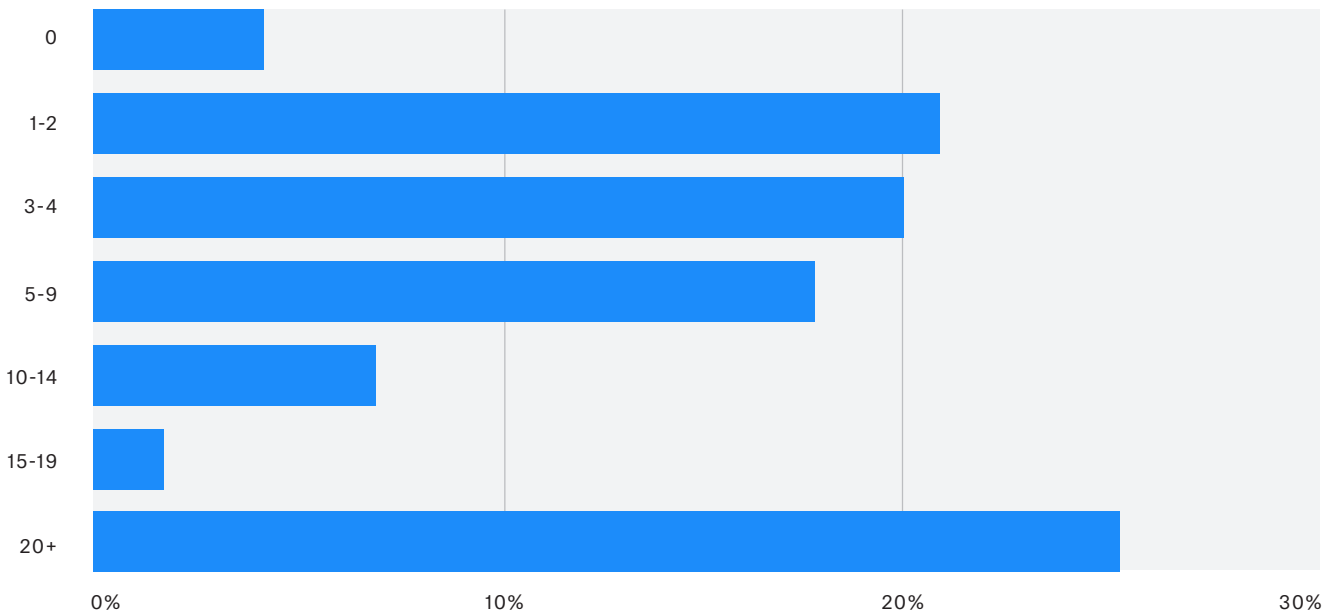


Data Science Teams

The size of the data science team varies, though 25% work on teams with 20 or more members. Combining the lower ranges, we see over 40% work on teams of fewer than five people.

Of users that are currently employed as a data scientist, 4% reported that they had a team size of zero. Either these respondents weren't counting themselves, or perhaps data science is only a portion of their responsibilities.

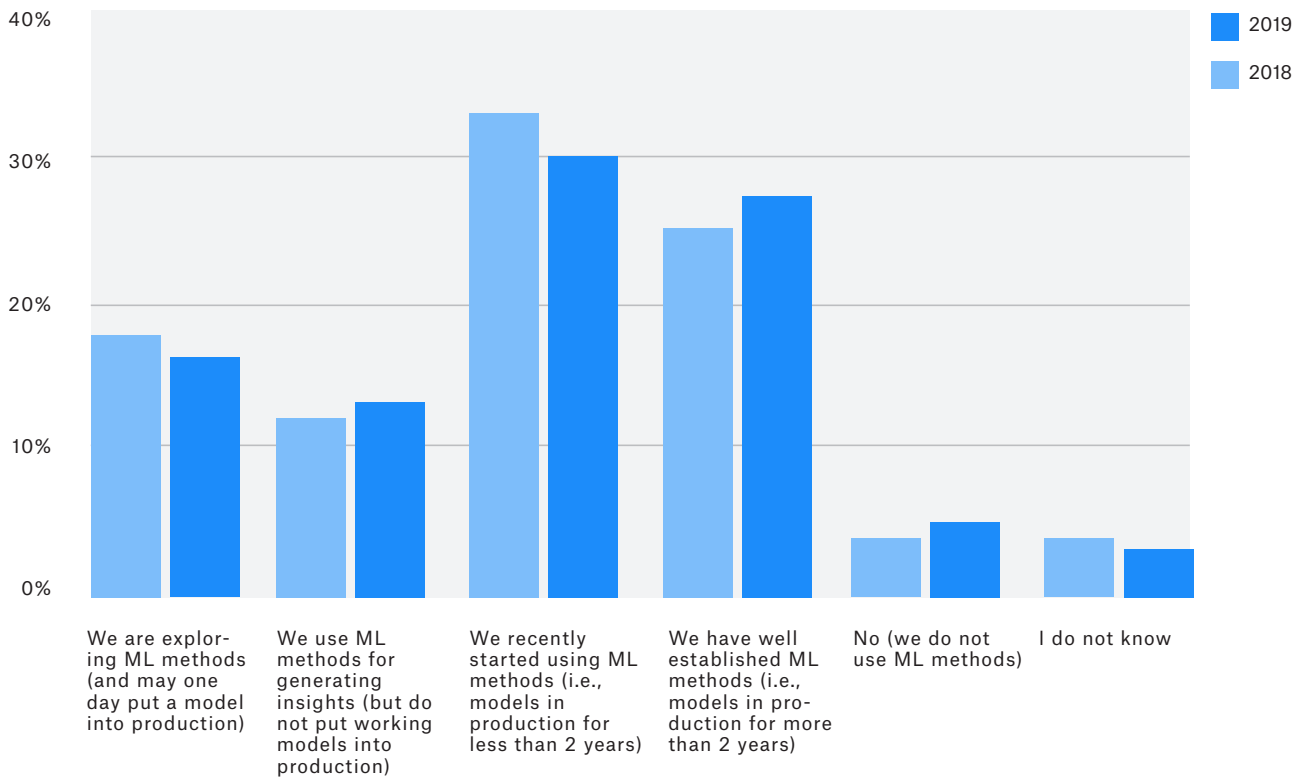
DATA SCIENCE TEAMS (# OF EMPLOYEES)



Enterprise Machine Learning Adoption

Matching other responses, machine learning is becoming more popular. Over 30% of users say their company has recently started using machine learning methods and 17% say they're exploring machine learning methods. The percentage of data scientists that work for companies with well established machine learning methods increased by 11% from 2018.

MACHINE LEARNING ADOPTION

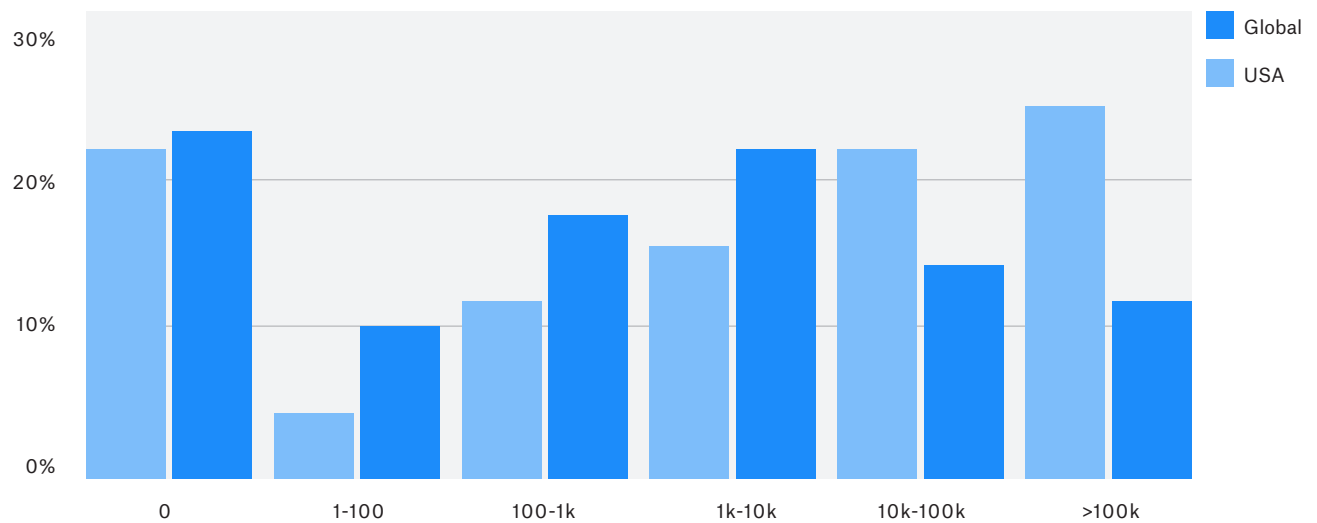


Spending

Internationally, the plurality of respondents (23%) didn't spend money on machine learning and cloud computing products at all. That being said, 11% of respondents have spent more than \$100,000.

The story is different in the United States, where a plurality (24%) have spent over \$100,000 on products in the past five years. Only 34% report having spent less than \$1000, compared to nearly 43% globally.

ENTERPRISE SPENDING IN PAST 5 YEARS (\$USD)

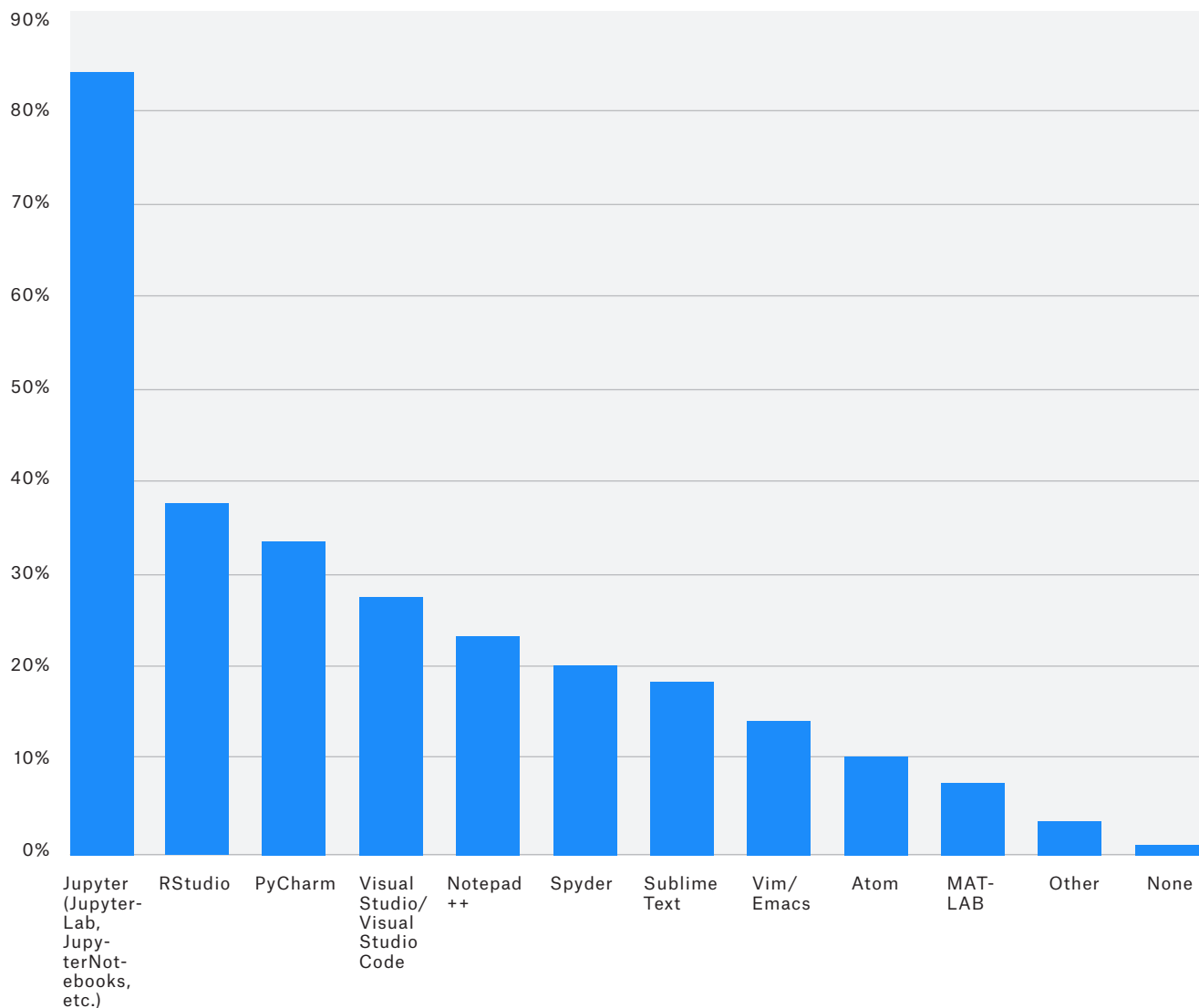


Technology

Interactive Development Environments

The most common analytics tools are by far local development environments. Out of those, **Jupyter-Lab and its offshoots are the most common**, with 83% of data scientists using it on a regular basis.

POPULAR IDE USAGE

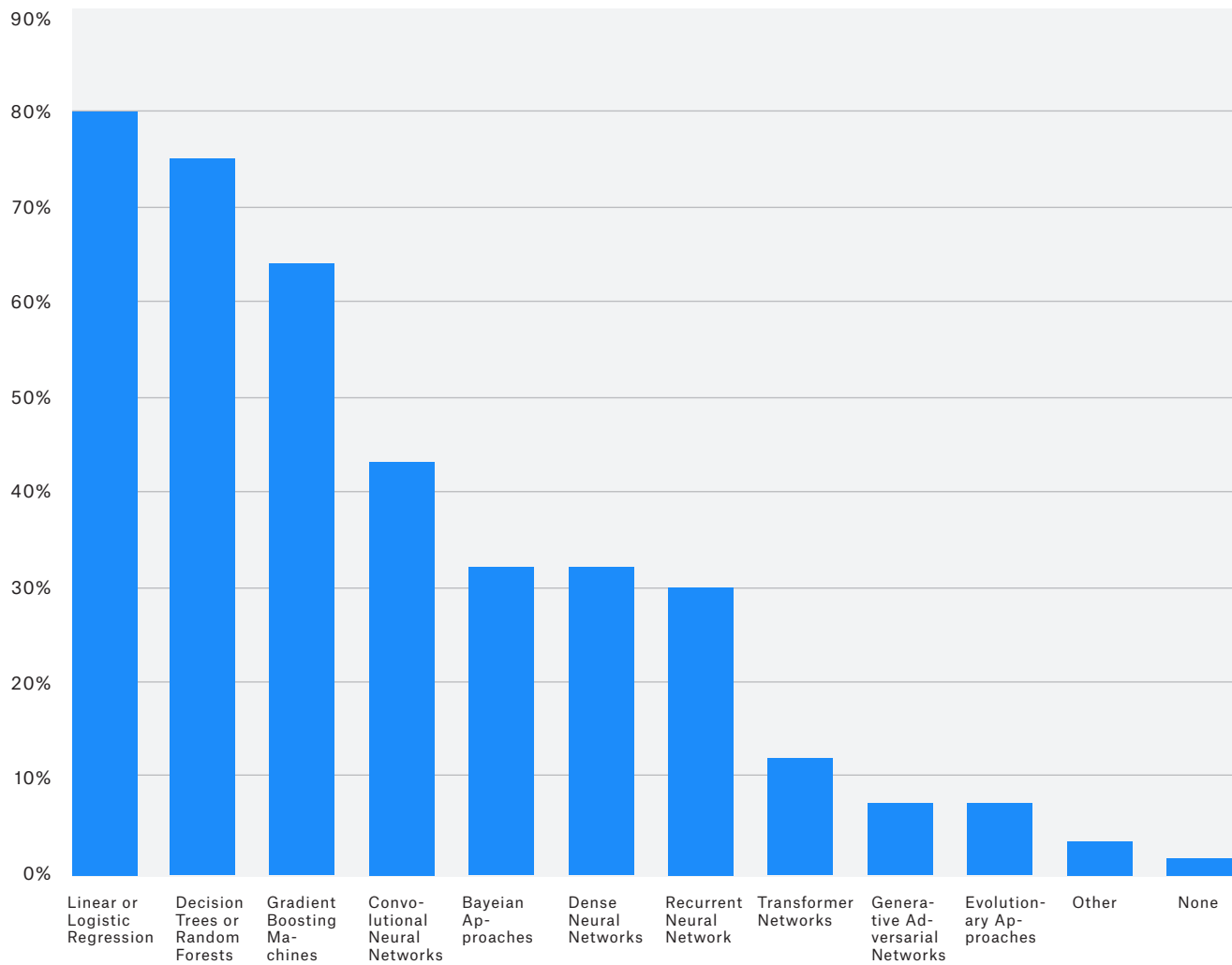


Methods and Algorithms

Respondents are big fans of keeping it simple.

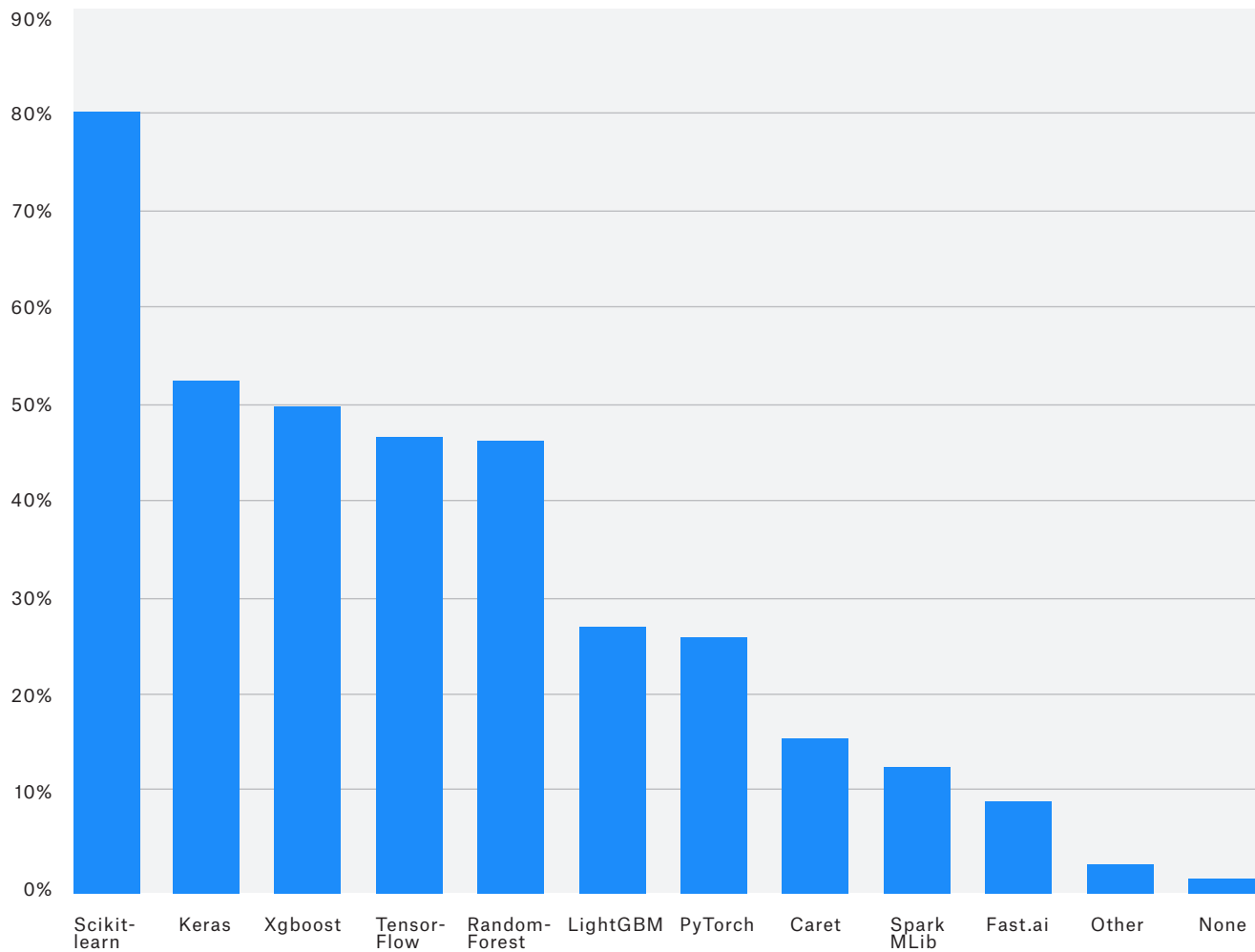
The most common methods are linear or logistic regression, followed by decision trees. While not as powerful as more complex techniques, they can still be quite effective and are easier to interpret.

METHODS AND ALGORITHMS USAGE



As for the machine learning frameworks used to employ their techniques, data scientists use multiple tools. **Over 80% use Scikit-learn**, a Python package containing popular data science algorithms. **TensorFlow and Keras**, often used in combination, continue to be the dominant deep learning framework.

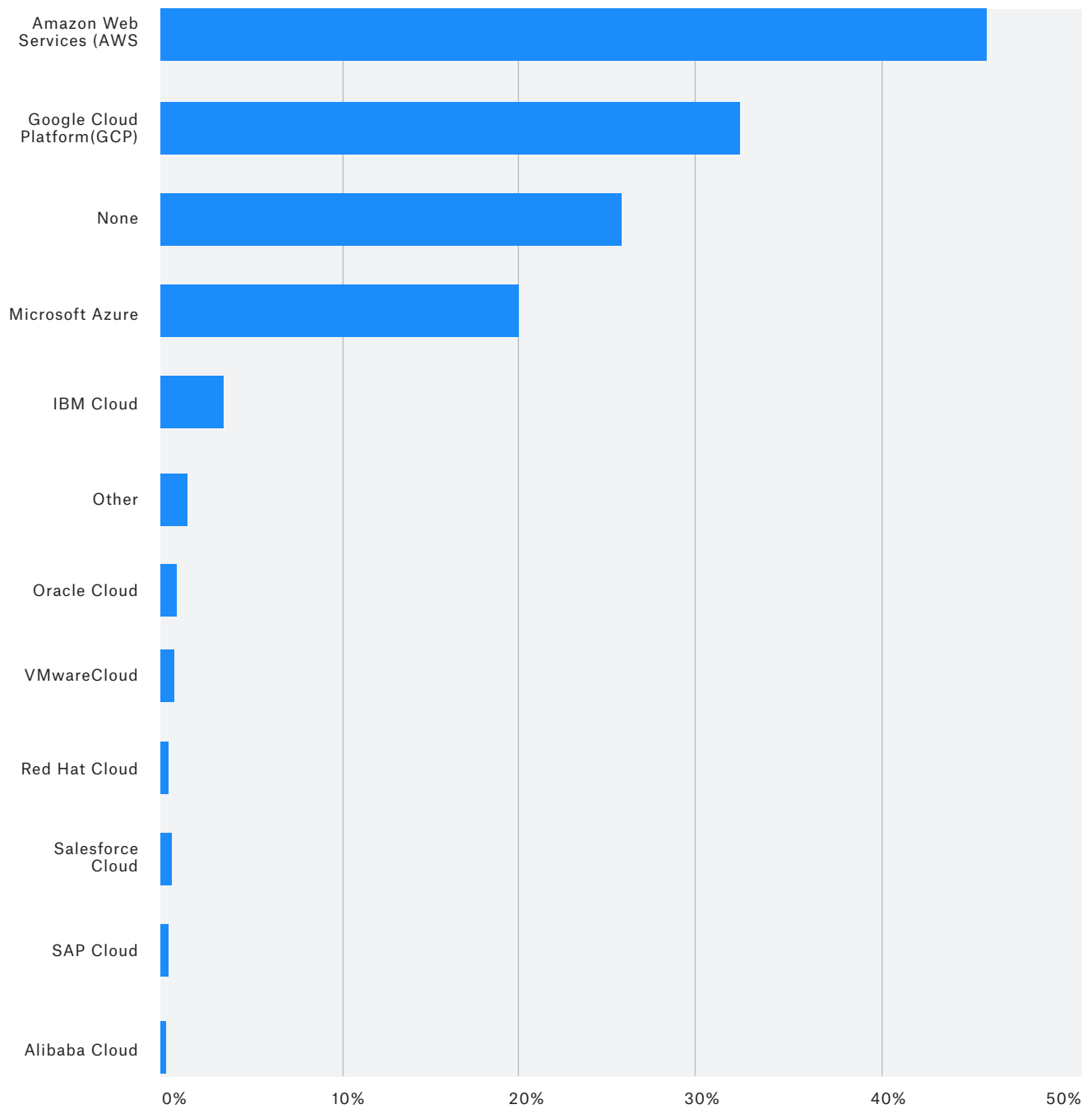
FRAMEWORKS USAGE



Enterprise Tools

Most professional data scientists are making use of cloud computing, though over 24% still are not. AWS, Google Cloud Platform, and Microsoft Azure are by far the top three choices among data scientists using cloud tools.

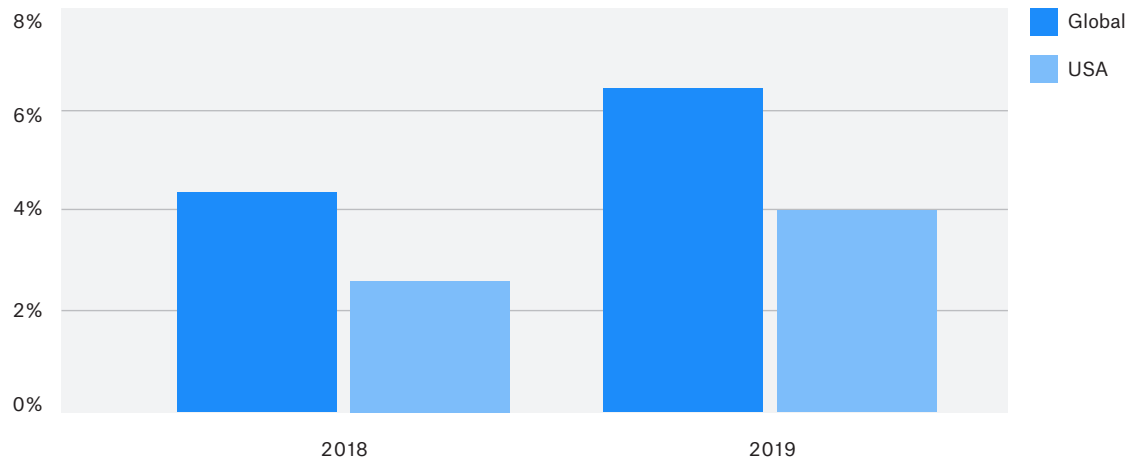
ENTERPRISE TOOLS USAGE



Automated Machine Learning

Particularly notable is the growth of Google Cloud AutoML since last year's survey. **The number of respondents using this machine learning platform nearly doubled overall**, with a similar rate of growth among US-based data scientists.

GOOGLE CLOUD AUTOML



Conclusion

This 2019 edition of the State of Machine Learning and Data Science includes insights gathered from a survey of 19,717 Kaggle members. Their answers covered demographic, education, employment, and technology usage.

Much of the charts and results are culled from professional data scientists (covering 21% of respondents). There's even more to uncover in the most comprehensive dataset available on the state of machine learning and data science today.

Kaggle has published the [complete dataset of responses](#) for the community to review, and we'll run a competition from November 11 to December 2nd to learn even more about data science practitioners in 2019.



kaggle™