Data-science-skills-study-2019-by-aim-imarticus-learning:

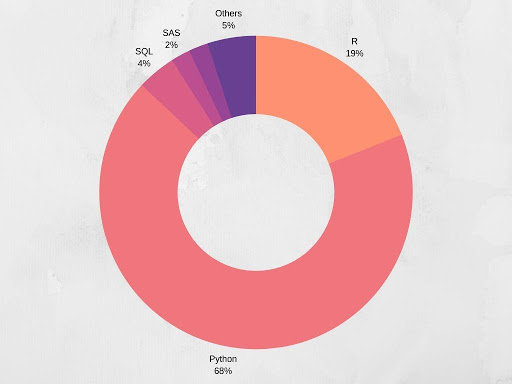
Key Highlights

* Python continues to be the most popular language in the industry in 2019 with its popularity growing to 68%. Last year, **44%** of the respondents had said it was their preferred language
* R has ceded ground to Python with 19% preferring R
* Other languages falling out of the analysts’ toolbox are SQL (4%) and SAS (2%)
* Another significant change seen this year is the increase in the use of GPUs at work. While most of the data scientists still use PCs and similar models, the second-favourite product is Nvidia GeForce GTX 9 Series GPU. The number of people using it has grown from a mere **8%** last year, to **28%** in 2019
* In terms of cloud as an infrastructure model, a majority of the respondents have said that they preferred Amazon Web Services. However, its popularity seems to have dropped slightly, since the numbers have gone down from **45%** to **43%**

Following are the key findings from the 2019 survey:

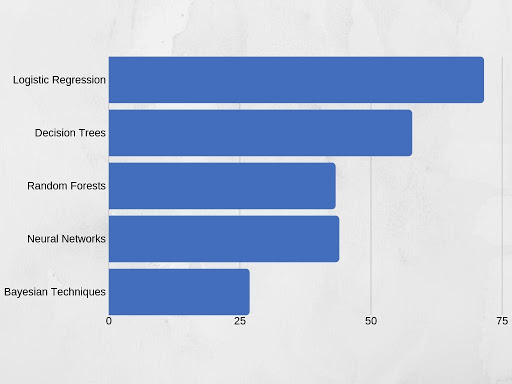
Which Language Do Data Scientists Prefer For Statistical Modelling?

* The favourite language for data scientists in today’s era is Python, as almost **68%** of the professionals use it the most. Last year, **44%** of the respondents had said it was their preferred language
* A distant second is R at **19%** — a versatile language, but which has clearly lost its popularity
* SQL (4%) and SAS (2%) claim only a minor share of the attention of the data scientists

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/2.jpg)

Which Data Science Methods Are The Most Popular At Work?

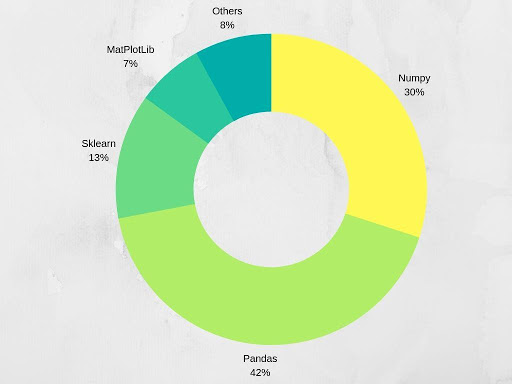
* **71%** scientists answered that they used Logistic Regression most at work
* This was followed by Decision Trees at **58%** and Neural Network at **44%**
* This data is almost unchanged compared to last year

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/3.jpg)

Which Is The Most Popular Python General Purpose Library?

Python is one of the largest programming communities in the world. There are plenty of libraries which a data scientist can use to analyse large amounts of data. But here are our readers’ favourites. (This was a multiple choice question).

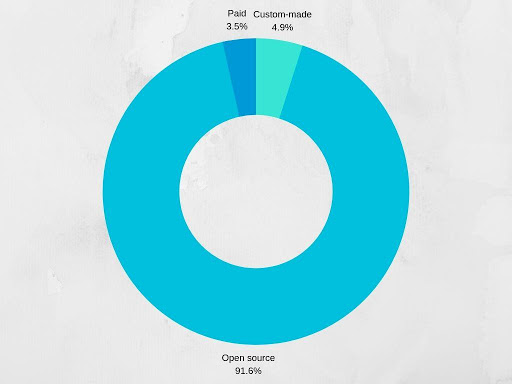
* Pandas emerged as a clear choice for most data scientists at almost **42%**
* Numpy was the second-favourite at **30%**, having seen a **6%** increase from last year’s numbers
* Sklearn and MatPlotLib followed at **13%** and **7%** respectively

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/4.jpg)

Which Tools Do Data Scientists Prefer?

With a plethora of data analytics tools available online, we asked data scientists if they were willing to use open-sourced tools at work. The answer was a resounding yes.

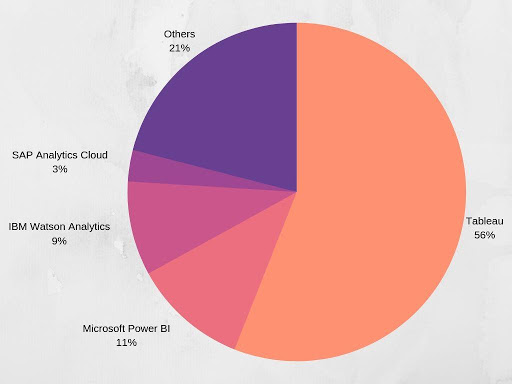
* Almost **92%** of the data scientists said that they preferred to work with open-sourced tools. This number was at **89%** last year, which shows that more and more people are now opting or open-sourced tools
* Less than **5%** of the data scientists said that they liked to work with custom-made tools which are tweaked and personalised for their particular projects

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/5.jpg)

Which Dashboard/Visualisation Tools Do Data Scientists Prefer?

Data Visualisation may be a tricky path for many data scientists. Crunching numbers is one thing, but telling a story with numbers is a whole different deal. When we asked about this to our readers they had one clear winner:

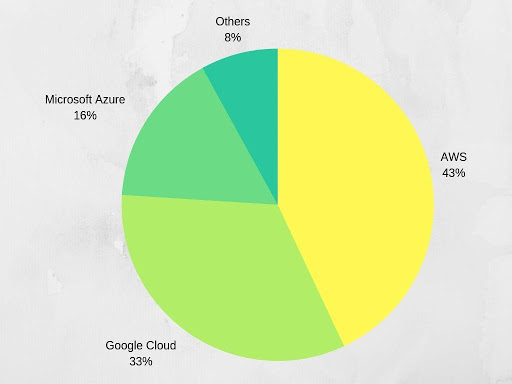
* More than half the respondents, **56%**, said that they preferred to use Tableau as a dashboard or visualisation tool
* Microsoft Power BI claimed only an **11%** share of the respondents’ interest, a distant second
* This data has remained relatively unchanged since last year

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/6.jpg)

Which Cloud Provider Do Data Scientists Prefer?

A smooth and flawless flow of information is a crucial part of data science and analytics. While data usage and storage are important, security and privacy of the data are also key to the job.

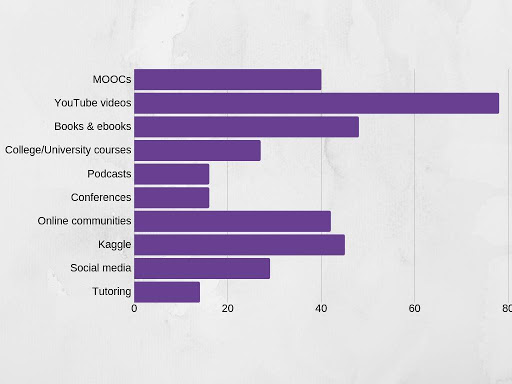
* Amazon Web Services is a clear winner here with over **43%** of the votes
* Google Cloud is the second favourite with over almost **33%** votes
* Microsoft Azure is a distant third at **16%**
* This trend has remained relatively unchanged since 2018

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/7.jpg)

What Kind Of Learning Resources Do Data Scientists Use To Keep Themselves Updated?

With the rapidly-evolving technology, it is crucial for data scientists to keep themselves updated. And they seem to have found out an interesting way to do so! (This was a multiple choice question).

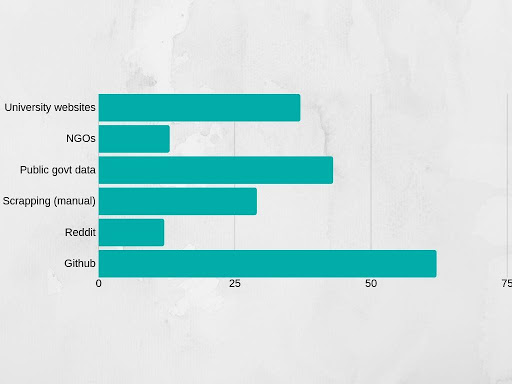
* **78%** of our readers said that they liked watching tutorials and videos on YouTube
* Almost **48%** of the data scientists said that they like learning the old-school way — through books and e-books
* **40%** of respondents also look at MOOCs as a way to upskill themselves
* Kaggle also seems to be a popular medium of upskilling at **45%**

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/8.jpg)

Where Do Data Scientists Find Open Data?

Finding open data is not that hard, but getting clean open data is often a trying experience. No data scientist wants to waste their time cleaning it. There were four clear popular options here about this multiple choice question:

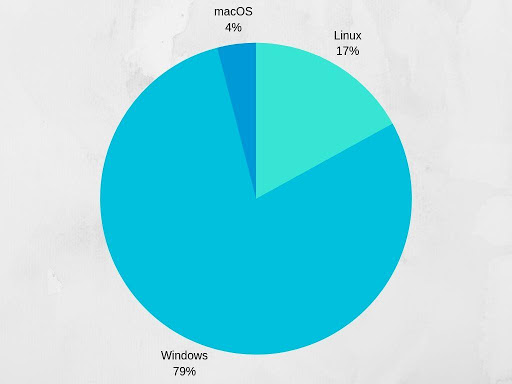
* **62%** respondents use GitHub.
* **37%** readers used university websites and the data uploaded by them for research
* **43%** data scientists also use data publicly uploaded on official government websites
* **29%** of the respondents source their data manually

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/11.jpg)

Which OS Do Most Data Scientists Use At Work?

Compatibility with the data scientists’ other tools and ease of use at work are the two key factors considered in any good operating system. For this question, the respondents had a clear liking for one OS. (This was a multiple choice question).

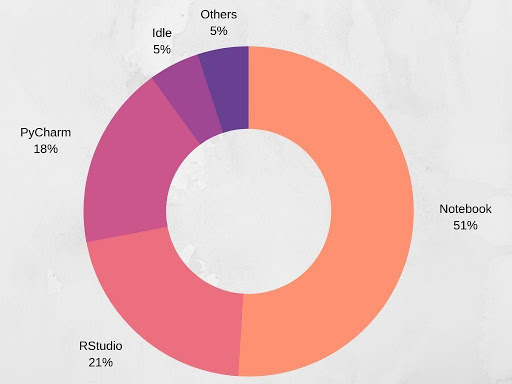
* Almost **79%** of data scientists use Windows OS. This number has seen a drastic increase since last year, where the number of data scientists who preferred Windows was at **69%**
* **17%** prefer Linux, a number which was considerably higher at **24%** last year
* And only **4%** prefer MacOS

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/10.jpg)

Preferred Development Environment

An integrated development environment (IDE) is very important to set up and streamline data science processes. Among the many options presented, the data scientists who took part in our survey chose:

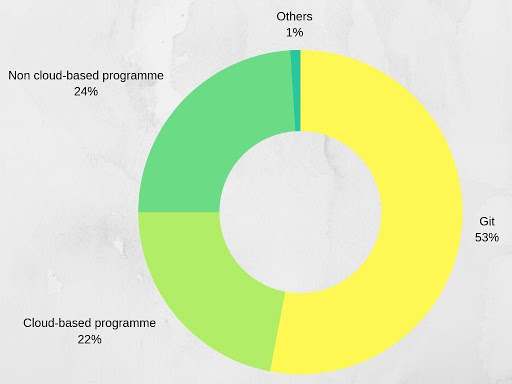
* Almost **51%** prefer using Notebook — a number which has seen a huge spike since last year’s **37%** share. This has clearly reflected in the shares of other IDEs
* And close to **21%** data scientists like using RStudio
* PyCharm ranks a distant third with only **18%** data scientists preferring to use it at work

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/12.jpg)

How Is Code Shared At Your Workplace?

Privacy, operational efficiency and security are of paramount importance in any organisation that deals with data. That’s why the method in which the codes are shared at work also play an important part. Here’s what we found out:

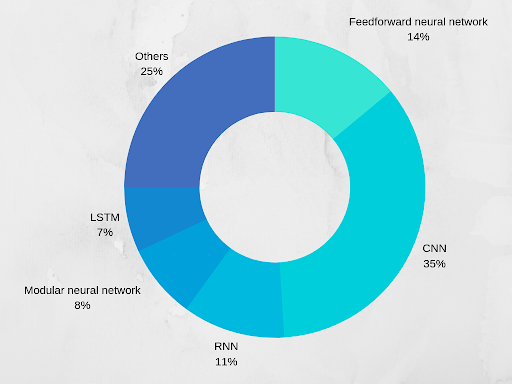
* Over **53%** of the respondents use Git to share codes at workplaces, a share which was at **45%** last year
* **24%**of the data scientists said that their organisations use non-cloud based programmes to share codes
* And **22%** of our readers shared codes over cloud-based programmes

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/13.jpg)

What Is The Neural Network Architecture Data Scientists Use Most Frequently?

Neural networks are a crucial part of programming as well as data science. We got a clear picture that the data scientists, as well as their organisations, use a variety of architectures.

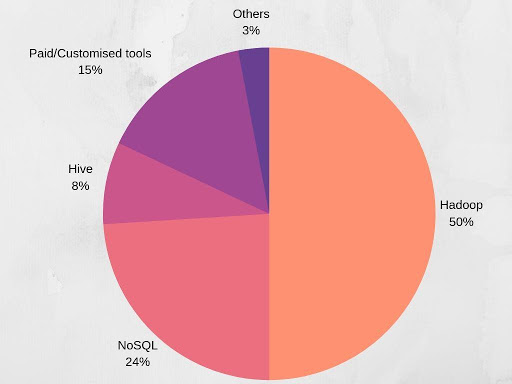
* According to our study, the convolutional neural network was the most frequently used NN at **35%**
* Feedforward neural network was a second favourite at a distant **14%**, a number which has fallen down considerably since last year’s**25%**

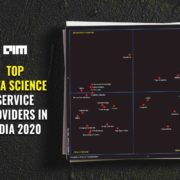
[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/14-2.png)

Which Big Data Tool Have You Used The Most?

From open-source tools to paid or customised ones, many professionals prefer different tools based on the projects or the organisation they are working for. Data scientists from our survey rated their most-favoured big data tools in the following order:

* **50%** of the users said they used Hadoop the most
* Almost **24%** data scientists used NoSQL
* And about 15% of the respondents said that they used specially paid for or customised tools the most
* These numbers are very similar to that of last year’s

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/15.jpg)

SEE ALSO[](https://analyticsindiamag.com/top-data-science-service-providers-in-india-2020/)

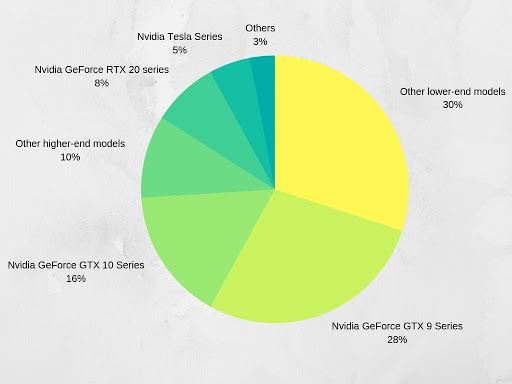
[RESEARCH](https://analyticsindiamag.com/category/research/)

[**Top Data Science Service Providers In India 2020**](https://analyticsindiamag.com/top-data-science-service-providers-in-india-2020/)

Which GPUs Do Data Scientists Use At Work?

GPUs come in handy, especially when data scientists have to work with areas of deep learning such as back-propagation, Natural Language Processing (NLP) and Artificial Neural Networks (ANN), among others, which are advancing gradually and are already catching up with traditional technologies. When we asked our respondents about their preferred models, the answer was:

* We saw a monumental change in the numbers for Nvidia GeForce GTX 9 Series GPU. The share of people using it has grown from a mere **8%** last year, to **28%** in 2019.
* **16%**of our respondents said that they preferred Nvidia GeForce GTX 10 Series

[](https://www.analyticsindiamag.com/wp-content/uploads/2019/08/16.jpg)

Which Skills Are Most Important?

With a large chunk of work being automated in sectors like BFSI, retail, e-commerce and IT, recruiters are looking for a newer set of contributors who are equipped with self-service reporting tools like Tableau and Python. That said, proficiency in R and SQL can help land positions in analytics teams. In addition to this, familiarity with neural networks – at least the most popular ones like CNN and RNN can bring a lot of additional value to teams.

Here’s a bird’s eye view:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **BFSI** | **IT products and services** | **Manufacturing** | **E-commerce** | **Retail** |
| **Language** | Python | Python | Python | Python | Python |
| **DS Method** | Logistic Regression | Logistic Regression | Logistic Regression | Logistic Regression | Logistic Regression |
| **Python Library** | Pandas | Pandas | Pandas | Nympy | Pandas |
| **Data Visualisation Tool** | Tableau | Tableau | Tableau | Tableau | Tableau |
| **Cloud** | AWS | AWS | AWS | Google Cloud | AWS |
| **IDE** | Notebook | Notebook | Notebook | RStudio | Notebook |
| **NN Architecture** | CNN | CNN | RNN | CNN | CNN |
| **Big Data Tool** | Hadoop | Hadoop | Hadoop | Paid/Customised tools | Hadoop |

As the demand for talent grows, companies look for two types of professionals — specialist and generalist skills. Adding neural networks to your toolkit will be a value addition. In the next few years, with data literacy becoming core to organisations, recruiters will look for a digitally-savvy workforce that can solve business problems through an analytical approach.

In terms of talent outlook, recruiters are developing new ways of hiring talent — from hackathons to internal employee training programmes, organisations and leaders are building a culture of continuous learning.

Conclusion

As the analytics industry grows at an astronomical speed, more professionals are expected to segue into the Data Science and Analytics sector. The number of jobs in the analytics industry is growing. With companies diversifying their product portfolio and new-age SaaS providers, fintechs and e-commerce players building niche solutions, the industry will need to integrate more analytics and machine learning specialists to work on their product roadmap.

With data growing and upstarts latching onto digital channels, companies have to work on new use cases related to fraud and data security. In the future, we will see more talent working on fraud detection, predictive analytics (especially forecasting spend behaviour) and even location-based analytics.

On the other hand, another interesting aspect is that professionals are aware of the importance of upskilling themselves. Most working professionals like to keep themselves updated by watching videos and reading books. Overall, the study reveals a positive picture of the Indian Analytics and Data Science sector.