SAS vs. R (vs. Python) – which tool should I learn?

We love comparisons!

From Samsung vs. Apple vs. HTC in smartphones; iOS vs. Android vs. Windows in mobile OS to comparing candidates for upcoming elections or selecting captain for the world cup team, comparisons and discussions enrich us in our life. If you love discussions, all you need to do is pop up a relevant question in middle of a passionate community and then watch it explode! The beauty of the process is that everyone in the room walks away as a more knowledgeable person.

I am sparking something similar here. SAS vs. R has probably been the biggest debate analytics industry might have witnessed. Python is another worthy candidate to put in the mix now. The reason for me to start this discussion is not to watch it explode (that would be fun as well though). I know that we all will benefit from the discussion.

This has also been one of the most commonly asked question to me on this blog. So, I thought I'll discuss it with all my readers and visitors!

Latest Rankings : Top Certifications on SAS, R, Python, Machine Learning, Big Data of 2015 – 2016

Hasn't a lot already been said on this topic?

Probably yes! But I still feel the need for discussion for following reasons:

- The industry is very dynamic. Any comparison which was done 2 years back might not be relevant any more.
- Traditionally Python has been left out of the comparison. I think it is a worthy consideration now.
- While I'll discuss global trends about the languages, I'll add specific information with regards to Indian analytics industry (which is at a different level of evolution)

So, without any further delay, let the combat begin!

Background

Here is a brief description about the 3 ecosystems:

- **SAS:** SAS has been the undisputed market leader in commercial analytics space. The software offers huge array of statistical functions, has good GUI (Enterprise Guide & Miner) for people to learn quickly and provides awesome technical support. However, it ends up being the most expensive option and is not always

enriched with latest statistical functions.

- **R:** R is the Open source counterpart of SAS, which has traditionally been used in academics and research. Because of its open source nature, latest techniques get released quickly. There is a lot of documentation available over the internet and it is a very cost-effective option.
- **Python:** With origination as an open source scripting language, Python usage has grown over time. Today, it sports libraries (numpy, scipy and matplotlib) and functions for almost any statistical operation / model building you may want to do. Since introduction of pandas, it has become very strong in operations on structured data.

Attributes For Comparison

I'll compare these languages on following attributes:

1. Availability / Cost
2. Ease of learning
3. Data handling capabilities
4. Graphical capabilities
5. Advancements in tool
6. Job scenario
7. Customer service support and Community

I am comparing these from point of view of an analyst. So, if you are looking for purchasing a tool for your company, you may not get complete answer here. The information below will still be useful. For each attribute I give a score to each of these 3 languages (1 – Low; 5 – High).

The weightage for these parameters will vary depending on what point of career you are in and your ambitions.

1. Availability / Cost

SAS is a commercial software. It is expensive and still beyond reach for most of the professionals (in individual capacity). However, it holds the highest market share in Private Organizations. So, until and unless you are in an Organization which has invested in SAS, it might be difficult to access one.

R & Python, on the other hand are free and can be downloaded by any one. Here are my scores on this parameter:

SAS – 2

R – 5

Python – 5

2. Ease of Learning

SAS is easy to learn and provides easy option (PROC SQL) for people who already know SQL. Even otherwise, it has a good stable GUI interface in its repository. In terms of resources, there are tutorials available on websites of various university and SAS has a comprehensive documentation. There are certifications from SAS training institutes, but they again come at a cost.

R has the steepest learning curve among the 3 languages listed here. It requires you to learn and understand coding. R is a low level programming language and hence simple procedures can take longer codes.

Python is known for its simplicity in programming world. This remains true for data analysis as well. While there are no widespread GUI interfaces as of now, I am hoping Python notebooks will become more and more mainstream. They provide awesome features for documentation and sharing.

SAS – 4.5

R – 2.5

Python – 3.5

3. Data Handling Capabilities

This used to be an advantage for SAS till some time back. R computes every thing in memory (RAM) and hence the computations were limited by the amount of RAM on 32 bit machines. This is no longer the case. All three languages have good data handling capabilities and options for parallel computations. This I feel is no longer a big differentiation. Also, I might not be aware of the latest innovation in each ecosystem and hence I see all 3 as equally capable.

SAS – 4

R – 4

Python – 4

4. Graphical Capabilities

SAS has decent functional graphical capabilities. However, it is just functional. Any

customization on plots are difficult and requires you to understand intricacies of SAS Graph package.

R has the most advanced graphical capabilities among the three. There are numerous packages which provide you advanced graphical capabilities.

Python capabilities will lie somewhere in between, with options to use native libraries ([matplotlib](#)) or derived libraries (allowing calling R functions).

SAS – 3

R – 4.5

Python – 4

5. Advancements in Tool

All 3 ecosystems have all the basic and most needed functions available. This feature only matters if you are working on latest technologies and algorithms.

Due to their open nature, R & Python get latest features quickly (R more so compared to Python). SAS, on the other hand updates its capabilities in new version roll-outs. Since R has been used widely in academics in past, development of new techniques is fast.

Having said this, SAS releases updates in controlled environment, hence they are well tested. R & Python on the other hand, have open contribution and there are chances of errors in latest developments.
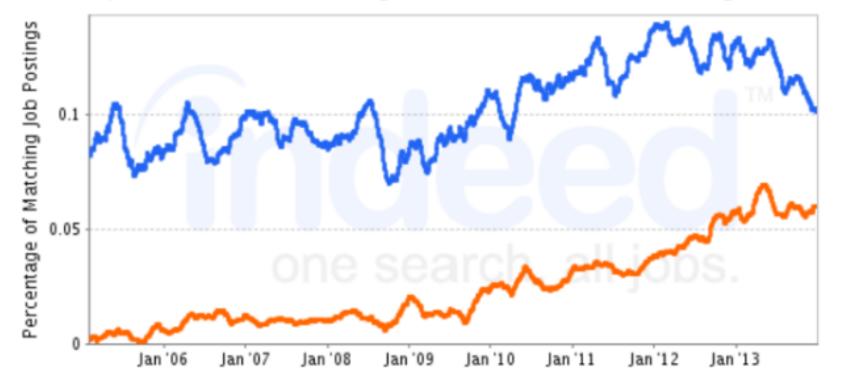
SAS – 4

R – 4.5

Python – 4

6. Job Scenario

Globally, SAS is still the market leader in available corporate jobs. Most of the big organizations still work on SAS. R / Python, on the other hand are better options for start-ups and companies looking for cost efficiency. Also, number of jobs on R / Python have been reported to increase over last few years. Here is a trend widely published on internet, which shows the trend for R and SAS jobs. Python jobs for data analysis will have similar trend as R jobs:

Job Trends from Indeed.com

— R !"R D" !"A R" !"H R" !"R N" !toys !kids !" R Walgreen" !walmart !"HVAC R" !"R Bard" and (
— SAS !"system administrator" !"school age" !sata !firmware !scsi !raid !samsung !scandinav

Source: r4stats.com

In India, specifically, the gap in SAS vs. R is bigger. My estimate is that SAS would have about 70% of market share, R around 15% and Python less than 5%. However, the trends are similar to global trends.

SAS – 4.5

R – 3.5

Python – 2.5

7. Customer Service Support & Community

R has the biggest online community but no customer service support. So if have trouble, you are on your own. You will get a lot of help though. Similar for python, though at a lower scale.

SAS on the other hand has dedicated customer service along with the community. So, if you have problems in installation or any other technical challenges, you can reach out to them.

SAS – 4

R – 3.5

Python – 3

Other Factors:

Following are some more points worthy to note:

- Python is used widely in web development. So if you are in an online business, using Python for web development and analytics can provide synergies
- SAS used to have a big advantage of deploying end to end infrastructure (Visual Analytics, Data warehouse, Data quality, reporting and analytics), which has been mitigated by integration / support of R on platforms like SAP HANA and Tableau. It is still, far away from seamless integration like SAS, but the journey has started.

Conclusion

Clearly, there is no winner in this race yet. It will be pre-mature to place bets on what will prevail, given the dynamic nature of industry. Depending on your circumstances (career stage, financials etc.) you can add your own weights and come up with what might be suitable for you. Here are a few specific scenarios:

- If you are a fresher entering in analytics industry (specifically so in India), I would recommend to learn SAS as your first language. It is easy to learn and holds highest job market share.
- If you are some one who has already spent time in industry, you should try and diversify your expertise be learning a new tool.
- For experts and pros in industry, people should know at least 2 of these. That would add a lot of flexibility for future and open up new opportunities.
- If you are in a start-up / freelancing, R / Python is more useful

Here is the final scorecard:

Parameter	SAS	R	Python
Availability / Cost	2	5	5
Ease of learning	4.5	2.5	3.5
Data handling capabilities	4	4	4
Graphical capabilities	3	4.5	4
Advancements in tool	4	4.5	4
Job scenario	4.5	3.5	2.5
Customer service support and Community	4	3.5	3

These are my views on this comparison. Now, its your turn to share your views through the comments below.

If you like what you just read & want to continue your analytics learning, [subscribe to our emails](#), [follow us on twitter](#) or like our [facebook page](#).