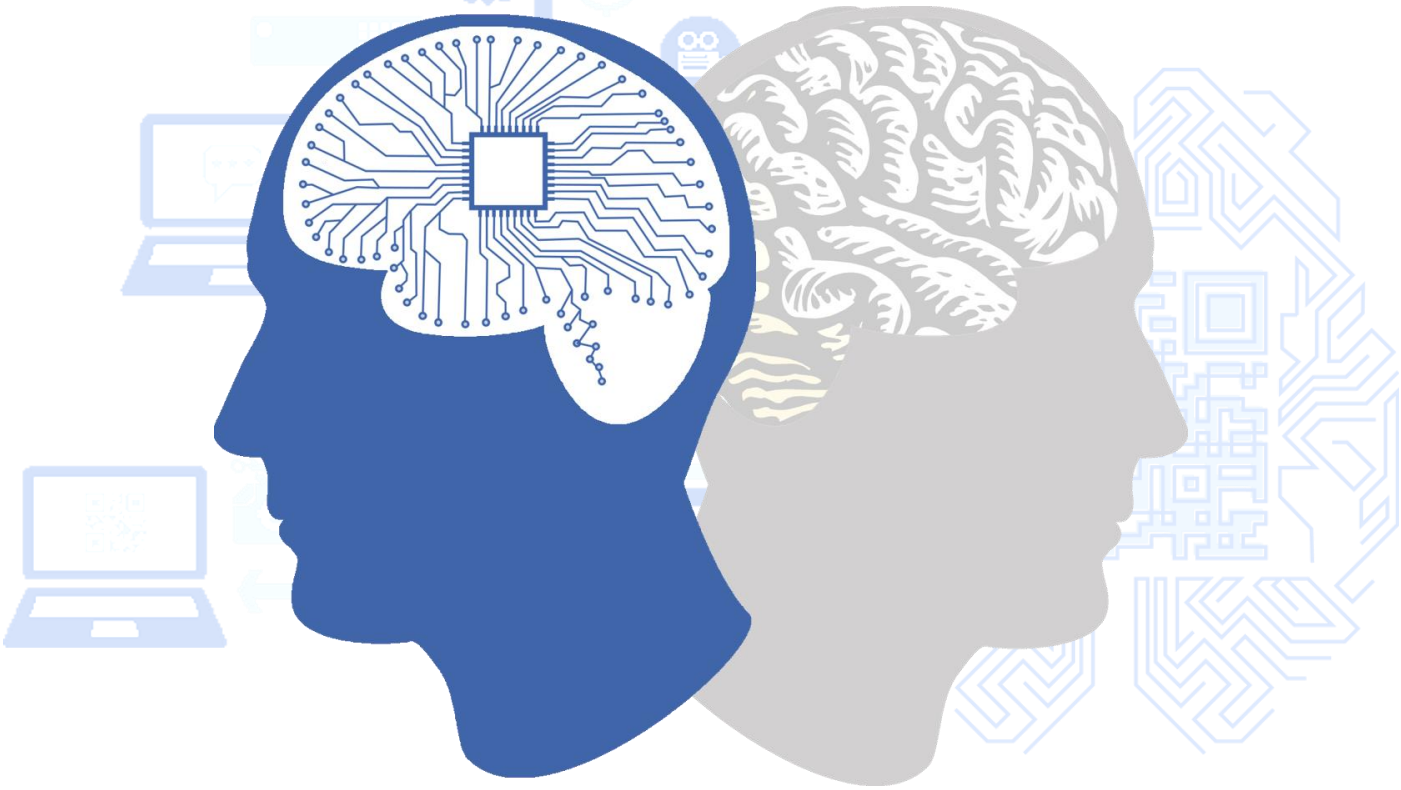
# DATA GEEK

## By BUSINESS ANALYTICS Specialization

Volume 1/Issue 2/ March 2019



## Analytical Tools: Some Interesting Insights

Institute Of Management,
CHRIST (Deemed to be University)

## From the Editor's Desk…

According to Reuters, the Business Analytics market is estimated to touch a whopping $71.1 billion in 2022. Analytical tools extend a very crucial contribution towards this. While the large businesses invest extensively on developing new tools and keep unleashing new dimensions, they serve as magic wands for the small businesses. After the overwhelming response for the first issue of DATA GEEK, we intend to bring about some interesting discussion on The Analytical Tools. This issue will throw light upon a few popular tools of Analytics and how businesses benefit from them.

The article 'A Deep Dive into Analytics Tools' is an introduction to the tools for Business Analytics, the issues and challenges related to the selection of right tools etc. The following article discusses the advantages of the powerful data visualization tool 'Tableau'. The popular tools 'R' and 'SAS' are the focus of discussion in the next two articles.

The interview with Mr. Ravi Kumar, Complete Analytics, the Crossword and the Glossary are the other attractions of the issue.

Congratulations to the entire crew of DATA GEEK for their sincere efforts to actualize this endeavour and wish this continue to impart knowledge and insights in the terrain of Business Analytics.

Please reach out to us for any queries or suggestions at datageek@mba.christuniversity.in

Happy Reading!

Prof. Reena Raj
Department of Business Analytics
Institute of Management
CHRIST (Deemed University)

# CONTENTS

# A DEEP DIVE INTO ANALYTICS TOOLS

"Without data you are just another person with an opinion"- Edward Deming. This statement may be of much relevance at present as we are muddled deep into the world of data. The scope and reach of data is blowing minds which is clearly illustrated through the fact that ninety percentage of the data has been generated in the diminutive and tiny duration of last two years. Moreover, this data has major importance to statisticians and analysts worldwide. Businesses have now realized how important it is to capitalize the data in hand in a correct manner through extracting the true sense of what each data depicts to solve the business problem. The period of 2016 to 2026 holds good promise for data analysts and other roles related because a research suggest that these jobs will show a positive growth of thirty-three percentage. Now, the billion dollar question is raised so as to why the data analysts are becoming the crusader of the movement led by the almighty king called data with weapons in the form of statistical tools. In the simplest terms, a data analyst takes crude information and transforms it into important bits of knowledge that businesses can use for progressively powerful decision-making. Doubtlessly, no data analyst can work in isolated and segregated silos. In order to fulfil the objective of deriving meaning from the data, the correct tools must be brought in to surf through the multi-fold dimensions of information.

The widespread usage and reach of the tools used by people working primarily in the field of data analysis have surpassed its limit and crossed the borders. The tumultuous change is spearheaded by the desire to grow in the momentum of functionality sans the help of a data scientist. The change is facilitated by Hadoop and NLP expanded as Natural Language Processing. These expansive changes have brought about another period of forefront business analytics tools. The might of these constructs and tools lies in the simplicity of usage, perceptive navigations and collection of assimilation points.

## Defining the Tools

The tools are solid examples of application software that brings and compiles data from different organizations into one repository. These organizations utilize more than one analytics tool to carry out the various statistical functions. These tools not only help to give the organization a better understanding to make smart decisions but also help in the conversion of end users and better facilitation of the project. And, this is just the tip of the iceberg.

Usually, usage of business analytics tools is helpful to gather and showcase the data but now the scope of it is shifting beyond just reporting. Now, the focus is assessing what went wrong, what was the reason of loss, how to predict uncertain events and foresee future results.

## Selecting the Right Tool

The success of getting the rift answer to any problem lies in how correct the methods are used to answer the question. Any company will be able to win the competitive rat

race to provide the right solution to the users through identification of need followed by the use of right tool. The different analytics tools made use in businesses should gather, review, and retrieve data. Yet, the sort of tools used decides the specific outcome of the analysis. Along these lines, a company ought to recognize their needs first to help them in getting the right data analytic tool to use. The following are variables to look at for while acquiring an analytic tool.

1. Collaboration: Most of the activities in business need collaboration among its stakeholders. And hence, an analytical tool must be efficient enough to offer a point that is without delay accessible on different types of platforms for all the activities.

2. Mining of Data: A good software tool must be having the capability to provide a deep analysis of the data, and thus give the results in a simpler way. One of the advantageous factor will be the less amount of time spend in comprehending the data.

3. Data Visualization: It deals with converting the data into a form which can be easily understood by everyone. This in turn facilitates in easier analysis and interpretation.

4. Predictive analysis: Any analytics tool requires the ability to predict the future results based on the limited present data. Moreover, these tools are expected to provide an optimal solution to address any business problem.

5. Cloud Connectivity: This is an icing on the cake for an analytic tool. It is mainly because it provides easier accessibility with any other device leading to better collaboration in business. Hence, the prediction is that by 2020, majority of the businesses will have adopted the cloud connectivity tool to improve their business.

Impact of Tools on Data Quality

With the passing time, enterprises have come to a realization that the data quality plays a major role. And thus evolved the concept of a new analytics tool. Just like any other process, in order to obtain a high level in terms of data quality, what matters is the order of operations. And the first step toward it is to analyse the data. The saying that you can't fix what you'd don't measure holds true in relation to data quality.


Data quality analytics tools are obtainable for a wide range of measurements. These tools fall into the following categories:

Duplicate data detection

One of the major issues in most business activities is the duplication of records. This particular problem is handled by this type of data analysis tool which assess the count on duplicate records which are at present occupying space inside the database.

Data Cleansing

Fixation and amendment of bad formats, incomplete and incorrect data from the data storages are the main object of the data cleaning analytics tools.

Data Monitoring

The impact of data to a conventional business rule is an important aspect to be taken care of. Data monitoring tools help us analyse and control this impact, which is a factor affecting the quality of data.

Data Enrichment

Enrichment of data helps us get over some serious issues like that of missing fields, outdated data and incorrect submissions. Using various data analytics practices, these kinds of tools can handle a lot beyond imagination.

Data Standardization

Getting all the data under a specific format is the task of a data standardization tool. This helps in a standardized way of dealing with different forms of data. The disintegration of text fields into components is also accomplished by this tool.

The Bright Future

The future of these tools is beyond doubt promising, considering the fact that the algorithms used are reliable in itself. Since the door towards data availability is wide open, these tools are definitely going to be a revolutionary change in the field of analytics. In just a few years, we will definitely be able to utilize these robust time-efficient tools to ease the analysis and visualization of huge data. After all, a tool is required by a carpenter to turn a pile of lumber into any desired outcome. Without those tools, the lumber will remain nothing but glorified firewood. Data analysts can be viewed in a similar sense. They will be unable to extract anything meaningful from the enormous data without the right tool in hand.

ALEENA CHACKO K

1828034

# ALL ABOUT TABLEAU

Business Analytics is all about analysing the data and making data driven decisions. What is the best way to analyse the data? We must know how the data behaves to understand and analyse it, which is necessary to make decisions that impact the business and eventually maximize revenue! The best way to understand the behaviour of data is to visualize it.

Why Tableau stands out?

The popularity of Tableau with organizations worldwide is because of the fact that it can extract huge amounts of data and present it in a format that is easy to understand and interpret. Also it provides over 40 data connectors such as MS Excel, Amazon Redshift, and Oracle DB to extract data. Tableau is not IT centric for designing visualizations i.e. it provides ease of functionalities with options like drag and drop.

## FORBES TOP 7 PAID FOR VISUALIZATION TOOLS

1. Microsoft Power BI
2. Tableau
3. QlikView
4. Sisense
5. Fusion Charts
6. Plot.ly
7. Carto

According to Forbes top 7 paid for visualization Solutions, Tableau stands on Number 2.

Products offered by Tableau

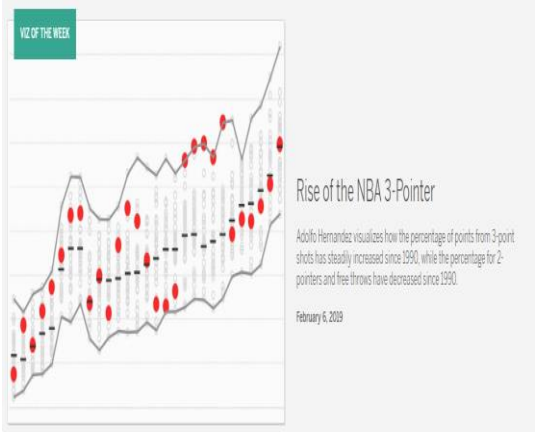
The Product Line includes different products based on requirement:

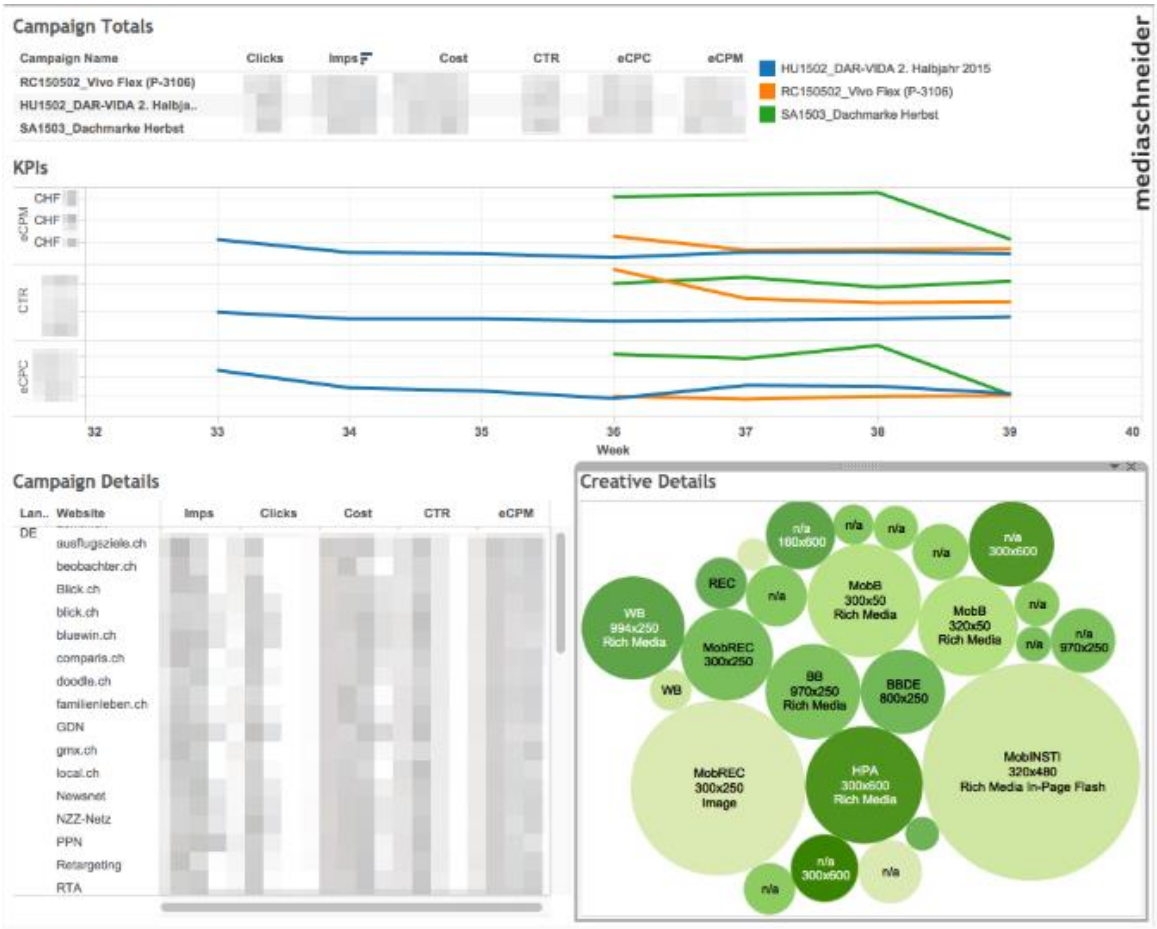
1. Tableau Desktop – Referred as “The Gold Standard” in Visual Analytics, a freeware for personal use.
2. Tableau Prep - Essentially designed for preparation of Data like Cleaning, combining, converting of Data for Analysis.
3. Tableau Online – Cloud based tool for visualization.
4. Tableau Server – Server provides collaboration within or different organizations to share the data and visualizations typically designed for Enterprises.

A Peek into Tableau:

Working on Tableau at an Organizational level involves – Gathering Business Requirements, Analysing the data needed and determining relationships between sources of data and extracting them, transformation of data: calculations, bins driven by requirements and finally creating visualizations that helps to solve business problems.

A few of the Data Visualizations from Tableau Public Gallery:




How Does Tableau Work?

The objective of the Tableau is to provide interactive data visualizations for large data sets very quickly. Tableau software is coded majorly on C++. All the libraries for visualization are in-built. While performing drag and drop on Tableau, in the backend it is converted to SQL (Structured Query Language) using VizQL technique, which is Tableau's proprietary programming language.

Tableau finds its application in every domain where data can be extracted. For marketing, it can be used for developing visualizations for Salesforce analysis, market segmentations, and product baskets, to create trend analysis, growth rate for finance, to analyse performance and rank candidates by proficiency in HR and so on. Along with the traditional business domains, Tableau is also used in many domains like HealthCare, Media and communication.

One such example is a campaign by an NGO –Path to eliminate Malaria in Zambia by 2021. The campaign was started in 2015, based on the historical data of Malaria cases like community, place, symptoms, a model was built on Tableau to tackle outbreaks and informed decisions were taken. With this advent, by 2018, Malaria cases were dropped by 93% in Zambia and the work still continues to achieve 100% drop.

## Companies Currently Using Tableau

Download CSV Sample (25 companies

| COMPANY NAME | SITE | HQ ADDRESS | CITY | STATE | ZIP | COUNTRY | TOP LEVEL INDUSTRY | SUB LEVEL INDUSTRY |
|---|---|---|---|---|---|---|---|---|
| Verizon | izon.com | 1095 Ave. of the Am... | New York | NY | 10036 | US | Telecommunications | Networking & Communication Products... |
| Honeywell | ieywell.com | 4D5, 115 Tabor Road | Morris Plains | NJ | 07950 | US | Retail | Consumer Electronics |
| WeWork | work.com | 115 W. 18th St | New York | NY | 10011 | US | Unclassified Establi... | Nonclassifiable Establishments |
| U.S. Bank | ank.com | 800 Nicollet Mall | Minneapolis | MN | 55402-7... | US | Finance | Banking |
| Conduent | iduent.com | 2828 N. Haskell Ave... | Dallas | TX | 75204 | US | Technical | Software Development & Technical Con... |
| IGT | com | 66 Seymour St | London | England | W1H 5BT | GB | Technical | Software Development & Technical Con... |

Most companies like Verizon, Honeywell, U.S. Bank uses Tableau to utilize their large datasets to create dashboards, build reports and henceforth resulting in informed decisions in a very less time.

According to Gartner's magic Quadrant, 2018 for Analytics and Intelligent Platforms, Tableau emerges as leader.


Drawbacks of Tableau

Although Tableau is user friendly, provides interactive visualizations, handles large amounts of data, there are few drawbacks. Since Tableau is designed for Data Visualization, there are very few options for data pre-processing and it can handle only structured data. Also visualizations cannot be customized in Tableau that is organization can use the data from dataset that fits the goals of business. Tableau is much expensive when compared to other BI tools like Qlikview, Power BI. (K.Gowthami, 2017)

NAMRATA MANGALGI

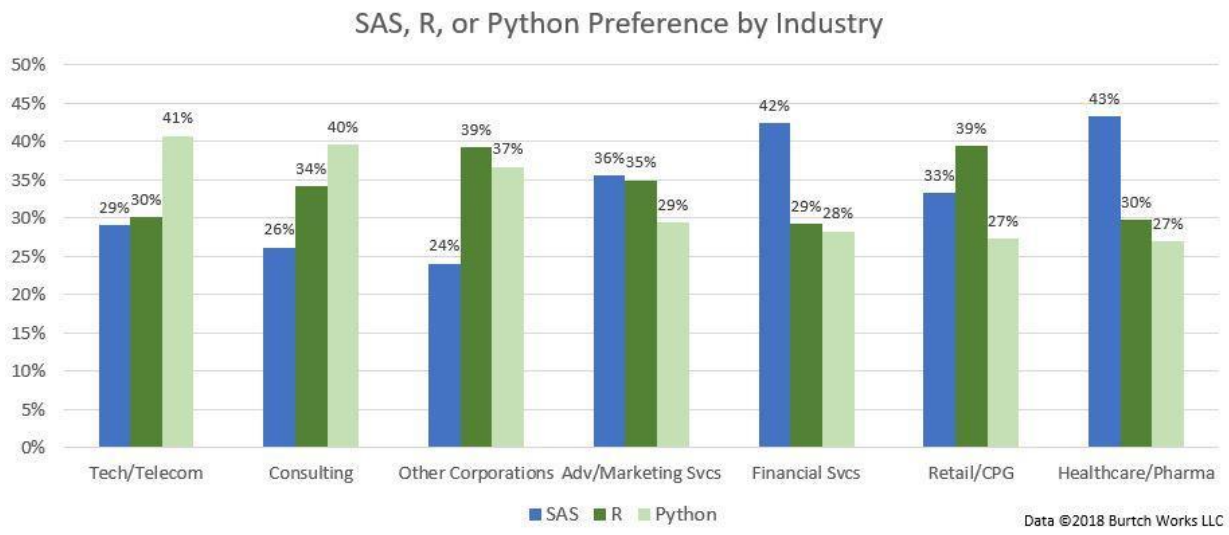
1827648

# R: THE "GOLDEN CHILD" OF DATA SCIENCE

In the year 1996, when Ross Ihaka and Robert Gentleman outlined their understanding in crafting and deploying the R software through their research paper, R was termed as a language for data analysis and visualization. Truly, to this date, the crux of R remains the same. However, it has de finitely outgrown the phrase. During the last decade, the R programming language has become one of the most widely used tools for statistics and data science and has over 3 million users worldwide. The number of R users has increased to 40% along with increased adoption of R in organizations for their day to day operations. Now the major reason for this shift can be attributed to the capability of the R software to handle a wide variety of tasks ranging from computational statistics, econometrics, and optimization to the hyped natural language processing (NLP) along with its gamut to perform data pre-processing, cleaning, web scraping and detailed visualization. Another, interesting aspect of R software that has attracted the business professionals towards it is its ability to generate business ready reports and machine learning powered web applications. This has given R a competitive edge over other tools. Besides these the various cutting edge algorithms available in R as well as successful community support namely, CRAN and R Community are yet other supporting factors for the adoption of R. While CRAN provides access to a store of more than 14000 packages R Community provides a cool and highly knowledgeable platform for learning and networking. Thus, flexibility, community and open source nature of R has led to its wide acceptance among the business professionals.

Comparison of R with Other Data Analytics Tools

Although several tools are available in the industry, according to the survey conducted by KDNuggets Analytics, Data Science, Machine Learning Software Poll, it is observed that Python and R are two trending software that has comparatively higher market share among the other tools. Again SAS is yet another tool that is still in the running. In order to understand the impact of these tools in the industry, the below-mentioned graph can be utilized. According to the ButchWorks Survey, it is observed that the usage and popularity of R



is much higher than SAS in all industries except the Healthcare/Pharmaceuticals industry. However, one cannot ignore the fact that Python is still leading in all industries whereas R is still in the growth phase.

In order to give a tough competition to Python, R has to work around 4 challenges that the researchers have identified, which are: the requirement to learn a new language, speed, memory management and efficiency. Once these areas are addressed R has the potential to overpower Python in the near future.

Future Scope of R Programming

R is one of the most popular and vibrant tools in the market and has indeed been a game changer in the analytics industry. In India, especially, large corporate companies like Genpact, Wipro and Accenture, MuSigma are looking at R and are encouraging their staff officials to build their expertise in R. Besides the corporates there is a huge demand for R among the start-ups as well as among various financial firms, banks and retail organizations. This indicates the increasing acceptance of R in the industry and the potential of new job openings for professionals among the organizations that have embedded or are in the process of embedding R into their system. Thus, the bottom line is that in this era where business analytics is ever growing R as a programming language is a “golden child” of data science that can uplift data exploration, clarification as well as presentation and getting up skilled in R will prove to give an edge to all the professionals across the industries.

AISHWARIA PAUL

1827834

# THE SASS ABOUT SAS

According to 2019 Gartner Magic Quadrant, SAS has placed itself as a leader for Data Science & Machine Learning Platforms. It has consequently sustained that position since last six years. Indeed, this is a moment of astonishment to know how one software has aligned that rank in a VUCA market. And that has been an eye opener for new learners as well as existing analysts to observe the paradigm shift occurring in the data world.

What Is SAS?

SAS (previously Statistical Analysis System) is an integrated software suite used for data manipulation. It was developed by the SAS Institute in the year 1966 which was further advanced in late 80s and 90s. It is easily accessible, manageable and data can be analysed from multiple sources. This product has good capabilities in data handling and options for parallel computations. SAS has an upper hand in deploying end to end infrastructure (Visual Analytics, Data quality and warehouse, reporting and analytics). It has been mitigated by integration and backing of R on platforms like SAP HANA and Tableau. The product offers a range of statistical functions spreading over descriptive measures like correlations, logistic regression to sophisticated methods such as modern model selection and Bayesian hierarchical.

Over the dialects of SAS are numerous brilliant nuggets in terms of functions, formats, and programming features holding on to inspire the analytics aspirants. A portion of the remarkable highlights of the software are mentioned in this article. SAS has been used to advance numerous fields and industries that include banking, communications, health care, travel and more.

Applications of SAS

One of the popular examples of SAS is the use case of Bank of America. The organization had a constant trouble with credit-card portfolio loss. The solution ahead was building a credit-risk model and forecasting it way ahead than a customer is involved into the system. However, to proceed with the solution, there was another hurdle of processing time reduction and availability for ad-hoc analysis. Indeed, with the implementation of a platform comprised of SAS Enterprise Risk Management on Grid Computing, IBM's XIV Storage System and SAS Scalable Performance Data Server on a 224 core IBM BladeCenter Grid, the processing time for ad hoc jobs has been reduced from a default time of 96 hours to mere 4 hours leaving the research investment as a profitable one.

Unlike any other tool, SAS does not demand for a prior knowledge in programming. It is not a typical programming and syntax-based language rather a procedure and function-based language. Based on this cause, SAS is easily deployed in the leading communication provider, Grameenphone of Bangladesh. While the issue was saturation of customer base due to competition, this business decided to do a churn-prevention analysis and micro campaigns based on the segments of customer base. Due to lack of quality resources, the company manifested upon going for a structured software suite that requires less of technical skills. As a result, they initiated with SAS Enterprise Miner and Forecast Server that generated impeccable outcomes. It has helped them to identify 10 broad segments constituting 27 million subscribers based on the usage patterns and demonstrated a positive impact on the marketing campaign.

By the same token, SAS is not an open source and has a diligent customer service along with the community. So, if a user encounters any technical challenge, their service is right at the door-step. For the similar reason, medical statisticians of Abbott Laboratories prefer SAS because if they encounter problems

due to software bugs and it comes to lawsuits, the organization recompenses the users which is definitely not the case in any other open-source software.

Likewise, not just firms, many government agencies are also relying on SAS. To name a few, Securities and Exchange Board of India (SEBI) utilizes SAS to detect and prevent fraud in India's securities markets. Similarly, Kansas Department of Wildlife and Parks, Wichita state, USA created an Analytics habitat by increasing the federal aid certifications, data forecasting along with generation of annual progress and management reports.

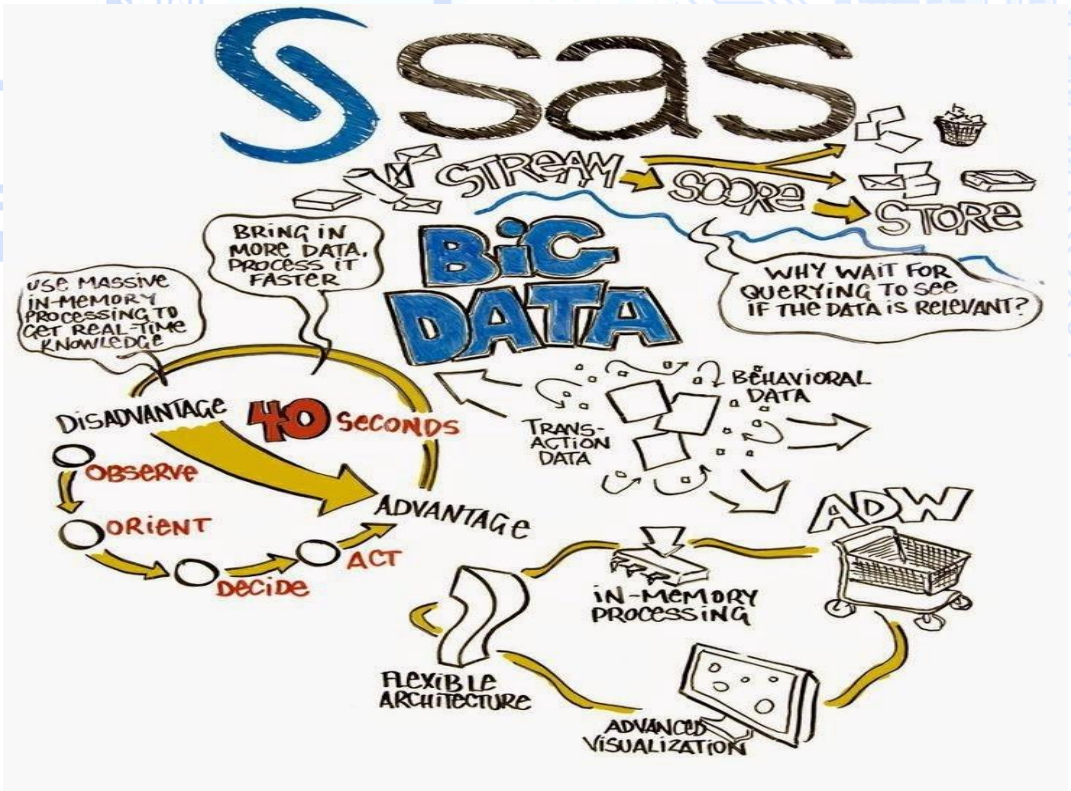
Regardless of all the success stories, there are some limitations of SAS that are overcome by R and Python.

Drawbacks of SAS

One of the notable limitations of this integrated software suite is Cost. SAS falls under a premium commercial software category putting itself as a staggering expense for a mid to small-scale organization. Though currently SAS has introduced a University Edition that is free to access but it has their own set of restrictions.

Apart from this, other disadvantages include lack of graphical representation and cumbersome text mining. In comparison with R and Python, SAS has a less stable GUI (Enterprise Guide & Miner) interface in its repository which hampers the exploratory factor of the tool. Also, every new package involves high price that brings down to the previous point of the software being expensive.

Despite the drawbacks, SAS, the 43-year-old company, has remarkable achievements in terms of deployment at 83,000 business, government and university sites across 147 countries till date. And with the onset of Artificial Intelligence, the software's core value proposition to provide enterprises with a centralized place to do data analytics along with a standardized workbench where all interested parties can access the same environment in a reliable as well as governed manner has become even more adamant which is visible in the leadership charts.




CHIMMANI SRIJA

1828629

# INDUSTRIAL INSIGHT

## Mr. Ravi Kumar, Complete Analytics

Mr. Ravi Kumar is a Data Science professional with more than ten years of experience building products using Machine Learning techniques to solve business questions around customer context building, recommendation systems, product data, predictive maintenance for factory IoT and Artificial Intelligence. Previously Mr. Ravi worked as a Sr. Data Analyst at Dell and now owns an educational institution named Complete Analytics. Mr. Ravi has vast experience with supervised and unsupervised Machine Learning techniques including Classification, Regression, Clustering and Deep Learning algorithms.

Complete Analytics

Complete Analytics is an educational institution where individuals are trained to think analytically, communicate efficiently, reason logically, make smarter choices and solve problems in a structured manner. Their training programs include SAS, SPSS, VBA, Advanced Excel, Predictive Modelling, Machine Learning and Deep Learning which empowers the analytical intelligence at every sphere of the career.


The following are some excerpts from the interview-

1. How progressive is the Analytics pitch in Indian business scenarios?

Currently, 40% of data science projects are outsourced in India. Not just that, most of the data science resource (in terms of knowledge) as well as start-ups are booming in India. Specifically, in Bengaluru, more than 45 data-science start-up companies have launched in 2018. Though there is a high demand in market, supply is inversely proportional. To be specific, talent gap is at the peak in the Indian business scenarios.

2. Could you throw some light on the projects that you have dealt with?

While I was working for Dell, we built predictive models to target the right customer at the right time for the right product to provide relevant discounts. We also developed a product for segmentation that artificially deploys data from large data sets and model validation using big-data on a prediction basis. Moreover, another prominent project I worked was in telecom industry where we did a churn analysis to find the root cause of customers leaving providers.

3. What are some of the major challenges faced by an Analytics professional?

One of the major difficulties faced by an Analytics professional is Data itself. In data, there is always a fear of Garbage in Garbage out (GIGO). Also, another challenge I have come across frequently is the large amount of data availability and its storage. Notable issue is that we are unable to capture the correct information from such huge volumes.

4. What are the popular tools used in business analytics?

Business analytics is a vast area to talk about. Primarily, we can classify the overall tools into two categories - Reporting tools and data science tools. Top reporting tools on charts are Tableau, PowerBI, MS

Excel, and VBA. Whereas data science tools can again be split into two divisions. First division is tools where most popular ones are R, Python, SAS, SPSS, Julia and the second one is techniques. Not just tools, storytelling aka output interpretation to the client also plays a prominent role.

5. Which tools conforming to you, are required to ensure the data quality?

Data quality is predominantly ensured by domain knowledge and business understanding. However, tools like Excel, R, process flow, quality control charts, and six sigma ads as a medium to channelize the data.

6. As per KDNuggets 2018 report, Python has overtaken R and placed itself as a top analytics tool. What could have been the reason and according to you, which amongst these is more dynamic?

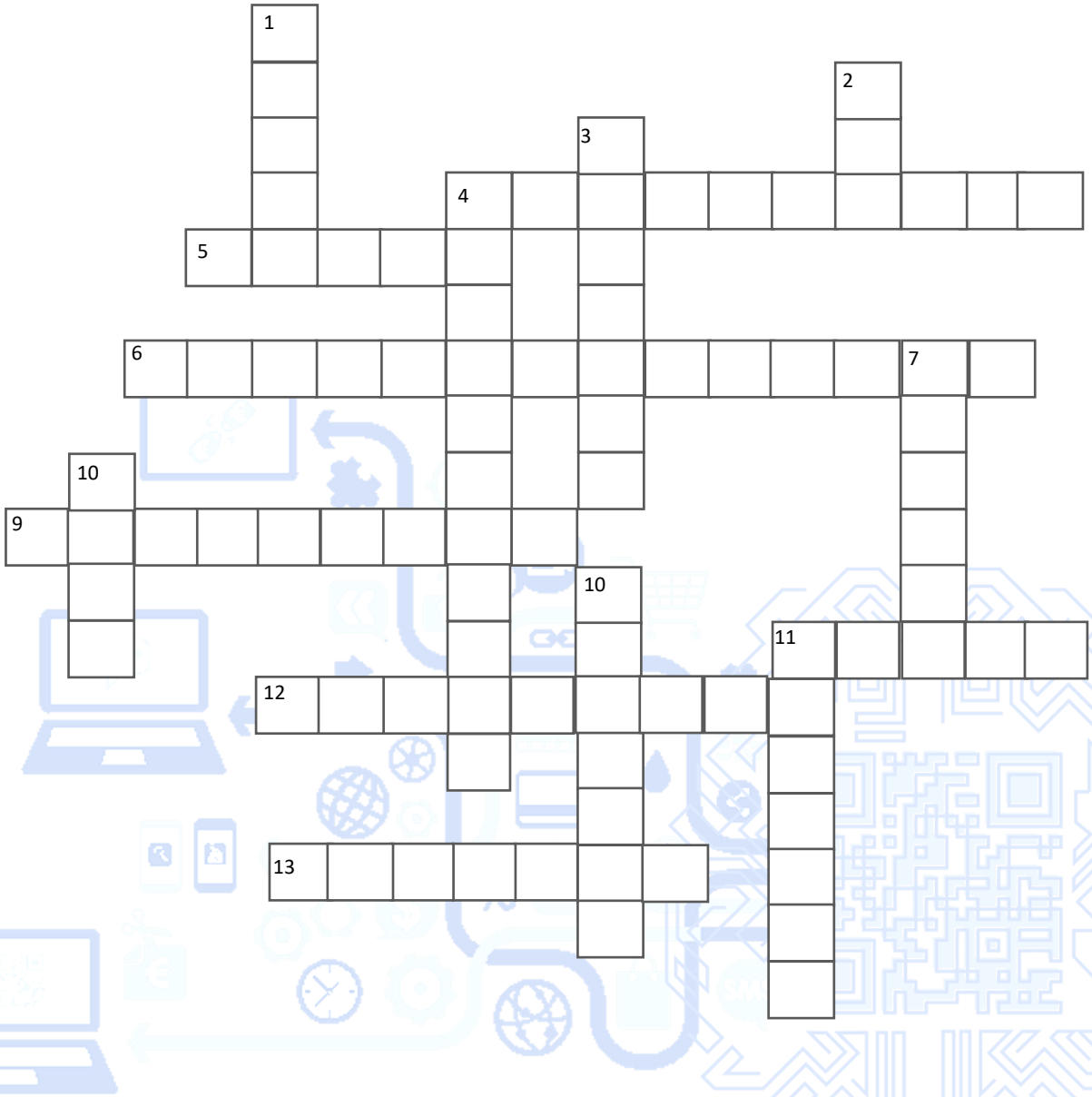
About four years ago, SAS was the industry leader. Until last two years, R was the leading software in the market and now Python has taken the position. According to me, Python is a frequently used language in Data Science domain. It is also great at handling unstructured data. As in most of the software, algorithm building in the backend uses another language but Python has revolutionized in such scenario. You can observe that all the packages in Python are in the same language. From development, production to deployment, everything can be done in Python with the concept of OOP.

7. Two years down the line, which software/tool do you think will most prominently be used by a data analyst?

Working on the unstructured data is going to resonate in the near future. Thereupon Julia, Pig, Hive, Parallel Processing, R and Python are probably going to lead the Data Science industry.

By large and sum, healthcare, banking and insurance domains are shifting its business towards Analytics in India in the coming years. Notably, Mr. Ravi Kumar has left us, the analytics aspirants, with an amazing piece of advice to explore the Data like how a kid pushed in lake automatically adapts and learns to swim, similarly an analytics aspirant should get started with any dataset and start exploring the horizons.

# CROSS WORD




Across

4. Digging Information.
5. Amazon's virtual personal assistant.
6. A Cloud concept related to pooling and sharing of resources.
9. Diagram of entities and relationships.
11. The method used by Facebook to tackle big data.
12. Basic building block of information.
13. Computer program or an artificial intelligence which conducts a conversation via auditory or textual methods.

Down

1. An Open source DBMS.
2. A cloud that manages the security of data.
3. computing that allows to pay per use.
4. Organizations appoint this individual to be responsible for data management and data quality from a business perspective.
7. A non-profit artificial intelligence (AI) research organization that aims to promote and develop friendly AI.
8. A distribution model in which applications are hosted by a service provider and made available to users.
10. The organization which has the world's largest Hadoop cluster.
11. Overriding unnecessary and irrelevant consideration in AI system.

SOLUTIONS:

Across:

4. DATAMINING
5. ALEXA
6. VIRTUALIZATION
9. DATAMODEL
11. PRISM
12. CHARACTER
13. CHATBOT

Down:

1. MYSQL
2. VPN
3. UTILITY
4. DATASTEWARD
7. OPENAI
8. SAAS
10. FACEBOOK
11. PRUNING

# JOURNOs

ALEENA CHACKO K

1828034

NAMRATA MANGALGI

1827648

AISHWARIA PAUL

1827834

CHIMMANI SRIJA

1828629

# DESIGN & CREATIVITY

MRUTHUNJAY ANKALI

1827520