

A Review Paper on Why Data Science isn't Producing Desired Results and How Can This Be Fixed

Dr. Mehul P. Barot
Assistant Professor,
Computer Department,
LDRP – ITR

Sanjeev Khatwani
Student 7th Sem,
I.T. Department,
LDRP-ITR

Keval Desai
Student 7th Sem,
I.T. Department,
LDRP-ITR

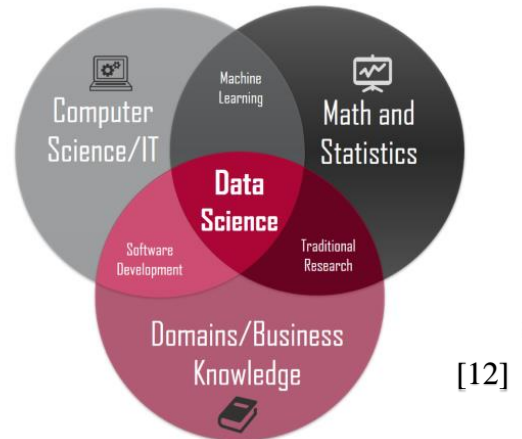
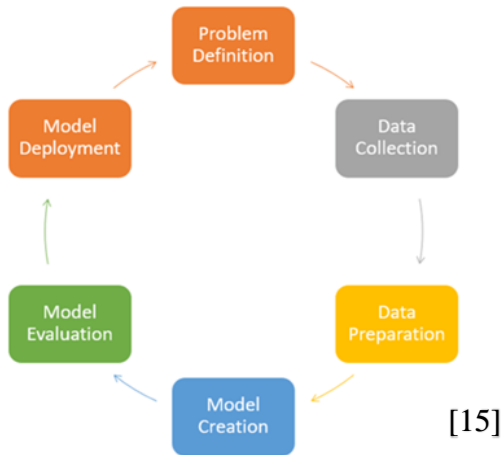
Abstract: A Data Scientist is someone who finds solutions to problems by analysing big or small data using appropriate tools and then tells stories to communicate his/her findings to the relevant stakeholders. Data Science is what a data scientist does. The goal of Data Science is to extract useful values, suggest conclusions and/or support decision making. Thus, Data science is being used in almost every department like Marketing, Finance, Human Resource, and IT. Firms like Google, eBay, LinkedIn, and Facebook were built around data from the beginning and are doing exceedingly well but many companies even after spending huge sums on Data Scientists are not able to achieve desired results. Through this article, we are going to discuss the problems being faced by such companies and how to fix them.

Key words: Data Science, Data Wrangling, Data Analysis, Storytelling, Statistics, Big Data.

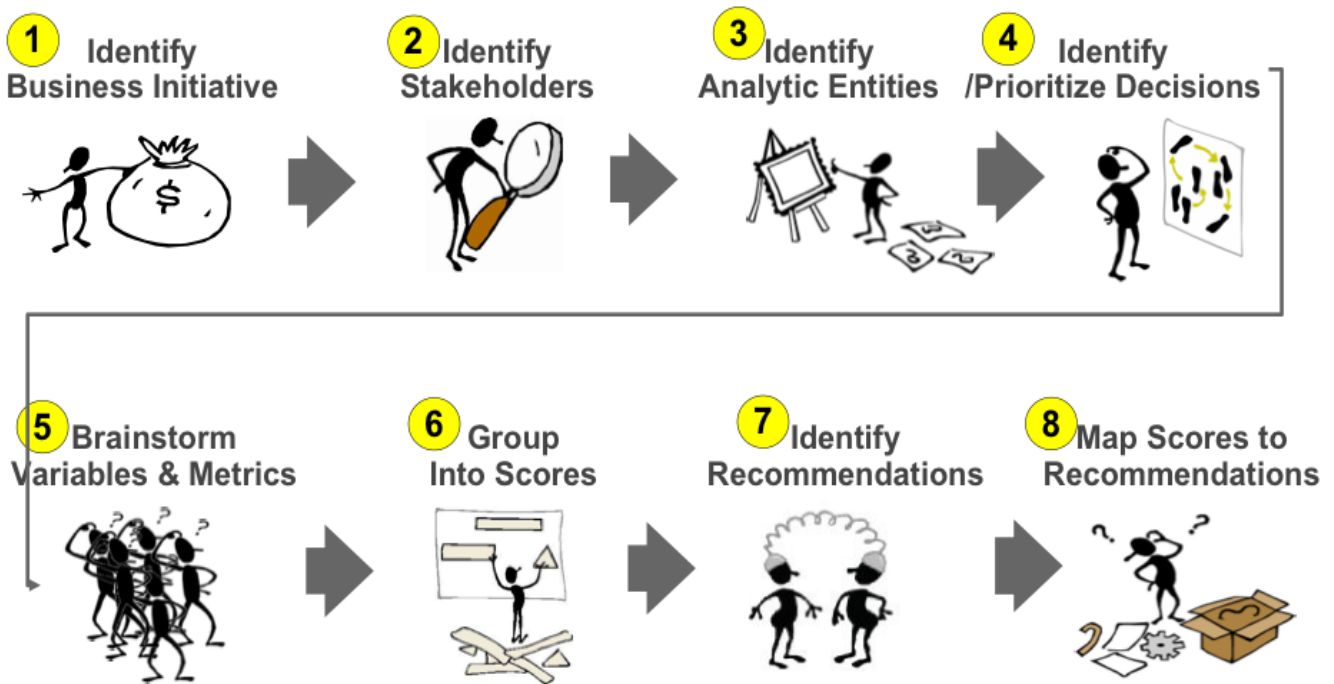
1. INTRODUCTION:

Data Science is a field that comprises of everything related to data cleansing, preparation, and analysis. Data Science is the combination of statistics, mathematics, programming, problem-solving, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing and aligning the data. In simple terms, it is the umbrella of techniques used when trying to extract insights and information from data [10]. Chief Data Scientist of the United States (2015-2017) DR. Patil told the Guardian newspaper in 2012 that a “data scientist is that unique blend of skills that can both unlock the insights of data and tell a fantastic story via the data [13].” Data science researchers utilize the capacity to discover and translate rich information sources; oversee a lot of information notwithstanding equipment, programming, and transfer speed imperatives; consolidate information sources; guarantee consistency of datasets; make representations to help in comprehension of information; construct scientific models utilizing the information; and display and impart the information experiences/discoveries [5]. The Steps involved in the process of Data Science are 1) Obtain Data 2) Scrub (Filter) Data 3) Explore Data 4) Model Data 5) Interpret Data [6]. Data science and analytics have taken the world by storm. Newspapers and television broadcasts are filled with praise of data and analytics. Business executives and government

leaders do not get tired of praising data and how they have embraced it to improve their respective bottom lines. The hype around data has created the demand for individuals skilled in analytics and data. Stories of six-figure starting salaries for data scientists are feeding the craze. By the time you are halfway through this paragraph, 2.5 million Facebook users would have exchanged contents online. Google would have received more than 4 million search requests. More than 200 million email messages would have flown over the Internet and some 275,000 tweets would have been heard. Never before in the history of humankind have, we been able to generate a living history of ourselves. In the process, we are creating new data of immense size and scope. It is indeed a transformative change to see that within a few decades we have moved from complaining about the lack of data to a data deluge. This makes data analytics even more exciting and indispensable [7]. A data science researcher needs to realize what will be the yield of the data science transform and have an unmistakable vision of this yield. A data science researcher needs to have a plainly characterized arrangement on in what manner this yield will be accomplished inside of the limitations of accessible assets and time. A data scientist needs to profoundly comprehend who the individuals are that will be included in making the yield. For a layman data science is mainly: collection and *preparation* of the data, alternating between running the *analysis* and *reflection* to interpret the outputs, and finally *dissemination*(*spreading*) of results in the form of written reports and/or executable code [5]. The Scope of Data Science and Big data analytics is going to increase further coming years. The future for Innovation and business can see as big data. Data science will come with great usability and development in the upcoming years. Digital experiences will completely relate to Human experiences. This type of trends is expected to get exposure and updating beyond measurements. One type of **big data analytics** will target on updating business operations. For the most part, they will be in understanding the total business workflows and moving Data to companies. Finally, all the above concepts will explain What is the Scope of Data Science in 2019. And **big data analytics salary in the USA** will be more in comparison with other technologies [8].



“Thinking Like A Data Scientist” Process



Source: Bill Schmarzo “Big Data MBA” Course Curriculum

2. THE PROBLEM:

In spite of The Companies hiring the best of data scientists very few have been able to unleash the true power of Data science. For an analytics project to create value, the team must first ask smart questions, wrangle the relevant data, and uncover insights. Second, it must figure out—and communicate—what those insights mean for the business. The ability to do both is extremely rare—and most data scientists are trained to do the first, not the second [1]. Becoming “data-driven” has been a commonly professed objective for many firms over the past decade or so. Whether their larger goal is to achieve digital transformation, “compete on analytics,” or become “AI-first,” embracing and successfully managing data in all its forms is an essential prerequisite. Consistent with these goals, companies have attempted to treat data as an important asset, evolve their

cultures in a more data-oriented direction, and adjust their strategies to emphasize data and analytics.

We knew that progress toward these data-oriented goals was painfully slow, but the situation now appears worse. Leading corporations seem to be failing in their efforts to become data-driven. This is a central and alarming finding of NewVantage Partners’ 2019 Big Data and AI Executive Survey, published earlier this month. The survey participants comprised 64 c-level technology and business executives representing very large corporations such as American Express, Ford Motor, General Electric, General Motors, and Johnson & Johnson.

Here are some of the alarming results from the survey:
 72% of survey participants report that they have yet to forge a data culture
 69% report that they have not created a data-driven organization

53% state that they are not yet treating data as a business asset
52% admit that they are not competing on data and analytics.
[3],[9]

Some of the reasons could be:

- I. You don't identify the problem you are trying to solve.
- II. You don't use the right metrics to gather insights.
- III. You don't have the right data and systems.
- IV. You don't have the right culture.
- V. You don't have the right people.

[4]

The above mentioned are the trivial reasons but one of the advanced reasons could be Data Quality:

Most managers know, anecdotally at least, that poor quality data is troublesome. Bad data wastes time, increases costs, weakens decision making, angers customers and makes it more difficult to execute any sort of data strategy. Indeed, data has a credibility problem [2].

In a question on Kaggle's 2017 survey of data scientists, to which more than 7,000 people responded, four of the top seven "barriers faced at work" were related to last-mile issues, not technical ones: "lack of management/financial support," "lack of clear questions to answer," "results not used by decision-makers," and "explaining data science to others." Those results are consistent with what the data scientist Hugo Bowne-Anderson found interviewing 35 data scientists for his podcast; as he wrote in a 2018 HBR.org article, "The vast majority of my guests tell [me] that the key skills for data scientists are....the abilities to learn on the fly and

to communicate well to answer business questions, explaining complex results to nontechnical stakeholders [11]."

But in the rush to grab in-demand data scientists, organizations have been hiring the most technically oriented people they can find, ignoring their ability or desire (or lack thereof) to communicate with a lay audience.

That would be fine if those organizations also hired other people to close the gap—but they don't. They still expect data scientists to wrangle data, analyze it in the context of knowing the business and its strategy, make charts, and present them to a lay audience. That's unreasonable, which takes me to the next and the most important reason:

P.T.O

"Explaining Data Science to Others."

Gaps between business and technology types aren't new, but this divide runs deeper. Consider that 105 years ago, before coding and computers, Willard Brinton began his landmark book *Graphic Methods for Presenting Facts* by describing the last-mile problem: "Time after time it happens that some ignorant or presumptuous member of a committee or a board of directors will upset the carefully-thought-out plan of a man who knows the facts, simply because the man with the facts cannot present his facts readily enough to overcome the

opposition....As the cathedral is to its foundation so is an effective presentation of facts to the data [14]".

A rare combination of skills for the most sought-after jobs means that many organizations will be unable to recruit the talent they need. They will have to look for another way to succeed. The best way is to change the skill set they expect data scientists to have and rebuild teams with a combination of talents [1].

3. THE SOLUTION:

A good data science team needs six talents: project management, data wrangling, data analysis, subject expertise, design, and storytelling. The right mix will deliver on the promise of a company's analytics [1].

Define talents, not team members. It might seem natural that the first step toward dismantling unicorn thinking is to assign various people to the roles the "perfect" data scientist now fills: data manipulator, data analyst, designer, and communicator [1].

Project management: Because your team is going to be agile and will shift according to the type of project and how far along it is, strong PM employing some scrum like methodology will run under every facet of the operation [1]. Here a person with good communication skills and leadership skills is required who can bind the team together and give them directions to work in. Most important he should help maintain the synergy and coordination of the team.

Data wrangling. Skills that compose this talent include building systems; finding, cleaning, and structuring data; and creating and maintaining algorithms and other statistical engines [1]. This person should have in-depth knowledge of statistics and its various methods and various latest technologies like Artificial Intelligence, Machine Learning, and Deep Learning Algorithms.

Data analysis. The ability to set hypotheses and test them, find meaning in data, and apply that to a specific business context is crucial—and, surprisingly, not as well represented in many data science operations as one might think. Some organizations are heavy on wranglers and rely on them to do the analysis as well. But good data analysis is separate from coding and math. Often this talent emerges not from computer science but from the liberal arts [1]. This is so because generally tech-savvy people aren't that good at explaining logic to the stakeholders and the liberal arts people bridge that gap.

Subject expertise. It's time to retire the trope that data science teams are stuck in the basement to do their arcane work and surface only when the business needs something from them. Data science shouldn't be thought of as a service unit; it should have management talent on the team [1]. These people have created a separate sub-section of Data Science Known as **Business Analytics** because they are able

to infer what results will be shown on the product due to the use of data science.

Design: This talent is widely misunderstood. Good design isn't just choosing colours and fonts or coming up with an aesthetic for charts. That's styling—part of design [1], but by no means the most important part because this 'Design' has more to do with how to present the data so that the explanation to the layman becomes easy via different types of visualizations.

Storytelling. Narrative is an extremely powerful human contrivance and one of the most underutilized in data science. The ability to present data insights as a story will, more than anything else, help close the communication gap between algorithms and executives [1]. This is the problem which helps us in getting out of Brinton's 'Last-Mile' Problem that is making a layman believe in our plan.

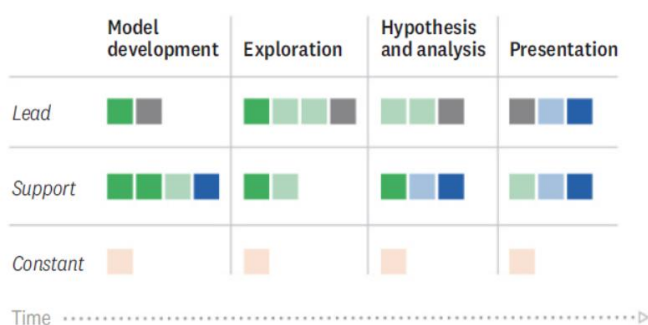
4. DISCUSSION:

Even after understanding the problem and the solution, there's a chance that an organization commits a blunder. This blunder is trying to find all these qualities/talents in a single person because most probably the people you find will have three or at max four out of the above-mentioned talents. So, our approach suggests rather than trying to find a person who has the above-mentioned talents and is almost impossible to find, we build a team comprising of all the talents according to a talent dashboard which will make life much easier for the HR department of the company. The Pictures below explain how and what one should include in the Talent Dashboard and how can it be used efficiently. This will also help in determining which project will be done more efficiently given the talents you have at hand and how much role will a particular person play in a project so that he/she can work on more than projects parallelly which will again increase the efficiency of the organization.

Put It to Use

Knowing what talents are available, managers can now assign units of talent to a project according to when it's needed. Typically, one cluster of talents will take the lead early in a project, and different clusters will do so in the later stages. Project management usually plays a big role throughout.

Upcoming project plan



[1]

5. CONCLUSION:

- I. **Assign a single, empowered stakeholder:** It's possible, or even likely, that not all the people whose talents you need will report to the data science team manager. Design talent may report to marketing; subject-matter experts may be executives reporting to the CEO.[1]
- II. **Assign leading talent and support talent:** Who leads and who supports will depend on what kind of project it is and what phase it's in.? For example, in a deeply exploratory project, in which large volumes of data are being processed and visualized just to find patterns, data wrangling and analysis take the lead, with support from subject expertise; design talent may not participate at all since no external communication is required.[1]
- III. **Co-locate:** Have all team members work in the same physical space during a project. Also, set up a shared virtual space for communication and collaboration. It would be undesirable to have those with design and storytelling talent using a Slack channel while the tech team is using GitHub and the business experts are collaborating over e-mail.[1]
- IV. **Make it a real team:** The crucial conceit in colocation is that it's one empowered team. The collaboration should be immaculate rather than being superficial. Things like Regular feedback can be of a lot of help for, say, a data person who needs help with storytelling, or a subject expert who needs to understand some statistical principle.[1]
- V. **Reuse and template:** Think of this as a group of people who combine their design talents and data wrangling talents to create reusable code sets for producing good data visualizations for the project teams. Such templates are invaluable for getting a team operating efficiently.[1]

THE PRESENTATION OF data science to lay audiences—the last mile—hasn't evolved as rapidly or as fully as the science's technical part. It must catch up, and that means rethinking how data science teams are put together, how they're managed, and who's involved at every point in the process, from the first data stream to the final chart shown to the board. Until companies can successfully traverse that last mile, data science teams will underdeliver. They will provide, in Willard Brinton's words, foundations without cathedrals.[1]

6. CONFLICTS OF INTEREST:

"The authors declare no conflict of interest."

7. REFERENCES:

- [1] Data Science and the Art of Persuasion - Harvard Business Review by Scott Brinatto, a senior editor at Harvard Business Review.
- [2] Only 3% of Companies' Data Meets Basic Quality Standards Tadhg Nagle, Thomas C. Redman, David Sammon.
- [3] Companies Are Failing in Their Efforts to Become Data-Driven by Randy Bean, Thomas H. Davenport.
- [4] 5 Reasons Why Your Company's Analytics Program Is Failing B Y Bill Petti and Sean Williams.

- [5] Challenges in Data Science: A Comprehensive Study on Application and Future Trends by Proyag Pal, Triparna Mukherjee, DR. Ashoke Nath.
- [6] 5-steps-of-a-data-science-project-lifecycle by DR. Cher Han Lau.
- [7] <http://aci.info/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic>.
- [8] <https://onlineitguru.com/blog/what-is-the-scope-of-data-science-in-2019>.
- [9] NVP Big Data and AI Executive Survey 2019 Executive Summary of Findings.
- [10] www.simplilearn.com
- [11] www.kaggle.com
- [12] [www.https://towardsdatascience.com](https://towardsdatascience.com)
- [13] Getting Started with Data Science by Murtaza Haider
- [14] Graphic Methods for Presenting Facts by Willard Brinton
- [15] www.mssqltips.com