
Riyadh Radhi

Iraqi Data Analysis to Action: The Journey Towards Data-Driven Community

Data Science's Current Situation

To have a clear perception of the current state of data science in general, we need to go back a few years in time. It has been almost ten years since Thomas Davenport and DJ Patil published their article in the Harvard Business Review describing data scientists as "the sexiest job of the 21st century".

The generous description was mainly due to two reasons; the vast amounts of data volume to harvest and the shortage of enough skilled people to tame these data for proper insights at that time.

The last two decades have witnessed an explosion in data volume across the globe, leading to what is called by some the data revolution. According to Domo, a business intelligence specialized company, people are generating 2.5 quintillions daily (quintillion has 18 zeros after one) which translates to around 143.43 GB per day for each one of us.

These estimates are expected to witness further rapid growth in the future. In fact, former Google's CEO, Eric Schmidt, stated, "There were 5 Exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days".

While I do not advise trusting the mentioned statistics blindly, what I can say for sure is that data is insanely expanding. This fact, coupled with the technological advancements in computer hardware, has led to the creation of the modernized version of data science and artificial intelligence.

The second reason that made data science very appealing is the talent shortage in a high-demand field. In fact, the over-excitement around getting this large amount of data into use has caused a gap between the organizations and the data community.

In an attempt not to be left out, organizations started competing in hiring "data scientists" with little knowledge of the added value that data science can bring. Similarly, the data science fuss hindered the growth of human capital too. After gaining a few certificates, many people started labeling themselves as "data scientists" which made it a burden to spot qualified talents from the crowd. These facts naturally made the demand for data scientists larger than the number of qualified candidates in the market.

You might wonder where Iraq stands in all of this, and I would like to take a more data-oriented approach to put this topic into perspective. According to Glassdoor, working as a data scientist was the best job for four consecutive years from 2016 to 2019 and is the second-best job of 2021 in America. Additionally, LinkedIn ranked data science as the fastest-growing job in 2007 with a growth rate of 650% since 2012.

Meanwhile, in Iraq, there was only one job opening of a data science position found in the Iraqi job matching platforms compared to 5,971 job openings found in Glassdoor this year in America. Additionally, according to research

published by KAPITA on the reality of information and communication technology in Iraq, the Iraqis ranked their data science skills as the lowest compared to the other digital skills they have.

The Perk of Joining Late is That We Can Join Right

Many Iraqi organizations have not started investing in data science yet, but this could be good news for us. In recent years, many international companies have tried to race for the leading roles in data science. While being the first is considered an advantage, it also involves higher risks of making mistakes.

As a result, many firms failed over and over until they made a success story. The good news is that we can learn from these mistakes while going through the same path ourselves in Iraq.

The following recommendations are geared towards organizations and individuals separately and based on the recent global insights about data science.

Organizations are recommended to:

1. **Invest in Low-hanging Fruit:** The world started realizing that the bubble created by the luxuries terms in data science is starting to burst. Organizations should know that they do not have to use deep learning or artificial intelligence to harness the power of data.

In fact, many of the market questions and problems can be addressed with simple visualizations, descriptive analysis, and automation rather than fancy mathematical models for which we barely have enough data. Therefore, instead of doing bootcamps and training about artificial intelligence and machine learning, we should start teaching individuals and organizations about ways with low entry barriers and high return value.

2. **Generate Questions First:** More thoughts and attention should be given to creating questions before taking any other step. Often, organizations make the mistake of allocating expensive resources to answering questions that yield very little value.

Instead of hiring machine learning engineers or statisticians that can take months and a fortune to answer a few questions, hire fast analysts to generate questions at first and then invest time in answering the ones related to your objectives.

3. **Define Your Organization's Needs:** Most of us often hear about a talent shortage in the market generally and data science particularly. However, when it comes to data science, a clear understanding of the wide range of available skills and the definition of precise organizational needs will make a faster and more efficient talent match process.

This is especially true for data science as the field is rapidly changing with new job titles, different technologies and algorithms, and continuous emerging data types.

4. **Do not Look for Unicorn Candidates:** A common practice for inexperienced organizations when hiring data scientists is searching for an "all-in-one" talent to save money.

This mission of finding unicorn candidates does not usually yield successful stories for data science teams. Instead, organizations should encourage interdisciplinary teams with diverse talents and mixed domain knowledge to build a coherent data science team.

Similarly, **individuals** are recommended to have the following in mind:

1. **Jack of All "Tools", Master of None:** Individuals who are interested in learning data science should know that this field is a never-ending ocean. There are plenty of tools to extract insights from data, many data science fields to focus on, and various domains to be involved in. This fact is causing a lot of distractions, scattered efforts, and stress in learning as many skills as possible to compete in the job market.

Nevertheless, I am here to tell you that scratching the surface of many tools without mastering any is counterproductive for your career and well-being too! Instead, individuals should find what they love in data science and dive deep into it, even if that means not learning the most hyped technology out there. There is no predefined set of tools to analyze data.

The tool you choose could involve anything from coding (ex: R, Python, Julia), drag and drop applications (ex: Power BI, Tableau), or could be designed for specific tasks or concepts in mind.

The main principle is to choose the tool that answers your questions, can get your job done efficiently, and makes you happy while using it.

- 2. Some Managers Need to Catch Up With the Game:** The gap between data teams with technical backgrounds and managers has immensely grown.

Some individuals in managerial positions lack a fundamental understanding of the working nature of data scientists and analysts.

This results in unrealistic demands, setting faulty timelines, and making wrong decisions when dealing with their data teams.

Therefore, it is time that managers start to be familiar with the values data teams can provide instead of undermining their efforts with confusing tasks leading to ineffective business decisions and team failures.

- 3. How Much Theory is Enough:** Often in data science, we see the differences between academia and industry are being confused.

There is a lot of pressure on data scientists to understand the theory behind each mathematical model that exists.

In a sense, this is true to provide correct insights. Yet, individuals and organizations should know when to draw the line between

academic knowledge and practical analysis tasks. It is better if decision-makers put less emphasis on pure theoretical knowledge, which is rarely needed in the day-to-day job, and focus more on grasping the practical side.

For example, an individual with technical skills and domain expertise is probably a better investment, in most times, for industrial organizations than an individual who knows the math behind convolutional neural networks but has no clue about the sector your organization operates in.

Having said this, I am not against theoretical knowledge, and this is not an invitation to neglect that aspect of data science. However, it is a piece of advice to focus on the relevant theoretical topics that are foundational for the technical.

The Next Step: iDATA

The extensive need for a data-driven community combined with the lack of expertise in this field is a recurrent challenge in Iraq. Hence, carefully planned actions are needed, and we believe that the iDATA initiative is one of these actions.

What is iDATA

iDATA, short for Iraqi Data Analysis To Action, is a non-profit initiative dedicated to activating the role of data and creating a data enthusiast community in Iraq through three main parts; data science training and education, implementation of data-driven projects, and demonstration and communication of the results to the public audience.

We expect that a large number of parties will benefit from and contribute to iDATA efforts and programs to change the data science scene in Iraq, yet we see the following groups as the key target:

1. Individuals with interests in starting their journey in data analysis.
2. Data analysts who wish to be more involved in the field and learn more advanced skills.

IRAQI DATA ANALYSIS TO ACTION: THE JOURNEY TOWARDS DATA-DRIVEN COMMUNITY



3. International NGOs and donors who aim to create job opportunities and support youth skills.
4. Institutions and organizations that wish to establish in-house or outsource data science teams.

Keeping the aforementioned points of the article in mind, iDATA aims to introduce a unique and healthy style in spreading data science across Iraq. The main goals of iDATA are to:

1. Support individuals by providing data analysis training through a series of workshops, bootcamps, and meetups.
2. Improve and boost individuals with existing data science skills through applied projects, competitions, and hackathons.

3. Offer help in bridging the community members with relevant jobs in the private sector through sharing job opportunities and internships.
4. Build analytical mindsets and raise awareness about the importance of data in decision-making.
5. Contribute to society by communicating insights on a broad range of social and environmental themes through the community members' projects.

Setting these goals, iDATA's mission is to create a data-driven community by providing a supportive space for Iraqis to grow through education, hands-on projects, and networking opportunities. Additionally, iDATA envision making data science simple, accessible, and beneficial across Iraq.