Is Data Science for All?  
Data Science Bootcamp



**Anish Mahapatra**

Contents

[Introduction 3](#_Toc105635751)

[Is Data Science for you? 3](#_Toc105635752)

[The intuition of Data in Data Science 3](#_Toc105635753)

[The roles that come under Data Science 4](#_Toc105635754)

[What role should you pick? 7](#_Toc105635755)

[How long does it take to become a Data Scientist? 8](#_Toc105635756)

[Mandatory skills you need to become a Data Scientist 9](#_Toc105635757)

[Interpersonal Skills 9](#_Toc105635758)

[Technical Skills 10](#_Toc105635759)

[Data Science Resources 10](#_Toc105635760)

[Conclusion 11](#_Toc105635761)

[Author 11](#_Toc105635762)

# Introduction

Data Science seems to be the buzzword for the decade. The internet is talking about it, companies want it, and your friends and even your family is in the loop about how Data Science is changing the world we live in. Your dad is noticing how Google’s news feed algorithm is getting updated based on the conversations that are happening around you, and your mom is enjoying food videos on YouTube based on recommendations.

Let’s now come to you. Irrespective of the position, industry and level of technical knowledge, you would like to understand where do you fit into all of this? Is it possible to leverage this “new oil” everyone keeps talking about? The title does say, “Is Data Science for All?” but I would like to focus on “Is Data Science for you?”. Let’s dive in.

# Is Data Science for you?

* The short answer:   
  **Yes**.
* The long answer:   
  Data Science has been around for a while, and it’s been the sexiest job around, according to a lot of people and companies, for over a decade. If you are willing to put in the time and effort, no matter what your nature, area of interest /expertise/domain is, **Data Science can be for you**.

### The intuition of Data in Data Science

Let me give you some context. If your grandparents have managed to learn and upskill themselves on how to use the internet, they probably enjoy a whole set of new privileges in life-like, like video calls with family, watching content online, participating in group conversations, and seeing digital photos of your vacation etc.

Now, on the other hand, if they do not know how to use the internet, they are almost digitally disabled, and they are missing out on a big part of the new world.

Similarly,

If you do not know how to leverage Data in this Digital Age, you become digitally disabled

Whether you are a student, professional or entrepreneur, let me explain to you a situation that you might have faced in your day to day. If I were to give you a dataset or an excel file with *one million rows* (ten lakh rows), you can open up excel and perform some sort of analysis.

The situation changes if I were to give you ten million rows – you would not be able to open it in Microsoft Excel. Now, all of a sudden, you can’t do much because Excel has a limit of one million rows. Learning how to work with data, perform analysis and even visualize data can help you make better decisions more efficiently. People who are done with their MBA are coming back to Data Science because it can help them make exponentially better decisions.

The most unique thing about Data Science is whether you are a fresher or an experienced professional in any field, you can become a Data Scientist.

There are two main reasons that anyone can enter Data Science with a little bit of effort are:

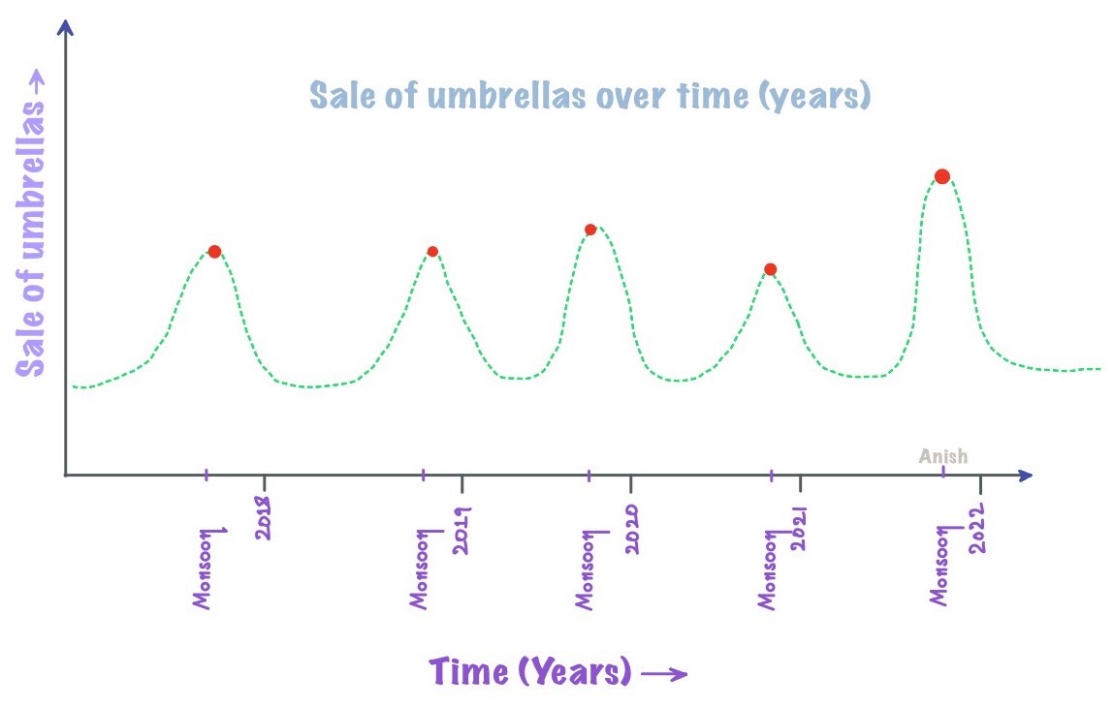
1. Data Science is an **industry-agnostic field**, which basically means that irrespective of the field you are in, you can add a “data-edge” to it, and you will also be able to leverage the knowledge and experience you already have
2. Data Science is **extremely practical**. Most of the problems being solved in companies across are understandable and doable.

Whatever your nature or experience is, there is a place for you in the end to end Data Science process. How can I be so sure? Let me show you exactly what I am talking about using the roles that come under Data Science.

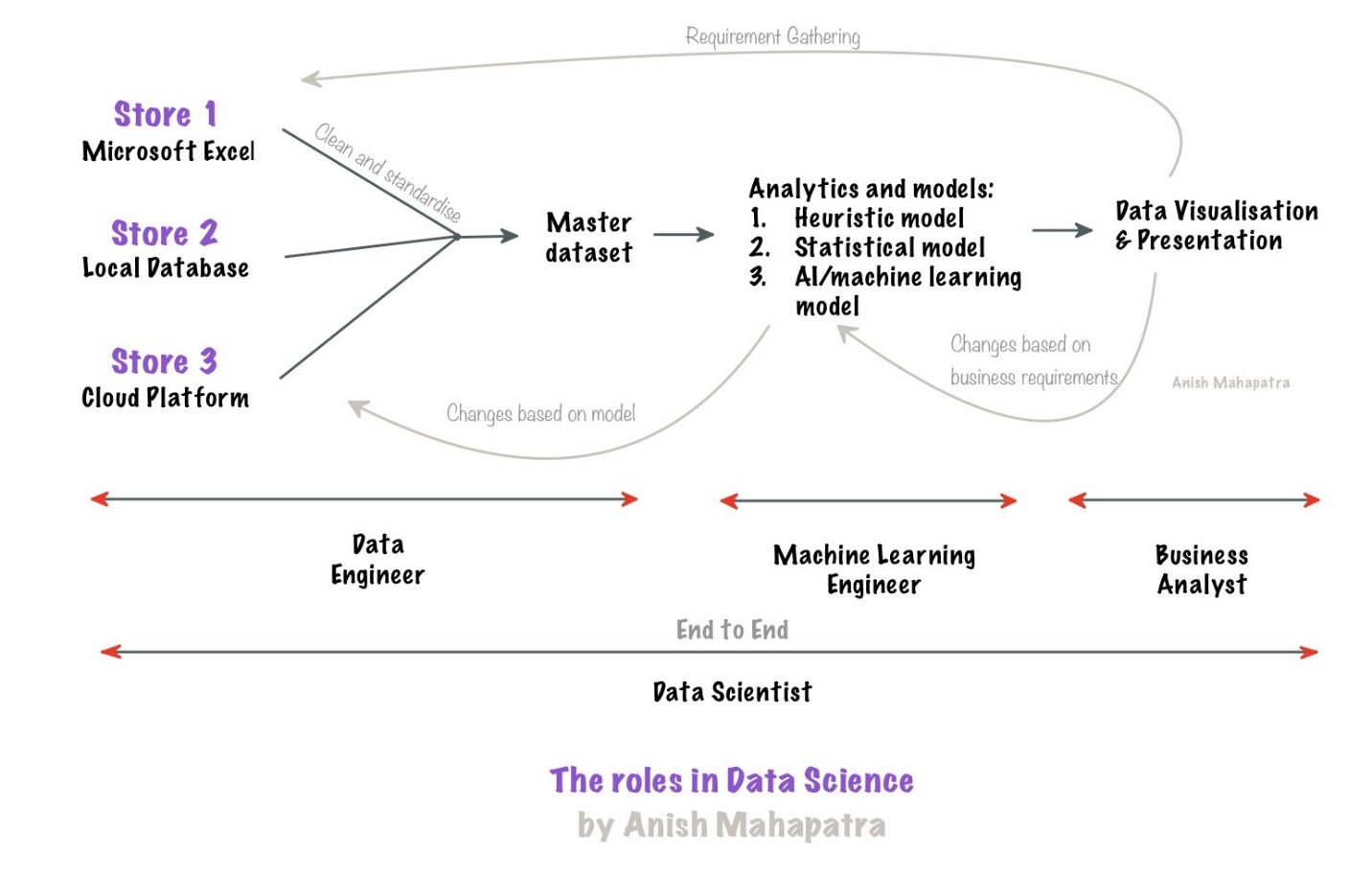
# The roles that come under Data Science

I could go about and just jot down the roles and tell you what we do in them. Instead, let’s try to understand with a solid example of how Data Science works in the real world. Follow along closely. I’ll also be talking briefly about the tools and technologies needed as well!

Let’s say we are a retail giant, Reliance Retail, and we are looking to **optimize products** that we have in our retail stores across India. For this, the business stakeholders will need to understand what are our best selling and worst selling products. The steps that we can take to understand what is Data Science all about are as follows:

* **Orchestrate Data**: Get data from the various different stores that we have. The data can be in excel, some sort of local database or on the Cloud.
* **Clean & Standardize Data**: Bring the data into the same format, clean and organize the data to make one big master dataset
* **Modelling & Analytics**: On this master data set, we need to perform our analytics of top and bottom performing products to be able to optimize. This can be done in three ways:
  + **Heuristic model**   
    A heuristic model is a fancy way of saying that business rules are used to decide the products that are performing well or badly. For instance, we can say that if a particular oil brand is selling over 1,700 packets in a month, it is a good product. Writing code for each of these business rules as if-else statements would entail a heuristic model, for example.
  + **Statistical model**Let’s say we are selling umbrellas and the sale of umbrellas is low throughout the year, except during the monsoon season. Does this mean that umbrellas are bad products? No. If we were to look at the distribution over five years, we would see that it peaks during the monsoon season. The seasonal sales can be inferred by looking at their statistical distribution.   
      
     So, in this case, umbrellas would not be a bad product but rather a seasonal product, based on what we can see from the statistical distribution over the years.
  + **AI/Machine Learning model**Think of a Machine Learning model as a black model. The input is the same that we have seen above, which is a big master dataset, and the output is the good and bad performing products. The only difference is instead of a manual, heuristic approach or just statistics, we let the model look at historical data and tell us what the good and bad products are.

Great job, you have an understanding of how the end to end process of how a Data Science process flow works. That was all about Data Science. Let’s have a look at what you have understood in the form of an image and discuss briefly the main roles that exist in the industry.

   
The roles that exist in the field of Data Science are traditionally distributed as follows:

* **Data Engineer**   
  Similar to how water flows from one end of a pipeline to the other, data also flows in from multiple sources to the master dataset. Data Engineers help make the data pipelines to get data from a plethora of sources and bring it to a common location. While getting the data, the data needs to be cleaned, standardized and bought to a common format.  
    
  Platforms to be familiar with (either one):   
  *Google Cloud Platform, Microsoft Azure, Amazon Web Services*  
  Languages & skillsets: *SQL, understanding of databases and python (fundamentals)*
* **Machine Learning Engineer**   
  As discussed, there are three primary kinds of models that one can use, which are the heuristic model, statistical and machine learning model. A machine learning engineer needs to work with the business team to be able to understand the requirements and work with the data engineering team to get the relevant data. A knack for experimental problem solving can aid an individual in fitting into this role.   
    
  Languages: *Python, SQL, MLOps fundamentals*   
  Other skillsets: *Machine Learning, Big Data, Business & domain understanding*
* **Business Analyst**   
  Once the model is made, and some sort of output has been drawn, the next task is to showcase it to the business in a manner where they are able to view the information (visually or otherwise) and come to a business decision that benefits the organization. This can be done with the help of a dashboard and presentations to quantify business impact.  
    
  Visualization & Dashboarding tools (either one or two): *Tableau, Power BI, Qlik Sense*  
  Tools & skillsets: *Business Understanding, stakeholder management, Microsoft excel, project management, strong communication skills, presentation skills*
* **Data Scientist**   
  The reason that the most sought after role in the market is that of a *full-stack* Data Scientist is that to be one, the candidate must understand everything mentioned above end-to-end. A Data Scientist needs to have the skills of a Data Engineer, Machine Learning Engineer and Business Analyst combined.   
    
  Skills: *All of the above (fundamental understanding) + Team management*

Now that you have gained a fair understanding of the roles that come under Data Science let’s explore the golden question – “What role should you pick?”

# What role should you pick?

This is a subjective question. Based on the five years of experience I have had in the industry, I would chalk out this decision to the knowledge of the person you are currently and the kind of person you would like to be. We would also like to take into consideration future growth over the next few years. One of the best pieces of advice I can give you is:

Talk to people who are where you want to be.

I have done the research for you, and here is my opinion that I am going to divide based on the roles in Data Science that we discussed above.

* **Data Engineer**: If you are the kind of person that is an introvert and is comfortable with extremely process-oriented work, this role is for you. In terms of growth path, you will learn and eventually master a particular cloud platform and implement it for a few projects.  
    
  As you grow up the ladder to Senior Data Engineer, Technical Architect etc., you will need to manage data pipelines for more projects, help team members/teams when they get stuck and eventually design the end to end pipelines for various teams.
* **Machine Learning Engineer**: Having a knack for experimentation with data, along with being good with mathematics, numbers and statistics, will put you in a good position to be a machine learning engineer. This is one of the roles in Data Science that has the word “engineer” in it, which indicates that there is some amount of Software engineering present in it as well.   
    
  An extremely respected and future-facing role in the market, the growth in this role can be in terms of business or technology, depending on the inclination of the individual.
* **Business Analyst**: With a heavy emphasis on efficient communication, presentation and visualization, this role is suitable for extroverts that have a presence of mind. In a role that has a relatively low amount of coding, business analysts are also extremely desired in the market.   
    
  Growth in this role is along the lines of Senior Business Analyst, Business Managers, VP, Director etc., as strong domain knowledge can help business analysts be effective decision-makers.
* **Data Scientist**: Given the current pace of technology, being a Data Scientist would entail a lifelong role of learning. Having the patience and know-how of how to manage technology, clients and teams. A strong growth mindset coupled with strong technical experience can help individuals make a big difference in the organizations that leverage Data Scientist folks! Growth can be along with technology or management.

**Disclaimer**: The Data Science market is in its nascent stages of growth. Although the terminology and work mentioned in the above sections would be ideal, different companies have different terminology, roles and responsibilities. The same has not yet been standardized across the industry. Before applying for any role(s), it is strongly recommended to speak to someone who is already doing the role you are looking out for.

Now that you have understood the “what”, you might be interested in getting to know more on the timeline for you to become one!

# How long does it take to become a Data Scientist?

The Short Answer: It depends. Typically six months to one year of dedicated effort, depending on the individual.

The long answer:

It would depend on how much passion you have for data/mathematics/business. Essentially, Data Science is a combination of Mathematics, Business and Technology. If you are willing to work towards being good at one of these and intermediate in another, it is more than enough to succeed as a data scientist.   
  
Focusing not just on how you can crack the interview but also on giving importance to other factors can make you stand out in a crowd.

* Doing a Data Science course/certification
* Writing articles on Data Science
* Being active, engaging and posting on LinkedIn
* Working towards building a strong GitHub profile

The more work you put into it, the more you get out of it. I have practised all that I preach; please feel free to Google my name. You can get into Data Science as fast as you would like to.

# Mandatory skills you need to become a Data Scientist

My major fear when I started out as a data scientist was my fear of code. The one thing I have learned is

The most effective way to get good at something is to keep doing/practising it on a daily basis.

Secondly, rather than just learning the theoretical way, put the knowledge into practice, and that can supremely enhance one’s skillset.

Having said that, here, I would like to break the mandatory skills into different domains of skills, i.e., interpersonal and technical.

### Interpersonal Skills

* **Effective and Efficient Communication skills**: Personally, I have enhanced this skill by observing, communicating and practising with different peers, team leads, higher-up management folks, multiple business stakeholders and IT managers across various industries.
* **Grit and the learning attitude**: You might wonder why I would have included this here. One thing that you will be committing to in this field is **lifelong learning** as this industry evolves on a daily basis. Almost every day, there is something new coming up in different technologies and tools. A learning attitude definitely takes you a long way.
* **Organizational Skills:** As one starts getting involved in different aspects of a project by portraying different roles, good organizational skills are a necessity. Putting structure into your thought process on a day to day basis can make almost everything you do – personally as well as professionally, much easier.

### Technical Skills

* + **Programs and Softwares**: Python, SQL, R, and Java are the go-to languages to start off with anything related to data sciences
  + **Cloud Computing**: Microsoft Azure, Google Cloud Platform and Microsoft Azure. The best way to pick up this skill would be to master one and draw parallels across the other.
  + **Machine Learning Skills**: Start learning from the few basic ML algorithms such as Linear Regression, Decision Tree, Random Forrest, etc. And you can go higher on the ladder by looking into Support Vector Machines, Neural Networks, Deep Learning and so on. But again, practice and application become key to learning ML. There are wide ranges of resource material available if one would want to deep dive into the math of the algorithm, a few of them which I have tagged below
  + **Data Visualization Tools**: Basic tools, to begin with, can be Microsoft Power BI and Tableau as these would be relatively easier for one to pick up. There are also a lot of [other tools](https://www.simplilearn.com/data-visualization-tools-article) one could look into
  + **Database Management:** One thing I can vouch for is that as a data scientist, you would spend about 80% of your time preparing your data for analysis and modelling in the next steps. So, database management through different skills such as MySQL, SQL Server, Oracle and NoSQL databases such as MongoDB, HBase amongst the others

To master all of these tools, let us discuss some of the more popular Data Science resources available.

# Data Science Resources

There are multiple resources that can be used for Data Science.

1. [Kaggle](https://www.kaggle.com/): The top machine learning and data science community on the internet, you can find like-minded people, code, data, courses and everything data science here
2. YouTube – From the same community of data scientists, YouTube can be a disorganized but extremely informative source to learn data science.
3. Udemy – Mini-courses can be done from here. Industry professionals/teachers take the effort to make entire courses here
4. [Knowledgehut Complete Data Science Bootcamp](https://www.knowledgehut.com/data-science/data-science-bootcamp-training) – An end to end Bootcamp for a beginner to understand end to end Data Science
5. Books & Self-Learning – If you are a self-motivated individual, it is tedious but possible to learn data science on your own from scratch

# Conclusion

Congratulations! You are the top 5% that have managed to come to the end of the article. This shows that you are here to really understand and know how Data Science works in the real world. Whatever you decide regarding your journey in Data Science may be, remember that data is the new future. Everyone is afraid of code initially, but eventually, we have to move in the direction that the world is moving. The faster we adapt, the more we will benefit from the boom of Data Science.

Thank you for taking the time to read my article. I am a Lead Data Scientist with the biggest Data Science firm in the country – you can connect with me on LinkedIn. In case you have any doubts, please feel free to reach out to the team or Google me and shoot me an email. I am always happy to help!

# Author

**Anish Mahapatra**   
*Senior Data Scientist*   
LinkedIn: [https://www.linkedin.com/in/anishmahapatra](about:blank) ([LinkedIn](about:blank))