

Data Scientist Careers: Applied Orientation for the Beginners

Mahesh Babu Pasupuleti

Data Analyst, Department of IT, TekSystems Inc, 200 S College St Suite 1200, Charlotte, NC 28202, USA

*Corresponding Contact:

Email: maheshbp.gs@gmail.com

ABSTRACT

A data scientist's job is making sense of complex, unstructured data that comes from a variety of sources, including smart devices, social media feeds, and emails, and that doesn't cleanly fit into a database structure. According to the findings of this study, Data Scientists require programming, mathematics, and database abilities, all of which may be learned by self-study or through formal education. Companies looking to hire for a Data Science team must be aware of the wide range of tasks that Data Scientists may fill, as well as the need of soft skills such as storytelling and connection building in addition to technical abilities and knowledge. The interpretation emphasizes that high school students interested in pursuing a career in Data Science should learn programming, mathematics, databases, and, most importantly, exercise their newfound knowledge. The findings of the study centered on data scientists as analytical specialists that employ their expertise in both technology and social science to discover patterns and manage data. The solutions to business difficulties are discovered via the use of industry expertise, contextual awareness, and skepticism of current assumptions.

Key Words: Data Scientist, Data Scientist Careers, Big Data, Data Scientist Skills

INTRODUCTION

Data scientists are tasked with sifting through massive amounts of organized and unstructured data and presenting their findings. In the function of a data scientist, computer science, statistics, and mathematics are all applied. Their work involves data analysis, processing, and modeling, followed by interpretation of the results in order to provide actionable strategies for businesses and other groups. There's a good chance that every large, successful company is backed by a data scientist. Data scientists crunch data and figures to find innovative solutions to issues and help their employers climb to the top — or at the very least compete with their competitors — in the marketplace.

A data scientist is an analytical data specialist who possesses technical capabilities as well as the capacity to tackle complicated issues in a fast-paced environment. The term "data scientist" refers to someone who works in both the technological and business sectors. A

data scientist is a combination of mathematicians, computer scientists, and trend-spotters. And what does it all come down to in the end? Someone who gathers, analyzes, and evaluates data in order to identify opportunities to assist a business in improving its operations and gaining an advantage over rivals (Adusumalli, 2016a).

When compared to some of the other positions discussed above, the title "data scientist" is a relatively recent one to the industry. Because all of the positions listed below are part of the broader area of data science, the particular job title of "data scientist" is frequently mistaken for a more prestigious version of the term "data analyst," although this is not necessarily the case. Consequently, data scientists must be fluent in numerous distinct fields at the same time. They must be proficient in mathematics, computer science, and business strategy, among other things. In order to successfully navigate this complicated skill set, data scientists must have a regular presence in both the information technology and the business worlds (Srivathsan and Arjun, 2015). That is one of the reasons why they are in such great demand.

Data scientists, like many of the other professions described, must have a mix of analytic, machine learning, data mining, and statistical abilities, as well as prior knowledge with algorithms and coding, among other things. Data scientists are also proficient in the following programming languages: R, SAS, Python, Matlab, SQL, Hive, Pig, and Spark, among others. A smart data scientist's ability to communicate the relevance of his or her findings in a way that others can understand is, nevertheless, one of the most crucial skills he or she possesses. In a manner that they may not be for many of the other roles described below, polished verbal and written communication skills are essential for this position.

The need to communicate results and analyses to supervisors, colleagues on different teams, and even corporate stakeholders who may (but not always) be able to follow technical language that data scientists are accustomed to hearing will arise frequently for data science professionals. According to the Harvard Business Review, one of the most difficult challenges facing the sector is a lack of clear communication between data teams and non-technical stakeholders, which is described as follows: "Data teams are aware that they are sitting on important insights, but they are unable to market them," according to HBR. "They claim that decision makers misinterpret or oversimplify their analyses and that they want them to perform miracles, such as providing the correct answers to all of their inquiries," they write. The failure to perceive real outcomes is due to the results not being presented in their language," says the CEO. This is at the heart of what a competent data scientist will be able to do. A complete and simply digestible explanation of the information that has been discovered, as well as what has to be done in response to it now that it has been discovered (Tomar and Agarwal, 2013).

HOW CAN SOMEONE BECOME A DATA SCIENTIST?

High-end data scientists are typically PhDs. They often originate from physics, statistics, computer science, arithmetic, databases, statistics, probability, and all that stuff. You need to know some algebra, at least up to analytics, geometry, and hopefully calculus, some basic probability, some basic statistics, I mean really understand the difference and different statistical distributions and databases if you're joining a data science team. In other words, one of the easiest places to start is relational databases, which contain a lot of our data, so individuals can learn about computers, databases, and how we store things before moving on to big data clusters. You don't need to fully grasp the MapReduce programming

methodology. The top end data scientists, the PhD data scientists, work with the intersection of computer science theory, statistics, and probability.

I am a lifelong learner. So, I learnt about Hadoop by myself, reading papers and videos, thinking and playing around with it. What is a Beowulf cluster? I mean, what is that? That's when we got a grant from Citi Bank and created a five-computer cluster, and I realized that's HPC. I had one of the university's first HPC clusters, it was modest but our researchers liked it since they could run everything 40-50 times quicker. Because of things like IPython and Jupyter Notebooks as well as Zeppelin, you can really go in and take some of these courses and do things right then. You can see, feel, and play with them, and then you'll start to understand what's going on. The essential issue in all of this is motivation, and I believe the Big Data University's badge system is one method to keep people engaged. But they can if they wish. It's up to them. So they must comprehend the aim.

It can't be under the Chief Information Officer. Many contemporary CIOs come from accounting or financial backgrounds and are ignorant. Sorry. But it needs to come from the research side. So you'll find data scientists in firms that do research, medicines, finance, technology, etc. In fact, we can't keep some of our PhD data scientists in our program because demand for PhD data scientists is so high. They make a lot of money and work on interesting problems. How do you plan Uber cars? You have a lot of data.

RECRUITING FOR DATA SCIENCE

When firms hire employees for a data science team, whether a data scientist, analyst, or chief data scientist, they tend to hire people who have all the skills and domain expertise. They can analyze organized and unstructured data well. They're also outstanding presenters and storytellers. Putting it all together, you're hunting for a unicorn. And finding a unicorn is unlikely. It can teach analytics skills to anyone willing to put in the time and effort. What counts is who is enthusiastic about your business. An excellent data scientist in retail may not be enthused about working in IT or with terabytes of weblogs (Adusumalli, 2016b). But if someone is passionate by those weblogs, or health statistics, they may offer much more to your productivity. Is that individual inquisitive about why this room is decorated a specific manner, why the bookshelves hold books, and what types of books? They must be curious about all they see. Second, do they have a sense of humour? You know, you have to be lighthearted. A person who takes things too seriously may not be able to see the lighter components. Knowing there is a tale there. They should be able to explain their findings. We often find that nobody cares about r-square or confidence interval. So you must be able to introduce and explain things convincingly. And they need to find someone approachable, because data science is new, which means the individual in that capacity must build relationships and collaborate across departments. Have a strong mathematical and statistical background. They must consider analytical and problem-solving skills. A scientist must be able to analyze difficulties. They should hire people who enjoy playing with data. Then they know how to visualize data. They are analytical. When a corporation hires someone to work on a data science team, they must consider their function (Pasupuleti, 2015c). A corporation must first establish what they expect from its data science staff. Then they must hire to begin. They need to know whether they require engineers, architects, or designers to work on visualization. You need additional individuals who can multiply big matrices. Let's focus on technical abilities first. What type of technological platform would you like to adopt? Assume you wish to work in a structured data environment and in market research. So your skillset will be significantly different from

someone who wants to work in large data scenarios. If you wish to deal with traditional market research data, you should know some statistics and fundamental statistical techniques, as well as certain machine learning methods. And these are the instruments you wish to create. If you want to work with big data, you also need to be able to store data. So you start with huge data storage expertise. Then you hunt for platforms that accept it. The next stage is to manipulate enormous volumes of data, and then apply algorithms to them. So three steps. But it probably starts with where you want to be, in what industry, in what domain. When working with standard predictive analytics and not big data, R, Stata, or Python are the tools of choice. Python is more suited for unstructured data than R. If you're working with large data, you'll be using Hadoop and Spark. So, first decide where you want to go and what type of job thrills you, then choose your tools. Aside from technical capabilities, data scientists must be able to communicate. Skills in communication or presentation. And the reader understands it right there. So the capacity to convey your results vocally, visually, or in writing. Communication and presentation abilities are as vital as technical skills. Imagine you're driving down a mountain and taking a sharp bend. And you can't see around the bend. Then you make that corner and suddenly see a huge valley front of you. So when you present your results and you have a terrific finding and you explain it properly, people feel surprised. Then it empowers them, giving them suggestions on how to use this new information and understanding. It's a big thrill. And you may share with your clients as a data scientist since you enabled it.

CAREERS IN DATA SCIENCE

Due to the rise of the Internet of Things and advancements in distributed computing, massive volumes of data, as well as the technological capabilities to analyze it, have become available. Now that we have the ability to extract important insights and fresh information from data, we must learn how to structure that data so that it can be focused on what we should do with it and what it can accomplish for us (Pasupuleti, 2015b). This is where data science comes in. Companies such as LinkedIn, Glassdoor, Indeed, and Dice track employment trends, which reveal that a career in data science has risen up the list of the most promising positions to become the number one job in the country since 2016. Despite this, it is still one of the top three employment options for 2020. Dice pointed out that job ads come from firms across a wide range of industries, not only the technology sector. GIA projects that the data science platform market would expand by \$314.8 billion US by 2025, with a compounded growth rate of 38.2 percent, according to Global Industry Analysts Inc. According to the McKinsey Global Institute, there will be a severe scarcity of data and analytics skills by 2018. According to Forrester Research Analyst Brandon Purcell, who stated in January 2019 that the need for data scientists would only increase as firms increasingly rely on data-driven insights, the demand for data scientists will only expand. We've now reached the end of that time, and recruiters are finding it difficult to meet the increased need for competent data scientists in the industry. What is it that drives someone to pursue a career in data science? For starters, data science may be used to practically every field of study. So, if you have the talent and drive to work with data, love coding, have no difficulty understanding math and statistics, and are a good storyteller, you can undoubtedly succeed in the area of data science and beyond. For the majority of individuals, this entails the acquisition of extra tools and abilities, as well as the ongoing learning of new tools and approaches in the field. A commitment has been made by the Stanford Institute for Computational and Mathematical Computing to inspire and educate data scientists throughout the world, regardless of gender, as well as to encourage women working in the

area of data science. When pursuing a career in data science, it is important to ensure that your skill set is appropriate for the position you are aiming for. The talents you have can be tailored to a specific area of expertise that you wish to pursue by taking advantage of one of the many good online training options available (Young, 2015). Then you'll be well-prepared for a job that is both exciting and fulfilling. So now is the time to pursue a career in this industry, especially given the wide range of options and educational resources available to make it a possibility.

HIGH SCHOOL STUDENTS AND DATA SCIENCE CAREERS

Learn a little bit about statistical analysis. Then it's time to have fun. Make something, or write something, or both. In other words, when I say "construct," "programme," and "build systems," I don't only mean "create things physically," do I? You can design computer systems, statistical systems, and whatever else you choose. But, when you attempt anything, you'll be able to determine what tools you'll want, right? And you'll think to yourself, "Oh, my God, what?" Among other things, there's the term "what does an inner product mean?" What exactly is it? "How do I, uh, well, that's something I can learn." They will therefore have a significant advantage over many other college students when they arrive at their respective institutions. As a result, when they graduate from college, they will have made an even greater leap and will be earning a substantial sum of money. And they'll be pleased as well. Isn't this thing a lot of fun? It's a lot of fun. For those in high school who are seeking a career in data science, I recommend being familiar with data bases, learning SQL, and thinking about other options such as computer science if it is something you are interested in. If your school offers a computer science course, you should definitely take it, as it is an important aspect of being a data scientist. Beyond that, there are undoubtedly certain things you can do to encourage your creativity, your curiosity, and so on. If you enjoy detective games, that's really fantastic, don't you think? And if you enjoy treasure hunts or other similar activities, I believe you will, and if you have the opportunity to do so, it will benefit you in your career as a data scientist because it is a really fantastic method to ensure that you can remain interested as you go about your daily activities in your job. (Music) Encourage the pursuit of knowledge and the pursuit of knowledge via experimentation. It's similar to science fairs; science fairs are excellent because they promote experimenting, learning from mistakes, asking a question and addressing it using the scientific method, but instead of vinegar volcanoes, they encourage data sets. It's similar to the previous point, but because we're in the midst of election season right now, there's a lot of information in the news about polls and survey results, and that's a great way to start a conversation about how the people who ran the polls know, and how they can predict what's going to happen in the election. In other words, that's another interesting approach to begin a talk about data science. (Music) If you are interested in data science, I would advise you to follow your passion because it is a rewarding career path that will undoubtedly be required in the near future. There are few few knowledge-based professions that are as crucial to enterprises all around the world as it is to this particular company. A data scientist's skill set, as well as the ability to assist businesses as they expand and learn how to do things more effectively or more intelligently, will always be required. And data scientists are the ones who do this. (Music) I would say that I understand what you're saying because we was never a particularly good student in mathematics myself. And I believe there are a number of data scientists, many of whom are really successful and well-liked, who are in the same position as I am. Knowing that there is some arithmetic involved, and that math in school is not necessarily the ideal topic for everyone, if it's used in conjunction with anything else,

anyone can be sure that these aren't simply hypothetical figures, or that they aren't merely issue statements that have no link to you (Pasupuleti, 2015a). I discovered that when you have a personal connection to a subject, it suddenly becomes lot simpler to apply mathematics to help you grasp it. And so, knowing the individuals who will benefit from the math that you perform, I think it's pretty wonderful that you know who they are.

SKILLS NEEDED TO BE A DATA SCIENTIST

To be a data scientist, you'll need a variety of talents, but the Bureau of Labor Statistics considers the following to be particularly important:

Analytical Capabilities: A data scientist, according to the Bureau of Labor Statistics, "must be methodical in their thinking and assess the outcomes of their study in order to make conclusions."

Communication and Interpersonal Skills: Data scientists don't operate in isolation; they collaborate with programmers and managers, and they must be able to communicate effectively with them, according to the Bureau.

Critical and Logical Reasoning Capabilities: Because data scientists work on difficult issues, critical thinking skills, as well as a natural aptitude for reasoning and relying on logic, are essential for success.

Mathematical Abilities: Perhaps unsurprisingly, data scientists "must have understanding of complex arithmetic and technical issues that are crucial in computers," according to the Bureau of Labor Statistics.

Ingenuity: When it comes to tackling complicated challenges, inventiveness is frequently required. According to the Bureau of Labor Statistics, data scientists must be able to come up with "new approaches to tackle issues, particularly when their ideas do not initially function as intended."

Hard Skills: Data Scientists must be fluent in a variety of programming languages and software packages in order to be successful. Despite the fact that there are many various types, Glassdoor study discovered that Python, R, and SQL are three of the most commonly used.

DATA SCIENTIST RESPONSIBILITIES

Among the tasks of a data scientist are the following:

- Consolidate and organize massive amounts of organized and unstructured data. They query structured data from relational databases using computer languages such as SQL, which allows them to do complex calculations. They collect unstructured data through scraping websites, using APIs, and conducting surveys.
- Use complex analytical approaches, machine learning, and statistical methods to prepare data for use in predictive and prescriptive modeling.
- Thoroughly clean data to remove extraneous information and prepare the data for preprocessing and modeling.
- Use complex analytical approaches, machine learning, and statistical methods to prepare data for use in predictive and prescriptive modeling.

- Thoroughly clean data to remove extraneous information and prepare the data for preprocessing and modeling. Explore data analysis (EDA) to figure out how to deal with missing data and to identify patterns and/or possibilities in a dataset
- Developing new algorithms to solve issues and creating programs to automate repetitive tasks; communicating forecasts and results to management and IT departments through efficient data visualizations and reports; and developing new algorithms to solve problems.
- Make recommendations for cost-effective modifications to existing procedures and policies.

Every organization will have a unique approach to data science job responsibilities. Some organizations approach their data scientists as data analysts or mix their responsibilities with those of data engineers; others want top-level analytics professionals who are adept in intensive machine learning and data visualization.

Whenever data scientists gain new levels of experience or shift professions, their roles and responsibilities alter as well. Example: A single individual working alone at an SMB may spend a significant chunk of his or her day cleaning and munging data. A high-level employee at a company that provides data-based services may be tasked with structuring large-scale data projects or developing innovative solutions.

CHARACTERISTICS OF A SUCCESSFUL DATA SCIENTIST PROFESSIONAL

They should be inherently interested about their surroundings, but they should do so through an analytical lens. Data scientists do not just need to grasp computer languages, database administration, and how to transform data into visuals. Data scientists may have psychological qualities that are similar to those seen in quality assurance departments, since they evaluate enormous volumes of data and look for patterns and solutions. They are also innovative in designing new algorithms to crawl data or in creating well-organized database warehouses, among other things.

Generally speaking, people in the data science sector must be able to communicate effectively in a variety of settings, including with their team, stakeholders, and clients. There may be many dead ends, wrong turns, and rocky roads along the way, but data scientists must have the determination and fortitude to stay afloat while exercising patience in their research (Archenaa and Anita, 2015).

In addition to having strong technical backgrounds, the top data scientists also have excellent intuition when it comes to the data they are working with. Does each of the characteristics have significance, and do they accurately represent what you believe they should mean? If your data is dispersed in the manner described above, which model should you employ? The absence of a value signifies something, but what you should do with it is unclear. The finest data scientists are also excellent communicators, with the ability to communicate effectively to both fellow data scientists and non-technical individuals. Airbnb's effectiveness depends on the quality of our studies, which must be both technically sound and presented in a clear and actionable manner to other members of the firm.

CONCLUSION

Data scientists use their analytical, statistical, and programming talents to massive data sets in order to collect, evaluate, and interpret the information. Later, they use this knowledge to the development of data-driven solutions for complex business situations. A bachelor's

degree in statistics, mathematics, computer science, or economics is typically required for data scientists. In terms of technical skills, data scientists have a wide variety of abilities, including statistics and machine learning, programming languages, databases, machine learning, and reporting technologies, among others.

REFERENCES

- Adusumalli, H. P. (2016a). Digitization in Production: A Timely Opportunity. *Engineering International*, 4(2), 73-78. <https://doi.org/10.18034/ei.v4i2.595>
- Adusumalli, H. P. (2016b). How Big Data is Driving Digital Transformation?. *ABC Journal of Advanced Research*, 5(2), 131-138. <https://doi.org/10.18034/abcjar.v5i2.616>
- Archenaa, J., and Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*, 50, 408-413.
- Pasupuleti, M. B. (2015a). Data Science: The Sexiest Job in this Century. *International Journal of Reciprocal Symmetry and Physical Sciences*, 2, 8–11. Retrieved from <https://upright.pub/index.php/ijrsps/article/view/56>
- Pasupuleti, M. B. (2015b). Problems from the Past, Problems from the Future, and Data Science Solutions. *ABC Journal of Advanced Research*, 4(2), 153-160. <https://doi.org/10.18034/abcjar.v4i2.614>
- Pasupuleti, M. B. (2015c). Stimulating Statistics in the Epoch of Data-Driven Innovations and Data Science. *Asian Journal of Applied Science and Engineering*, 4, 251–254. Retrieved from <https://upright.pub/index.php/ajase/article/view/55>
- Srivathsan, M., and Arjun, K. Y. (2015). Health monitoring system by prognostic computing using big data analytics. *Procedia Computer Science*, 50, 602-609.
- Tomar, D., and Agarwal, S. (2013). A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241-266.
- Young, S. D. (2015). A big data approach to HIV epidemiology and prevention. *Preventive medicine*, 70, 17-18.

--0--