

Data-science-14-self-examination-questions?

1. What is data science?

Data Science is such a broad field that includes several subdivisions like data preparation and exploration, data representation and transformation, data visualization and presentation, predictive analytics, machine learning, deep learning, artificial intelligence, etc. It is possible to consider three levels of data science competency (3 levels defined based on topics covered in one of the best machine learning textbooks out there: *Python Machine Learning* by Sebastien Raschka, 3rd Edition), namely: **level 1** (basic level); **level 2** (intermediate level); and **level 3** (advanced level). Competency increases from level 1 to 3, as shown in **Figure 1** below.



Figure 1. Three levels of data science competency. Image by Benjamin O. Tayo

2. What does a data scientist do?

A data scientist works with data to draw out meaning and insightful conclusions that can drive decision making in an institution or organization. Their job role includes data collection, data transformation, data visualization and analysis, building predictive models, providing recommendations on actions to implement based on data findings. Data scientists work in different sectors such as healthcare, government, industries, energy, academia, technology, entertainment, etc. Some top companies that hire data scientists are Amazon, Google, Microsoft, Facebook, LinkedIn, Twitter, Netflix, IBM, etc.

3. What is the job outlook for data scientists?

The job outlook for data scientists is very positive. [IBM predicts](#) the demand for data scientists to soar 28% by 2020. A [recent study](#) using the LinkedIn job search tool shows that a majority of top tech jobs in the year 2020 are jobs that require skills in data science, business analytics, machine learning, and cloud computing (see **Figure 2** below).

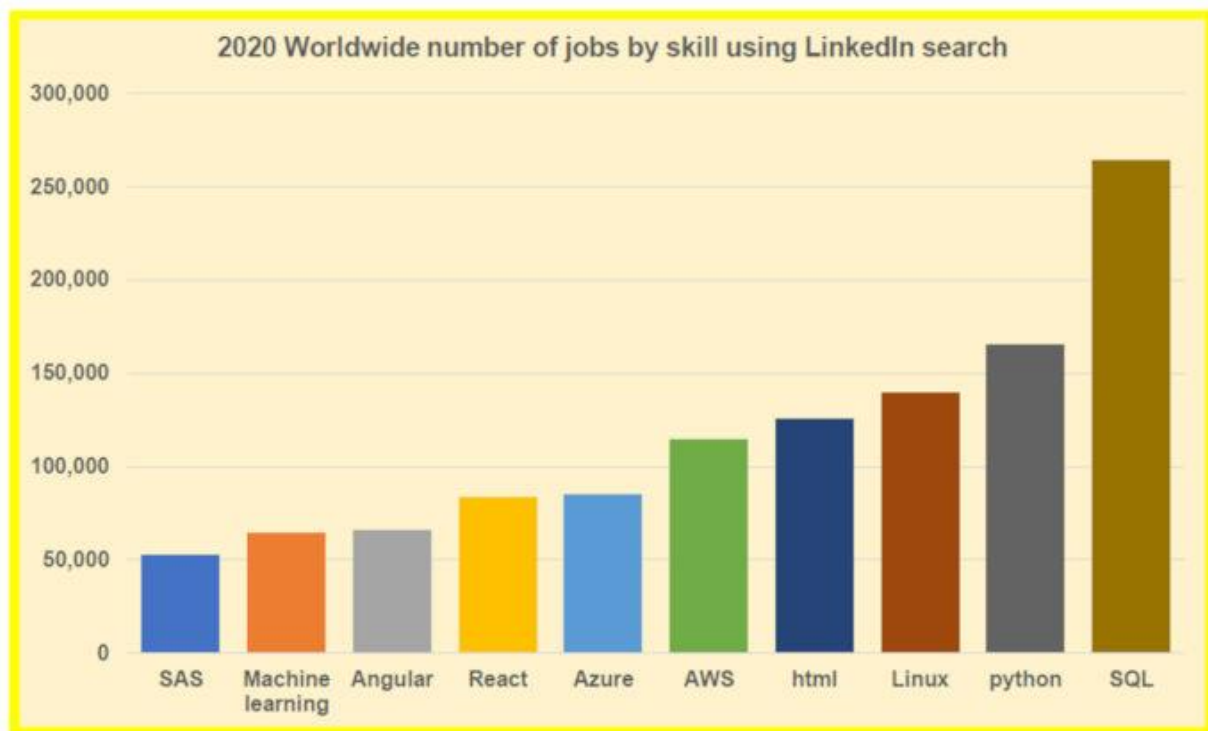


Figure 2. The worldwide number of jobs by skill using LinkedIn's job search tool. Image by Benjamin O. Tayo.

4. How much do data scientists make?

How much you make as a data scientist depends on the organization or company you are working for, your educational background, your number of years of experience, and your specific job role. Data scientists make anywhere from \$50,000 to \$250,000, with the median salary being about \$120,000. This [article](#) discusses more the salaries of data scientists.

5. How can I prepare for a career in data science?

Most data science or business analytics programs require the following:

1. a) A high level of quantitative ability
2. b) A problem-solving mindset
3. c) Programming proficiency
4. d) The ability to communicate effectively
5. e) Ability to work in a team

Hence to prepare for a career in data science, you may start by pursuing a bachelor's degree in a quantitative discipline such as **science, technology, engineering, mathematics, business, or economics**.

6. What programming languages should I focus on?

If you are interested in learning the fundamentals of data science, you need to start from somewhere. Do not be overwhelmed by the [ridiculous list of programming languages mentioned in data scientist job ads](#). While it is important to learn as many data science tools as possible, it is recommended to start with just one or two programming languages for a start. Then once you have built a solid background in data science, you can then challenge yourself to learn about different programming languages or different platforms and productivity tools that can enhance your skillset. According to this [article](#), Python and R are still the top two programming languages used in data science. I would recommend starting with Python as more and more academic training programs and industries are using it as the default language for data science.

7. How long does it take to become a data scientist?

If you have a solid background in an analytical discipline, such as **physics, mathematics, engineering, computer science, economics, or statistics**, you can basically teach yourself the basics of data science. You may start by taking free online courses from platforms like edX, Coursera, or DataCamp. **Level 1** competency (see **Figure 1**) can be achieved within 6 to 12 months. **Level 2** competencies can be achieved within 7 to 18 months. **Level 3** competencies can be achieved within 18 to 48 months. The amount of time required to gain a certain level of competence depends on your background and how much time you are willing to invest in your data science studies. Typically, individuals with a background in an analytic

discipline such as physics, mathematics, science, engineering, accounting, or computer science would require less time compared to individuals with backgrounds not complementary to data science.

8. Am I patient enough to keep on working even when a project seems to have hit a roadblock?

Data science projects could be very long and demanding. From problem framing to model building and application, the process could take weeks and even months, depending on the scale of the problem. As a practicing data scientist, hitting a roadblock with a project is something inevitable. Patience, tenacity, and perseverance are key qualities essential for a successful data science career.

9. Do I have the business acumen that would enable me to draw out meaningful conclusions from a model that can lead to important data-driven decision making for my organization?

Data science is a very practical field. Remember that you may be very good at handling data as well as building good machine learning algorithms, but as a data scientist, the real-world application is all that matters. Every predictive model must produce meaningful and interpretable results in real-life situations. A predictive model must be validated against reality in order to be considered meaningful and useful. Your role as a data scientist is to draw out meaningful insights from data that can be used for data-driven decisions that can improve the efficiency of your company or improve the way business is conducted, or help increase profits.

10. Do I have good communication skills?

Data scientists need to be able to communicate their ideas with other members of the team or with business administrators in their organizations. Good communication skills would play a key role here to be able to convey and present very technical information to people with little or no understanding of technical concepts in data science. Good communication skills will help foster an atmosphere of unity and togetherness with other team members such as data analysts, data engineers, field engineers, etc.

11. Am I a lifelong learner?

Data science is a field that is ever-evolving, so be prepared to embrace and learn new technologies. One way to keep in touch with developments in the field is to network with other data scientists. Some platforms that promote networking are LinkedIn, GitHub, and medium ([Towards Data Science](#) and [Towards AI](#) publications). The platforms are very useful for up-to-date information about recent developments in the field.

12. Am I a team player?

As a data scientist, you will be working in a team of data analysts, engineers, administrators, so you need good communication skills. You need to be a good listener, too, especially during early project development phases where you need to rely on engineers or other personnel to be able to design and frame a good data science project. Being a good team player would help you to thrive in a business environment and maintain good relationships with other members of your team as well as administrators or directors of your organization.

13. Am I Ethical?

Ethics and privacy considerations are a must in data science. You need to understand the implication of your project. Be truthful to yourself. Avoid manipulating data or using a method that will intentionally produce bias in results. Be ethical in all phases, from data collection and analysis to model building, analysis, testing, and application. Avoid fabricating results for the purpose of misleading or manipulating your audience. Be ethical in the way you interpret the findings from your data science project.

14. What are some resources for learning about data science?

You may pursue a master's degree in data science or in business analytics if your circumstances allow you to do that. If you cannot afford a master's degree program, you may pursue the self-study route for learning about data science. Generally, if you have a solid background in an analytic discipline such as physics, mathematics, economics, engineering, or computer science, and you are interested in exploring the field of data science, the best way is to begin with massive open online courses (MOOCs). Then after establishing a solid foundation, you may then seek other ways to increase your knowledge and expertise, such as studying from textbooks, engaging in projects, and networking with other data science aspirants.

Find below are recommended MOOCs and textbooks that can help you master the fundamentals of data science.

Recommended MOOCs:

[Professional Certificate in Data Science \(HarvardX, through edX\)](#)

[Analytics: Essential Tools and Methods \(Georgia TechX, through edX\)](#)

[Applied Data Science with Python Specialization \(the University of Michigan, through Coursera\)](#)

Recommended Books:

Learning from a textbook provides a more refined and in-depth knowledge beyond what you get from online courses. This book provides a great introduction to data science and machine learning, with code included: “Python Machine Learning,” by Sebastian Raschka. <https://github.com/rasbt/python-machine-learning-book-3rd-edition>