What is data science?

And what is it not?

Have this question too? Request Answers:

Request From Quora

We will distribute this question to writers, and notify you about new answers.

54 Answers

As this is a very open ended generic question, so would like to state a very broad answer for the same from my blog here: What is Data Science? by Pronojit Saha on Journey to planet Datum & Beyond[1]

Data Science is the **practice** of:
  1. Asking questions (formulating hypothesis), answers to which solve known problems or unearth unknown solutions that in turn drive business value,
  2. Defining the data needed or working with an existing data set and employing tools (computer science based) to collect, store and explore such data generally in huge volume & variety (often more than 1 TB and 1000s of dimensions),
  3. Identifying the type of analysis to be done to get to the answers and performing such analysis by implementing various algorithms/tools (statistics based), often in a distributed and parallel architecture,
  4. Communicating the insights gathered from the analysis in the form of simple stories/visualizations/dashboards (the Data Product) that a non-data scientist can understand and build conversation out of it. (It should be kept in mind that a product can also be an piece of code that is internal to a company and is used by various departments. The presentation, maintenance, scalability, etc of the code are then the product features, which is often not practiced in many organizations)
  5. Building a higher level abstraction that does steps 2-3-4 in an autonomous way, analyzing & taking actions on new data as they are fed to the system.

24.8k ViewsView Upvotes[2]

Data scientist performs research and analysis on data and helps companies to improve business by predicting growth, trends and business insights based on huge amounts of data.

Armed with data and analytical results, a top-tier data scientist will then communicate informed conclusions and recommendations across an organizations leadership structure.

Successful big data scientists will be in high demand and will be able to earn very nice salaries. But in order to be successful, big data scientists need to have a wide range of skills that until now did not even fit into one department.

Learning how to become data scientist can be quite costly, with an average cost of $9,600 (according to Harvard Extension School[3]). But if you know which skills employers are looking for you can find many free resources online. That is exactly what we did for you!

Below is the required skills set for becoming a data scientist with top 2-3 free resources to learn each skill online.

1. Python[4]

Learn Python Programming From Scratch by Udemy[5]

Learn to program in Python by CodeCademy[6]

LearnPython.org interactive Python tutorial[7]

2. Machine Learning[8]

Machine learning online[9]

Operational Intelligence and Machine Data with Splunk[10]

3. R Language[11]

R Basics – R Programming Language Introduction by Udemy[12]

Introduction to R at DataCamp[13]

Learn R at Code school[14]

4. Big Data[15]

Big Data University[16]

Big Data and Hadoop Essentials by Udemy[17]

Basic overview of Big Data Hadoopby- Udemy[18]

5. Statistics[19]

Statistics One by Coursera[20]

Statistics and Probability[21]

Probability & Statistics[22]

6. Data Mining[23]

Data Mining and Web Scraping: How to Convert Sites into Data by Udemy[24]

Data Mining by Coursera[25]

7. SQL[26]

Interactive Online SQL Training for Beginners[27]

Sachin Quickly Learns (SQL) – Structured Query Language by Udemy[28]

SQL Tutorial by w3schools[29]

8. Java[30]

Learn Java: The Java Programming Tutorial For Beginners by Udemy[31]

Learn Java – Free Interactive Java Tutorial [32]

Learn Java Programming From Scratch – Udemy[33]

For more information: How to become data scientist for free and from scratch[34]

ViewsView Upvotes[35]

Luis Otavio Martins[36], Mechatronics Technician, Engineer and Jr. Data Scientist

Updated Oct 28[37]

I just left an interview where they asked me the same question. After reading the other 41 answers, I will try to adress a simple and more correct one:

WHAT IS

1. It is a little bit of a misnomer and a buzz word that media is using to describe everything. However, it's good to have this dicussion to come into an agreement.
2. The questions is about Data science. So I will not talk about Data Scientists. Go to What is a data scientist?[38] if you are interested.
3. The biggest error that I found in most of the answers was some sort of "Data Science is when you are dealing with Big Data, large ammounts of data". That is not true, Data Science can be applied to a data set with one thousand lines, there is no problem with this.
4. If we are goig to call as "science" we need to consider the Science and Scientific Method[39] definition. According to this, Data Science is not only about the practical or empirical methods, it needs scientific foundations.
5. No one talked about the difference between Data and Information.
    1. **Data** is a raw, unorganized set o things that need to be processed to have a meaning.
    2. **Information** is when data was processed, organized, structured or presented in a given context so as to make it useful
    3. Based on this, we would have Data science and Information science. Right now, people have a bias to talk about Data science including Information science.

6. It was clearly being used in a lot of fields for the past years:

1. Statistics/Mathematics
2. Business analytics
3. Market intelligence
4. Strategic Consulting
5. Many others…
6. The craziest part is that you see professionals of these areas updating their resumes with something like "I worked with Data Science…"

7. The creation of data science in a simple way. Two sides that were not totally connected, but with the new fast paced and technological world would have to merge together:
   1. **Statistics/mathematics:** formulate proper models to generate insights.
   2. **Computer science**: make the bridge between the models and the data in a feasible time to come with the result.

8. Topics/tools that a person neeed to understand or have some knowledge when working with Data Science:
   1. Linear algebra
   2. Non-linear systems
   3. Analytical geometry
   4. Optimization
   5. Calculus
   6. Statistics
   7. Programming language (R, Python, SAS)
   8. Softwares: Excel, SPSS by IBM
   9. General platforms: Watson Anlytics by IBM, Azure Machine Learning, Google Cloud machine learning,
   10. Data visualizations: Power BI, Tableau, R/Python using plotly/ggplot
   11. Machine Learning (supervised, unsupervised and reinforcement learning)
   12. Big Data
   13. Big Data Frameworks (Hadoop and Spark)
   14. Hardware (CPU, GPU, TPU, FPGA, ASIC)

9. One Picture Worth Ten Thousand Words. The Drew Conway's Data Science Venn Diagram[40] . The Substantive expertise (or Domain expertise) is the specific knowledge of the area that you are applying Data Science. To know more about the lack substantive expertise in data science: What's Missing in Data Science Talks - As Risky As It Gets[41]
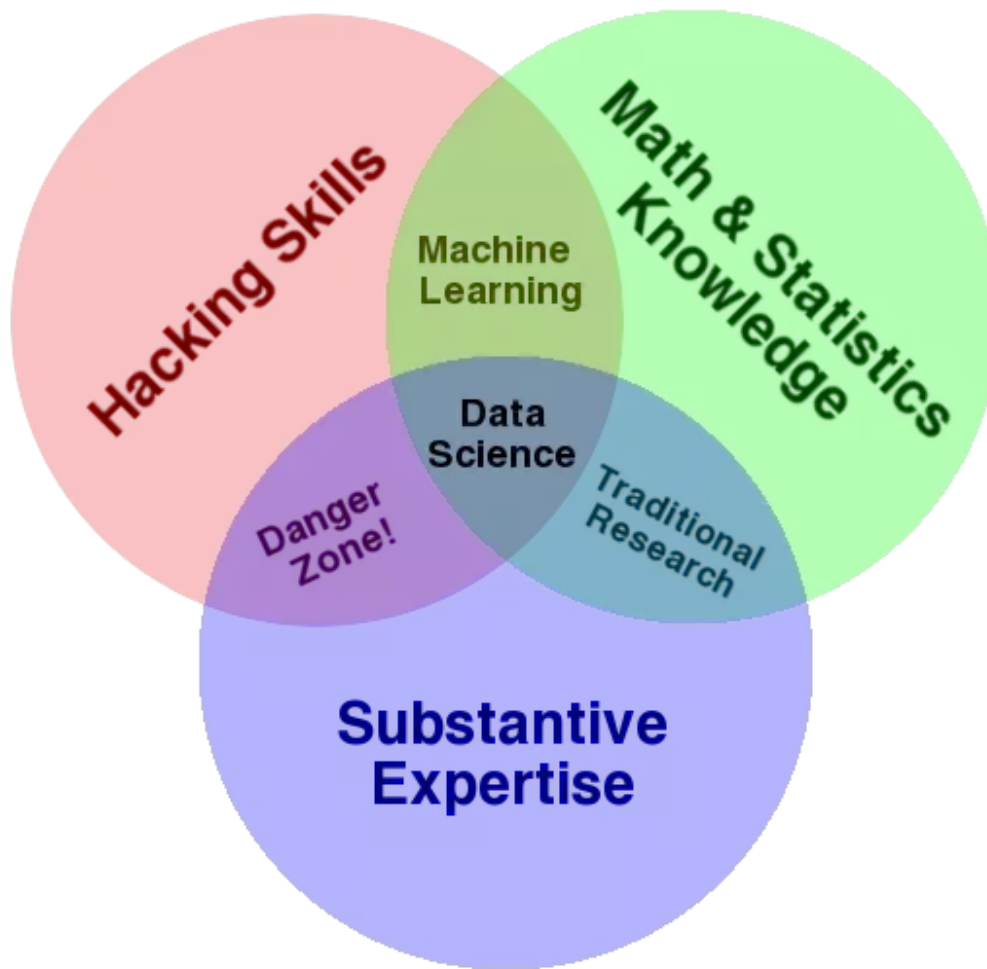
WHAT IS NOT

1. Machine Learning is not a branch of Data science. Machine Learning originated from Artificial Intelligence. Data science is only using ML as a tool. The reason is that it produces amazing and autonomous results for specific tasks
2. It's not the salvation of companies that never measured anything and now want to get insights from their data. "Garbage in, garbage out" Data science will be as good as the data generated on the following years.
3. Just present data using some Excel charts without any insight about the data.

Always upvote answers that you find useful. Everyone can be wrong so be respectful and polite.

ViewsView Upvotes[42]

Drew Conway[43], PhD student in Politics at NYU; studying networks, terrorism, conflict and technology

Written Aug 23, 2010[44]Upvoted by William Chen[45], Data Scientist at Quora

First, the term "data science" is a misnomer with respect to what most people consider endeavors classified as such. Fundamentally, "science" is about formalizing a hypothesis given a reasonable set of observations and assumptions, designing an experiment around that

hypothesis, testings it and analyzing the data generated through that process to either confirm or falsify the hypothesis. Therefore, "data" is simply a natural byproduct of science. Very (very) rarely are things labeled as data science actually scientific.

Rather, data science most often refers to the tools and methods used to analyze large amounts of data. As such, the discipline is an amalgamation of many bits from other areas of research. For tools, the influence primarily comes from computer science, where issues of algorithmic efficiency and storage scalability form the main focus. For analysis, however, the influences are much more varied. Modern methods are borrowed from both the so-called hard sciences (physics, statistics, graph theory) and the social sciences (economics, sociology, political sciences, etc). Specific classes of techniques that are naturally interdisciplinary are also very popular, such as machine learning.

Mike Loukides had an interesting article in O'Reilly that posited the same question, and is worth a read: http://radar.oreilly.com/2010/06...[46]

35.2k ViewsView Upvotes[47]

Michael Hochster[48], Head of Research at Pandora

Written Jan 17, 2014[49]Upvoted by Jerrod Lowmaster[50], LinkedIn Data Scientist, Boxun Zhang[51], Data Scientist at Spotify; PhD in Computer Science, and Alon Amit[52]

Data Scientists are people with some mix of coding and statistical skills who work on making data useful in various ways. In my world, there are two main types:

Type A Data Scientist: The A is for Analysis. This type is primarily concerned with making sense of data or working with it in a fairly static way.The Type A Data Scientist is very similar to a statistician (and may be one) but knows all the practical details of working with data that aren't taught in the statistics curriculum: data cleaning, methods for dealing with very large data sets, visualization, deep knowledge of a particular domain, writing well about data, and so on.

The Type A Data Scientist can code well enough to work with data but is not necessarily an expert. The Type A data scientist may be an expert in experimental design, forecasting, modeling, statistical inference, or other things typically taught in statistics departments. Generally speaking though, the work product of a data scientist is not "p-values and confidence intervals" as academic statistics sometimes seems to suggest (and as it sometimes is for traditional statisticians working in the pharmaceutical industry, for example). At Google, Type A Data Scientists are known variously as Statistician, Quantitative Analyst, Decision Support Engineering Analyst, or Data Scientist, and probably a few more.

Type B Data Scientist: The B is for Building. Type B Data Scientists share some statistical background with Type A, but they are also very strong coders and may be trained software

engineers. The Type B Data Scientist is mainly interested in using data "in production." They build models which interact with users, often serving recommendations (products, people you may know, ads, movies, search results).

At Google, a Type B Data Scientist would typically be called a Software Engineer. Type B Data Scientists may use the term Data Scientist to refer just to themselves, and since the definition of the field is very much in flux, they may be right. But I see the term being used most often in the general way I am proposing here.

This categorization is crude. Many Data Scientists are some mix of A and B. But this answer is long enough already.

65.1k ViewsView Upvotes[53]Answer requested by William Chen[54]

Shlomo Argamon[55], Director of the Master of Data Science Program at Illinois Institute of Technology

Written Jul 17, 2014[56]Upvoted by Alex Blocker[57], Statistician at Google, PhD in statistics from Harvard, computational biologist much of the time.

Data Science is a real thing. It is more than just a collection of tools and methods for analyzing large and complex data sets, more than just collaboration between computer scientists and statisticians, and far more than just a fancy term for "statistics in San Francisco" or the like.

At its core, data science is about **making sense** of the world using data. This involves the use of various tools for data cleaning, structuring, querying, analysis, visualization, and the like; involves statistical modeling to some extent; often involves building new computational tools; involves figuring out the right relevant questions to ask the data, and how to explain what the data say (and don't say); and critically involves application of the scientific method to ensure that results are solid and properly understood (hence the word "science" in the name).

In a sense, this is a new word for work that scientists and statisticians have done for a long time, but the new abundance of data that carry information about virtually everything, together with the availability of tools (computational and statistical) to work with the data, makes the task of empirical sensemaking (i.e., data science) more broadly applicable and more important than before.

ViewsView Upvotes[58]

Ji Li[59], data scientist

Updated Mar 23, 2015[60]Upvoted by Shlomo Argamon[61], Director of the Master of Data Science Program at Illinois Institute of TechnologyRishabh Agnihotri[62], Masters Candidate, Masters of Data Science

I have been a data scientist for about two years. Here are some quick thoughts on what I think data science is. Or, why don't we start with what data science is not.

First, data science is not a software engineering piece of work. That is, data science is not about building products or product features or systems or any related fancy things.

Second, data science is not a visualization piece of work. Creating the cool visual is neither the end goal nor the beginning part of how a data scientist works. Needless to say, data science is not about creating visually impactful infographics.

Third, data science is not a scientific piece of work. In particular, data scientists don't work in the academia. It is the industry's particular requirements and the business markets' call that makes the job of data scientist needed. Data scientists usually don't publish papers, and neither is the paper or book publishing business part of any data scientists' daily concerns.

Last but not least, I don't agree with the public view that data science is, at least mostly, statistics. Just to cite a quick story of myself. Once I was asked to hire someone to assist my work and ended up interviewing lots of applicants through phone. Many of the applicants came from the filed of statistical analysis and most of these applicants tended to sound really confident that he or she would be more than qualified for the role. However, I didn't end up calling any of them on-site. One thing I realized at that time was that statistical knowledge alone doesn't make a person qualified for assisting me effectively on the kind of data science work that I needed to do, for reasons I'll mention in a short while.

Now, we are ready to talk about what data science is. It's a thing that encapsulates some programming skills, some statistical readiness, some visualization techniques, and, last but not least, a lot of business senses. The kind of business sense that I in particular care about is the ability and willingness, sometimes eagerness, to translate any business questions into questions answerable using currently or forthcomingly available data within one's reach. In fact, it takes a special way of connecting all the dots in the random world full of data most of which you may not find immediately useful to make a working data scientist.

A data scientist, based on my current understanding, is the person who connects the dots between the business world and the data world. Similarly, data science is the craft that a data scientist utilizes to make this happen.

I'm going to share a favorite analogue of mine about data science. Doing data science is like preparing a meal. One starts with data munging, which includes but is not restricted to ETL (extract, transform, and load), data cleansing, data debugging, etc. This is the step similar to preparing the food source, where you rinse cleans the vegetables, the meat, and the rice, chop the food source into reasonably sized pieces, and put them aside. After that is done, you are ready to cook the food source, which corresponds to data exploration, feature construction, feature reduction, running and ensembling the algorithms, etc. This is when you cook the vegetable and meat in a step-by-step fashion, adding ingredients and sources on particularly

calculated timing, and watching the raw material turn into edible pieces. The last step is to serve the food, when you arrange the cooked food in artistic ways and serve them in a particular sequence of first course, second course, etc, to customers who ordered the food to begin with. This is when you prepare your data mining results in artistic visualization and create reports or data stories to send to the business users who wanted this piece of data science work to be done on the first place.

Summarizing the above, the process of data science consists of data munging, data mining, and delivering actionable insights. Based on my own experience, a common toolset to get all or part of these done include Python, R, Tableau, SQL, etc.

Python is particularly handy as an all-purpose tool especially great for data munging. It can also be used for data mining, thanks to the almighty scikit-learn package, and even insight delivering based on its fast growing graphing abilities.

R is a bit shy on data munging compared to Python. However, because of its nature of being "statistically complete" - a word I just made up, meaning that any statistical thingie you have ever heard of is most likely already represented by a R package, or two - R is great for exploring the data and running algorithms on different parameter settings. This makes R a great tool for prototyping data science - for example, to identify the key feature set as well as a good enough machine learning algorithm with parameter setting, before you start to write complicated production code for "real". In addition to the above, R is also powerful with its visualization packages and can be used to turn a repeatable data mining piece into a shiny report. Talking about data visualization, Tableau is one of the best commercial software for visually explore your data. It is also handy for creating interactive visualization reports or data stories.

Besides Python, R, Tableau, there's one more data science tool that I want to mention before finishing this post. SQL is the language of English in the world of data munging, or at least have been so for a very long time. It is powerful in integrating different data sources, and handy for data exploration and data debugging.

These are just my two cents on what data science is. I hope it make sense to you so far. I'm still a learner, and merely a beginner in this field, and I expect to pick up a lot more and deeper understanding on this subject matter in the near future.

14.5k ViewsView Upvotes[63]

Akhil Sakhardande[64], Avid Reader. Enthusiast. Passionate. Always learning.

Written May 13, 2015[65]

Originally Answered: What is data science?[66]

Data science is a vast field which is primarily focused on making sense of the data you possess. It is the extraction of knowledge[67] from data[68]. It employs techniques and theories drawn from many areas of mathematics[69], statistics[70], information theory[71] and information technology[72], including probability models[73], statistical learning[74], data engineering[75], pattern recognition[76], visualization[77], predictive analytics[78], data warehousing[79], data compression[80] and high performance computing[81]. Methods that scale to Big Data[82] are of particular interest in data science, although the discipline is not generally considered to be restricted to such data.

The development of machine learning[83], a branch of artificial intelligence[84] used to uncover patterns in data from which predictive models can be developed, has enhanced the growth and importance of data science.

And some famous people has quoted '**Data Science is more art than science.**'
ViewsView Upvotes[85]Answer requested by Nilesh Jha[86]

Claudia Gold[87], Data Scientist at ClassDojo; Former Analytics Specialist at Airbnb

Written Jan 15[88]

Originally Answered: What the heck is data science really?[89]

It is a bit like digging in Excel, on some level, but yes much bigger and also more rigorous. Excel is limited to a few million rows, and even then it's very slow. Overall Excel offers a tiny subset of the things you can do to manipulate and analyze data.

In general data science and analytics are about looking for trends in data. The problem is, sometimes the datasets are too large (too many observations and/or too many pieces of information - variables - about each observation). Also, our brains want to find patterns naturally, so it's easy to think we've found a pattern when it's really just noise.

That means we need to use special tools and techniques to find patterns and separate them from the noise and handle the large quantities of data. You can also build more complex models to predict how different variables influence each other.

Using these kinds of tools also allows automation that is not possible in Excel, so you can update your analysis automatically over time. You can also connect it to the rest of the system. For example, data scientists might work on problems like recommending products someone might want to buy on Amazon. You couldn't do that in Excel, because it's not integrated with the rest of Amazon's code.

If you want to learn more about these kinds of tools and methods, R is a good place to start.

ViewsView Upvotes[90]

Michael E Driscoll[91], I ♥ data, analytics, & visualization

Updated Feb 15, 2011[92]Upvoted by Peter Skomoroch[93], Sr. Data Scientist @ LinkedInRodrigo Rivera[94], Data Science Lead at Rocket Internet

Data science, as it's practiced, is a blend of Red-Bull-fueled hacking and espresso-inspired statistics.

But data science is not merely hacking, because when hackers finish debugging their Bash one-liners and Pig scripts, few care about non-Euclidean distance metrics.

And data science is not merely statistics, because when statisticians finish theorizing the perfect model, few could read a ^A delimited file into R if their job depended on it.

Data science is the civil engineering of data. Its acolytes possess a practical knowledge of tools & materials, coupled with a theoretical understanding of what's possible.

Some worthwhile articles on the topic elsewhere:

Finally, it's worth noting: the job title 'Data Scientist' came into its recent vogue when Jeff Hammerbacher[95] coined the term at Facebook in 2007.
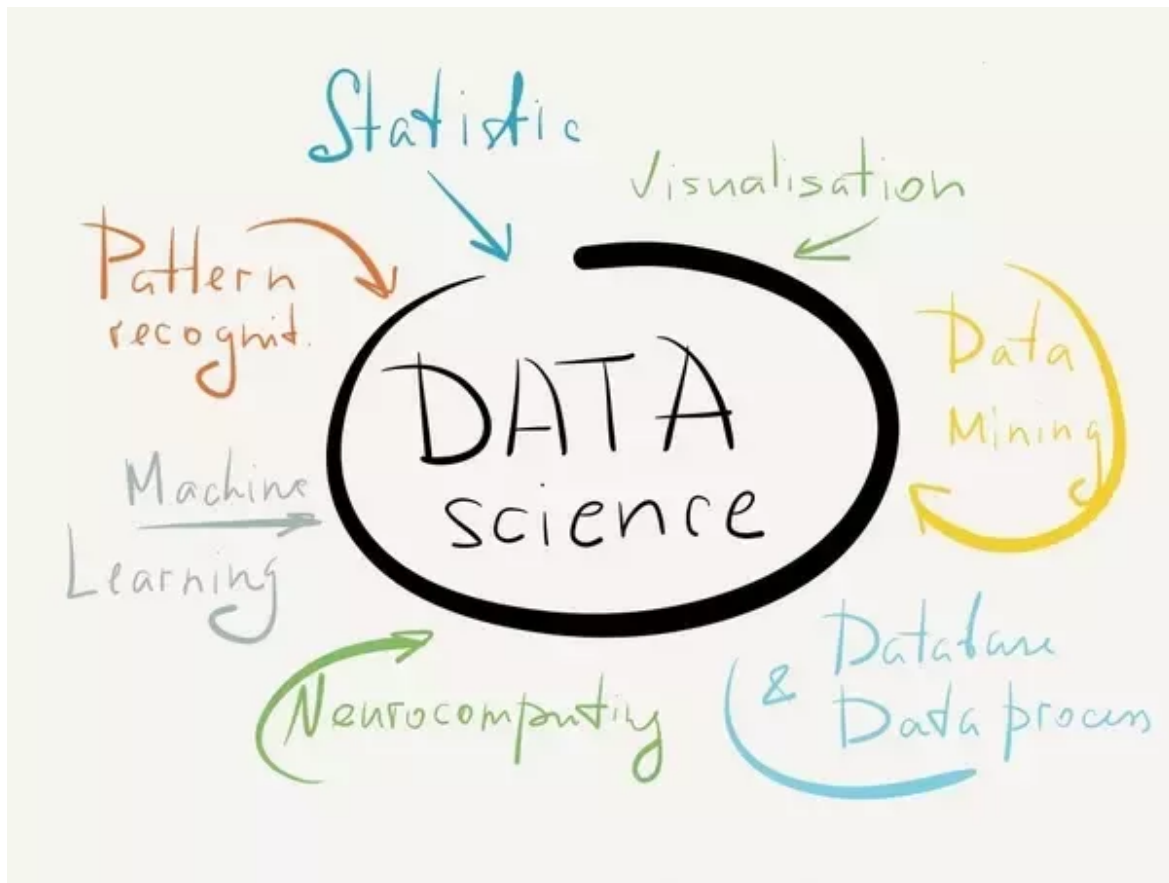23.2k ViewsView Upvotes[96]

I think the answers gives a great idea about Data Science , I would like to add different answer. 60 seconds answer

Some Quotes about Data Sciences:

* "Data scientists are involved with gathering data, massaging it into a tractable form, making it tell its story, and presenting that story to others." - **Mike Loukides, VP, O'Reilly Media**
* "A significant constraint on realizing value from Big Data will be a shortage of talent, particularly of people with deep expertise in statistics and machine learning, and the managers and analysts who know how to operate companies by using insights from Big Data." Big data: The next frontier for innovation, competition, and productivity[97], **McKinsey report**
* "Data is the new science. Big data holds the answers." – **Pat Gelsinger, CEO, EMC**, Big Bets on Big Data[98], Forbes

Ref : Quotes about Data Science[99]

Sum-up in image

ViewsView Upvotes[100]

During my work at McKinsey I have seen many companies in which the analytical and statistical roles were put in their own silo separate from the "business" side of the company. This makes some sense if you have experts for a certain aspect of your business, like actuaries computing the price of an insurance or statisticians evaluating the phase III trial of a drug. But if you try to inform day-to-day business decisions using data, a lot is lost in translation.

As data and technology become more and more available and companies can profit more and more by informing their decisions by it, there is a growing need to bridge that gap. For this you need a different kind of employee who combines strong analytical skills with strong business acumen. And this is exactly the role I filled during my McKinsey projects and now at Airbnb and how I generate an enormous amount of value.

Now given the description of the role it is useful to have a label which allows you to talk about it. I wasn't called a data scientist during my time at McKinsey, but what I did exactly matches the definition of a data scientist by IBM (Bringing big data to the enterprise[101]), so I happily picked this label up and applied it to myself.

If you look at the other answers to this question you will discover that there are a lot of people who use this label in a different sense, so it may already have outlived its usefulness. But even if the label turns out to be a fad, the role is here to stay.

ViewsView Upvotes[102]

---- Links ----

1. https://hail-data.quora.com/What-is-Data-Science

2. https://www.quora.com/api/mobile_expanded_voter_list?key=33coFRbeeeQ&type=answer

3. http://extension.harvard.edu/

4. https://www.reskill.me/skills/python

5. https://www.udemy.com/learn-python-programming-from-scratch/?dtcode=Y8tHdUo2GWxI

6. https://www.codecademy.com/tracks/python

7. http://www.learnpython.org/

8. https://www.reskill.me/skills/machine-learning

9. https://www.coursera.org/learn/machine-learning

10. https://www.udemy.com/operational-intelligence-and-machine-data-with-splunk/

11. https://www.reskill.me/skills/r

12. https://www.udemy.com/r-basics/?dtcode=GLwzuIc2GWEW

13. https://www.datacamp.com/courses/free-introduction-to-r

14. https://www.codeschool.com/courses/try-r

15. https://www.reskill.me/skills/big-data

16. http://bigdatauniversity.com/

17. https://www.udemy.com/big-data-and-hadoop-essentials-free-tutorial/?dtcode=Z0cAVQ02GWtu

18. https://www.udemy.com/overview-of-big-data-hadoop/?dtcode=TCMfbcJ2GWK6

19. https://www.reskill.me/skills/statistics

20. https://www.coursera.org/course/stats1

21. http://stattrek.com/

22. http://oli.cmu.edu/courses/free-open/statistics-course-details/

23. https://www.reskill.me/skills/data-mining

24. https://www.udemy.com/scraping-and-data-mining-for-beginners-and-pros/?dtcode=OiThMBJ2GWAW

25. https://www.coursera.org/course/mmds

26. https://www.reskill.me/skills/sql

27. http://www.sqlcourse.com/

28. https://www.udemy.com/sachin-quickly-learns-sql/?dtcode=35dX3aC2GWAW

29. http://www.w3schools.com/sql/

30. https://www.reskill.me/skills/java

31. https://www.udemy.com/java-tutorial/?dtcode=aCbOkF12GW2n

32. http://www.learnjavaonline.org/

33. https://www.udemy.com/learn-java-programming-from-scratch/?dtcode=quVznwK2GWwL

34. http://blog.reskill.me/how-to-become-data-scientist-for-free/

35. https://www.quora.com/api/mobile_expanded_voter_list?key=FSr8xxYrvFm&type=answer

36. https://www.quora.com/profile/Luis-Otavio-Martins

37. https://www.quora.com/What-is-data-science/answer/Luis-Otavio-Martins

38. https://www.quora.com/What-is-a-data-scientist-3

39. http://www.livescience.com/20896-science-scientific-method.html

40. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

41. http://www.ohadsamet.com/2012/08/26/whats-missing-in-data-science-talks/

42. https://www.quora.com/api/mobile_expanded_voter_list?key=MlIqPCp8oiA&type=answer

43. https://www.quora.com/profile/Drew-Conway

44. https://www.quora.com/What-is-data-science/answer/Drew-Conway

45. https://www.quora.com/profile/William-Chen-6

46. http://radar.oreilly.com/2010/06/what-is-data-science.html

47. https://www.quora.com/api/mobile_expanded_voter_list?key=LTovHvQx7H8&type=answer

48. https://www.quora.com/profile/Michael-Hochster

49. https://www.quora.com/What-is-data-science/answer/Michael-Hochster

50. https://www.quora.com/profile/Jerrod-Lowmaster

51. https://www.quora.com/profile/Boxun-Zhang

52. https://www.quora.com/profile/Alon-Amit

53. https://www.quora.com/api/mobile_expanded_voter_list?key=biC7VcNQHNY&type=answer

54. https://www.quora.com/profile/William-Chen-6

55. https://www.quora.com/profile/Shlomo-Argamon-1

56. https://www.quora.com/What-is-data-science/answer/Shlomo-Argamon-1

57. https://www.quora.com/profile/Alex-Blocker

58. https://www.quora.com/api/mobile_expanded_voter_list?key=86EmfYSBILi&type=answer

59. https://www.quora.com/profile/Ji-Li-1

60. https://www.quora.com/What-is-data-science/answer/Ji-Li-1

61. https://www.quora.com/profile/Shlomo-Argamon-1

62. https://www.quora.com/profile/Rishabh-Agnihotri

63. https://www.quora.com/api/mobile_expanded_voter_list?key=q98JH0bcISF&type=answer

64. https://www.quora.com/profile/Akhil-Sakhardande

65. https://www.quora.com/What-is-data-science/answer/Akhil-Sakhardande

66. https://www.quora.com/What-is-data-science-3?no_redirect=1

67. http://en.wikipedia.org/wiki/Knowledge

68. http://en.wikipedia.org/wiki/Data

69. http://en.wikipedia.org/wiki/Mathematics

70. http://en.wikipedia.org/wiki/Statistics

71. http://en.wikipedia.org/wiki/Information_theory

72. http://en.wikipedia.org/wiki/Information_technology

73. http://en.wikipedia.org/wiki/Probability_models

74. http://en.wikipedia.org/wiki/Statistical_learning_theory

75. http://en.wikipedia.org/wiki/Data_engineering

76. http://en.wikipedia.org/wiki/Pattern_recognition_and_learning

77. http://en.wikipedia.org/wiki/Data_visualization

78. http://en.wikipedia.org/wiki/Predictive_analytics

79. http://en.wikipedia.org/wiki/Data_warehousing

80. http://en.wikipedia.org/wiki/Data_compression

81. http://en.wikipedia.org/wiki/High_performance_computing

82. http://en.wikipedia.org/wiki/Big_Data

83. http://en.wikipedia.org/wiki/Machine_learning

84. http://en.wikipedia.org/wiki/Artificial_intelligence

85. https://www.quora.com/api/mobile_expanded_voter_list?key=R7Kn0YqaBX2&type=answer

86. https://www.quora.com/profile/Nilesh-Jha-8

87. https://www.quora.com/profile/Claudia-Gold

88. https://www.quora.com/What-is-data-science/answer/Claudia-Gold

89. https://www.quora.com/What-the-heck-is-data-science-really?no_redirect=1

90. https://www.quora.com/api/mobile_expanded_voter_list?key=1z2yVFKCAQP&type=answer

91. https://www.quora.com/profile/Michael-E-Driscoll

92. https://www.quora.com/What-is-data-science/answer/Michael-E-Driscoll

93. https://www.quora.com/profile/Peter-Skomoroch

94. https://www.quora.com/profile/Rodrigo-Rivera

95. https://www.quora.com/profile/Jeff-Hammerbacher

96. https://www.quora.com/api/mobile_expanded_voter_list?
key=ah7EVOVP4mA&type=answer

97. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

98. http://www.forbes.com/sites/ciocentral/2012/06/22/big-bets-on-big-data/?
goback=.gde_2013423_member_127364385

99. http://www.statistics.com/landing-page/data-science/data-science-quotes/

100. https://www.quora.com/api/mobile_expanded_voter_list?key=aaETVgJe34D&type=answer

101. http://www-01.ibm.com/software/data/infosphere/data-scientist/

102. https://www.quora.com/api/mobile_expanded_voter_list?key=Ypk0tk9HQy1&type=answer