

Different roles in data science field

Leiver Jiménez, Adrián Solís

*Computer Science student, Tecnológico de Costa Rica.
Santa Clara, San Carlos, Alajuela, Costa Rica.
leiver38hfh@gmail.com*

*Computer Science student, Tecnológico de Costa Rica.
Santa Clara, San Carlos, Alajuela, Costa Rica.
adriansolisz@outlook.com*

Abstract — This document summarizes the different roles in Data Science, which were established until now by the professionals of this field. Tasks for these roles will be explained in depth, in order to provide a general idea of what are the responsibilities for each professional, what are the skills required and some of the technology-based tools they use.

Keywords—Big Data, Data Science, Data Science Jobs, Data Science Career Path, Data skills.

I. INTRODUCTION

“This new field seems to be aiming to revolutionize industries from business to government, health care to academia”, says the *New York Times*. People have woken up to the fact that without analyzing the massive amounts of data that is at their disposal and extracting valuable insights, there really is no better way to get the most of the valuable information that is constantly in front of them, so close but so invisible.

Gaining fame as the most promising profession of the century, data science needs business savvy people who have listed data literacy and strategic thinking as their key skills. Anjul Bhambri, VP of Architecture at Adobe, says, “A Data Scientist is somebody who is inquisitive, who can stare at data and spot trends. It’s almost like a Renaissance individual who really wants to learn and bring change to an organization.”

Now, how is it that we get value from these tons of data in every sector in the economy? Well, we get persistent and data-mad personnel skilled in math, stats, and programming to weave magic using combinations of letters and numbers.

Over the last few years, people have moved away from what was considered a unique term: data scientist. Companies now are talking about and looking for a completely new set of job roles such as data engineers, data architects, business analysts,

MIS reporting executives, statisticians, machine learning engineers, and big data engineers.

Becoming a data scientist varies depending on industry, but there are common skills, experience, education and training that will give you the leg up in starting your data science career.

In this article, you will get a quick overview about these exciting positions in the field of analytics. However, do remember that companies often tend to define job roles in different ways based on the inner workings rather than market descriptions.

II. ROLES

i. BUSINESS ANALYST

In this role, many of the tasks make it a bit similar to Data Analyst; nevertheless, their specialty is the domain they work in. Their main objective is to narrow the gap between business and IT, now, how do they do it (or at least try it)? Well, they perform the task of understanding the business change needs, assessing the business impact of those changes, capturing, analyzing and documenting requirements, then they provide (mostly technology-based) solutions aiming to enhance business processes such as distribution or productivity.

Organizations need these “information bridges” for tasks like gap analysis, requirements gathering, knowledge transfer to developers, defining scope using optimal solutions, test preparation, and software documentation.

An academic profile which often postulates and actually performs properly in this position is the combination of a Computer Science major with additional courses that include statistics, organizational behavior, and quality management

Skills: Data Visualization, Data Modelling, web intelligence, business objects design.

Tools: Tableau, dashboard tools, SQL, MS Office.



ii. DATA ANALYST

Unofficially called “junior data scientists”, they perform somewhat similar tasks to them; however, they are more of generalists, and, being honest, their role is significantly wide: from acquiring massive amounts of data to processing and summarizing it.

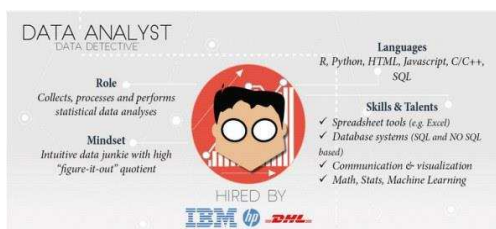
This profile requires to master a lot more technology-based tools than Business Analyst, Data Analysts are supposed to be more than familiar with data retrieval and storing systems, data visualization and warehousing using ETL tools, Hadoop-based analytics and business intelligence concepts. Besides programming, these data miners usually have a strong background in math, statistics and machine learning.

They are in charge of data that is scraped, assuring the quality and managing it. They have to interpret data and effectively communicate the findings.

Optimization is a must-know skill for a data analyst. Designing and deploying algorithms, culling information and recognizing risk, extrapolating data using advanced computer modeling, triaging code problems, and pruning data are all in a day’s work for a data analyst.

Skills: statistics, data munging, data visualization, exploratory data analysis.

Tools: R, Python, SQL, C++, Javascript.



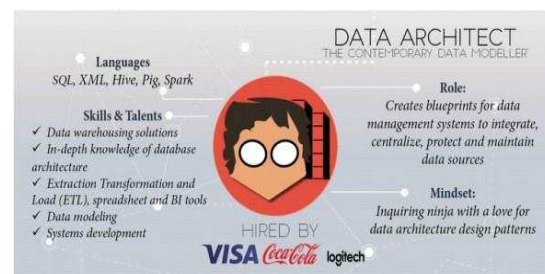
iii. DATA ENGINEER/ARCHITECT

Data engineers are precisely the ones who prepare the “Big Data” infrastructure to be analyzed by Data Scientists, in other words, software engineers who design, build and integrate data from different sources and manage big data. Besides, they are capable of writing complex queries on that, while trying to make sure data is easily accessible, works smoothly, with the main goal of optimizing their companies’ big data “environment”, meaning keep the Big Data ecosystem running without glitches for data scientists to carry out the analysis. .

They are fully equipped with knowledge of hardware, databases, data processing at scale and computer engineering, which allows them to not only use, but also implement productions tools. It is important as well, to mention that because Data Engineers focus more on the design and architecture, the most common scenario is that they are not expected to know any machine learning or analytics for big data.

Skills: Hadoop, MapReduce, Hive, Pig, Data streaming,

Tools: DashDB, SQL, MongoDB, Cassandra NoSQL, programming.



iv. STATISTICIAN

Statisticians collect, organize, present, analyze, and interpret data to reach valid conclusions and make correct decisions. They are key players in ensuring the success of companies involved in market research, transportation, product development, finance, forensics, sport, quality control, environment, education, and in governmental agencies. Many statisticians continue to enjoy their place in academia and research.

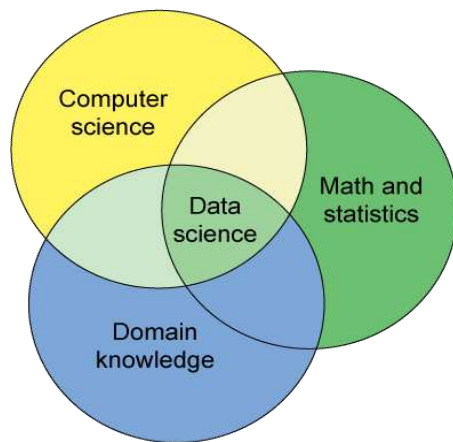
Typically, statisticians who want to immerse in Data Science world, should go for higher degrees in statistics, mathematics, or any quantitative subject. They need to be “mini-experts” of the industries they choose to work in. And is highly recommended to be familiar with R programming, MATLAB, SAS, Python, Stata, Pig, Hive, SQL, and Perl.

Their responsibilities are among using statistical analysis software tools, statisticians analyze collected or extracted data, trying to identify patterns, relationships, or trends to answer data-related questions posed by administrators or managers. They interpret the results, along with strategic recommendations or incisive predictions, using data visualization tools or reports.

In some cases, statisticians who do not know that much about the previously mentioned technologies, play an essential role as Data Modeling consultants for Data architects with not so strong statistics background.

Skills: Higher degrees in statistics suggested. DBMS, Statistical Software.

Tools: SQL, R, Python, MATLAB, SAS.



v. DATA SCIENTIST

A data scientist is the alchemist of the 21st century, someone who is capable of turning raw data into purified insights. These amazing thinkers apply statistics, machine learning and analytic approaches to solve critical business problems (or not problems). Their primary function is to help organizations make their volumes of big data become valuable and actionable insights.

This title holds stronger statistics and presentation skills than a data analyst and data engineer. A data scientist would have strong skills of Inferential Statistics, Machine Learning, Data Analysis, Data Communication.

Indeed, data science is not necessarily a new field itself, but it can be considered as an advanced level of data analysis that is driven and automated by machine learning and computer science. In another word, in comparison with 'data analysts', in addition to data

analytical skills, Data Scientists are expected to have strong programming skills, an ability to design new algorithms, handle big data, with some expertise in the domain knowledge.

Moreover, Data Scientists are also expected to interpret and eloquently deliver the results of their findings, by visualization techniques, building data science apps, or narrating interesting stories about the solutions to their data/business problems.

The problem-solving skills of a data scientist go beyond all they seem to reach, it requires an understanding of traditional and new data analysis methods to build statistical models or discover patterns in data. For example, creating a recommendation engine, predicting the stock market, diagnosing patients based on their similarity, or finding the patterns of fraudulent transactions.

Sometimes, Big Data can be presented to a Data Scientist without a particular business problem in mind. In this case, expectations are that this curious friend, come up with the right questions, and provide interesting findings. This is tricky because, in order to analyze the data, a strong Data Scientists should have a very broad knowledge of different techniques in machine learning, data mining, statistics and big data infrastructures.

Now, having in mind that Data Science is precisely the main topic in this paper, it is relevant to be more specific about some aspects that, no matter if you want them or not, come together with the Data Scientist charge.

Data scientist requirements

Each industry has its own big data profile for a data scientist to analyze. Here are some of the more common forms of big data in each industry, as well as the kinds of analysis a data scientist will likely be required to perform, according to the Bureau of Labor Statistics (BLS).

Business: Today, data shapes the business strategy for nearly every company — but businesses need data scientists to make sense of the information. Data analysis of business data can inform decisions around efficiency, inventory, production errors, customer loyalty and more.

E-commerce: Now that websites collect more than purchase data, data scientists help e-commerce businesses improve customer service, find trends and develop services or products.

Finance: In the finance industry, data on accounts, credit and debit transactions and similar financial data are vital to a functioning business. But for data scientists in this field, security and compliance, including fraud detection, are also major concerns.

Government: Big data helps governments form decisions, support constituents and monitor overall satisfaction. Like the finance sector, security and compliance are a paramount concern for data scientists.

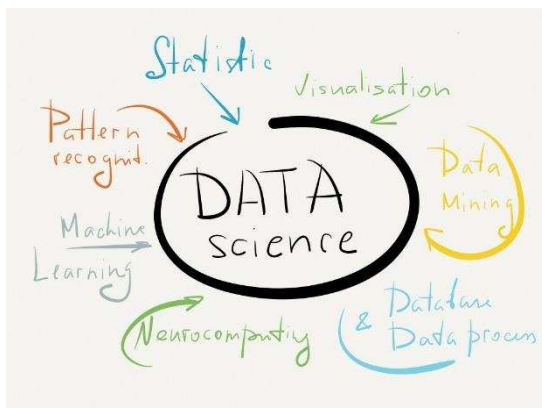
Science: Scientists have always handled data, but now with technology, they can better collect, share and analyze data from experiments. Data scientists can help with this process.

Social networking: Social networking data helps inform targeted advertising, improve customer satisfaction, establish trends in location data and enhance features and services. Ongoing data analysis of posts, tweets, blogs and other social media can help businesses constantly improve their services.

Healthcare: Electronic medical records are now the standard for healthcare facilities, which requires a dedication to big data, security and compliance. Here, data scientists can help improve health services and uncover trends that might go unnoticed otherwise.

Telecommunications: All electronics collect data, and all that data needs to be stored, managed, maintained and analyzed. Data scientists help companies squash bugs, improve products and keep customers happy by delivering the features they want.

Yes, we can say that basically ANY industry or field can get a lot of benefits from Data Science, and yes, it seems to be a pretty cool job this is completely the opposite to what many would like to think: something far away from “easy to achieve”, just like we can see in the requirements description above.



Data scientist skills

William Chen, Data Scientist at Quora (a platform similar to Yahoo! Answers, long story short), states there are five top skills a Data Scientist must have, and also that these five skills are a mix of soft and hard skills:

Programming: The most fundamental skill set. It improves your statistics skills, helps you “analyze large datasets” and provides you with the ability to develop your own tools.

Quantitative analysis: Quantitative analysis will improve your ability to run experimental analysis, scale your data strategy and places you closer to machine learning.

Product intuition: Understanding products will help you perform quantitative analysis. It will also help you predict system behavior, establish metrics and improve debugging skills.

Communication: Possibly the most important soft skills across every industry, strong communication will not only let you express yourself more clearly, but will also provide you with a much more powerful resource to learn and find “good teachers” among your co-workers, friends, relatives and anyone around you, faster.

Teamwork: Much like communication, teamwork is vital to a successful data science career. It requires being selfless, properly accepting and addressing feedback and sharing your knowledge with your team.

How about now that you have read this far? Would you like to become a Data Scientist?

In any case, it is important to talk about the suggested path to follow once you have received and processed the essential education and training.

There must always be an initiation, so start by applying for open positions with requirements that match and not exceed your level, meaning it is always good to aim high, but keep your feet firm on the ground. Then, keep yourself in constant learning, as any other discipline, keeping updated is one of the keys to excellence, and finally, never give up on your goal, and if you want a Data Scientist job but feel you are missing some skills, map your own strategy to get there, no matter how long it takes, never quit.

III. CONCLUSIONS

- Data Science is an amazingly promising field; nevertheless, it takes some hard work to get there, having in mind all the academic preparation, the cross-discipline topics to learn about, and even the long interview processes, it is worth the effort, but it is some real work in there.
- Playing a role in this field, requires more than computer science, or in some cases, it does not even require any computer science knowledge. This means that the fact that you have a computer science or any related major cannot be understood as you are ready to be a data scientist, the same applies for statisticians, there is a large amount of work for these professionals after their major before starting themselves in Data Science.
- About the roles and the different ways to define each of them, and the fact that there is still a lot of distinct perspectives, is understandable, mainly because it is clear that the confusion stems from two reasons: One, different companies have different ways of defining the roles. And second, data science is yet a newborn field and not everyone is familiar with the inner workings of the industry, so it is still an early time to ask for highly accurate definition.

IV. BIBLIOGRAPHY

Team Machine Learning. (2017, March 6). 8 different job roles in Data Science/Big Data. Retrieved October 15, 2017, from <http://blog.hackerearth.com/8-different-job-roles-data-science-big-data-industry>

Kalyan. (2017, May 12). Different job roles in Data Scienc and Analytics Industry. Retrieved October 15, 2017, from <http://makemeanalyst.com/different-job-roles-in-data-science/>

Huang, R. (2016, March 28). Data Science Career Paths: Different Roles in the Industry. Retrieved October 15, 2017, from <https://www.springboard.com/blog/data-science-career-paths-different-roles-industry/>

Heller, M. (2017, August 7). 10 hot data analytics trends — and 5 going cold. Retrieved October 25, 2017, from https://www.cio.com/article/3213189/analytics/10-hot-data-analytics-trends-and-5-going-cold.html#tk.cio_rs

Aghabozorgi, S. (2016, June 06). Data Scientist vs Data Engineer, What's the difference? Retrieved October 25, 2017, from <https://cognitiveclass.ai/blog/data-scientist-vs-data-engineer/>

Huang, R. (2016, August 16). What we learned analyzing hundreds of data science interview. Retrieved from <https://www.springboard.com/blog/data-science-interviews-lessons/>