ANALYST NOTE

DIAZ RESEARCH

This PDF illustrates the Analyst Note format.

Actionable analysis from Diaz Research

Data Scientist jobs

Diaz Research Ref: 7087
Filename: 7087 Data Scientist jobs
Date: July 2016

Data scientists have attracted much publicity in recent years. Many of the accounts of what they do have revealed a worryingly wide range of interpretations, raising questions about the whole field.

Now, in 2016, things have started to stabilize and job ads have seen a degree of convergence towards something that is credible. This Note presents work started in 2015, intended to set out what these jobs are about, and which has recently been tested and adapted in the light of the kinds of data scientist positions advertised by large companies today. These observations job profiles, and other material will be helpful to those looking to set out their requirements or set up a career structure, or simply to test their own ideas against external sources.

Diaz Research Ltd, 3rd Floor, Regent's Place, 338 Euston Road, London NW1 3BT, U.K.
T: +44 (0)20 7544 8692 E: research@diazresearch.com W: diazresearch.com

DIAZ RESEARCH

The rediscovery of numeric data

OVERVIEW

The **Lead Data Scientist** is a Google-level job.

The **Data Scientist** is a very solid mainly statistics-based job that also requires some broader applied mathematical knowledge.

The **Data Analyst** is concerned with statistics and does not need a wide range of applied mathematical knowledge.

This note sets out three job profiles for the kind of jobs that are typically labelled as data scientists. These are the jobs needed when conducting advanced analytics, including Big Data, pattern recognition and extending as far as machine learning. The box on the left provides an overview of the set of three jobs.

This Note is intended to form a useful resource when generating, or critiquing, job profiles. But the profiles may also be used to benchmark pay, develop career structures, etc. In addition, some explanatory material relating to data science and data science jobs is included.

The data science discipline

'Data' has long been a core part of the IT scene, but not in the sense used when talking about data scientists. The use of computers to perform tough calculations, so important in the 1950s and 1960s, was soon secondary to business management applications. So 'data' has long meant what is stored and manipulated in databases, much of it text-based, not numerical. Numerical complexity has been rare in corporate IT's systems.

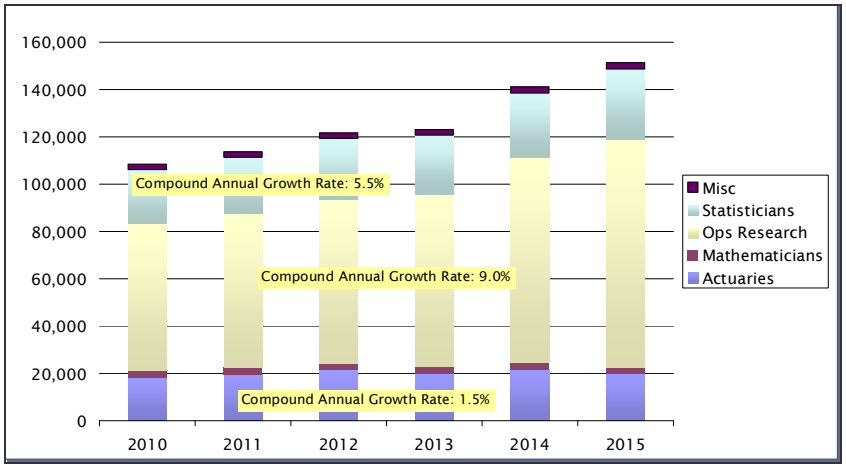
Today's data science is about advanced calculation and analysis, much of which (not all) is statistical in nature. The French term 'numérique' more accurately captures the essence of the discipline. This discipline fuelled the rise of search engines (Google is just one example), predictive analytics as practised by Amazon and many others, revenue maximization through dynamic pricing (e.g. by airlines and hotels), trading strategies, and more. As more and more data becomes available, more companies are realizing the opportunity that data science represents.

Data science headcount is growing

It is not easy to establish how many data scientists there are. The term does not appear in occupational statistics, and of the three countries providing accessible occupation-based headcounts (Switzerland, UK and USA) only the last has a category that approximates to this kind of job. There, the occupational category is **Operations Research**, a term that may be unfamiliar to readers. But the US Bureau of Labor Statistics' definition fits data science rather well:

> *'Formulate and apply mathematical modeling and other optimizing methods to develop and interpret information that assists management with decision making, policy formulation, or other managerial functions. May collect and analyze data and develop decision support software, service, or products. May develop and supply optimal time, cost, or logistics networks for program evaluation, review, or implementation.'*

The US job numbers for this and other mathematical occupations is very revealing:



US job statistics for mathematical occupations

Source: US Bureau of Labor Statistics, 2016

In short, actuaries and statisticians numbers are very stable, but the number of those engaged in Operations Research, the application of mathematical science to business operations, is growing steadily at about 9% a year. There are now 100,000 more such jobs than there were 5 years ago in the US.

This firmly corroborates all the other evidence that the data science is indeed taking off. It may be some time before better estimates can be collected but this data gives assurance, if any were needed, that Data Science is indeed a growing and significant area.

The job profiles

The three profiles are given in the following pages, but these comments should be helpful in using them.

Pitching your job profiles vs benchmark job profiles

The three benchmark jobs (labelled A, B and C) are very distinct and as with all benchmarks there will be some real-world jobs that match them closely, and others that are pitched between them. Such intermediate job matches can be labelled as A-, or B+, etc. if that is helpful.

Typical occurrence of these jobs in today's companies

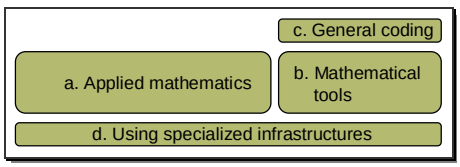
It is not easy to say exactly how common these jobs are today. Using the US figures we could deduce that a company with an IT function might *on average* have about one such person for every 30 or so IT people. But the dominance of analytic data science in some sectors means that most companies will have far fewer. Some will have many more. Many companies can be expected to have jobs mostly at Job C level but with one or two at Job B level. The existence of even one full Job A position is indicative of some very clever stuff underway in the company concerned, e.g. High Frequency Trading, image recognition and processing, etc.

Content of benchmark job profiles vs your job profiles

Real job ads for technical jobs tend to mention many technologies. This may be because hiring managers hope, against all experience, that perfect-fit candidates will, this time round, materialize. (The fact that such over-specification may put off some excellent candidates is overlooked.) To some readers, then, the profiles may look somewhat light on detail. Feel free to add essential requirements, but with too many, good applicants will be lost.

The specialized knowledge and skills areas

These, in the three job profiles, are listed in decreasing order of importance. They may be understood better by reference to the picture below, where the size of the boxes reflect the general importance of the knowledge:



The core mathematical knowledge (a.) has to be educated in; the rest can be trained in or picked up in the job.