What are the different skill sets required to work as a Data Analyst or a Data Scientist?

Author, Mayank Tripathi

A Data Science Foundation Blog

February 2020

----

www.datascience.foundation

If you have a mathematical background, can write code or have an analytical brain, then you might find a rewarding career working with data. You will often have heard the term, data scientist and will have read that they are in high demand and are paid well. Well data scientists do not work alone and there are many roles, including junior ones, that need to be filled before the data can do her job. However, in this article we will focus on just the roles of data scientist, a senior position and data analyst, a mid-ranking position.

One must be good at Statistics, Mathematic, Machine Learning and Visualizing, as well as being a good communicator and business operator be a good Data Scientist. Before you have developed all these skills you might want to aim lower and allow yourself time to develop by working as a Data Analyst. Being a data analyst is a great career move and a rewarding and well paid position.

Below are the some of the skills which are required to become a Data Analytics. Skills which all Data Scientists are expected to have mastered. Once you are in the role of Data Analyst, then you can create a pathway to become a Data Scientist.

What a Data Analyst is expected to do

1. How to manage and collect data
2. How to clean and prepare data
3. How to process and evaluate data
4. How to create unbiased insight
5. How to use Visualization tools

Some of the Skills required to become a Data Analyst.

- **Excel / Spreadsheet** Allows you to do a quick data analysis without having to write code. Some of the basic operations required frequently are
    - Using Formulas
    - Vlookup
    - Aggregate functions (sum; counts etc)
    - Data filter
    - Writing own function.
    - Charts and plots
    - Transpose data into Table
    - Pivot table

- **Database Knowledge** This is required to store and analyze data.

SQL As a Data Analyst, one has to run many SQL queries. Most of the time data resides in multiple databases like Microsoft SQL Server; Oracle; DB2; PostgreSQL; dBASE; MySQL; TeraData etc. You will need SQL to interact with the data held in these databases. In SQL its good if you have

- Basics of Relational Database (RDS)
- Basics of No-SQL
- Clauses
- JOINS
- Views

- **Programming Language** Knowing one programming language is beneficial, but not necessary. Although I recommend having at least basics of Python, so that in near future one can begin to consider the role of Data Scientist. Python is easy to start with compared to other programming languages. Python supports many basic functions and has data analysis libraries; scientific libraries; read data from file; write data into file; or connect to Database and get data; data visualization etc.
  - SAS
  - Python
    - Variables, Datatypes
    - Strings; List; Dictionary
    - Loops
    - IF; IF ELSE conditions
    - Functions
    - Read or Write into a file. File can be a text; csv; excel or any other format.
    - Basic understanding of various libraries like numpy; pandas.
  - R
  - JAVA
  - SCALA

- **Data Visualization** A Data Analyst uses data visualization to identify patterns in data; distribution of data; to help with a general understanding the of data. Below are the tools we can use.
  - Tableau
  - Microsoft Power BI
  - Qlik
  - Libraries like matplotlib and Seaborn to be used in Python for data visualization.

- **Communication** A good Data Analyst or Data Scientist does not just play with data, they have to explain the insights to business and present meaning in the data in a way the audience understands. For this one has to be a good at communicator.

One has to be effective communicator and know how to present the data's story well.

- Story Telling
- Showcase the data

Some of the Skills required to become Data Scientist.

Along with the above skills discussed for Data Analysts below are additional skills required of for Data Scientists.

- **Statistics** Learn statistics, probability and mathematical analysis. These skills will help Data Scientists to interpret the data.
- **Programming Tools** Already discussed above, Data Scientists must be skilled in multiple programming languages
  - R is a free software environment for Statistical computing and graphics. R supports most Machine Learning Algorithms for Data Analytics like Regression, Association, Clustering etc.
  - Python is an open source general purpose programming language. Python libraries like Numpy; Pandas and SciPy are used by most Data Scientists.
  - SAS also support statistical analysis on data and supports data retrieval from a variety of sources.
- **Data Wrangling** Data Scientist must know how to wrangle the data. At times Data Analysts may not be involved in the Data Wrangling process, thus its Data Scientist responsibility, if not to perform this, then to manage the team that does. However, most data scientists will tell you that this work takes up a lot of their time.

Data Wrangling involves

  - Cleaning Data
  - Manipulating Data
  - Organizing Data.

Some of the tools which one can use for data wrangling are R; Python; etc, thus having knowledge of python or R is very important.

- **Data Visualization** Visualizing data transforms a hard to comprehend set of numbers into something more pleasing to the eye and much easier to understand. Visualizations could be various sorts of diagrams, charts or graphs. There are many easy to use software packages on the market, these include
  - Tableau
  - QlikView
  - PowerBI
  - Google Data Studio

- **Big Data** Data Scientists should have a good knowledge of working with massive data sets and the tools required to handle Big Data Big Data. Large and complex data sets can't be processed by traditional data software. Some of the more popular big data tools are
  - Spark

- Hadoop
- Pentaho
- Talend
- Tableau
- Splunk
- Cassandra

- **Machine Learning** One of the principle skills of a data scientist. Once the data has been prepared, models created and algorithms trained. Computer systems can perform tasks without specific instructions which lead to the detection of patterns in data and the creation of insight.

Machine Learning can be achieved through various algorithms such as Regression; Naïve Bayes; SVM; Decision Tree etc.

- Deep Learning
- Predictive Modeling to increase and optimize customer experiences, revenues generation, ad targeting etc.

Becoming a data scientist is the goal of many working in the data sector. It is highly skilled and requires dedication and experience. If you are new to the sector, set your sites on the role of data analyst, master this role and then see where it takes you.

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: admin@datascience.foundation
Telephone: 0161 926 3641
Atlantic Business Centre
Atlantic Street
Altrincham
WA14 5NQ
web: www.datascience.foundation