Data scientist roles and responsibilities include:

* Data mining or extracting usable data from valuable data sources
* Using [machine learning tools](https://www.simplilearn.com/best-machine-learning-tools-article) to select features, create and optimize classifiers
* Carrying out preprocessing of structured and unstructured data
* Enhancing [data collection](https://www.simplilearn.com/what-is-data-collection-article) procedures to include all relevant information for developing analytic systems
* Processing, cleansing, and validating the integrity of data to be used for analysis
* Analyzing large amounts of information to find patterns and solutions
* Developing prediction systems and machine learning algorithms
* Presenting results in a clear manner
* Propose solutions and strategies to tackle business challenges
* Collaborate with Business and IT teams

## Data Scientist Skills

To pursue data scientist careers, you need to master the skills required for data scientist jobs in various industries and organizations. Let’s take a look at the must-have data scientist qualifications.

Key skills needed to become a data scientist:

* Programming Skills – knowledge of statistical programming languages like R, [Python](https://www.simplilearn.com/why-learn-python-a-guide-to-unlock-your-python-career-article), and database query languages like [SQL](https://www.simplilearn.com/tutorials/sql-tutorial/what-is-sql), Hive, Pig is desirable. Familiarity with Scala, [Java](https://www.simplilearn.com/tutorials/java-tutorial/learn-java), or [C++](https://www.simplilearn.com/c-plus-plus-programming-for-beginners-article) is an added advantage.
* Statistics – Good applied statistical skills, including knowledge of statistical tests, distributions, regression, maximum likelihood estimators, etc. Proficiency in statistics is essential for data-driven companies.
* [Machine Learning](https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article) – good knowledge of machine learning methods like k-Nearest Neighbors, Naive Bayes, SVM, Decision Forests.
* Strong Math Skills (Multivariable Calculus and Linear Algebra) - understanding the fundamentals of Multivariable Calculus and Linear Algebra is important as they form the basis of a lot of predictive performance or algorithm optimization techniques.
* Data Wrangling – proficiency in handling imperfections in data is an important aspect of a data scientist job description.
* Experience with [Data Visualization Tools](https://www.simplilearn.com/data-visualization-tools-article) like matplotlib, ggplot, d3.js., Tableau that help to visually encode data
* Excellent Communication Skills – it is incredibly important to describe findings to a technical and non-technical audience.
* Strong Software Engineering Background
* Hands-on experience with data science tools
* Problem-solving aptitude
* Analytical mind and great business sense
* Degree in Computer Science, Engineering or relevant field is preferred
* Proven Experience as [Data Analyst](https://www.simplilearn.com/what-does-a-data-analyst-do-article) or Data Scientist