

# Data Analytics Basics

A Beginner's Guide

---



# TABLE OF CONTENTS

Introduction	01
An overview of data analytics	02
The applications of data analytics	03
Real-world use cases of data analytics	05
A glossary of data analytics terms	07
Begin your journey in data analytics	10
Where do I start?	12

# AN OVERVIEW OF DATA ANALYTICS

In today's digital environment, a wide variety of sources - be it documents, electronic media, social platforms, organizations, or individuals - generate massive amounts of unstructured or raw data.

Systematically analyzing the data, and extracting actionable insights from it, is in recent times imperative for modern enterprises to strengthen their decision-making capabilities and maintain competitiveness. This is where data analytics comes in.

Data analytics involves multiple processes and applications that observe, transform, cleanse, and model data to derive meaningful information from large datasets.

The professionals tasked with collecting, analyzing, cleansing, and profiling data are data analysts. They are also responsible for testing and developing analytical models centered on the data obtained and analyzed.

The **basics of data analytics** revolve around the collection, refining, and structuring of bulk data into a well-defined analytical model, using different programming languages such as Python, R, Scala, or SQL.

Data analysts leverage the results gained from analytical models to identify groupings, trends, or relationships between various data types, which help organizations make informed decisions and meet business targets.

# THE APPLICATIONS OF DATA ANALYTICS

The major industries that are implementing advanced analytical technologies include -



## Retail

The retail sector most likely sees the maximum application of cutting-edge data analytics techniques. With the industry steadily shifting to a digital ecosystem, an increasing number of retailers are using data analytics to understand consumer behavioral patterns, which helps the designing of customized services that enhance the buying experience. From engineering product recommendations to forecasting demand, and crisis control, global retail brands are deploying analytics for almost everything.



## Healthcare

The dramatic advent of innovative technologies is transforming the healthcare industry worldwide. Data analytics is playing a vital role in helping healthcare professionals find medical breakthroughs, deliver hyper-personalized treatment, and improve the patient's quality of life. The medical industry relies on data analytics not to increase profits, but rather to improve the standard of healthcare by proactively identifying diseases and reducing risk factors.



## Media and Entertainment

An early adopter of data analytics technologies, the digital entertainment and media industry implements analytical tools and techniques for predicting viewer interests, personalizing content delivery, optimizing media streams, targeting advertisements, and gaining useful insights from audience reviews.



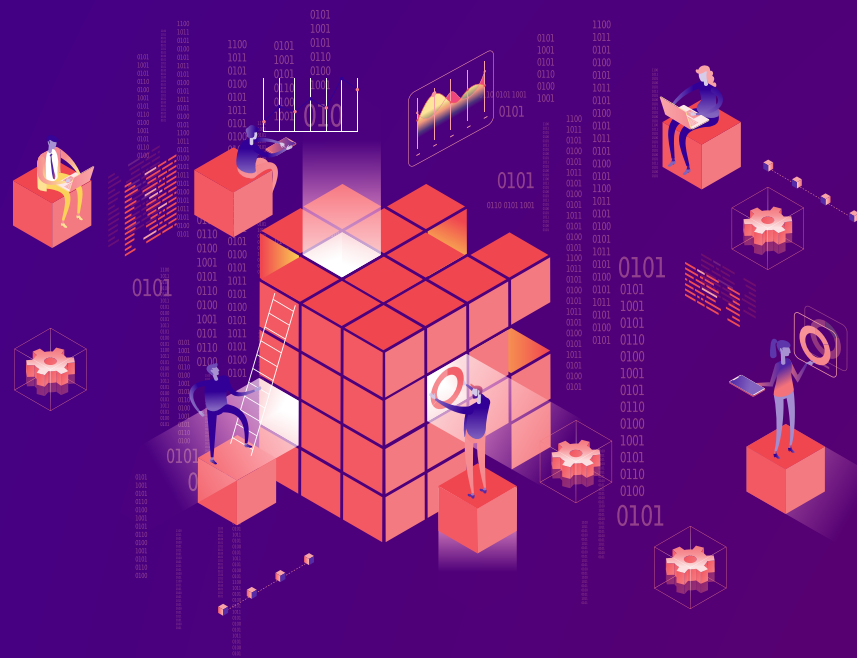
## Banking

After retail, the banking sector makes the most active use of data analytics. Analytical modeling allows banks to track down credit card misuse, detect fraudulent activities, and eliminate system loopholes. Besides empowering banks to create personalized products, other data analytics applications in the financial sector include risk management, performance monitoring, and improved compliance reporting.



## Transportation

Over the past few years, data analytics has been crucial for reforms in the transport industry. Using a variety of historical trends, technical data, and real-time information, data analytics helps the transport industry effectively manage assets, predict traffic congestion, and focus on everyday occurrences while minimizing operating costs.



# REAL-WORLD USE CASES OF DATA ANALYTICS

As digital technologies advance at a rapid pace, all organizations are getting fast access to a wealth of data. However, possessing enormous amounts of data makes no sense if businesses do not analyze the information to gain key insights. Here's how successful brands are using data analytics to address a diverse range of business-critical needs.



American multinational food, snack, and beverage corporation PepsiCo leverages data analytics to manage its supply chain efficiently.

The company aims to ensure that its retail store shelves never run out of products.

The organization's clients submit reports that contain Point of Sale (POS) and warehouse inventory data, which PepsiCo uses to forecast and reconcile the shipment and production needs.

In this way, the firm ensures that its retailers get correct quantities of PepsiCo products on time.



Amazon Fresh, a subsidiary of the global e-commerce company Amazon.com, is an excellent instance of how data analytics can drive product development and innovation.

Amazon implements data analytics to gain the expertise it needs to create and achieve greater value for its grocery delivery service brand—Amazon Fresh.

By putting an emphasis on data analytics, Amazon reviews how consumers buy their products and how the supplier interacts with them.

The results procured from analytical reasoning help Amazon make changes wherever and whenever needed.



Singapore-based multinational banking organization, United Overseas Bank, applies big data analytics to manage its risks.

The financial institution, in the recent past, conducted a test of its risk management system using big data analytics.

UOB found that the new risk management system significantly reduced the time needed to calculate the total value of assets at risk.

Previously, it took approximately 18 h, but after implementing data analytics, calculating risks is now a matter of few minutes.

## NETFLIX

Netflix, Inc., an American technology company and media services provider, is a brilliant example of how brands can use data analytics in targeted advertising.

With well over 100 million subscribers, Netflix collects vast amounts of data, which they analyze to determine what a subscriber is more interested in.

The world's #1 internet entertainment service, Netflix, achieves this through data analytics, which provides them with smart insights based on a subscriber's past search data.



Intercontinental beverage corporation, Coca-Cola, deploys data analytics for driving customer retention.

The 130-year-old company, in 2015, launched a data-driven customer loyalty program to bolster its digital strategy.

While consumers made the most of the rewards, Coca-Cola collected crucial personal information for data analytics, which helped the organization improve customer interactions.

Data analytics not only empowered the super-brand to keep its customers, but it also enabled the company to upsell its new products while boosting the consumption of existing product lines.

# A GLOSSARY OF DATA ANALYTICS TERMS

To become a part of the data analytics industry, it is important to be familiar with the key terminologies related to the field. Here is a list of terms and jargons that surround data analytics.



## Data Science

An interdisciplinary application, data science uses various processes, scientific methods, and algorithms to derive insights and knowledge from an array of structured and unstructured data. Researchers link this field to big data analytics, data mining, and machine learning.



## Data Mining

Data mining refers to the methods and techniques used for identifying patterns within big datasets. It leverages database systems, statistics, and machine learning to detect trends and patterns.



## Hadoop

The Hadoop framework incorporates a set of open-source software tools that allow distributed computing across several computer clusters. It enables data professionals to solve various problems involving large datasets. Apache Hadoop facilitates scaling up from a single server to hundreds of machines, where each machine offers local storage and computation.





## Predictive Modelling

Predictive modeling uses statistical analysis for predicting outcomes using data models. The models can predict whatever organizations want, from corporate earnings to technological advances, television ratings, or sports results.



## MapReduce

Software for distributed processing of datasets, MapReduce algorithms for data analytics comprises a 'map method' for filtering, and a 'reduce procedure' for summary operation.



## NoSQL Database

NoSQL databases provide storage and retrieving mechanisms for data, modeled in a way different from tabular relations seen in relational database applications. These databases have been around since the 1960s, but at the start of the 21st century, Web 2.0 companies coined the name "NoSQL".



## Python

A high-level, interpreted, general-purpose computer language developed by Guido van Rossum, Python incorporates a design philosophy that focuses on readable-code with significant use of whitespace.



## R

A free-to-use software environment, data professionals use the R programming language for computational statistics. Data miners, data analysts, and statisticians use R for data analysis and statistical software development.



## Recommendation Engine

A recommendation engine, also called a recommendation system, recommends products, services, and content to end consumers, based on data analytics. The recommendations rely on several factors, including customer behavioral patterns, or user search history.



## Spark

Apache Spark is a general-purpose, open-source, cluster-computing software framework. The widely-adopted computer program, which integrates modules for SQL, graph processing, machine learning, and streaming, offers a solid programming interface for cluster computers with inherent fault tolerance and data parallelism.



## Structured Data

Structured data refers to “clean” data, organized as tables or charts containing columns, rows, or multidimensional arrays. The easy-to-understand, readily available structured data, expedites analytical, and data mining processes.



## Unstructured Data

Unstructured data is unorganized, raw information, not arranged or structured in a defined format, such as tables or charts. Raw data is heavy, incomplete, inconsistent, or inaccurate, until analysts cleanse and refine the unstructured information to draw meaningful insights.



## Visualization

Visualization is the visual representation of datasets in the form of lists, graphs, charts, or maps. The display of data in a graphical or pictorial format makes recognizing patterns and gaining insights easier.

# BEGIN YOUR JOURNEY IN DATA ANALYTICS

In this section, we'll talk about how you can launch your career in data analysis.



## The Skills You'll Need

A career in the field of data analytics requires you to master the following skills:



## Programming

An aspirant must have a comprehensive knowledge of programming languages, including R, Python, and SQL. Other useful languages include Java, Oracle, SAS, MATLAB, Tensorflow, Scala, and Julia.



## Math

Those in data analyst roles must be strong in basic math skills, specifically statistics. That's because they are faced with many day-to-day situations when these skills comes in handy, for example, when dealing with small datasets, which are best tackled with the statistical capabilities of Microsoft Excel.



## Data Processing Platforms

Data analysts need to be familiar with big data processing platforms like Hadoop and Apache Spark in order to handle large datasets. The working knowledge and understanding of these frameworks allow data analysts to query data across multiple devices, and scrub, model, and interpret it to gain more in-depth insight into data patterns and identify relationships and trends.



## Visualization

Along with extracting insights out of datasets, it is equally important to be able to present them clearly, particularly for stakeholders. For this reason, a data analyst must know how to produce graphical representations of the findings using visualization tools, such as Tableau.



## Machine Learning

Being at the heart of any large-scale data analysis, automation is a crucial part of this field. This requires a data analyst to have considerable machine learning skills to be able to create, apply, and train the most fitting models and algorithms to datasets to find solutions for specific problems.

Besides, a data analyst must also possess:

- ✓ Skills to interpret, analyze, and model data
- ✓ Problem-solving skills
- ✓ Interpersonal skills
- ✓ Communication skills

# WHERE DO I START?

You should start by making the most of online resources and books that are available for free. There are some exciting stuff on the Internet that can help you build a solid foundation.

These days, employers are not emphasizing on formal education to fill talent gaps, however, having an academic degree in computer science, mathematics, or statistics will help you climb up the career ladder quickly.

The easiest way to establish yourself as a data analyst is to get certified by a top-rated institution, such as **Simplilearn**, which is one of the world's leading certification providers.

This **data analytics handbook** suggests Simplilearn because of the wide variety of courses it offers. From **Data Analytics Course for Beginners** to **Data Analyst Master's program** in collaboration with IBM, and **Post Graduate Program in Data Analytics**, Simplilearn positions itself uniquely to cater to the needs of every aspiring data analyst. **Begin Now!**

