Introduction to Data Analytics

Data Analytics, Data Science, and Machine Learning

Learning Objectives

By the end of this lesson, you will be able to:

- Define data science and machine learning
- Differentiate between data science, machine learning, and data analytics

DATA AND
ARTIFICIAL INTELLIGENCE

Introduction to Data Science

Data Science

Data science is the study of data, which involves gathering, storing, analyzing, and plotting data, to effectively extract useful information.

Aim: Gain meaningful insights from both structured and unstructured data.

Data Science

Data cleansing

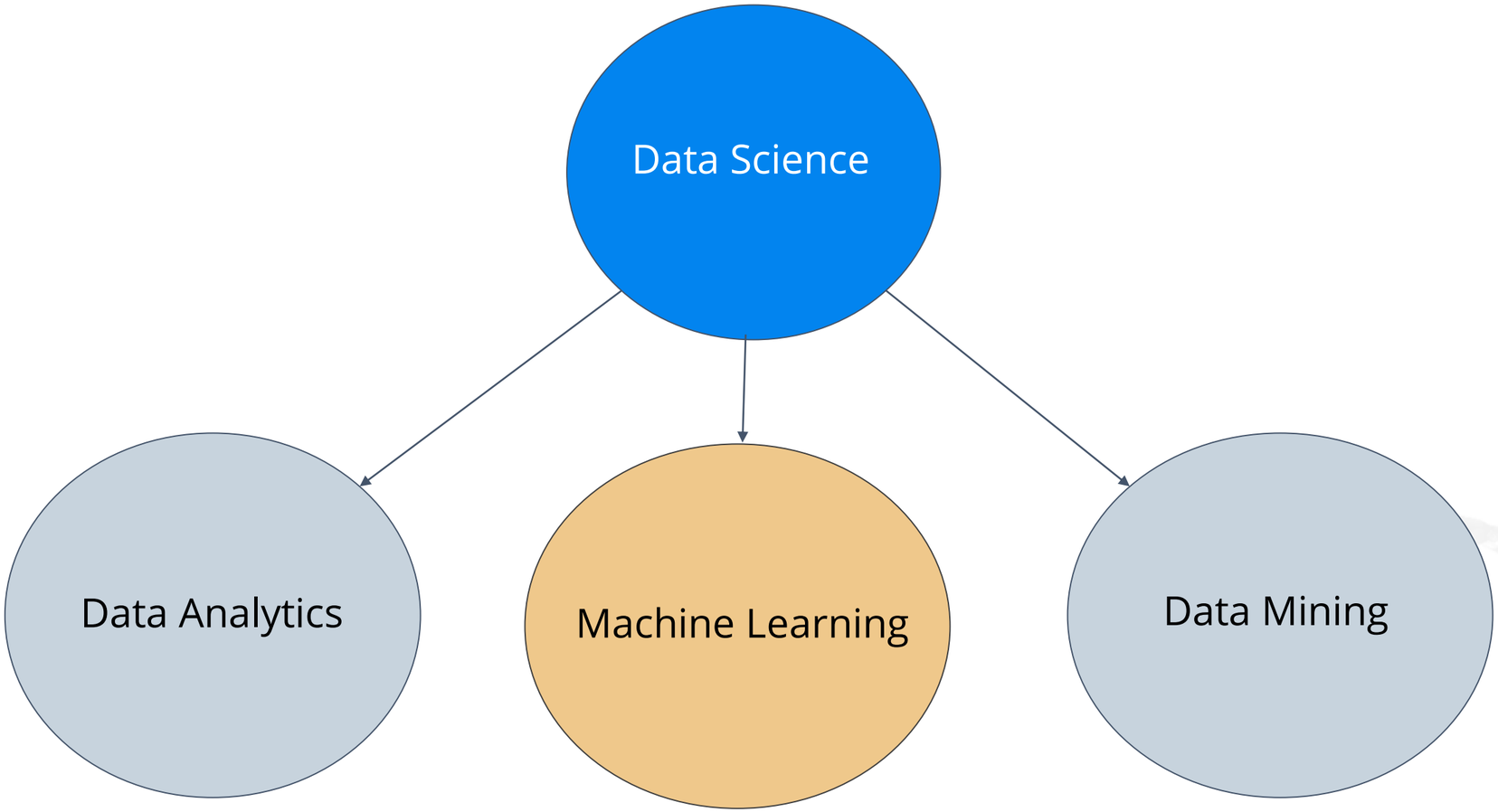
Preparation and analysis

Data Science

Trend forecast

Machine learning and data analytics

Types of Data Science



Data Analytics

Data analytics is the process of examining and analyzing raw data sets to:

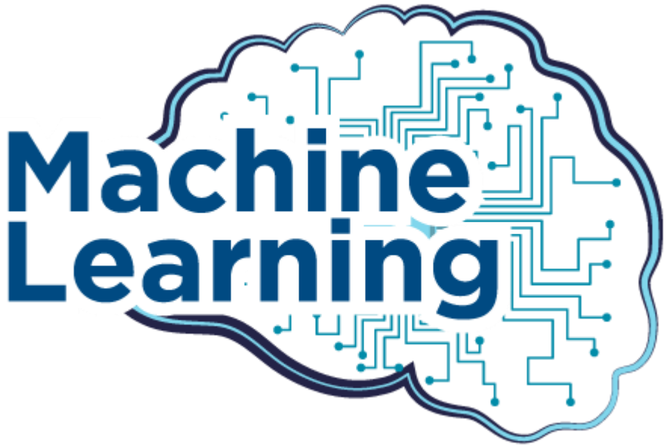
Draw conclusion

Derive information

Derive insights from raw data sources

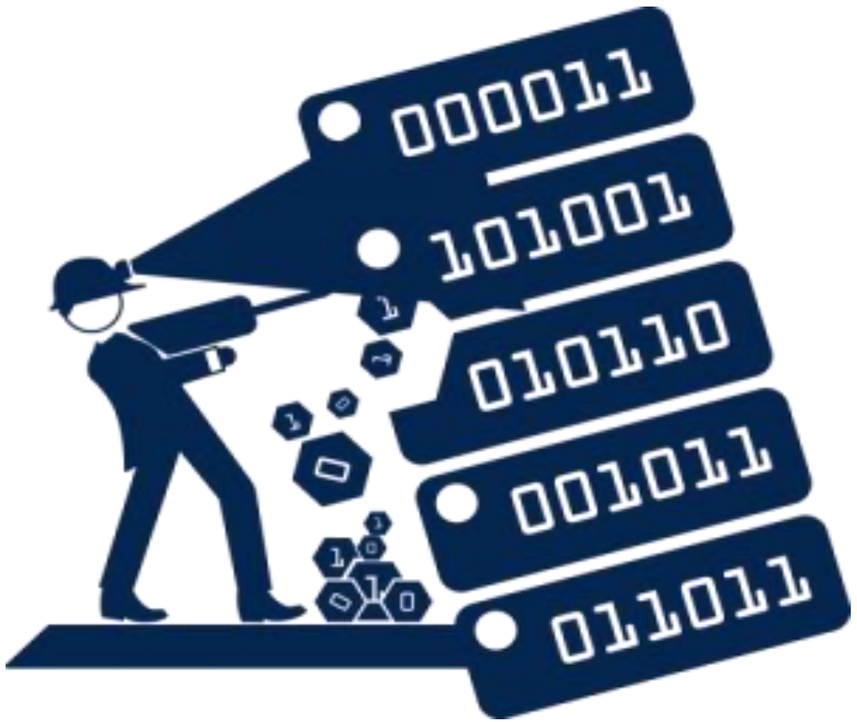
Machine Learning

Learns from patterns in the past using a set of algorithms



Predicts outcomes accurately

Data Mining

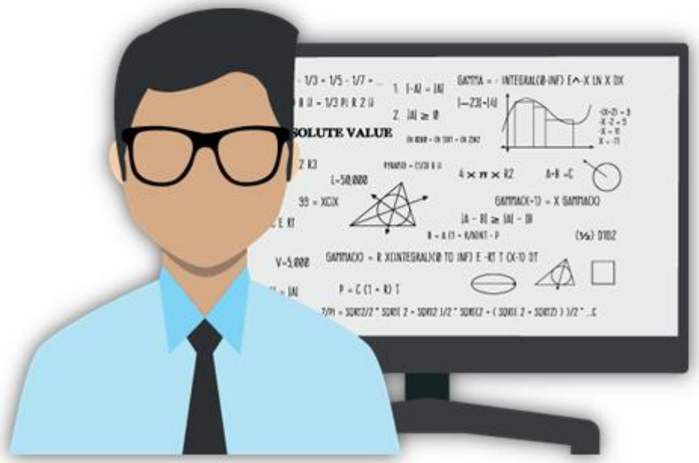


- Data mining is the process of analyzing data from different perspectives.
- It summarizes data into useful information.
- It helps increase revenue and cut costs.

DATA AND
ARTIFICIAL INTELLIGENCE

Data Science, Data Analytics, and Machine Learning

Data Science and Data Analytics



Forecasts the future based on past patterns

Data Scientist

Extracts meaningful insights from various data sources

Data Analyst

Machine Learning

Machine learning creates systems that can learn from the data.

It is the ability of machines to predict outcomes based on patterns in the past.

Machine Learning

Leverages various algorithms to train the machine

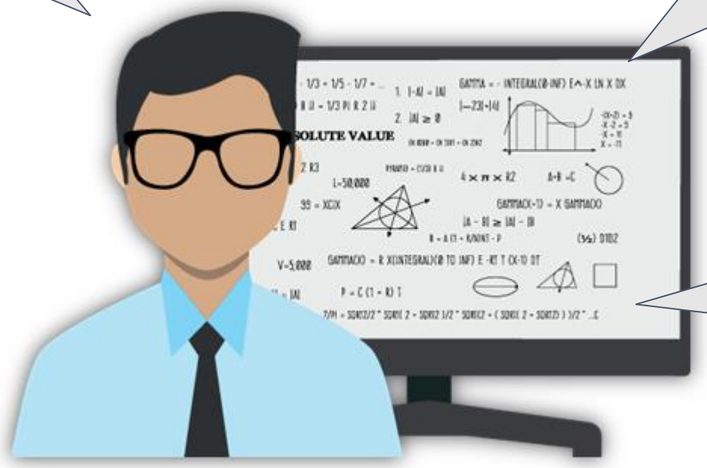
ML Engineer

Data Science and Machine Learning

Extracts useful information from collected data sets

Understands data from a business point of view

Gathers data from various sources

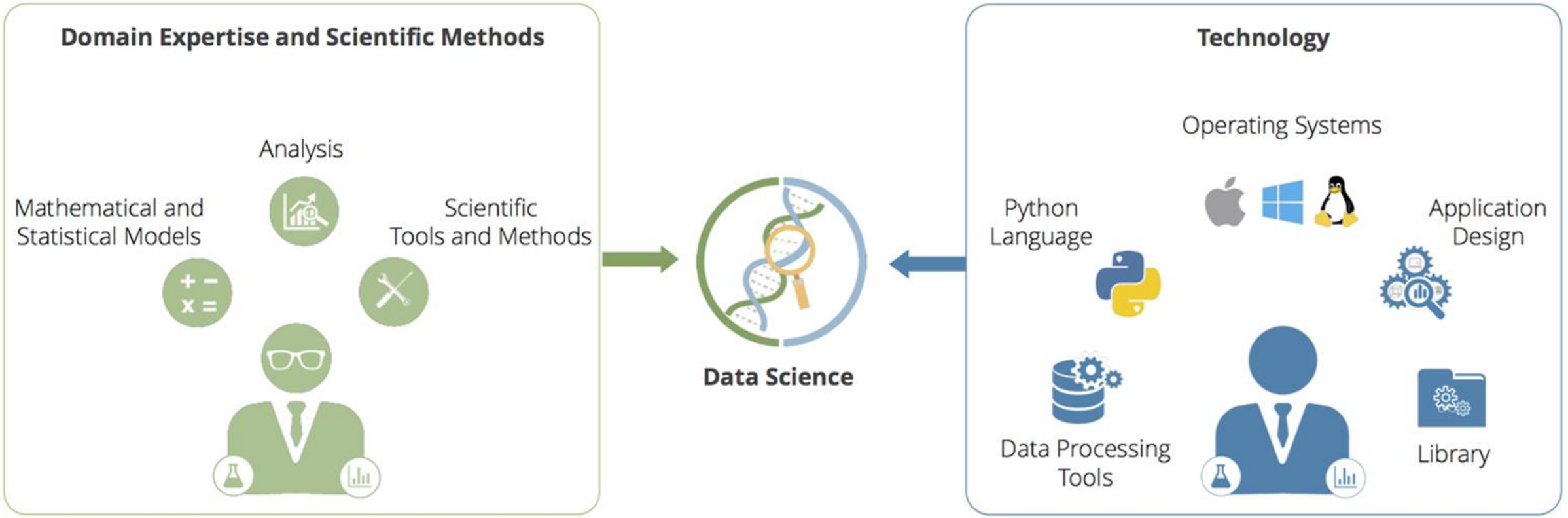


Provides accurate predictions to improve key business decisions

Data Scientist

Understanding Data Science

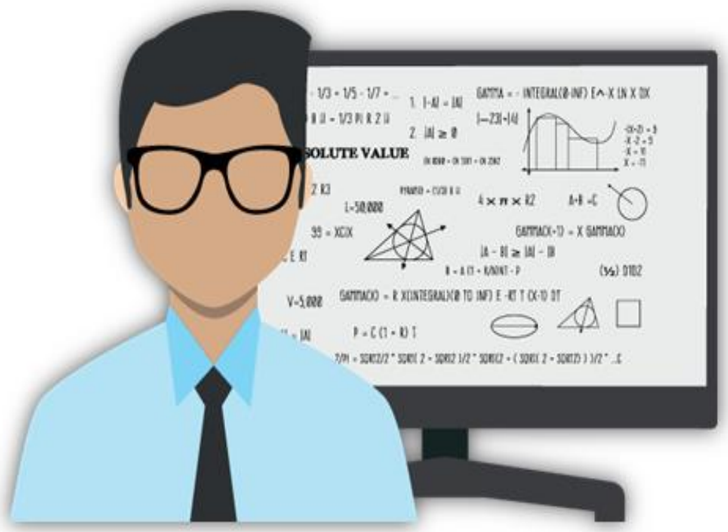
Understanding Data Science



A data scientist combines both domain and technology perspectives.

Understanding Data Science

Knows multiple analytical functions



Works with data from video and social media sources

Data Scientist

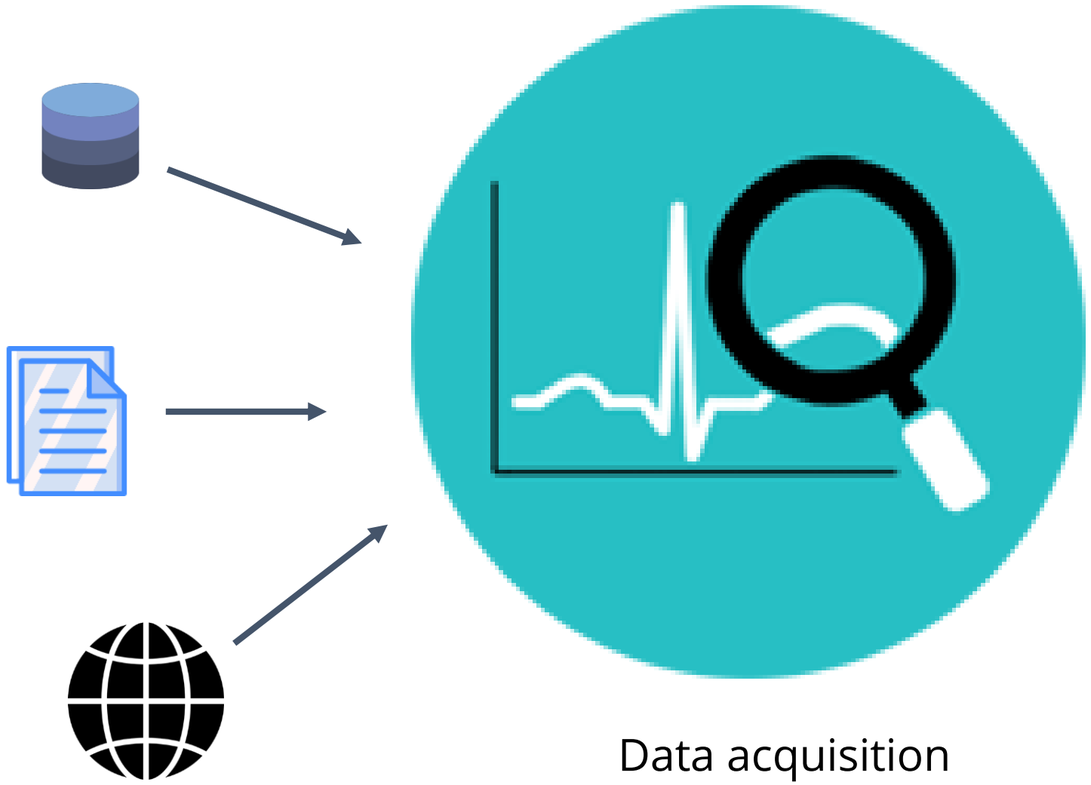
Has a sound knowledge of technologies such as Python, SAS, R, Scala, visualization libraries, SQL database, and machine learning

Data Science: Process Flow

How car insurance costs less if you pay bills on time?

Data scientists found that the people who pay bill promptly are less prone to the accidents

Data Science: Process Flow



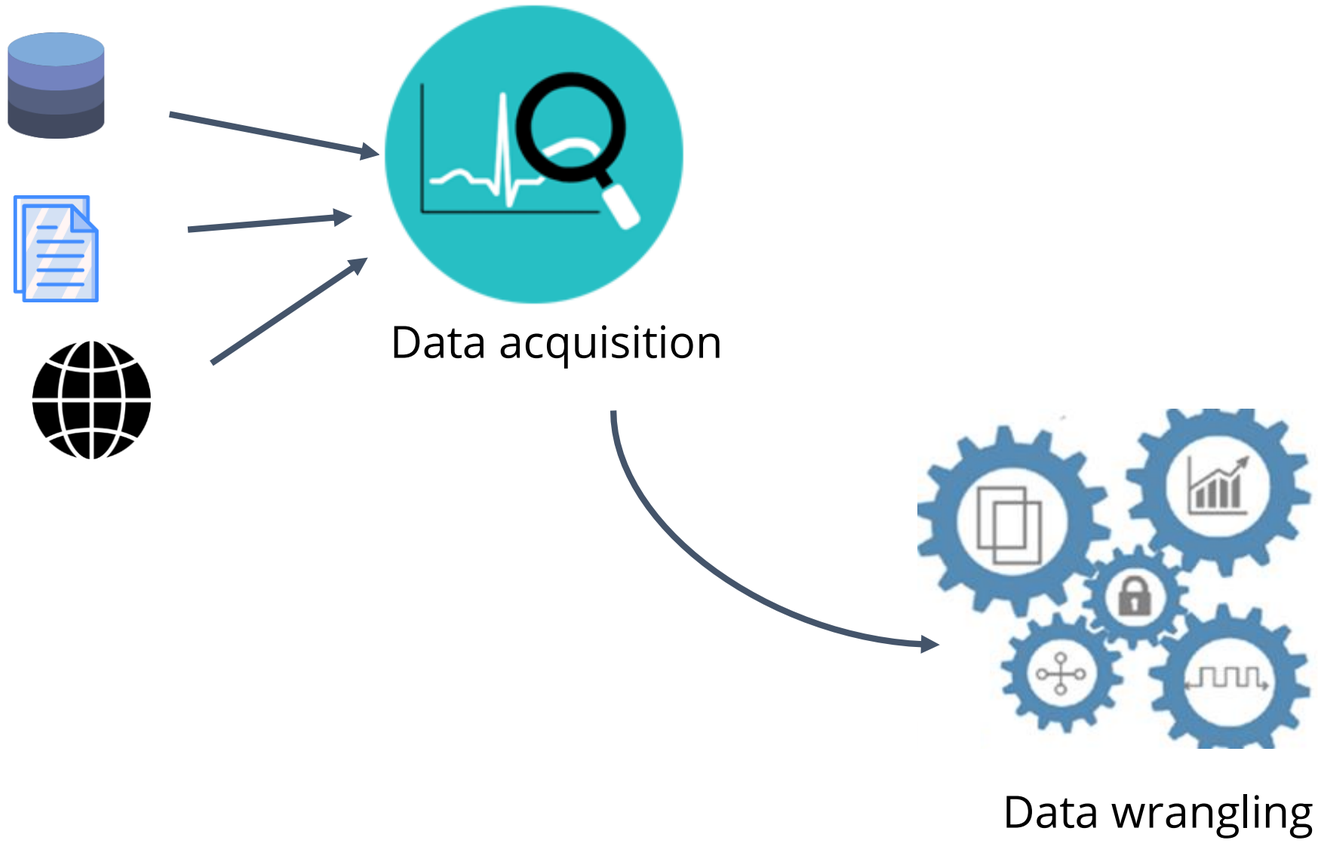
Data acquisition

Step 1: Data acquisition

Data scientists work with existing data sets or gather them from various sources.

The most important part of the whole process is to have the correct data.

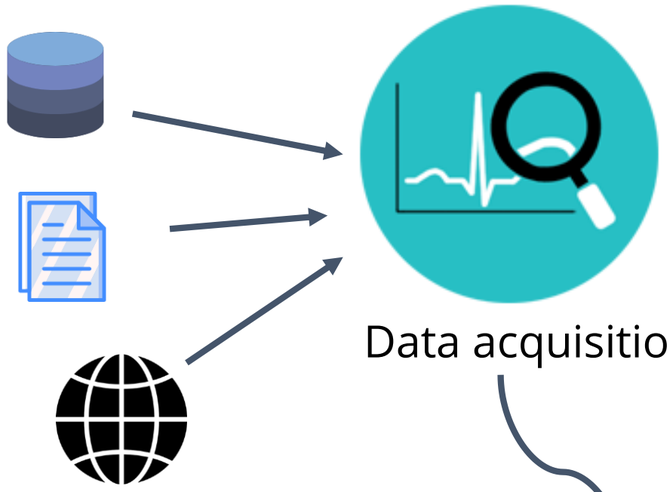
Data Science: Process Flow



Step 2: Data wrangling

- Choose the right tools from Python, R, and SQL
- Derive a clean data set
- Apply pick-and-shovel algorithms
- Obtain meaningful data

Data Science: Process Flow



Data acquisition

Data wrangling

Machine learning

Step 3: Machine learning

- Validate the model
- Perform necessary statistical analysis
- Apply machine learning or recursive analysis
- Run regression testing
- Compare results against other techniques or sources

Challenge of a Data Scientist

The most challenging part of being a data scientist is taking the results and presenting them to the stakeholders in an easy and consumable manner.

DATA AND
ARTIFICIAL INTELLIGENCE

Data Science and Business Strategy

Data Science and Business Strategy

Business owners used to measure their success based only on the Profit and Loss Statement.

Current era of technology leverages data science for efficient prediction on what will work.

Data Science and Business Strategy

The process flow of a data-driven decision-making process:

Data Scientist: Asset to the Business

Empowers management to make better decisions

Provides insights on various KPIs and parameters

Enables strategic changes for better results

Identifies and refines the target audience

Identifies areas of improvement

Identifies opportunities

Data Scientist

Companies Using Data Science

Successful Companies Using Data Science

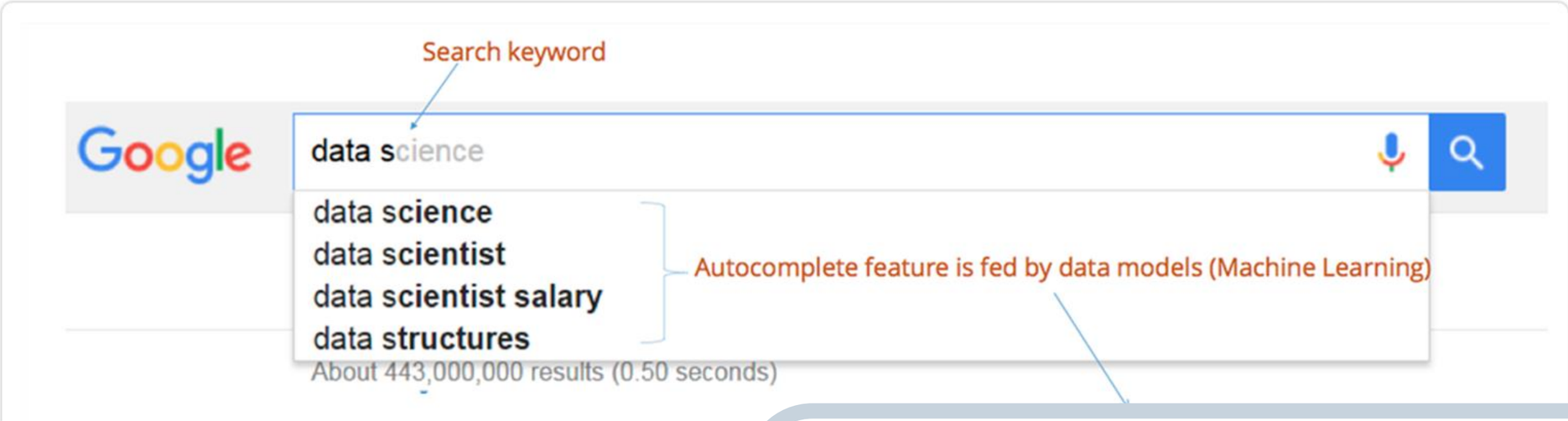
Few successful companies that use data science

DATA AND
ARTIFICIAL INTELLIGENCE

Google Search Engine

Google Search Engine

Google uses data science to provide relevant search recommendations.



The influencing factors include:

- Query volume: unique and verifiable users
- Geographical locations
- Keyword or phrase matches on the web
- Scrubbing for inappropriate content

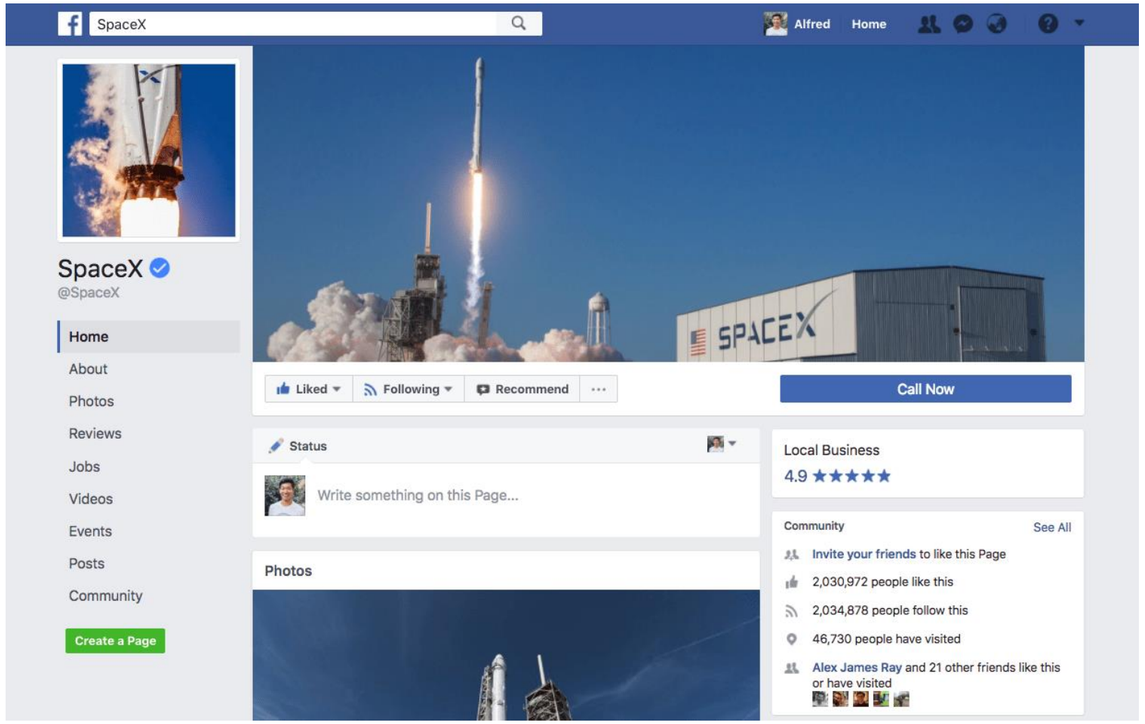
DATA AND
ARTIFICIAL INTELLIGENCE

Facebook Tags

Facebook Tags

Facebook uses machine learning in every aspect including:

Scrolling the news feed



Browsing images or videos

Facebook Tags

Uses clustering algorithm to:

- Find mutual friends
- Send friend suggestions

Alibaba

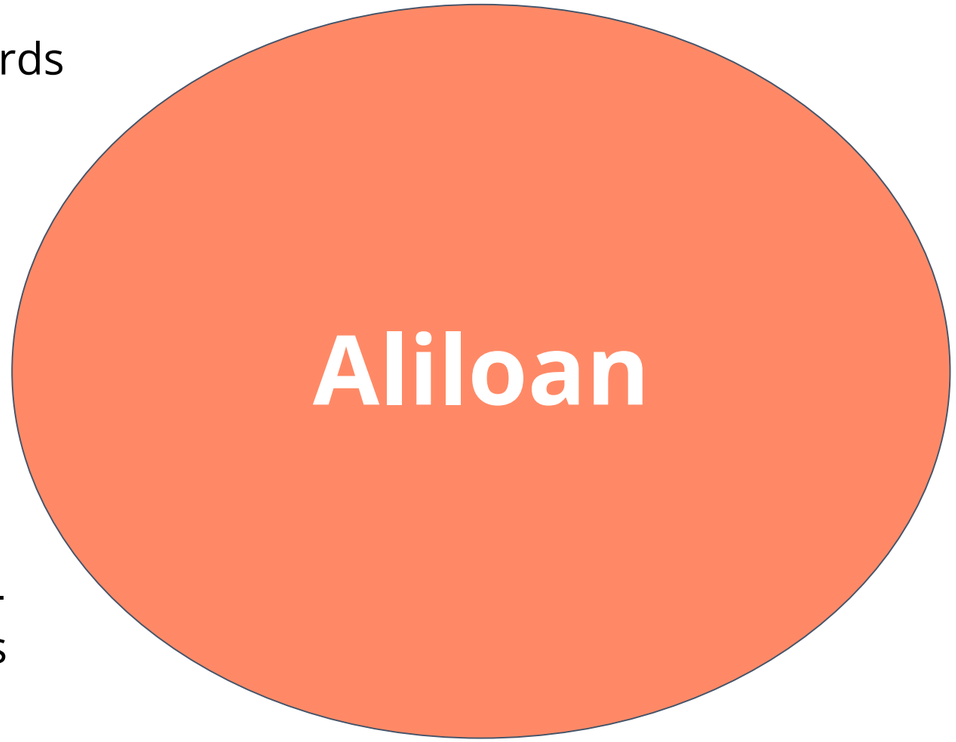
Alibaba’s Aliloan

Aliloan is an automated online system that provides flexible microloans to entrepreneurial online vendors.

Alibaba's Aliloan

Analyzes trading records and evaluates risk

Uses predictive models to analyze transaction records



Collects data from e-commerce platforms

Determines merchants' creditworthiness

DATA AND
ARTIFICIAL INTELLIGENCE

Travel Industry

Travel Industry

Travel companies use datasets from social media, itineraries, predictive analytics, and location tracking to arrive at the 360-degree view.

The sensors from different modes of transport provide real-time data on various parameters to predict and prevent problems.

Travel Industry

Integrates historical data to ensure maximum yield

Offers deals based on the user's preferences or recommended local attractions

Predictive algorithms help drivers predict fuel needs, ETAs, and delays.

DATA AND
ARTIFICIAL INTELLIGENCE

Retail

Retail

RFM analysis is a marketing technique that leverages data to determine the target customer.

Recency

Frequency

Monetary

Retailers use data science to segment customers into RFM groups and target marketing and promotions.

DATA AND
ARTIFICIAL INTELLIGENCE

E-Commerce

E-Commerce

Amazon is an e-commerce giant that leverages data science to the fullest extent.

Amazon prefers an *everything under one roof* model.

E-Commerce

E-commerce companies use data science to upsell through their websites.

Amazon's *People who viewed that product, also liked this* functionality uses sophisticated mining techniques and boosts business.

DATA AND
ARTIFICIAL INTELLIGENCE

Crime Agencies

Crime Agencies

Analytics keeps crime in check by:

- Using identified patterns to derive prediction techniques
- Analyzing previous data to prevent future burglaries

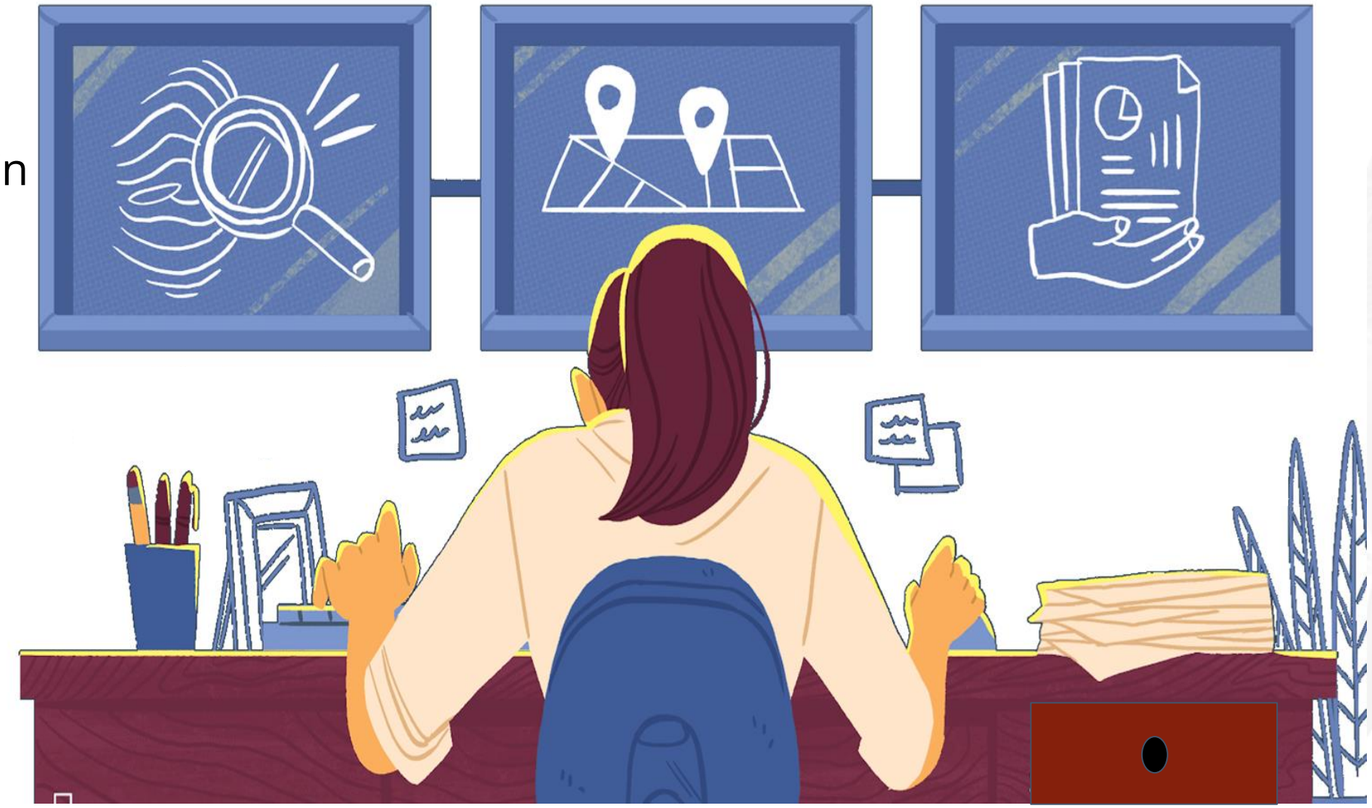
Crime Agencies

- Data mining can help identify pattern in from domestic violence to terrorism.
- Advanced analytics helps prevent crime by using information from social media.

Crime Agencies

Crime prevention agencies use data science in deciding:

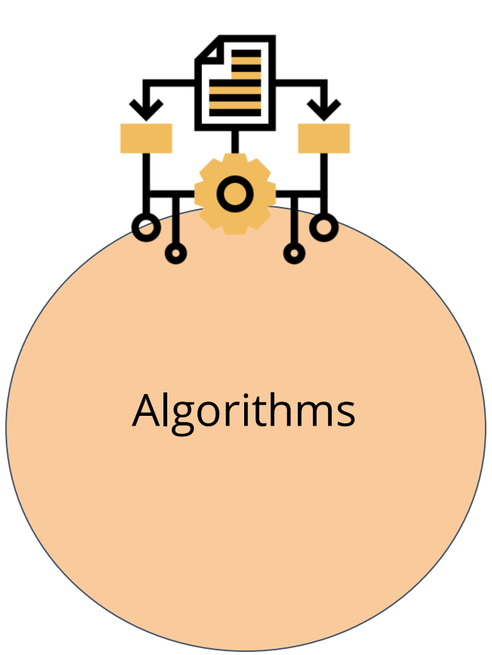
- Where to deploy police manpower?
- Who to search at a border crossing?
- Which intelligence to consider in counter-terrorism activities?

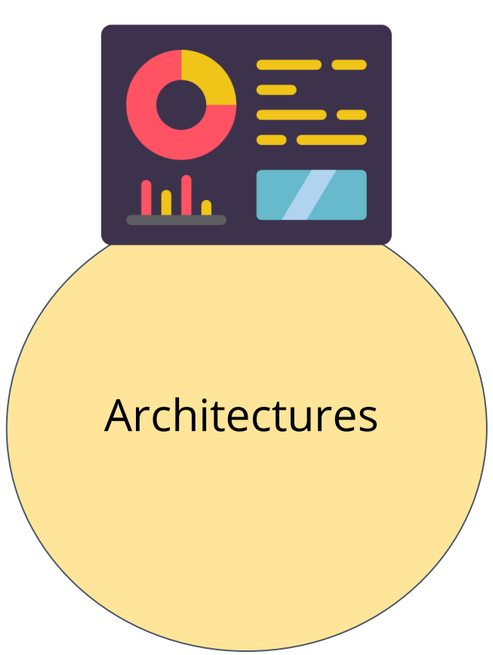


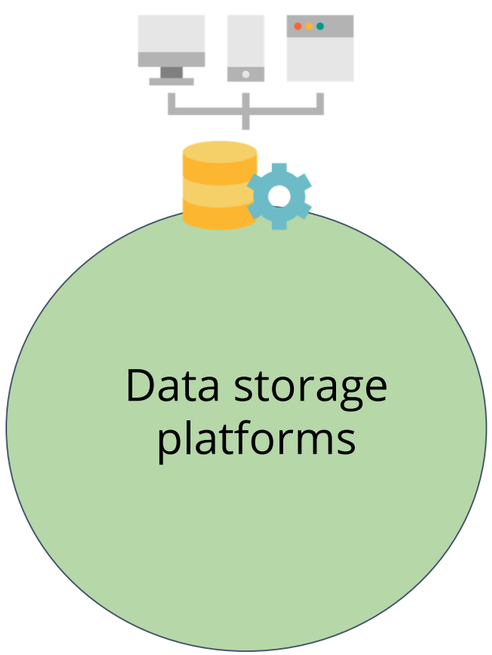
DATA AND
ARTIFICIAL INTELLIGENCE

Analytical Platforms across Industries

Analytical Platforms across Industries



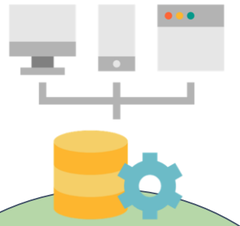




Analytical Platforms across Industries

Machine learning algorithms

Architectures



Data storage platforms

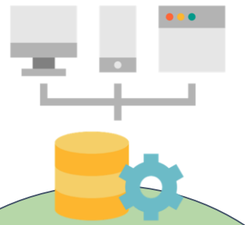
Tools

- Forecasting
- Regression
- Bayesian network
- Vector autoregression

Analytical Platforms across Industries

Machine learning algorithms

Deep learning architectures

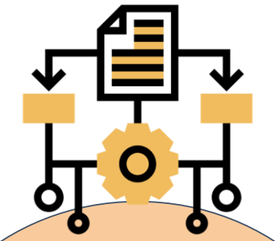


Data storage platforms

Tools

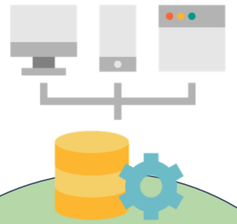
- Deep Belief Network (DBN)
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN)

Analytical Platforms across Industries



Machine learning algorithms

Deep learning architectures



Cloud storage platforms

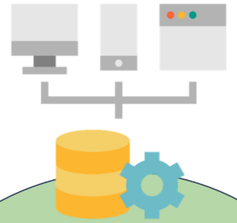
Tools

- Amazon AWS
- Microsoft Azure
- Lambda

Analytical Platforms across Industries

Machine learning algorithms

Deep learning architectures



Cloud storage platforms

Tools

Analytics tools

- Spark
- Python
- R
- Apache Pig

Reporting tools

- Tableau
- Splunk
- Power BI
- Kibana

Key Takeaways

- Data science is the study of data, which involves gathering, storing, analyzing, and plotting data, to effectively extract useful information.
- Data science is an umbrella that contains data analytics, data mining, and machine learning.
- Data science is used by many successful companies such as Google, Facebook, and Alibaba.
- Analytical platforms across industries include algorithms, architecture, data storage platforms, and tools.