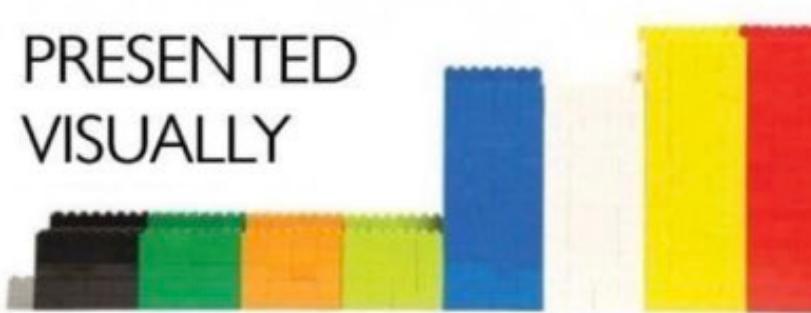


Data Science

for Managers and Executives

By: John Rey Goh

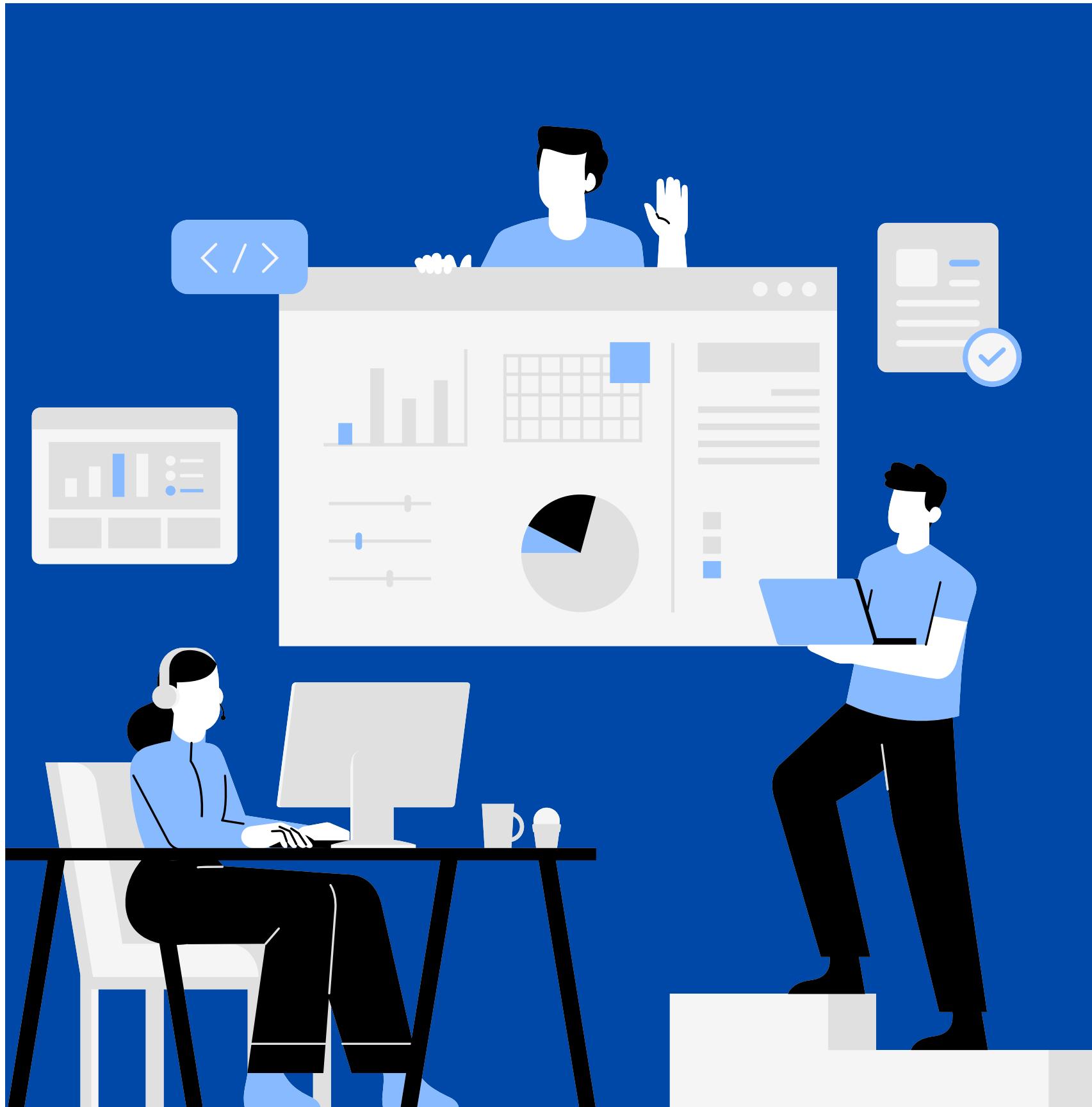
Data Science helps organizations unlock the value of their data, which can lead to improved performance, better decision-making, and a competitive advantage in the marketplace.



Data Science?

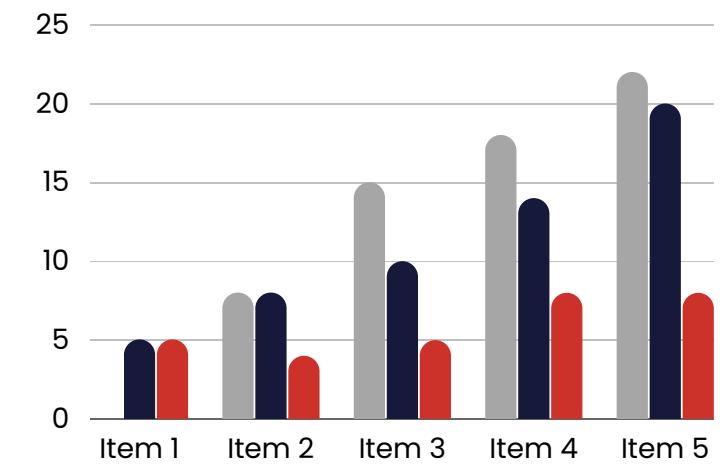
Data science is a multidisciplinary field that involves using statistical and computational techniques to extract insights and knowledge from data.

It combines elements of statistics, mathematics, computer science, and domain expertise to analyze and interpret complex data sets.



Goal of data science:

1. Extract meaningful and actionable insights from data
2. Develop new products and services
3. Improve operations, and;
4. Solve complex problems





Different Sectors Using Data Science



industries benefit from data science and analytics



Retailers

Retailers need to correctly anticipate what their customers want and then provide those things. If they don't do this, they will likely be left behind the competition. Big data and analytics provide retailers the insights they need to keep their customers happy and returning to their stores.



Medical Industry

The medical industry is using big data and analytics in a big way to improve health in a variety of ways. For instance, the use of wearable trackers to provide important information to physicians who can make use of the data to provide better care to their patients. Wearable trackers also provide information like whether the patient is taking his/her medication and following the right treatment plan.



Bank

The banking industry is generally not looked at as being one that uses technology a lot. However, this is slowly changing as bankers are beginning to increasingly use technology to drive their decision-making. For instance, the Bank of America uses natural language processing and predictive analytics to create a virtual assistant called Erica to help customers view information on upcoming bills or view transaction histories.

The assistant will eventually study their customers' banking habits and suggest relevant financial advice at appropriate times.



Construction

Construction companies track everything from the average time needed to complete tasks to materials-based expenses and everything in between. Big data is now being used in a big way in the construction industry to drive better decision-making.



Travel

There is always a need for people to reach their destinations on time and data science and analytics can be used by transportation providers, both public and private, to increase the chances of successful journeys. For instance, most Travel Road Guide Apps and uses statistical data to map customer journeys, manage unexpected circumstances, and provide people with personalized transport details.



Social Media and Online Content

Consumers now expect rich media in different formats as and when they want it on a variety of devices. Collecting, analyzing, and utilizing these consumer insights is now a challenge that data science is stepping in to tackle.

Data science is being used to leverage social media and mobile content and understand real-time, media content usage patterns.

With data science techniques, companies can better create content for different target audiences, measure content performance, and recommend on-demand content.



Education

One challenge in the education industry where data science and analytics can help is to incorporate data from different vendors and sources and use them on platforms not designed for varying data.

Schools have developed a learning and management system that can track when a student logs into the system, the overall progress of the student, and how much time is spent on different pages, among other things.

Big data can also be used to measure teachers' effectiveness by fine-tuning teachers' performance by measuring against subject matter, student numbers, student aspirations, student demographics, and many other variables.



Manufacturing

The manufacturing industry also generates huge amounts of data that has so far gone untapped. The increasing demand and supply of natural resources, such as oil, minerals, gas, metals, agricultural products, etc. has led to the generation of huge amounts of data that is complex, difficult to handle, and a prime candidate for big data analytics.



Government

In the Government, big data has many applications in the public services field. Places where big data is/can be used include in financial market analysis, health-related research, environmental protection, energy exploration, and fraud detection.

One specific example is the use of big data analytics by the Social Security System (SSS) to analyze large numbers of social disability claims that come in as unstructured data. Analytics is being used to rapidly process medical information and detect fraudulent or suspicious claims.

Another example is the use of data science techniques by the Food and Drug Administration (FDA) to identify and analyze patterns related to food-related diseases and illnesses.



Energy and Utilities

The energy and utilities industry generates and will continue to generate huge amounts of data that can be analyzed using big data analytics. This data can be used to better study the consumption of utilities, which in turn allows for better control of utility use and improved customer feedback. The use of big data by utility companies also allows for improved asset and workforce management and is useful for identifying and correcting errors as soon as possible.



Energy and Utilities

The energy and utilities industry generates and will continue to generate huge amounts of data that can be analyzed using big data analytics. This data can be used to better study the consumption of utilities, which in turn allows for better control of utility use and improved customer feedback. The use of big data by utility companies also allows for improved asset and workforce management and is useful for identifying and correcting errors as soon as possible.



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

1. Predictive Maintenance

By using sensors and data analysis techniques, mining and power corporations can predict when equipment might fail and proactively schedule maintenance, reducing downtime and improving safety.

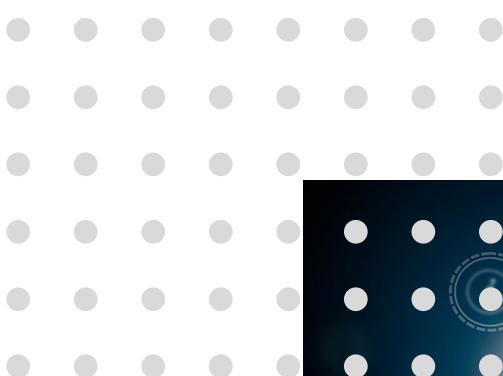


ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

2. Energy Efficiency

Data science can help identify opportunities for energy savings by analyzing energy consumption patterns and optimizing equipment performance.



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

3. Resource Optimization

Mining companies can use data science to optimize the extraction of minerals, identify new sources of minerals, and reduce waste and environmental impact.

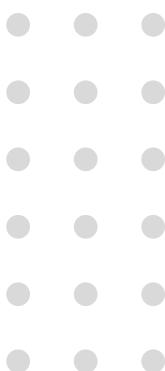


ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

4. Safety and Risk Management

Data science can help identify potential safety hazards and risks, and improve risk management strategies.

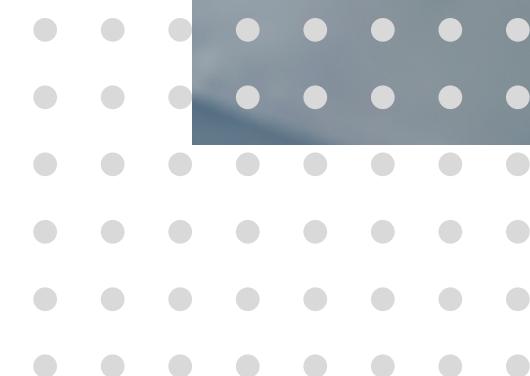


ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

5. Improved Decision Making

By analyzing large volumes of data, mining and power corporations can make more informed decisions about resource allocation, equipment maintenance, and investment.



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

6. Cost Reduction

Data science can help reduce costs by improving operational efficiency and reducing downtime.



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

7. Demand Forecasting:

Use data science to predict energy demand and adjust power generation accordingly, improving operational efficiency and reducing costs.



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

8. Environmental Monitoring

Use data science to monitor environmental conditions in and around mining sites, detecting and mitigating potential impacts on air and water quality.

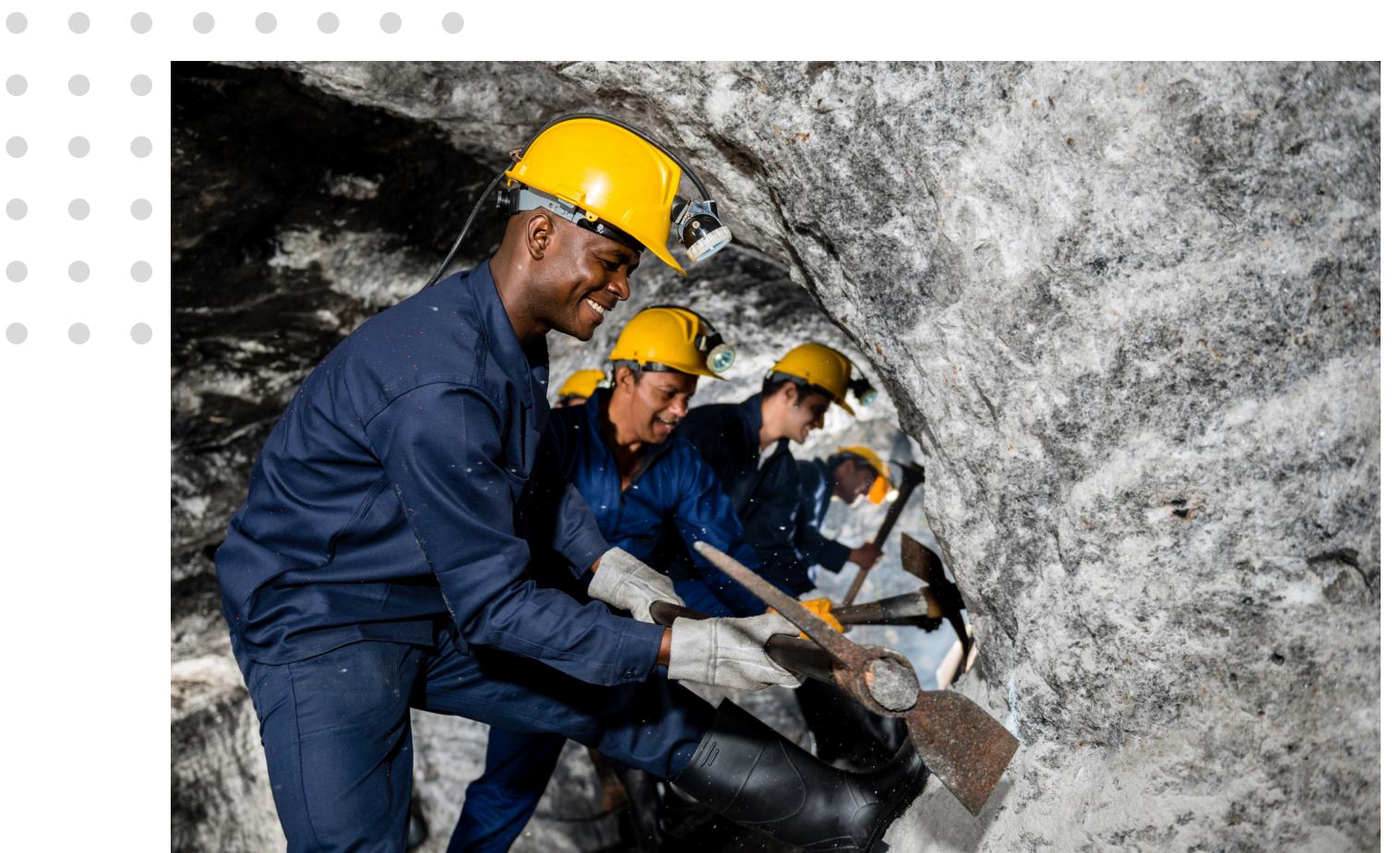


ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

9. Exploration and Surveying

Use data science to analyze geospatial data to identify potential mining sites and optimize exploration and surveying activities.



ADVANTAGES AND USES OF DATA SCIENCE

in Mining and Power Corporation

10. Supply Chain Management

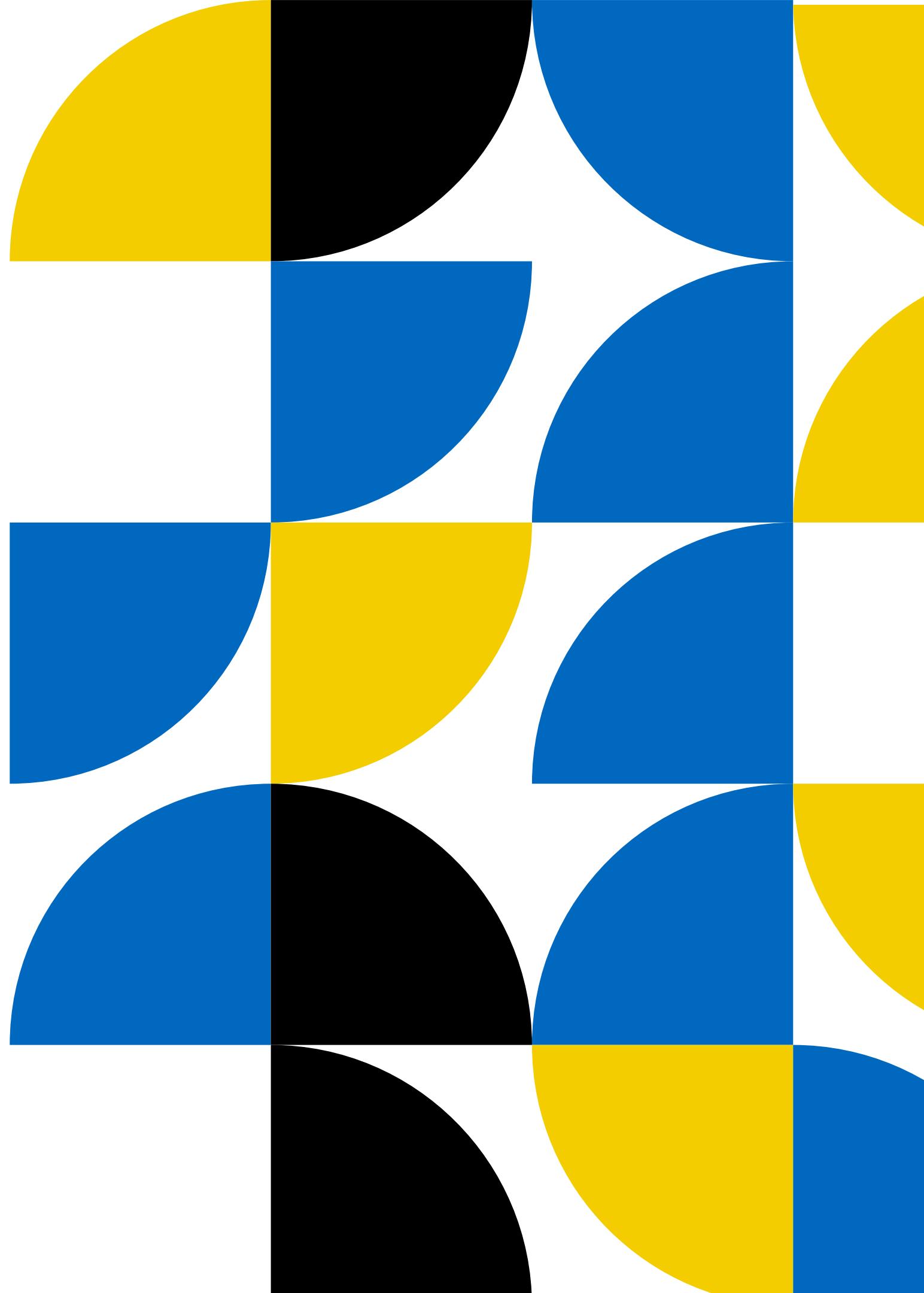
Use data science to optimize the supply chain, improving logistics, reducing waste, and enhancing sustainability.





Data Science Skills and Roadmap

Learn. Develop. Integrate.



DATA SCIENCE SKILL SET

BASIS

CORE

ENABLEMENT

TECHNICAL SKILLS

- CODING SKILLS
- HANDLING DATA
- COMPUTATIONAL TOOLS
- BASIC SOFTWARE DEVELOPMENT
- BIG DATA
- HIGH PERFORMANCE COMPUTING
- PARALLEL COMPUTING

ANALYTICAL SKILLS

- ADVANCED STATISTICS & INFERENCE
- MODELLING & SIMULATION
- MACHINE LEARNING
- COMPUTER SCIENCE
- ADVANCED MATH
- DATA VISUALIZATION
- EXPERIMENT DESIGN
- RESEARCH EXPERTISE

BUSINESS SKILLS

- EVALUATION AND DEVELOPMENT OF BUSINESS CASES
- PROJECT MANAGEMENT
- BUSINESS PROCESSES
- CHANGE MANAGEMENT
- COMMUNICATION SKILLS
- LEADERSHIP SKILLS



**Computer
Science/IT**

Machine
Learning



**Math and
Statistics**

**Data
Science**

Software
Development

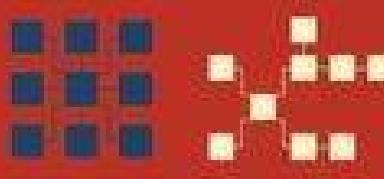
Traditional
Research

**Domains/Business
Knowledge**



ENGAGEMENT PROCESS

Step 1: Build the Data Model



Step 2: Define The Report



Step 3: Generate SQL commands



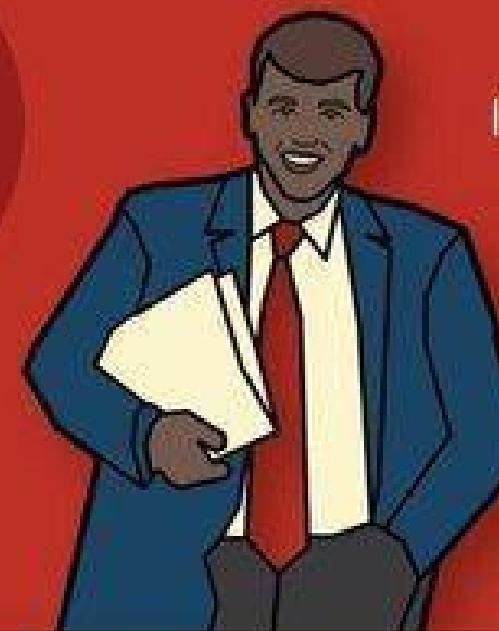
Step 4: Create Report



The data warehouse is a "schema-on-load" approach because the data schema must be defined and built prior to loading data into the data warehouse. Without an underlying data model, the BI tools will not work.

Business Intelligence Questions

What happened?



Business
Intelligence
Analyst

Data Science Questions

Why? What will happen?
What should I do?

Data
Scientist



DIFFERENCE BETWEEN BUSINESS INTELLIGENCE AND DATA SCIENCE

CHARACTERISTICS

Focus	Reports, KPIs, trends	Patterns, correlations, models
Process	Static, comparative	Exploratory, experimentation, visual
Data Sources	Pre-planned, added slowly	On the fly, as-needed
Transform	Up front, carefully planned	In-database, on-demand, enrichment
Data quality	Single version of truth	"Good enough," probabilities
Data model	Schema on load	Schema on query
Analytics	Retrospective, Descriptive	Predictive, Prescriptive, Preventative

ENGAGEMENT PROCESS

Step 1: Define Hypothesis to Test



Step 2: Gather Data



Step 3: Build Data Model



Step 4: Explore the Data



Step 5: Build and Refine Analytic Models



Step 6: Ascertain Goodness of Fit



DATA SCIENCE ROADMAP



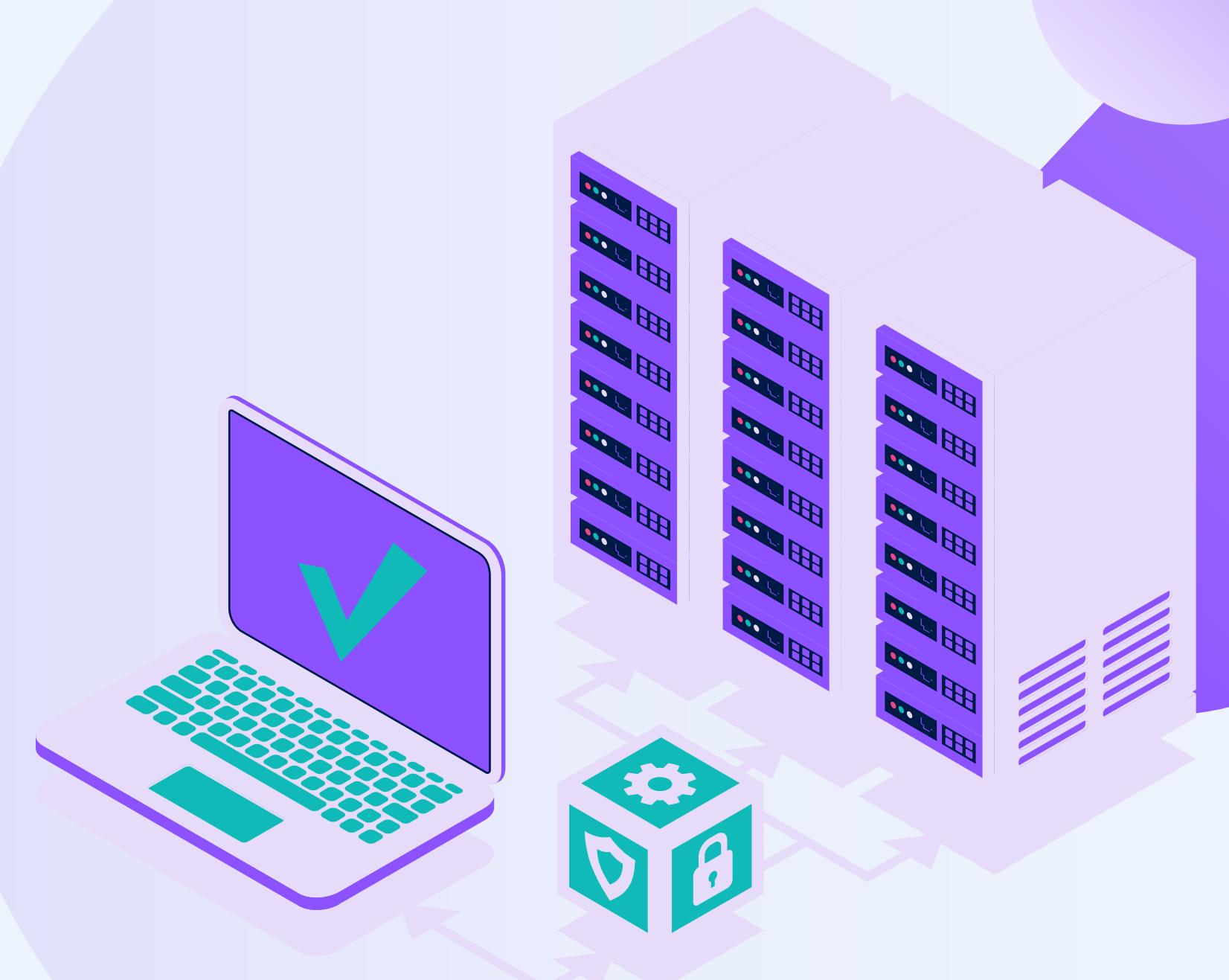
Data Science Lifecycle

Data scientists use a variety of tools and techniques to perform these tasks, including machine learning, statistical analysis, data visualization, and natural language processing.



Activity:

Data Analytics and
Visualization Demonstration
with Microsoft PowerBI





Data Science Trends

The field of data science is rapidly evolving, and there are several new and emerging trends that are shaping the future of the industry. Here are some of the latest trends in data science:



Data Science Trends

1. Artificial intelligence (AI) and machine learning (ML):

AI and ML are becoming increasingly important in data science as organizations seek to automate complex processes and extract insights from vast amounts of data. This includes the use of deep learning algorithms and neural networks to analyze images, videos, and other unstructured data.



Data Science Trends

2. Internet of Things (IoT):

The proliferation of connected devices is generating vast amounts of data, which can be used to inform business decisions and improve operations. Data scientists are increasingly using IoT data to develop predictive models and real-time analytics to improve efficiency and reduce costs.



Data Science Trends

3. Natural language processing (NLP):

NLP is a branch of AI that focuses on enabling computers to understand and interpret human language. Data scientists are using NLP techniques to analyze text data, including customer feedback, social media posts, and news articles, to gain insights into consumer behavior and sentiment.



Data Science Trends

4. DataOps:

DataOps is a set of practices and methodologies that combines the principles of DevOps with data science. DataOps aims to streamline the data management process, reduce time to insights, and improve collaboration between data scientists and IT teams.



Data Science Trends

5. Data privacy and ethics:

With the increasing importance of data in business, there is also a growing concern about data privacy and ethics. Data scientists are focusing on developing ethical frameworks for data use, including ensuring the privacy and security of personal data and avoiding biases in data models.



Data Science Trends

6. Cloud computing:

Cloud computing is becoming increasingly popular in data science as organizations seek to leverage the scalability and flexibility of cloud-based platforms to store, process, and analyze data. Cloud platforms also offer advanced analytics tools and machine learning capabilities that can help data scientists extract insights from large and complex data sets.

COMMON TERMINOLOGIES IN DATA SCIENCE



1. Data Mining: The process of discovering patterns in large datasets.
2. Machine Learning: The practice of using algorithms and statistical models to enable computers to learn from data without being explicitly programmed.



3. Predictive Analytics: The practice of using data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data.
4. Natural Language Processing (NLP): The ability of computers to understand, interpret, and generate human language.



5. Artificial Intelligence (AI): The ability of machines to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and language translation.
6. Big Data: A term used to describe large, complex datasets that cannot be easily processed using traditional data processing techniques.



7. Data Science: The interdisciplinary field that involves the use of statistical and computational methods to extract knowledge and insights from data.
8. Deep Learning: A subset of machine learning that involves neural networks with multiple layers, allowing for more complex data processing and analysis.



9. Data Visualization: The graphical representation of data and information to facilitate understanding and communication.

10. Data Cleansing: The process of detecting and correcting or removing corrupt or inaccurate records from a dataset.

