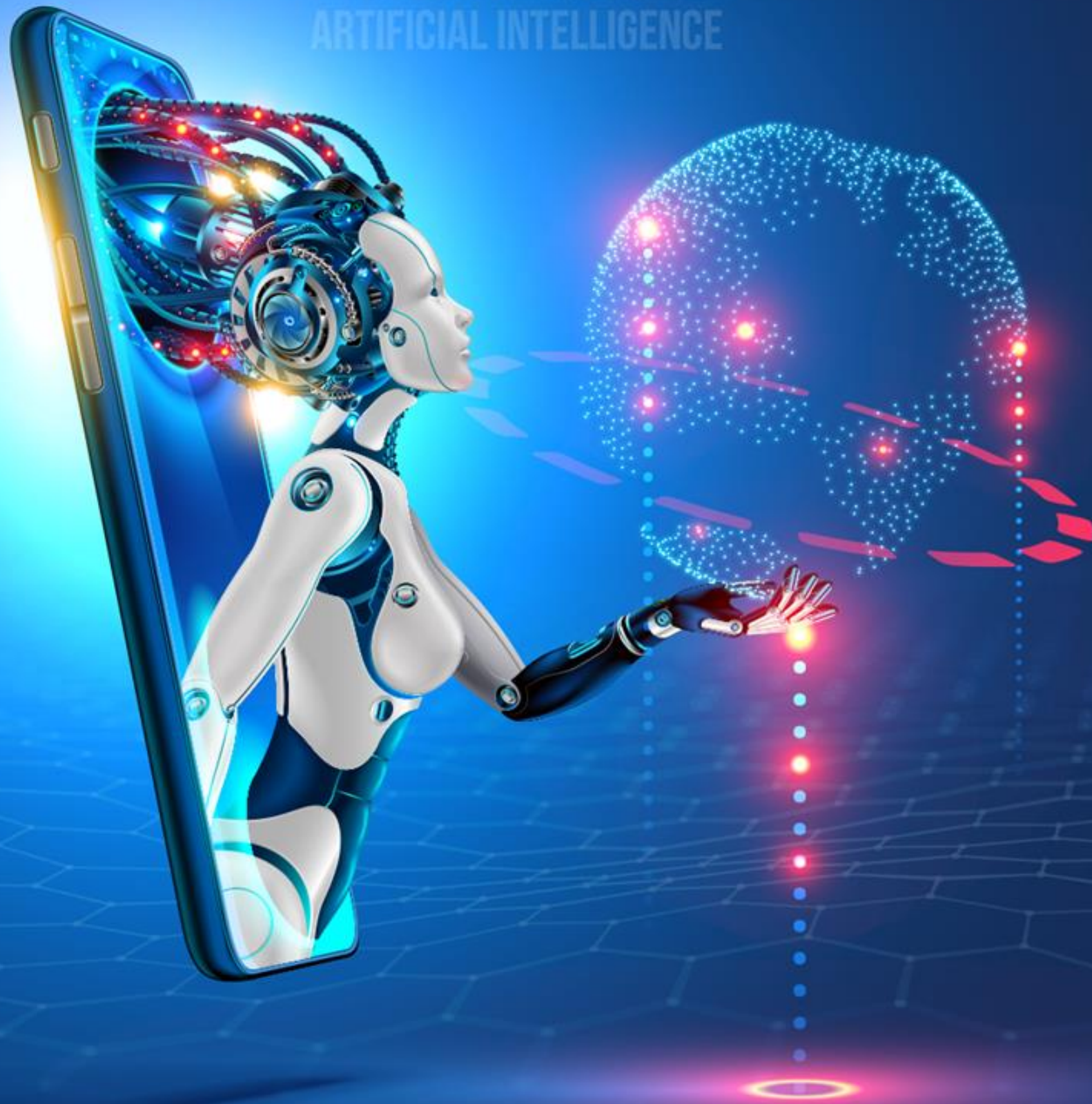


DATA AND ARTIFICIAL INTELLIGENCE



Data Science with Python

DATA AND ARTIFICIAL INTELLIGENCE



Data Science Overview

Learning Objectives

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

- 🕒 Define Data Science
- 🕒 Discuss the roles and responsibilities of a Data Scientist
- 🕒 List various applications of Data Science
- 🕒 Explain Data Science importance
- 🕒 Describe Python and its importance



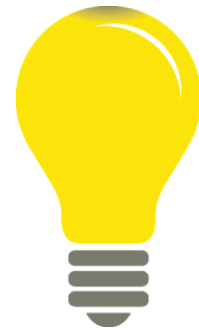
DATA AND ARTIFICIAL INTELLIGENCE

Data Science

What Is Data Science?

Some common definitions of Data Science are:

A powerful new approach to make discoveries from data



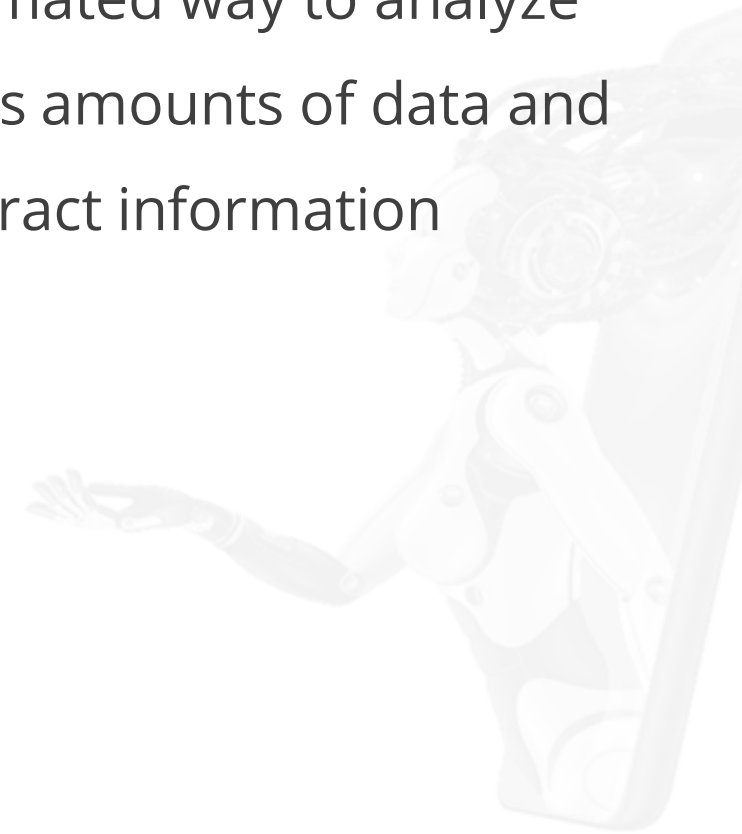
Data Science



An automated way to analyze enormous amounts of data and extract information

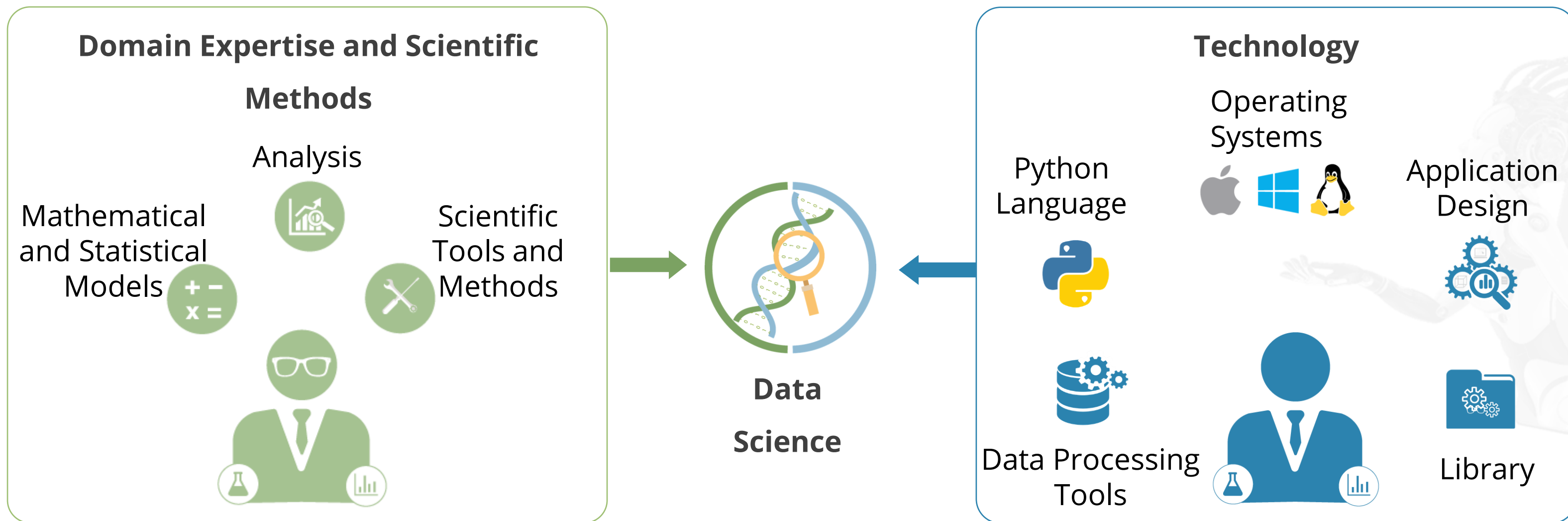


A new discipline that combines the aspects of statistics, mathematics, programming, and visualization to turn data into information



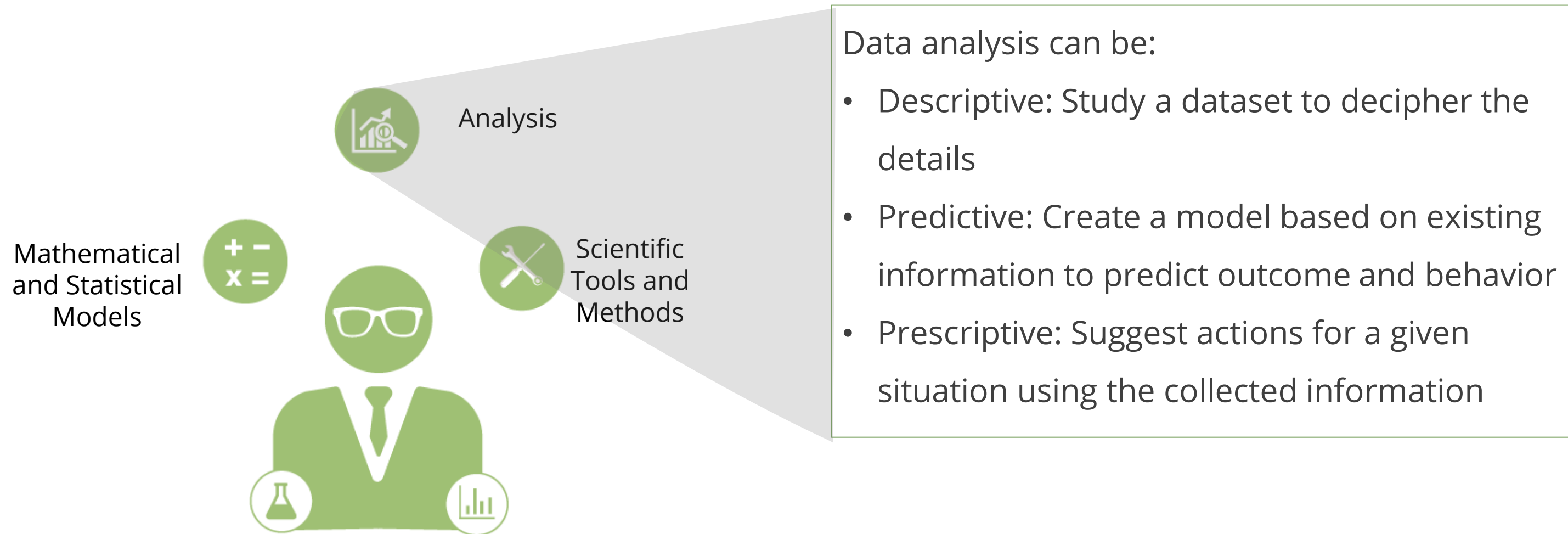
Components of Data Science

When you combine domain expertise and scientific methods with technology, you get Data Science.



Domain Expertise and Scientific Methods

Data Scientists collect, explore, analyze, and visualize data. They apply mathematical and statistical models to find patterns and solutions in the data.

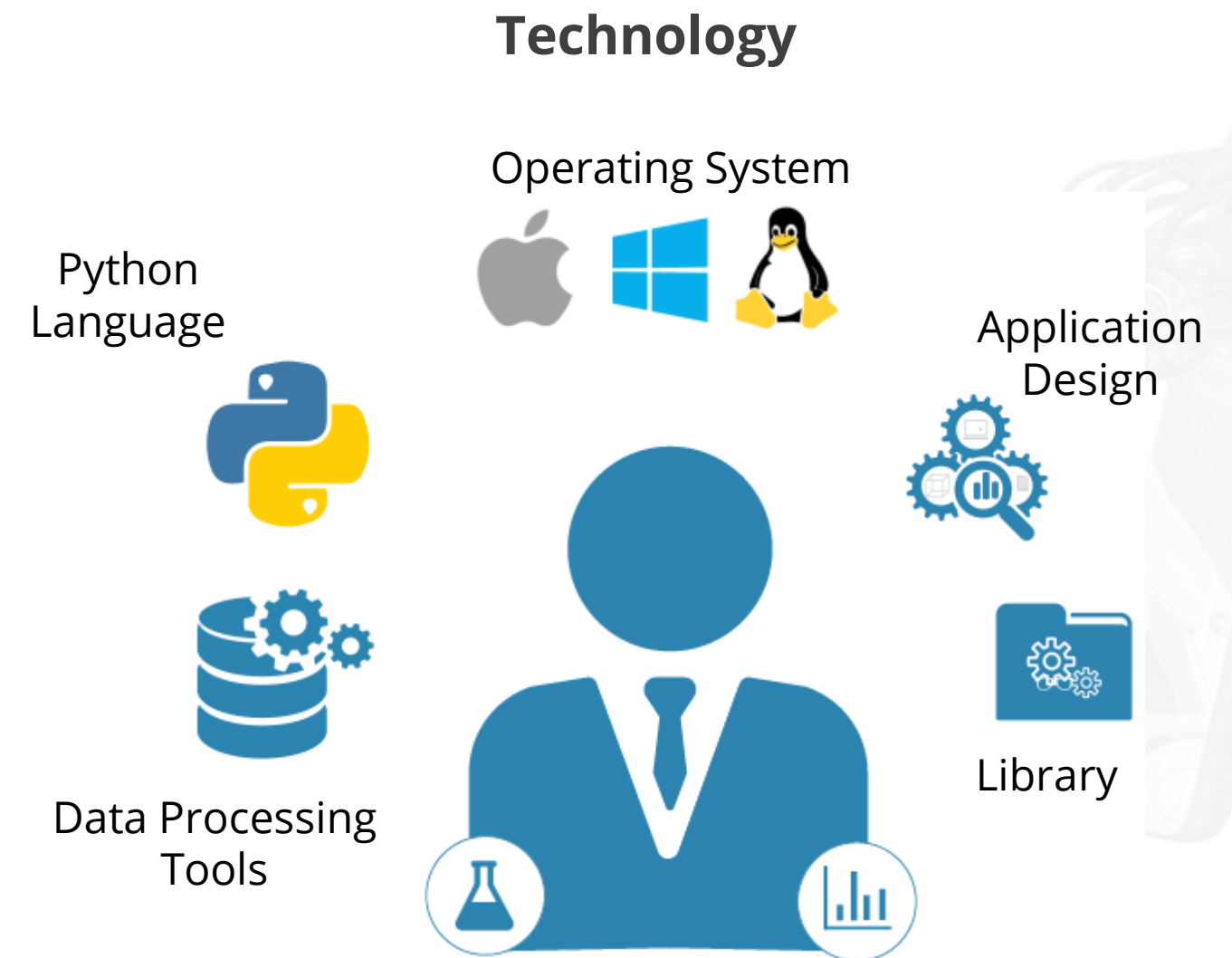


Data Processing and Analytics

Modern tools and technologies have made data processing and analytics faster and efficient.

These technologies help Data Scientists to:

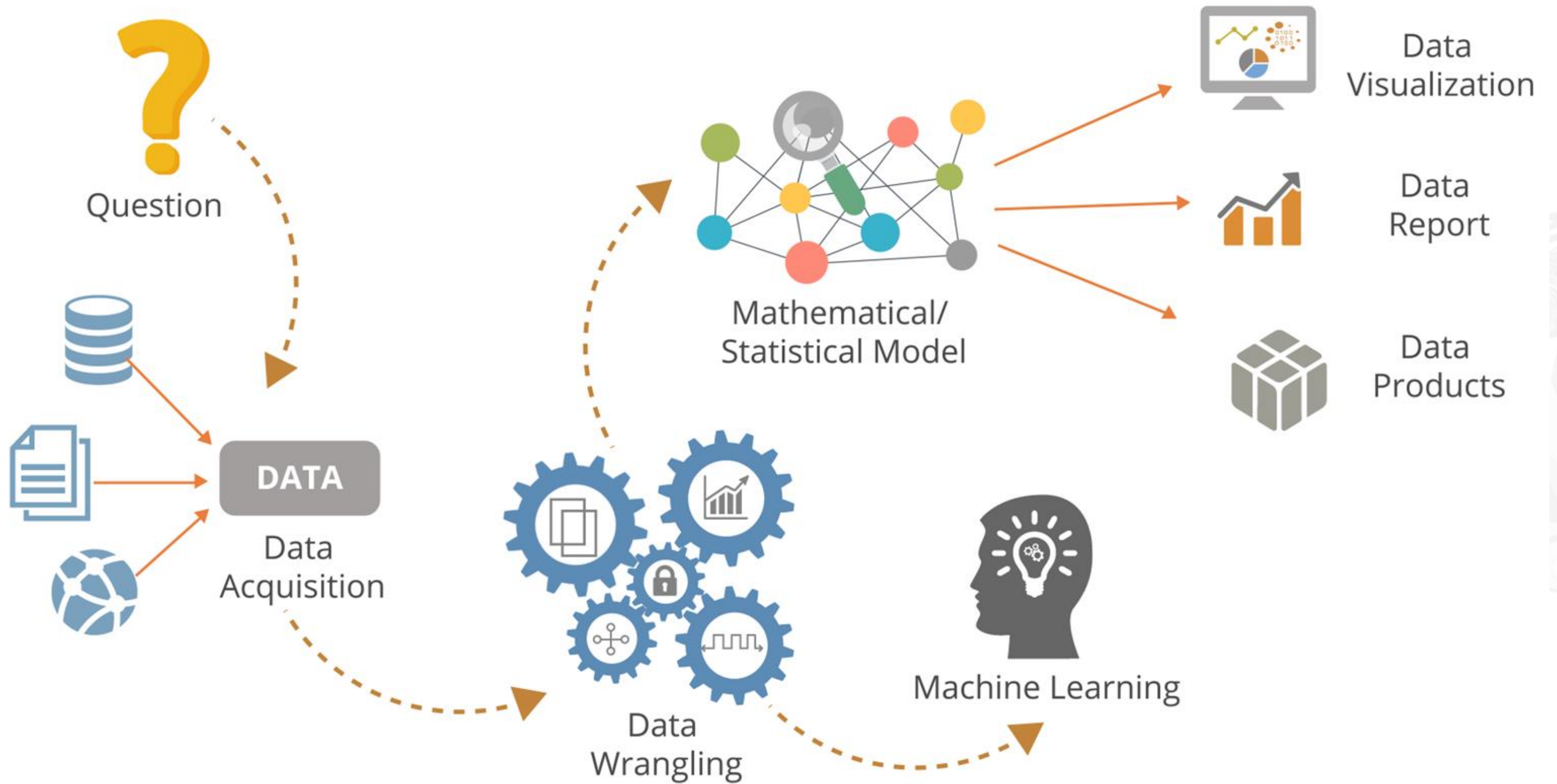
- Build and train machine learning models
- Manipulate data with technology
- Build data tools, applications, and services
- Extract information from data



Data analysis that uses only technology and domain knowledge without mathematical and statistical knowledge often leads to incorrect patterns and wrong interpretations. This can cause serious damage to businesses.

Roles and Responsibilities of a Data Scientist

Role of a Data Scientist



Basic Skills of a Data Scientist

A Data Scientist should be able to:

- Ask the right questions
- Understand data structure
- Interpret and wrangle data
- Apply statistical and mathematical methods
- Visualize data and communicate with stakeholders
- Work as a team player



Sources of Big Data

Data Scientists work with different types of datasets for various purposes. Now that Big Data is generated every second through different media, the role of Data Science has become more important.



3 Vs of Big Data

Volume

Enormous amount of data generated from various sources

Velocity

Large amount of data streaming in at great speeds, which requires quick data processing

Variety

Different formats of data: Structured, Semi-structured, and Unstructured

Big Data is a huge collection of data stored on distributed systems/machines popularly referred to as Hadoop clusters.

Data Science helps extract information from the data and build information-driven enterprises.

Applications of Data Science

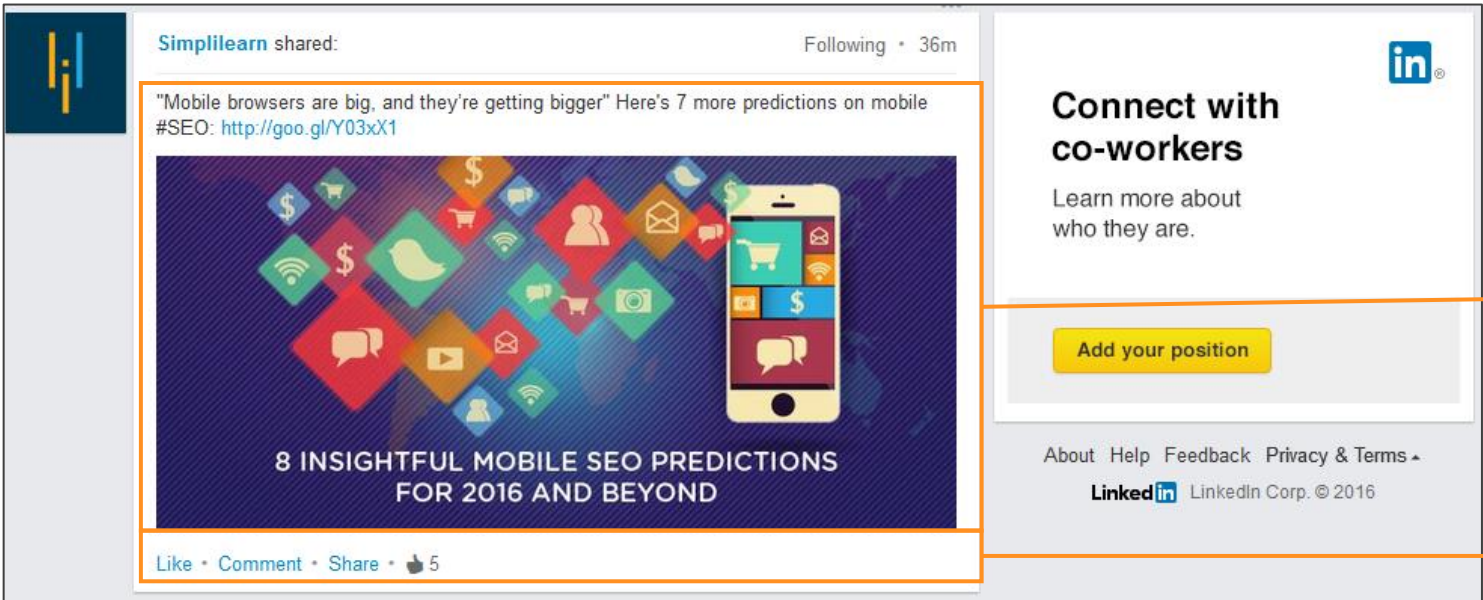
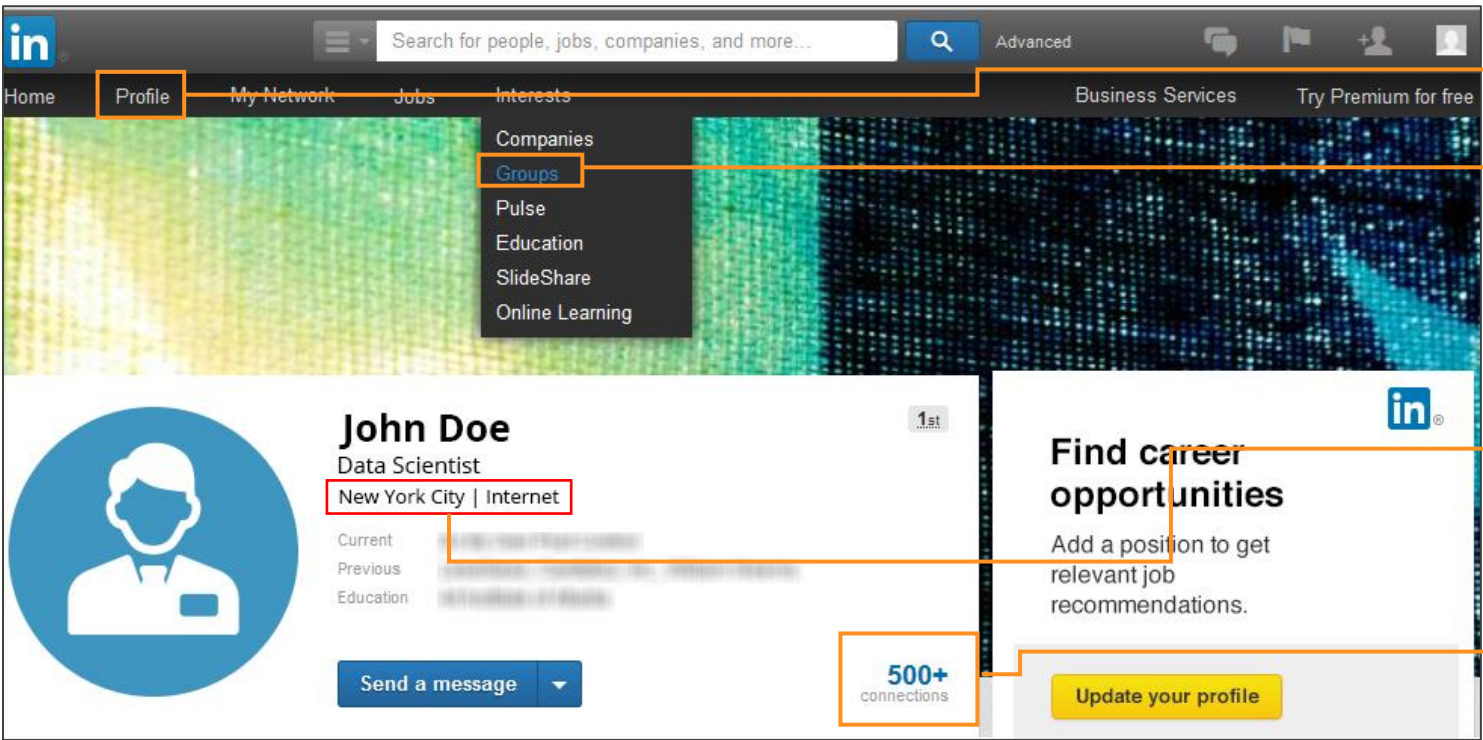
Different Sectors Using Data Science

Various sectors use Data Science to extract the information they need to create different services and products.



Using Data Science: Social Network Platforms

LinkedIn uses data points from its users to provide them with relevant digital services and data products.



Profile

Groups

Location

Connections

Post

Likes

Data Points



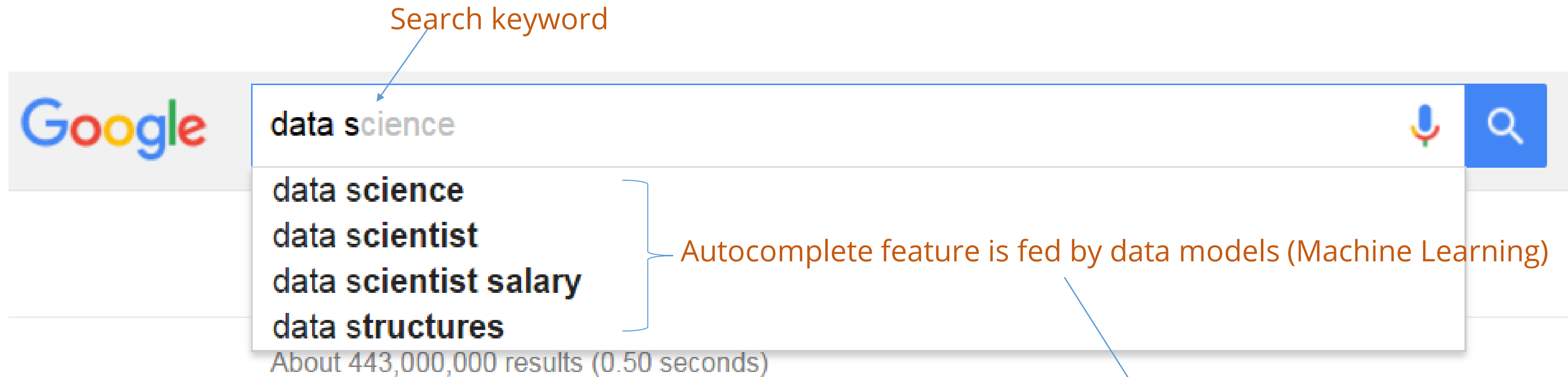
Information

Digital Services

Data Products

Using Data Science: Search Engines

Google uses Data Science to provide relevant search recommendations as the user types a query.



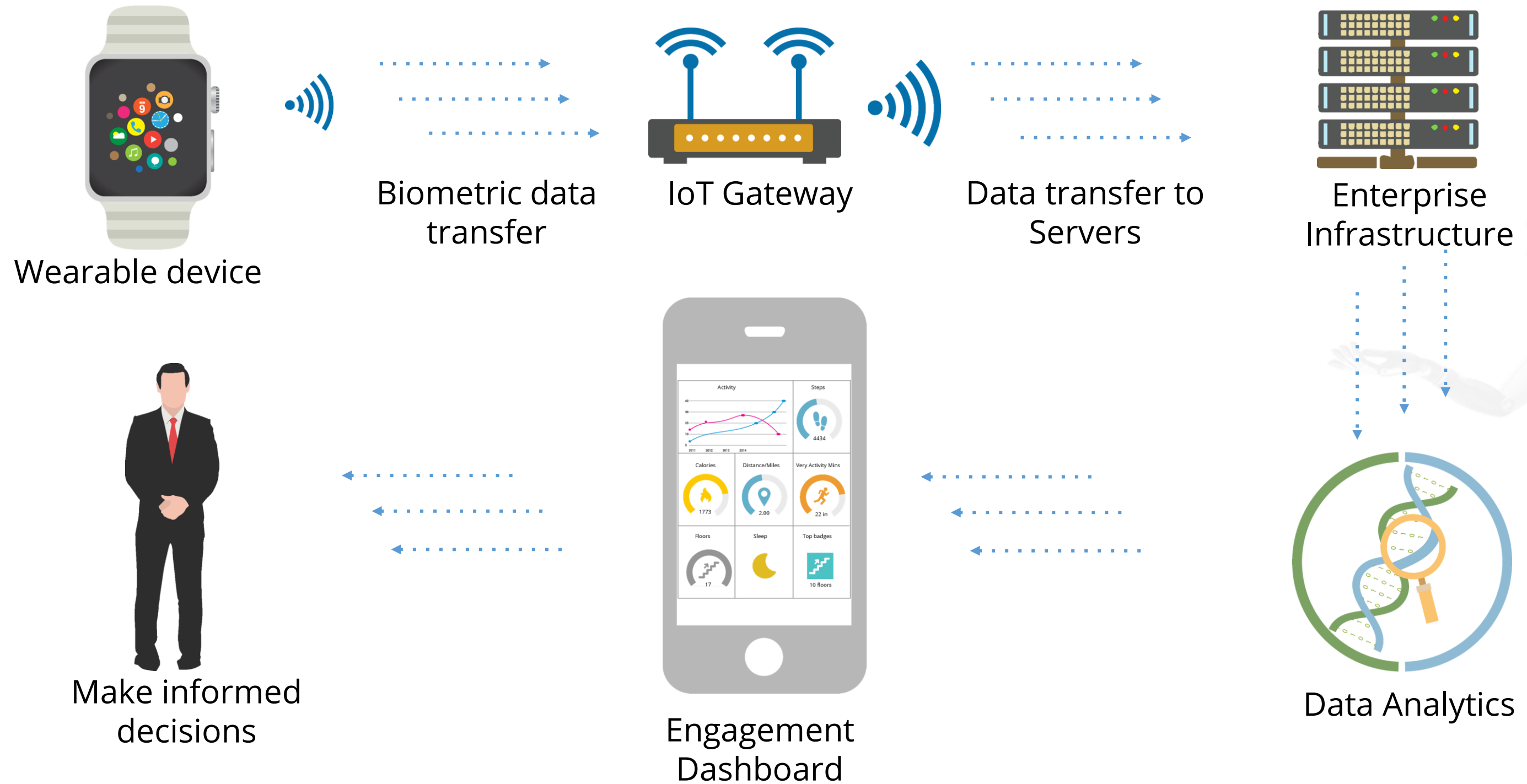
Fast and real-time analytics is made possible by modern and advanced infrastructure, tools, and technologies

Influencing Factors

1. Query Volume – Unique and verifiable users
2. Geographical locations
3. Keyword/phrase matches on the web
4. Some scrubbing for inappropriate content

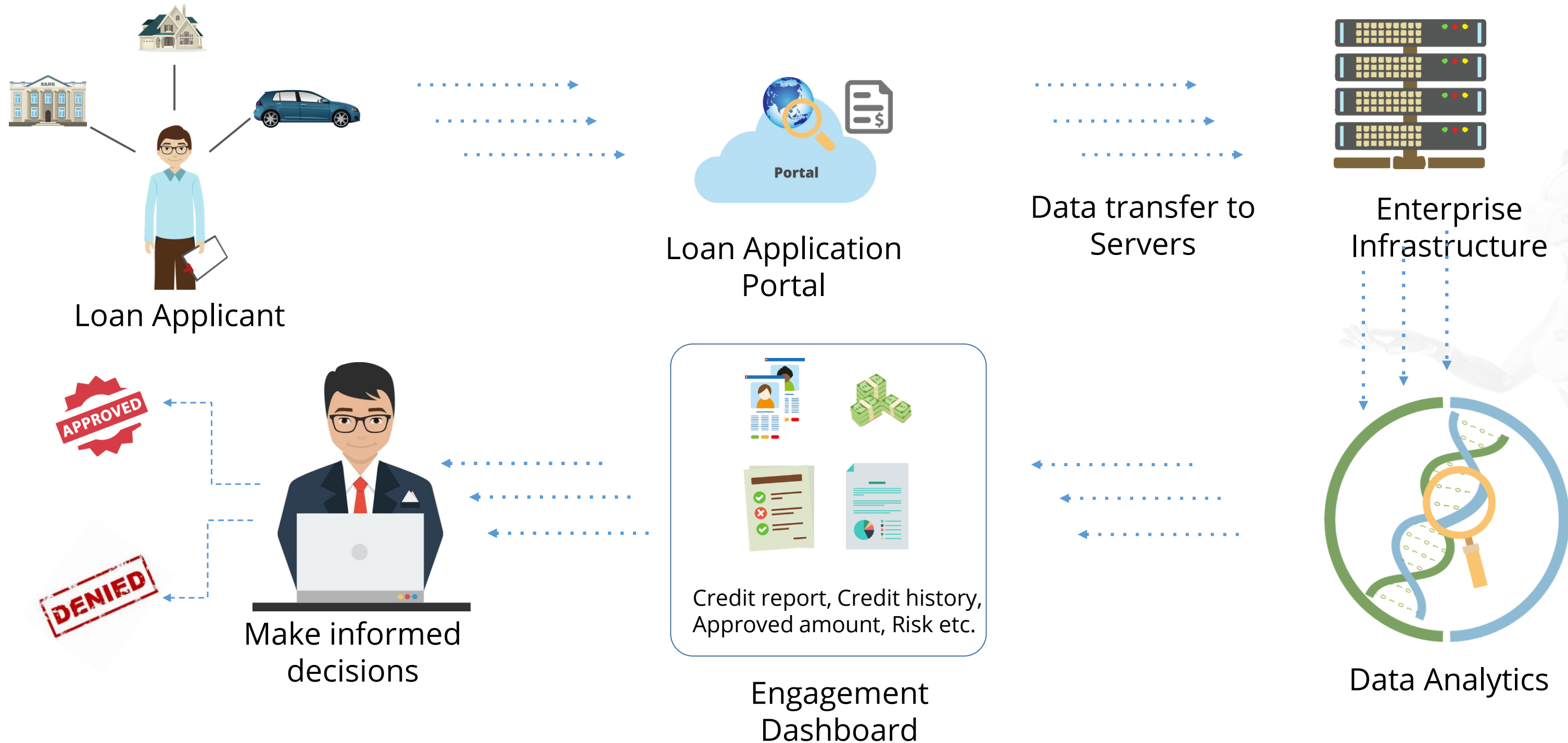
Using Data Science: Healthcare

Wearable devices use Data Science to analyze data gathered by their biometric sensors.



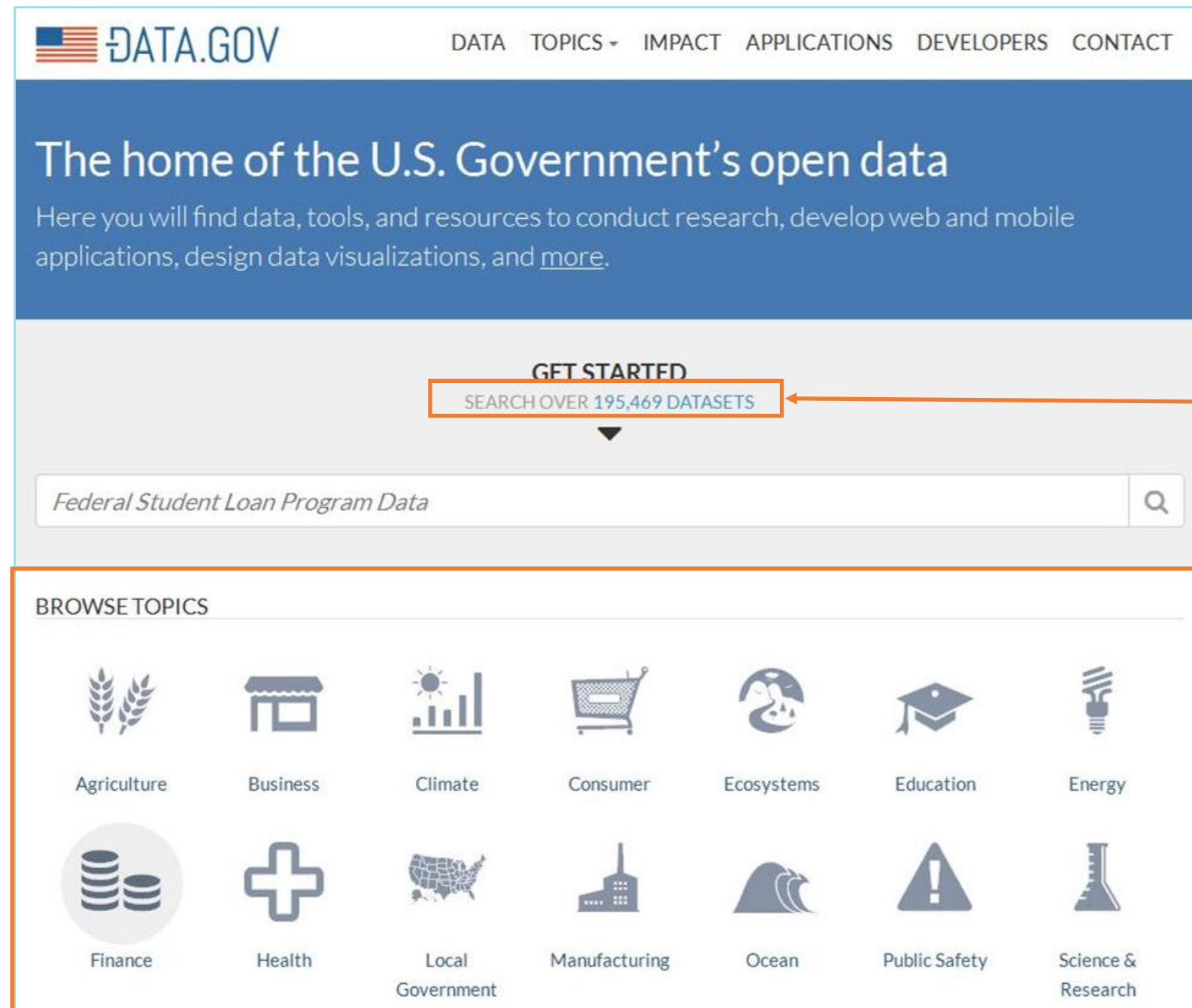
Using Data Science: Finance

A loan manager can easily access and sift through a loan applicant's financial details using Data Science.



Using Data Science: Public Sector

The governments in different countries share large datasets from various domains with the public. Data.gov is a website hosted and maintained by the U.S. government.



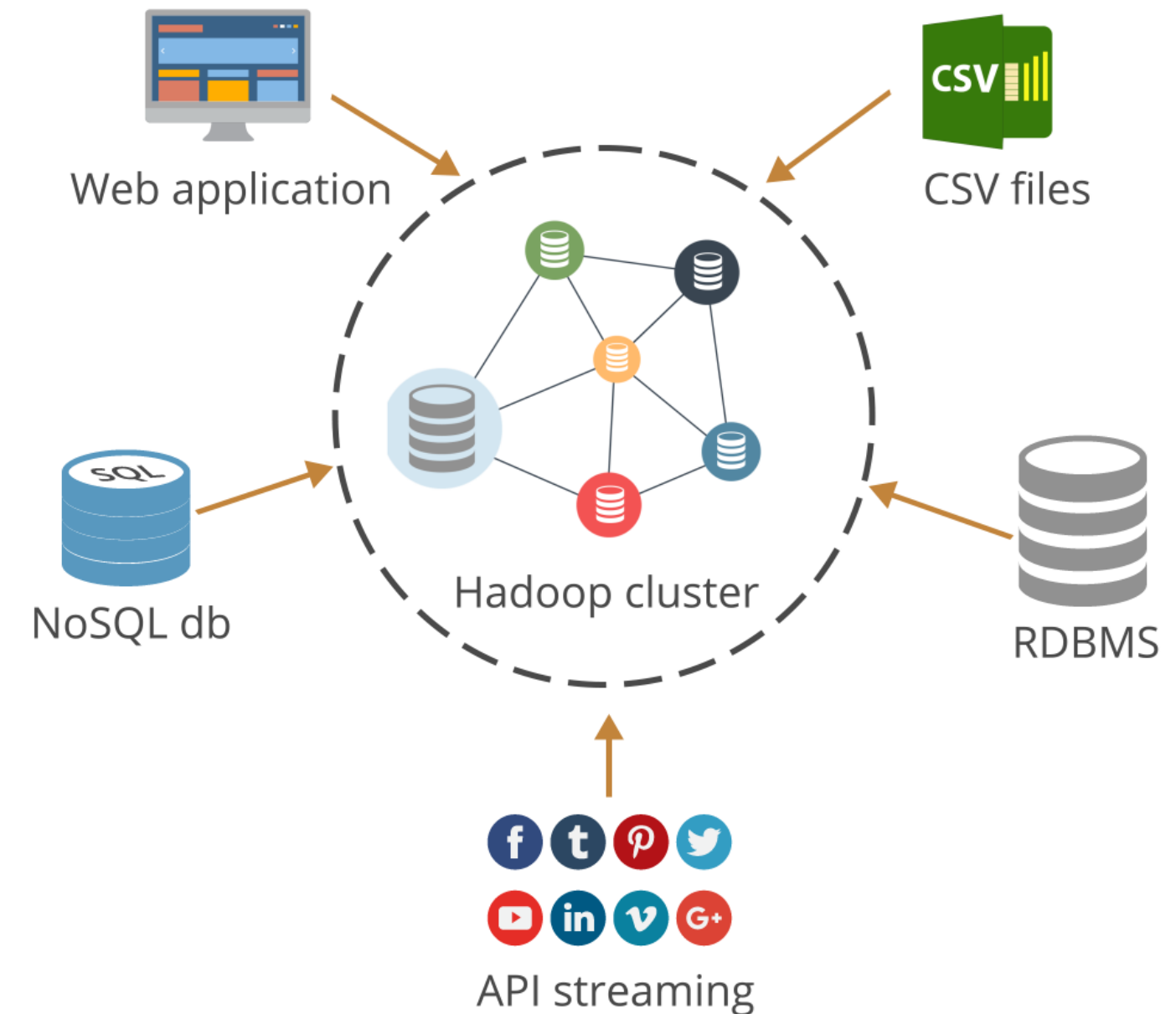
Large collection of datasets

Sectors/Domains

The Real Challenge

Some of the challenges Data Scientists face in the real world are:

- Data quality doesn't conform to the set standards
- Data integration is a complex task
- Data is distributed into large clusters in HDFS, which is difficult to integrate and analyze
- Unstructured and semi-structured data are harder to analyze

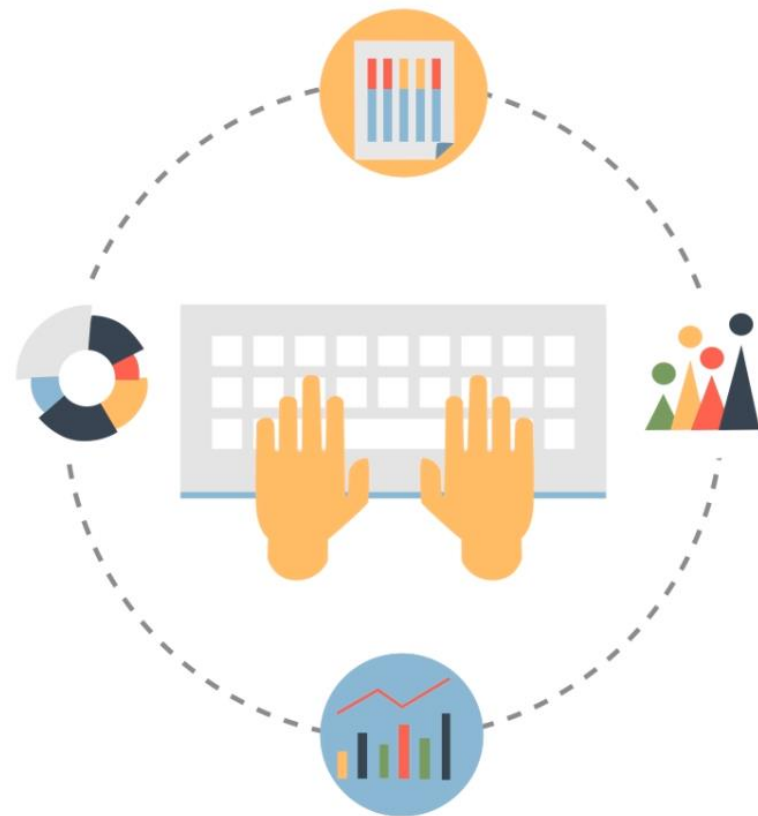


DATA AND ARTIFICIAL INTELLIGENCE

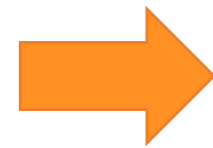
Python

Data Analytics and Python

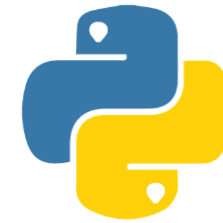
Python deals with each stage of data analytics efficiently by applying different libraries and packages.



Data
Analytics



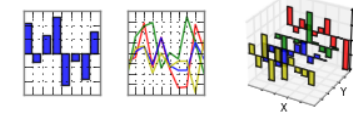
Acquire



Wrangle

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Explore

matplotlib

Model



Visualize

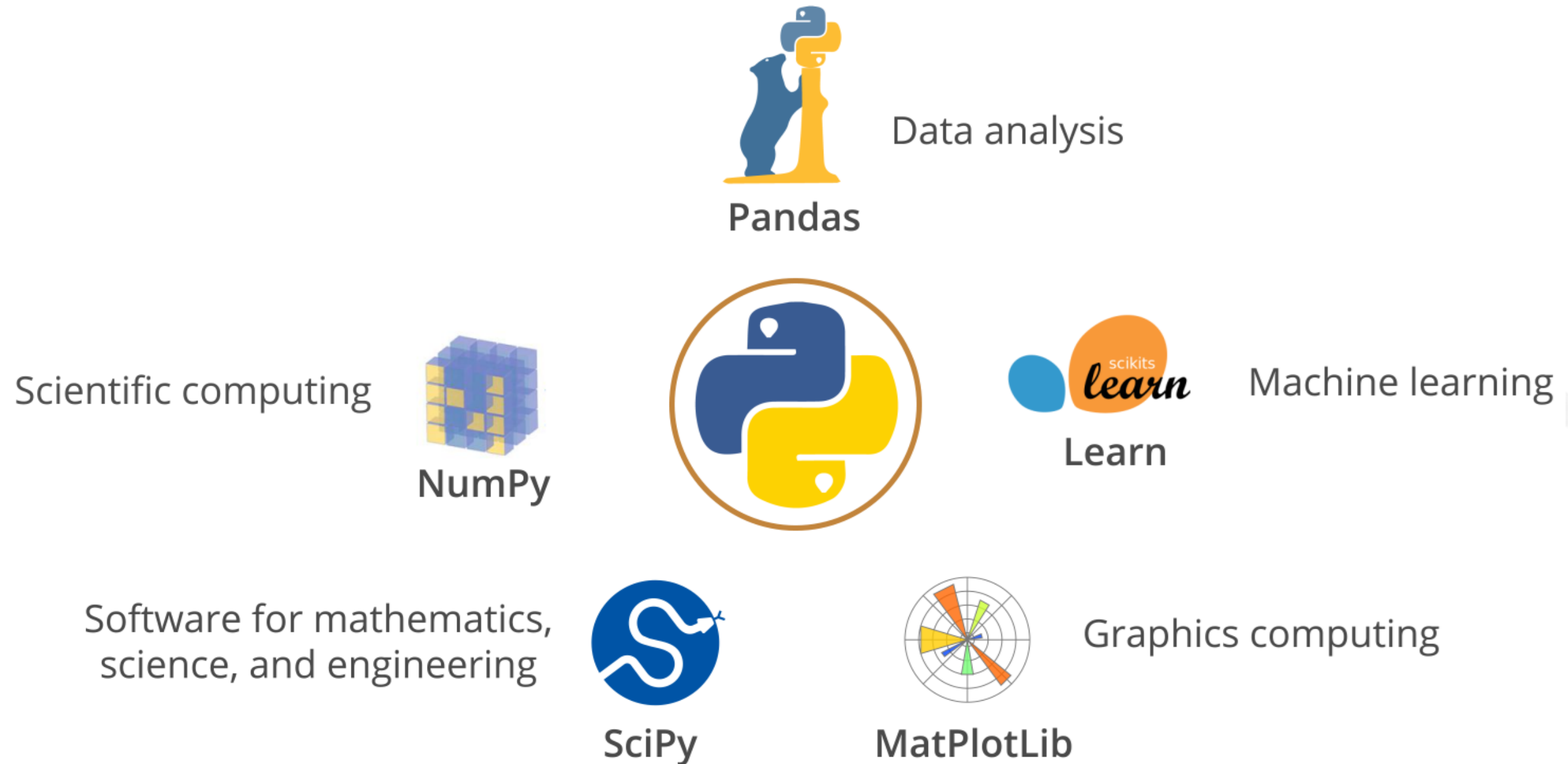


Bokeh

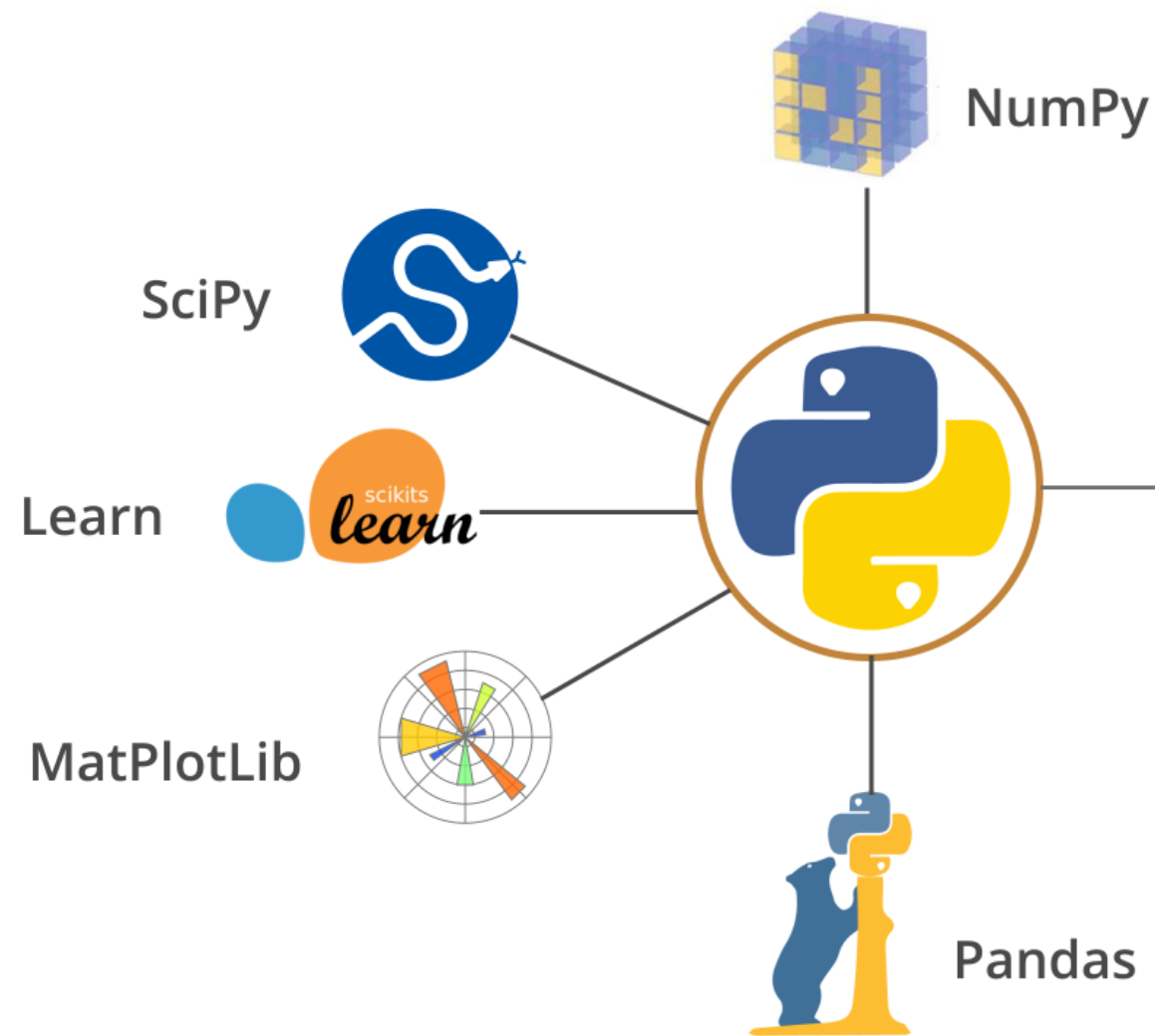


Python Tools and Technologies

Python is a general purpose, open source programming language that lets you work quickly and integrate systems more effectively.



Benefits of Python



Easy to learn

Open source

Efficient and multi-platform support

Huge collection of libraries, functions, and modules

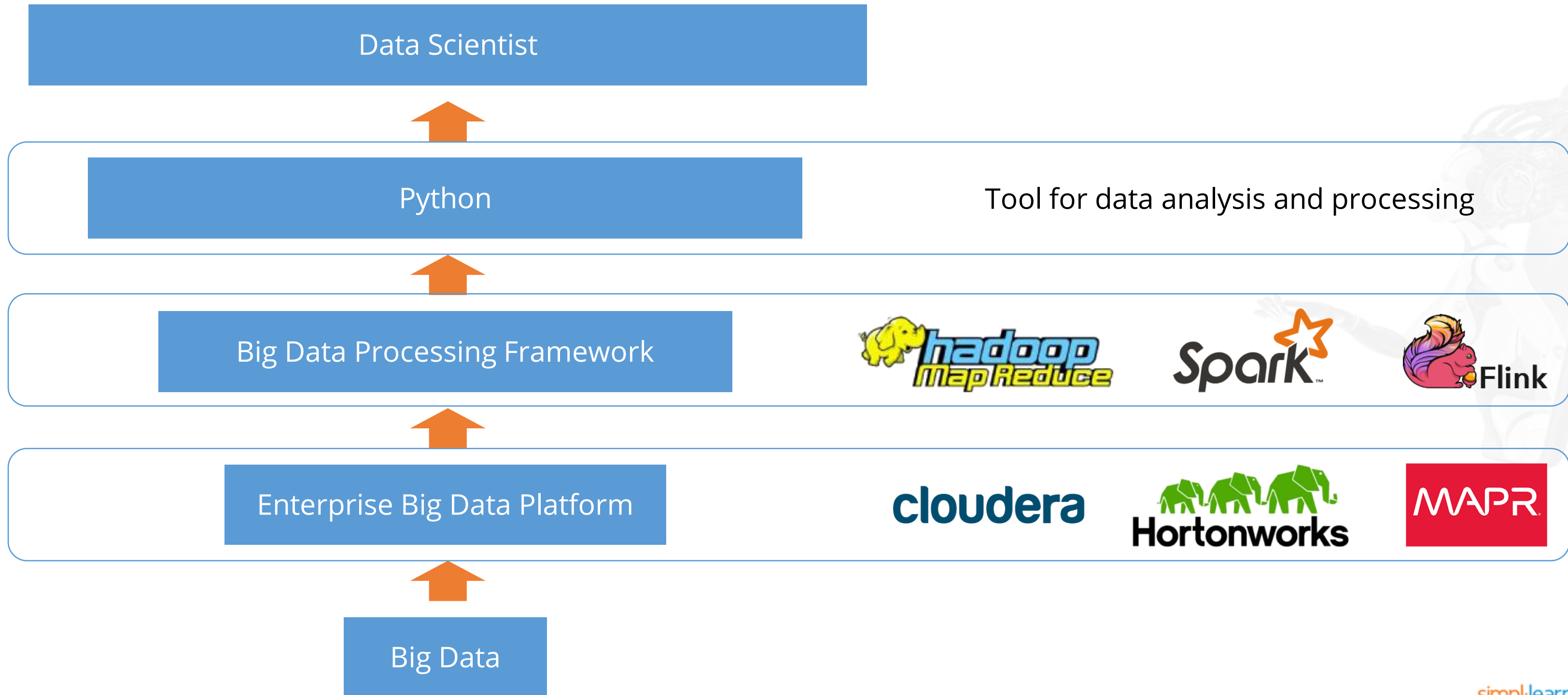
Big open source community

Integrates well with enterprise apps and systems

Great vendor and product support

Big Data Platforms and Processing Frameworks for Python

Python is supported by well-established data platforms and processing frameworks that help analyze data in a simple and efficient way.



DATA AND ARTIFICIAL INTELLIGENCE



Knowledge Check

Knowledge Check

1

A Data Scientist ____.

- a. Asks the right questions
- b. Acquires data
- c. Performs data wrangling and data visualization
- d. All of the above



Knowledge Check

1

A Data Scientist ____.

- a. Asks the right questions
- b. Acquires data
- c. Performs data wrangling and data visualization
- d. All of the above



The correct answer is **d**

A Data Scientist asks the right questions to the stakeholders, acquires data from various sources and data points, performs data wrangling that makes the data available for analysis, and creates reports and plots for data visualization.

**Knowledge
Check**

2

The search engine's autocomplete feature identifies unique and verifiable users who search for a particular keyword or phrase to _____. Select all that apply.

- a. Scrub inappropriate content
- b. Build a query volume
- c. Tag the location to a query
- d. Find similar instances on the web



Knowledge
Check

2

The search engine's autocomplete feature identifies unique and verifiable users who search for a particular keyword or phrase to _____. Select all that apply.

- a. Scrub inappropriate content
- b. Build a query volume
- c. Tag the location to a query
- d. Find similar instances on the web



The correct answer is **b, c**

The search engine's autocomplete feature identifies unique and verifiable users who search for a particular keyword or phrase to build a query volume. It also helps identify the users' locations and tag them to the query, enabling it to be location-specific.

What is the sequential flow of Data Analytics?

- a. Data wrangling, exploration, modeling, acquisition, and visualization
- b. Data exploration, acquisition, modeling, wrangling, and visualization
- c. Data acquisition, wrangling, exploration, modeling, and visualization
- d. Data modeling, acquisition, exploration, wrangling, and visualization



Knowledge Check

3

What is the sequential flow of Data Analytics?

- a. Data wrangling, exploration, modeling, acquisition, and visualization
- b. Data exploration, acquisition, modeling, wrangling, and visualization
- c. Data acquisition, wrangling, exploration, modeling, and visualization
- d. Data modeling, acquisition, exploration, wrangling, and visualization



The correct answer is **c**

In Data Analytics, the data is acquired from various sources and is then wrangled to ease its analysis. This is followed by data exploration and data modeling. The final stage is data visualization, where the data is presented and the patterns are identified.

Key Takeaways

You are now able to:

- Define Data Science
- Discuss the roles and responsibilities of a Data Scientist
- List various applications of Data Science
- Explain Data Science importance
- Describe Python and its importance



Thank You