BETES
Discover Technologies, expand opportunities!

Bootcamps for Emerging Technologies and Essential Skills

Topic 5: Data Analytics

Developed by

Digital Idea
Scientific Association

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.
Project Number: 2022-1-CY01-KA210-VET-000082706

Co-funded by the European Union

Bootcamps for Emerging Technologies and Essential Skills

BETES
Discover Technologies, expand opportunities!

Consortium

Co-ordinator: ECECT

Partners: Digital Idea Scientific Association, SKILLSZONE EMPOWER YOURSELF

Co-funded by the European Union

What will you learn from this bootcamp?

BETES
Discover Technologies, expand opportunities!

- Data Analysis Fundamentals: Definitions, Characteristics, Digital Footprint
- Data Analysis Importance: Why and How, Applications and Implications
- Data Analysis Careers: Ethics, Skills, Job opportunities
- Advanced Data Science: Machine learning, Artificial Intelligence

Co-funded by the European Union

Data Analysis Fundamentals

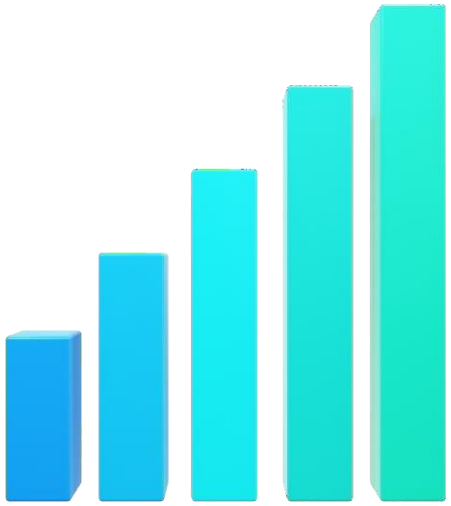
Definitions, Characteristics, Digital Footprint

BETES
Discover Technologies, expand opportunities!

Digital Idea
Scientific Association

Co-funded by the European Union

Introduction to Data Analysis



What?

- Examining, cleaning, transforming, and modelling data

Why?

- Discover useful information, Draw conclusions, Support decision-making

How?

- **Interpret** and **communicate** data findings using *statistical*, *algorithmic*, and *visualization* techniques

Where?

- Business, Science, Healthcare, Technology

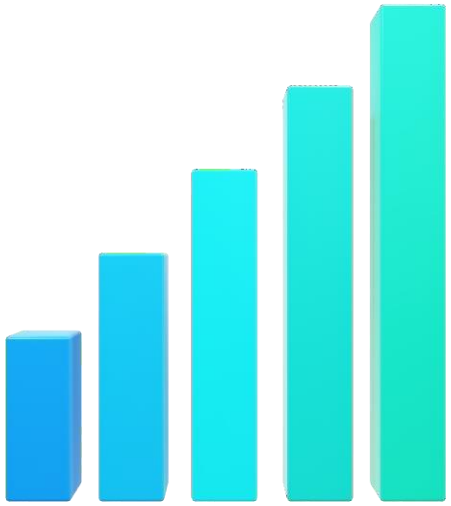
Who?

- Individuals, Organizations

Make data-driven decisions by turning raw data into insights, identifying trends, patterns, and anomalies, and providing a factual basis for strategic planning

Co-funded by the European Union

Introduction to Data Analysis



The importance of data analysis in the modern digital world

1. Drive Decision-Making
2. Enhance Customer Experience
3. Predict Trends
4. Improve Efficiency
5. Innovate and Compete
6. Mitigate Risk
7. Personalization

Integral to navigating the complexities of the digital world, enabling smarter strategies, better performance, and stronger competitive positioning

Co-funded by the European Union

Origins of DATA I

BETES
Discover Technologies, expand opportunities!

Internal Company Data:

Sales records

Customer databases

Inventory logs

Financial documents

Employee records

Operational data

Customer-Generated Data:

Online purchases

Social media interactions

Customer feedback and reviews

Survey responses

Customer service records

Co-funded by the European Union

Origins of DATA II

BETES
Discover Technologies, expand opportunities!

Machine-Generated Data:

Sensors and IoT devices

Telematics from vehicles

Smart appliances

Industrial machines

Web and Social Media:

User-generated content

Web traffic data

Clickstream data

Co-funded by the European Union

Origins of DATA III

BETES
Discover Technologies, expand opportunities!

Third-Party Data Providers:

Credit bureaus

Market research firms

Data aggregators

Mobile Data:

Location data from smartphones and GPS devices

App usage statistics

Co-funded by the European Union

Origins of DATA IV

BETES
Discover Technologies, expand opportunities!

Transactional Data:

Point of sale systems

E-commerce platforms

Research and Academic Databases:

Scholarly articles

Research reports

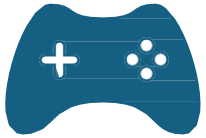
Thesis and dissertations

Co-funded by the European Union

Origins of DATA V

Multimedia and Entertainment:

Streaming services data



Gaming platforms

Biometric Data:

Health monitoring devices

Fitness trackers

Co-funded by the European Union

Origins of DATA VI

BETES
Discover Technologies, expand opportunities!

Log Files:

Server logs

Network logs

Application logs

Public Data and Open Data Sources:

Government databases

Public records

Open data platforms

Co-funded by the European Union

Data Categories

BETES
Discover Technologies, expand opportunities!

- Structured Data
- Unstructured Data
- Semi-structured Data
- Quantitative Data
- Qualitative Data
- Time-series Data
- Cross-sectional Data
- Big Data
- Transactional Data
- Web Data

Co-funded by the European Union

Structured Data

Highly organized and formatted in a way that makes it easily searchable and understandable.

It typically resides in relational databases (RDBMS).

Characteristics:

- It follows a schema
- The data is often tabular with rows and columns.
- Easy to enter, query, and analyze.
- Examples: Excel files, SQL databases, and CSV files.

Advantages:

- Easier to search and organize.
- Efficiently processed and analyzed by traditional data analysis tools.
- Ideal for operations like sorting, filtering, and aggregation.

Disadvantages:

- The rigidity of the structure > less adaptable to changes.
- May not be well-suited to capturing complex or nuanced information.

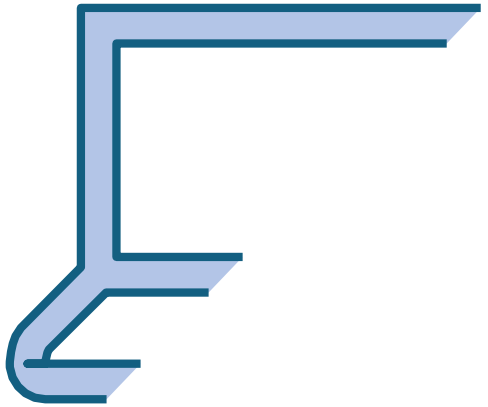
Co-funded by the European Union

Unstructured Data

Does not follow any specific format or structure, making it more complex and less easy for traditional data management practices to handle.

Characteristics:

- It does not fit neatly into traditional relational databases.
- Can be textual or non-textual, and includes data formats like email messages, videos, photos, social media posts, and webpages.
- Requires more storage.
- Difficult to analyze and process using conventional tools and techniques.



Advantages:

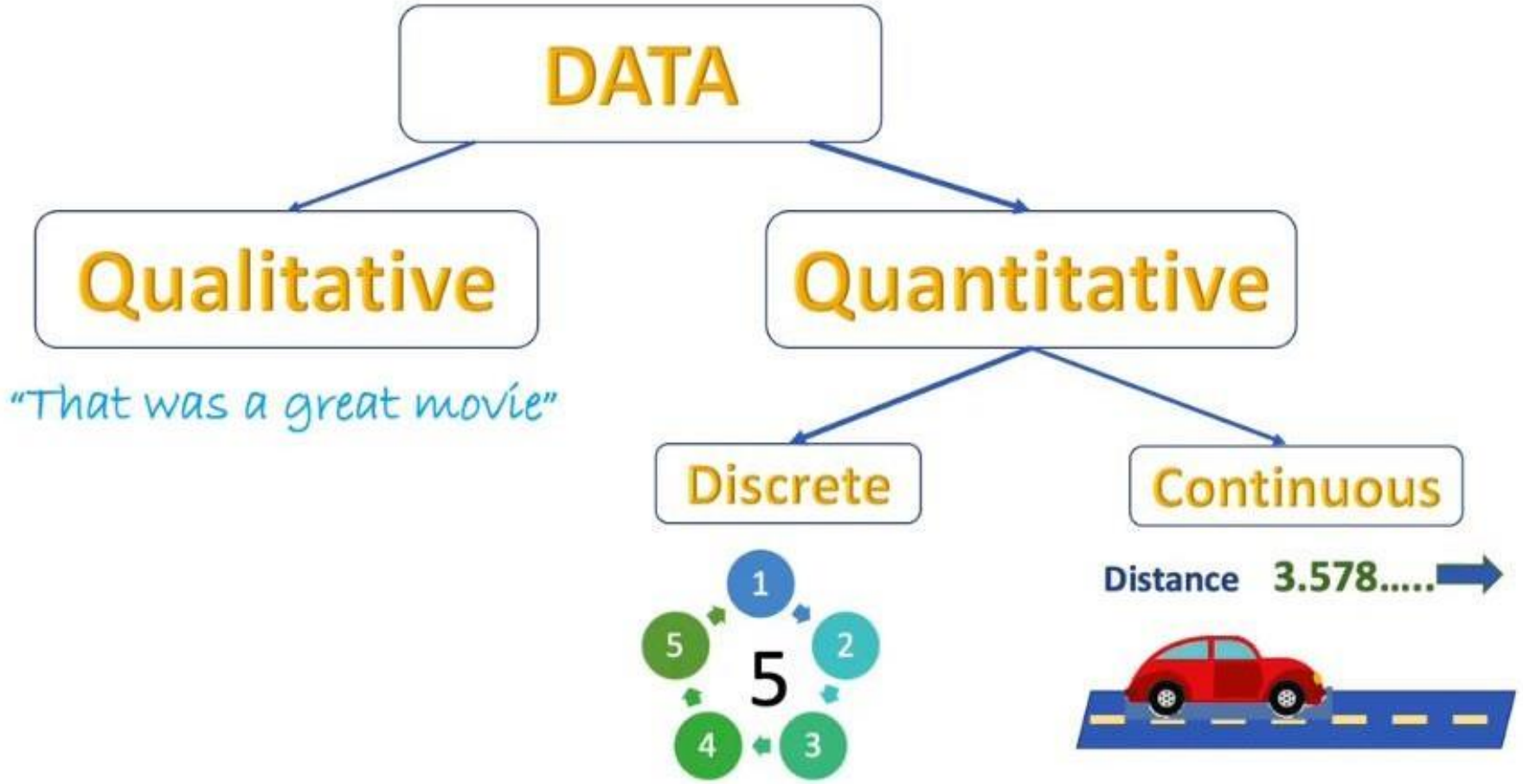
- Capable of capturing a wider and more varied range of information.
- Reflects the nature of data in the real-world.
- Provides rich data sources that can be mined for valuable insights using modern analytical techniques.

Disadvantages:

- Challenging to categorize and analyze without advanced tools like natural language processing (NLP) and machine learning algorithms.
- Requires more complex and resource-intensive processes to turn into usable data.

Co-funded by the European Union

Discrete vs Continuous Data



Co-funded by the European Union

Discrete Data:

- **Nature:** Distinct, separate values. It can be counted and typically involves integers.
- **Examples:**
  - The number of employees in a company.
  - The number of cars passing through a toll booth.
  - The number of questions on a test.
  - Shoe sizes of individuals.
- **Characteristics:**
  - Often represented as counts.
  - Can take on a finite number of values in a range.
  - Bar graphs and pie charts are common visual representations of discrete data.
- **Use Cases:** Countable items (the number of occurrences of an event).

BETES
Discover Technologies, expand opportunities!
Co-funded by the European Union

Continuous Data:

- **Nature:** Take on any value within a given range, infinitely divisible. Include fractions and decimals. It typically involves measurements.
- **Examples:**
  - The height of students in a classroom.
  - The amount of time it takes to complete a task.
  - Temperature readings over a week.
  - The weight of produce in a grocery store.
- **Characteristics:**
  - Usually represented by measurements.
  - Can take on an infinite number of values within a range.
- Histograms and line graphs are common visual representations of continuous data.
- **Use Cases:** Things that can be measured and where precision is important, such as time, distance, and temperature.

BETES
Discover Technologies, expand opportunities!
Co-funded by the European Union

Common Data Types

BETES
Discover Technologies, expand opportunities!

Primitive Data Types:

- **Integer:** Whole numbers, both positive and negative.
- **Float (Floating Point):** Numbers with a fractional part, represented with decimals.
- **Double:** Double-precision floating-point numbers, which are similar to floats but with more precision.
- **Char (Character):** A single character, letter, number, or symbol.
- **Boolean:** Represents two possible values: true or false.

Composite Data Types:

- **Array:** A collection of elements of the same type placed in contiguous memory locations.
- **String:** A sequence of characters, typically used to represent text.
- **Struct (Structure):** A composite type that includes a set of variables under one name in a block of memory, allowing different variables to be accessed via a single pointer.

Co-funded by the European Union

Common Data Types

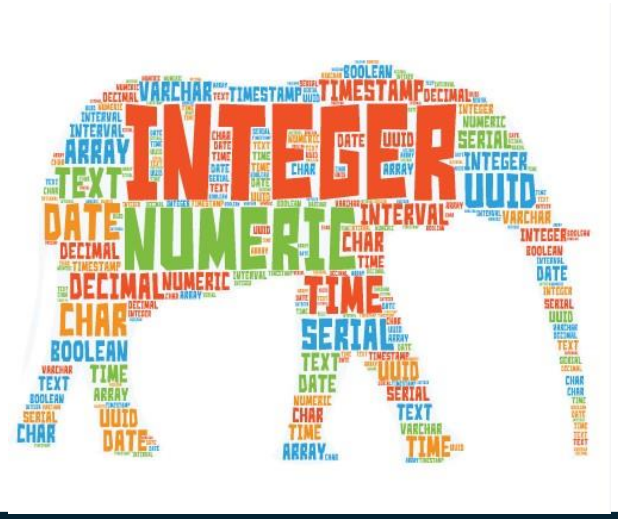
BETES
Discover Technologies, expand opportunities!

Abstract Data Types (ADT):

- **List:** An ordered sequence of items.
- **Map / Dictionary:** A collection of key-value pairs, with unique keys and associated values.
- **Set:** A collection of unique items where the order is not guaranteed.
- **Stack:** A collection that supports adding and removing elements in a last-in-first-out (LIFO) fashion.
- **Queue:** A collection that supports adding and removing elements in a first-in-first-out (FIFO) fashion.

Specialized Data Types:

- **Date/Time:** Represent specific moments in time.
- **Enumerations (Enum):** A type that consists of a fixed set of constant values (usually strings or numbers).
- **Object:** A type that represents instances of a class in object-oriented programming.



Co-funded by the European Union

Common Data Types

BETES
Discover Technologies, expand opportunities!

Pointer and Reference Data Types:

- **Integer:** Whole numbers, both positive and negative.
- **Float (Floating Point):** Numbers with a fractional part
- **Pointer:** Stores the memory address of another variable.
- **Reference:** An alias for another variable.

Data Structures:

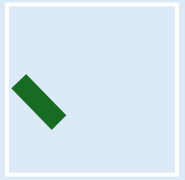
- **Linked List:** A linear collection of data elements where each element points to the next.
- **Graph:** A set of nodes connected by edges.
- **Tree:** A hierarchical structure with a single root value and subtrees of children.

File Types:
Used to represent and store data files, such as text files, binary files, image files, etc.

Co-funded by the European Union

Characteristics of Data Analysis I

Accuracy



Reliability

Relevance

Timeliness

Completeness

Co-funded by the European Union

Characteristics of Data Analysis II

- Consistency
- Granularity
- Accessibility
- Interpretability
- Integrity

Co-funded by the European Union

Digital Footprint

Definition: The trail of data left by interactions in a digital environment

Types: Active and Passive

Importance: Provide raw data for personalized marketing, targeted advertising, and enhanced user experience

Impact: Privacy and ethical considerations

Co-funded by the European Union

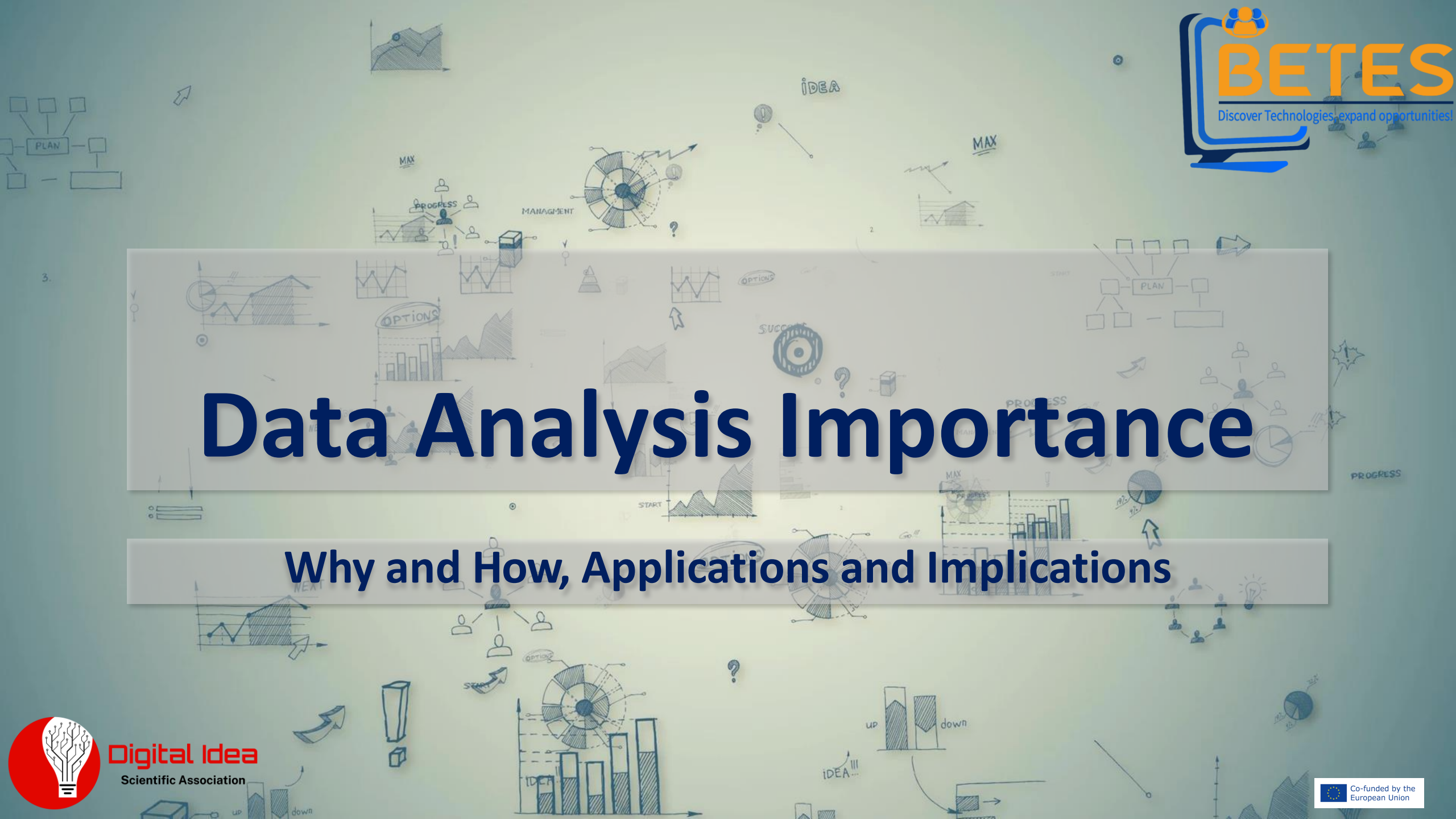
Managing Digital Footprint

- Regular reviews
- Being aware of shared information
- Privacy tools

Co-funded by the European Union

Challenges in Data Analysis

- **Data Quality and Integrity**
- **Volume and Complexity**
- **Skills Gap**
- **Integration and Siloed Data**


BETES
Discover Technologies, expand opportunities!
Data Analysis Importance
Why and How, Applications and Implications
Digital Idea
Scientific Association
Co-funded by the European Union

Evolution of Data Analysis I

BETES
Discover Technologies, expand opportunities!

Early Beginnings:

- Origins in Statistics
- Census Data

Computational Advancements:

- Mid-20th Century
- Database Systems

The Age of Software:

- Statistical Software
- Spreadsheets

Internet and the Data Explosion:

- Digital Footprints
- E-Commerce

Co-funded by the European Union

Evolution of Data Analysis II

BETES
Discover Technologies, expand opportunities!

Big Data and Analytics:

- 21st Century
- Machine Learning

Data Science Emergence:

- Multidisciplinary Approach
- Open-Source Movement

Modern Decision-Making:

- Strategic Asset
- Real-Time Analytics

Looking Ahead:

- Continuous Evolution
- Ethical and Regulatory Landscape

Co-funded by the European Union

BETES
Discover Technologies, expand opportunities!

Why Data Analysis is Critical?

- **Informed Decision-Making**
- **Predictive Power**
- **Efficiency and Cost Savings**

Co-funded by the European Union

The Rise of Big Data

Definition:

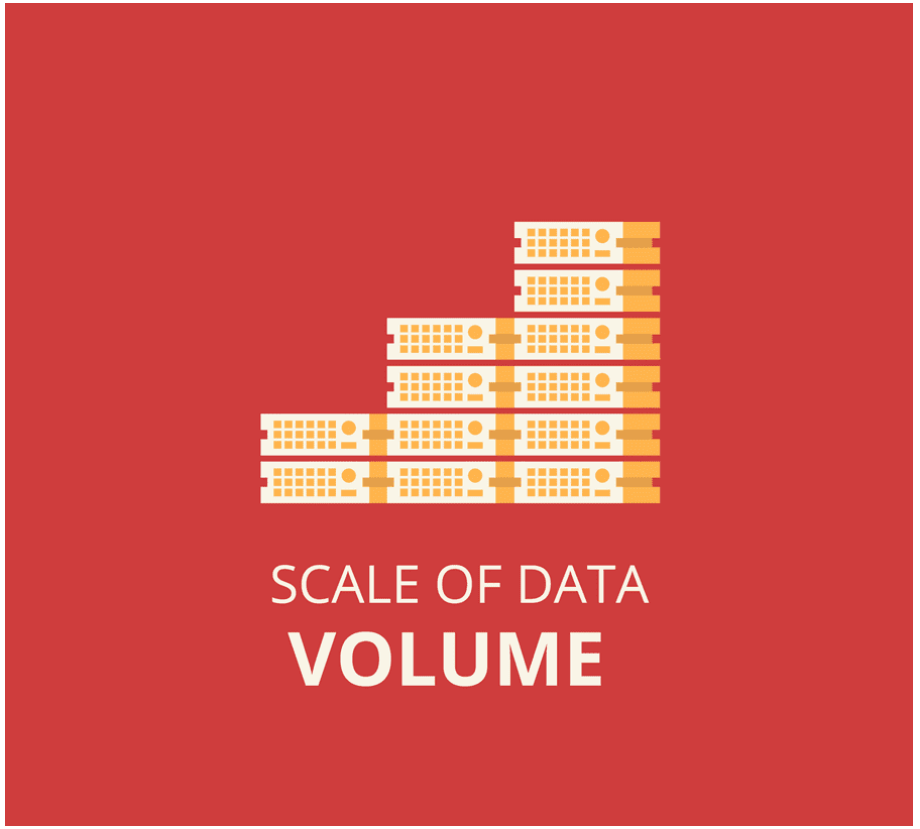
Large volumes of data, both structured and unstructured, that inundate businesses on a day-to-day basis

- Volume
- Velocity
- Variety
- Veracity

Co-funded by the European Union

Volume

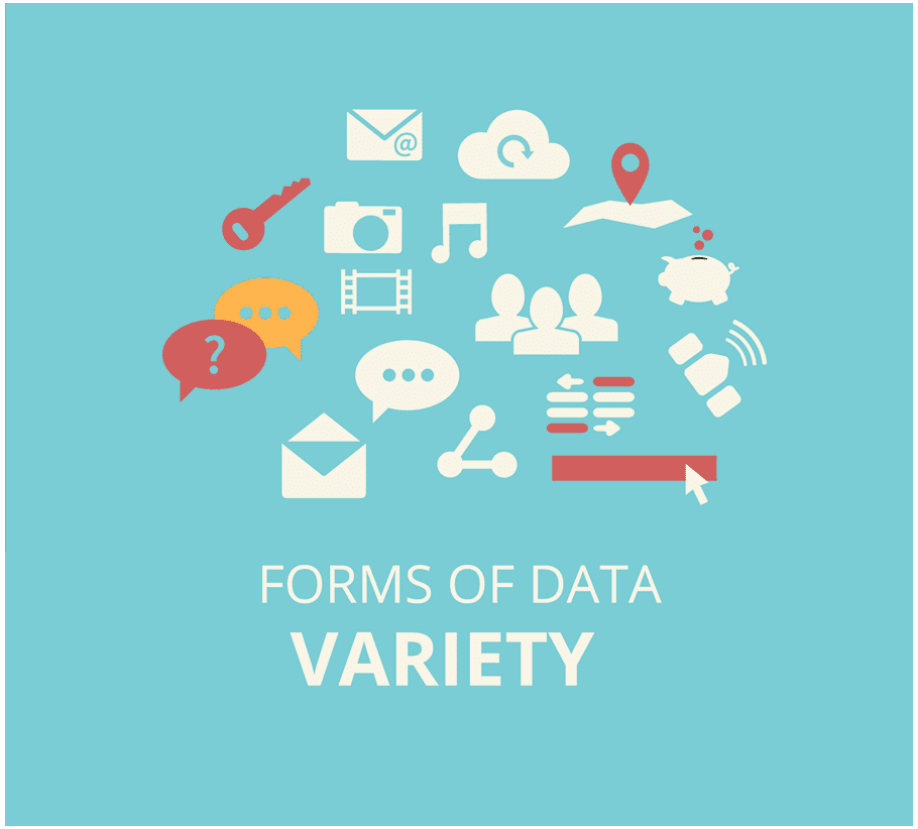
Volume describes the amount of data transported and stored. According to International Data Corporation (IDC) experts, discovering ways to process the increasing amounts of data generated each day is a challenge. They predict data volume will increase at a compound annual growth rate of 23% over the next five years. While traditional data storage systems can, in theory, handle large amounts of data, they are struggling to keep up with the high volume demands of big data.



Co-funded by the European Union

Variety

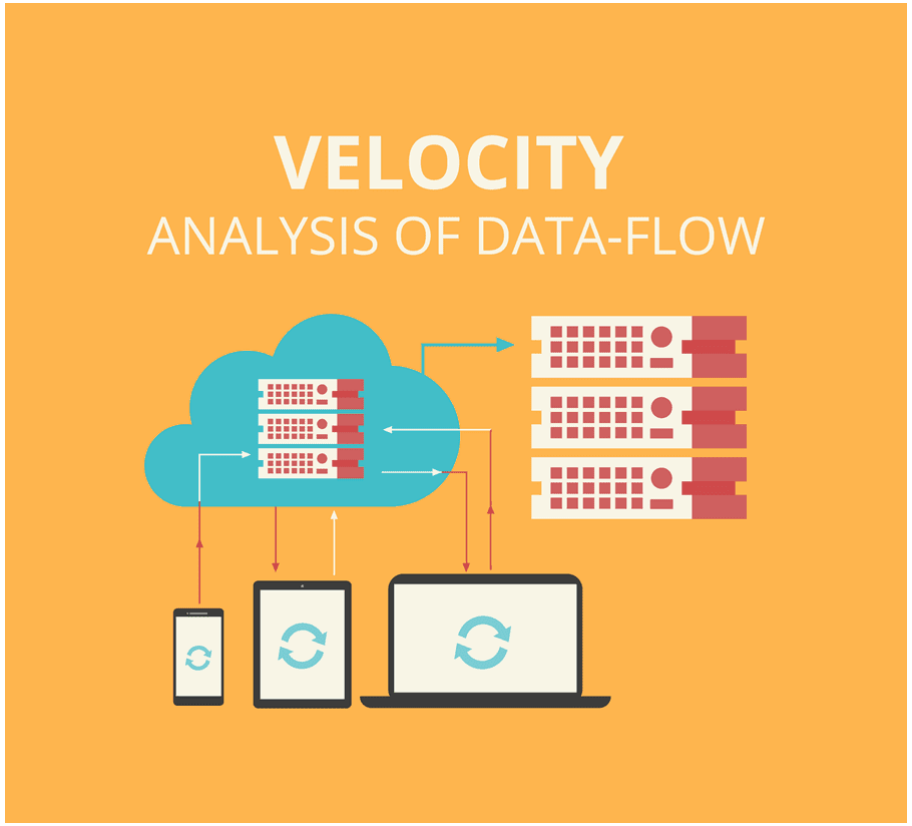
Variety describes the many forms data can take, most of which are rarely in a ready state for processing and analysis. A significant contributor to big data is unstructured data, such as video, images and text documents, which are estimated to represent 80 to 90% of the world's data. These formats are too complex for traditional data warehouse storage architectures. The unstructured data that makes up a significant portion of big data does not fit into the rows and columns of traditional relational data storage system.



Co-funded by the European Union

Velocity

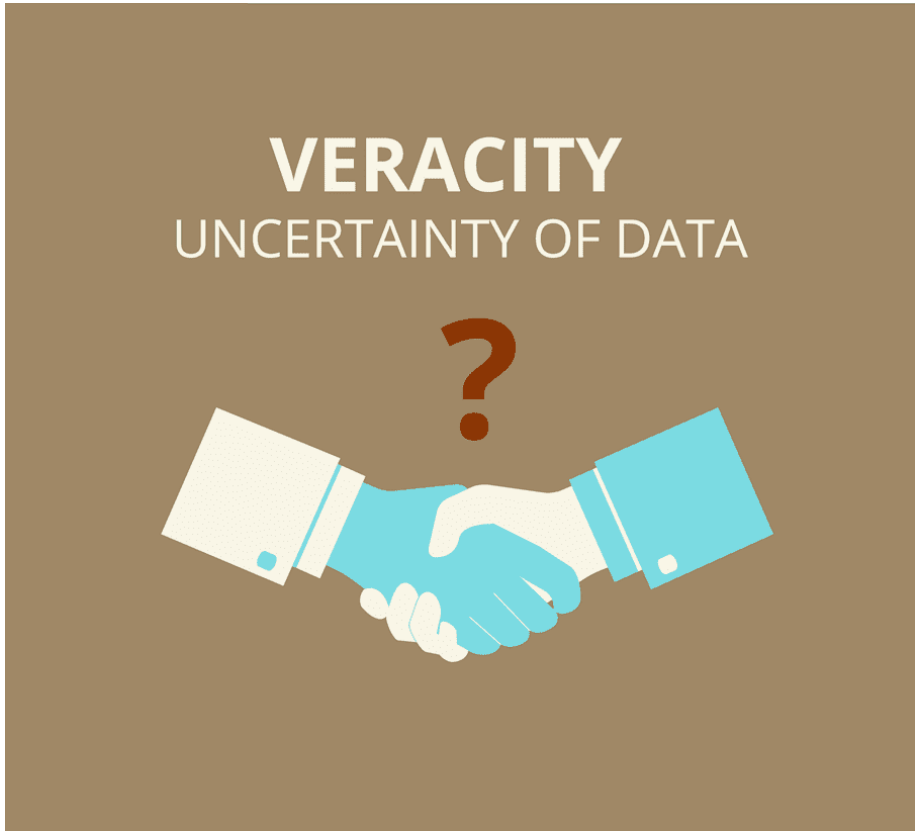
Velocity describes the rate at which this data is generated. For example, New York Stock Exchange generated data by a billion sold shares cannot just be stored for later analysis. It must be analyzed and reported immediately. The data infrastructure must instantly respond to the demands of applications accessing and streaming the data. Big data scales instantaneously, and research often needs to occur in real time.



Co-funded by the European Union

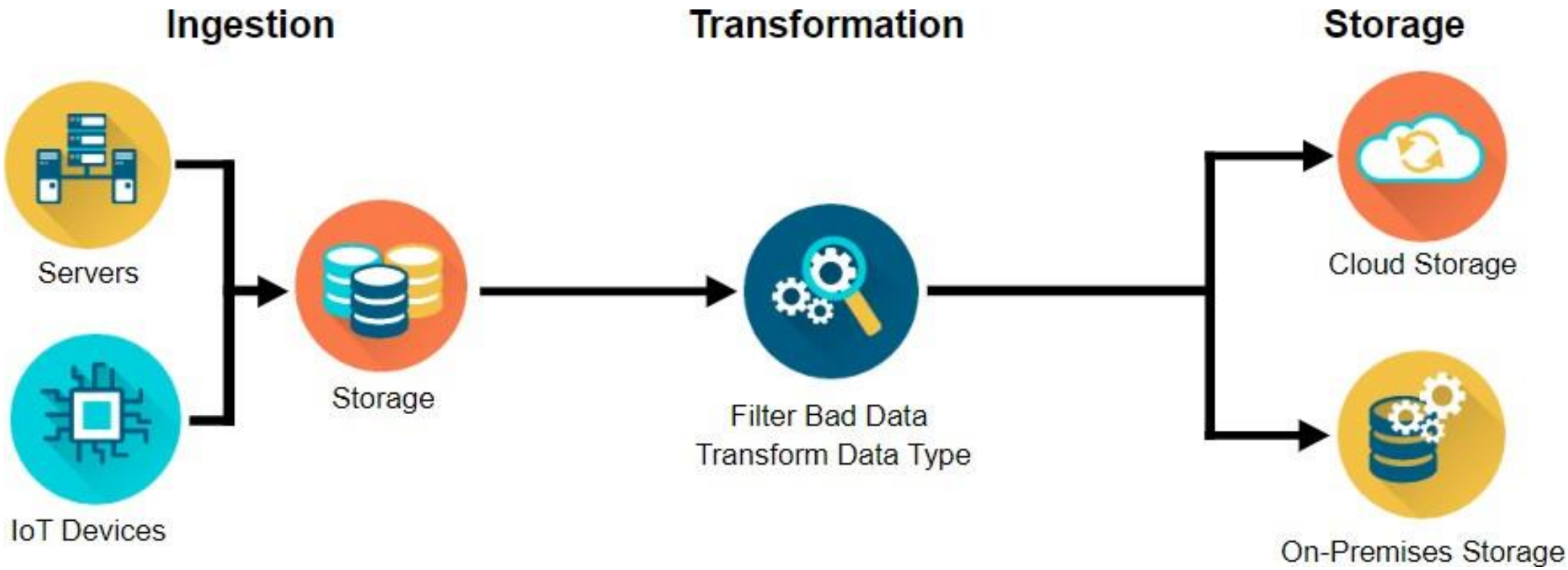
Veracity

Veracity is the process of preventing inaccurate data from spoiling your data sets. For example, when people sign up for an online account, they often use false contact information. Much of this inaccurate information must be “scrubbed” from the data before use in analysis. Increased veracity in the collection of data can reduce the amount of data cleaning that is required.



Co-funded by the European Union

Data Pipelines



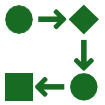
Co-funded by the European Union

The Lifecycle of Data Analysis

1. Data Collection
2. Data Processing/Cleaning
3. Data Analysis
4. Data Visualization
5. Interpretation
6. Decision making
7. Action
8. Feedback and Refinement

Analysis of the Lifecycle:

Cyclical Nature



Iteration and Adaptation

Dependency on Quality

Skill and Interpretation

Integration with Business Processes

Technological Infrastructure

Ethical and Legal Considerations

Co-funded by the European Union

Popular Tools in Data Analysis

- Spreadsheets
- Statistical Software
- Programming Languages
- Database Querying Languages
- Data Visualization Software

Co-funded by the European Union

Excel

What is Excel?

Excel is a powerful tool suitable for small datasets and quick data analysis. With Excel, you can manipulate data, summarize it with pivot tables, visualize it, and perform quick statistics to summarize it.

Why it's important to know:

Excel is powerful and very popular for performing small-scale data analysis, calculations, data summaries, and data visualizations.

Excel skills you will learn in this course:

- Perform data cleaning by removing blank spaces as well as incorrect and outdated information
- Format and adjust data using conditional formatting
- Perform data calculations using formulas
- Organize data using sorting and filtering
- Create visualizations using graphing and charting
- Calculate, summarize, and analyze data using pivot tables
- Aggregate data for analysis

Co-funded by the European Union

SQL

What is SQL?

SQL, which stands for Structured Query Language, is a powerful database management tool that allows data analysts to retrieve and interact with selections of data that are stored in relational databases. Relational databases have a defined structure and contain multiple interrelated data tables that need to be queried with a language like SQL to be useful. SQL is fast and can handle data sets much larger than Excel can. As a data analyst you will use SQL to access, read, manipulate, and analyze the data stored in a relational database to generate useful insights to drive a data-informed decision-making process.

Why it's important to know:

Popular big data systems make use of SQL for maintaining relational databases and processing structured data. It is used for carrying out data analytics with data stored in relational database management systems such as Oracle, Microsoft SQL, and MySQL.

SQL skills you will learn in this course:

- Create tables
- Retrieve data using SQL index
- Retrieve data using SQL queries
- Aggregate data with SQL joins

Co-funded by the European Union

Tableau

What is Tableau?

Tableau is one of the most used data analytics and visualization tools on the market. Visualizations are an important way to present data in a format that can easily be understood by non-technical decision-makers and stakeholders.

Why it's important to know:

- Tableau is a data analytics market leader due to the depth and quality of its data visualizations.
- Tableau can extract and combine data from multiple sources including Excel spreadsheets and SQL databases. It can also access large data storage locations, known as data warehouses, as well as cloud-based data repositories.

Tableau skills you will learn in this course:

- Compare data from multiple views using Tableau dashboards
- Create visualizations using Tableau visualization tools

Co-funded by the European Union

Dataset resources

Co-funded by the European Union

Big Data Technologies

BETES
Discover Technologies, expand opportunities!

Hadoop Ecosystem:

Core Components:

- **Hadoop Distributed File System (HDFS)**
- **MapReduce**
- **YARN (Yet Another Resource Negotiator)**

Advantages:

- **Scalability**
- **Fault Tolerance**
- **Cost-Effective**

hadoop.apache.org

Apache Spark:

Core Features:

- **In-Memory Computing**
- **Lazy Evaluation**
- **MLlib**

Advantages:

- **Speed**
- **Ease of Use**

spark.apache.org

Co-funded by the European Union

Analytical Techniques in Data Analysis

Descriptive Analysis

Diagnostic Analysis

Predictive Analysis

Prescriptive Analysis

Machine Learning

Co-funded by the European Union

Real-World Applications

Applications in Various Fields

Business and Finance:	Market Analysis Financial Forecasting Risk Management
Healthcare:	Patient Data Analysis Epidemiological Studies
Education:	Learning Analytics Educational Policy and Planning
Retail:	Customer Segmentation and Targeting Inventory Management
Technology and Social Media:	User Behavior Analysis Content Recommendation
Environmental Science:	Climate Change Analysis Resource Management

Co-funded by the European Union

How data analysis shapes decision-making I

BETES
Discover Technologies, expand opportunities!

Informed Strategy Development:	• Identify successful strategies and patterns • Set realistic goals
Enhancing Customer Understanding:	• Gain insights into customer preferences, behaviour, and feedback • Effective marketing and increased customer satisfaction
Optimizing Operations:	• Identifying bottlenecks, inefficiencies, and best practices • Optimal use of resources, reducing waste and increasing productivity
Risk Assessment and Mitigation:	• Predict potential risks and take preemptive measures to mitigate them • Scenario analysis and risk modelling for strategic planning and crisis management

Co-funded by the European Union

How data analysis shapes decision-making II

BETES
Discover Technologies, expand opportunities!

Financial Planning:	• Budgeting, forecasting, and investment decisions. • Understand cash flows, manage capital, and ensure financial stability
Policy Making:	• Informs policy decisions by providing evidence, ensuring that policies are grounded in reality • Evaluate the potential impact of policies
Advancing Healthcare:	• Leads to better diagnosis, treatment plans, and patient outcomes • Informs public health decisions and the management of healthcare systems
Driving Innovation:	• Identify opportunities for innovation • Development of new products, services, and business models
Tailoring Education:	• Tailor learning experiences to individual students' needs • Improve overall educational strategies

Co-funded by the European Union

Implications of Data Analysis

1. **Business Strategy and Competition:**
    - Strategic Insight
    - Innovation
2. **Economic Impact:**
    - Job Market Evolution
    - Resource Allocation
3. **Social and Cultural Effects:**
    - Behavioral Insights
    - Personalization
4. **Ethical and Legal Considerations:**
    - Privacy Rights
    - Data Ownership
6. **Data Governance:**
    - Security Measures
    - Transparency and Accountability
7. **Global Dynamics:**
    - International Relations
    - Access and Equity
8. **Technological Advancements:**
    - AI and Machine Learning
    - Quantum Computing

Co-funded by the European Union

Data Analysis Careers

Ethics, Skills, Job opportunities

BETES
Discover Technologies, expand opportunities!

Digital Idea
Scientific Association

Co-funded by the European Union

Ethical Considerations

BETES
Discover Technologies, expand opportunities!

PRIVACY

CONSENT

BIAS AND FAIRNESS

TRANSPARENCY

DATA SECURITY

Co-funded by the European Union

Ethical and Privacy Concerns

Data Privacy

Ethical Use

Co-funded by the European Union

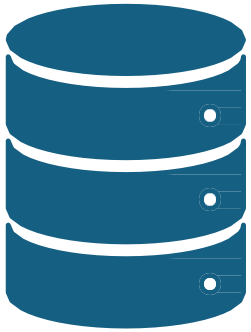
Security Vulnerabilities

Data Breaches

Cybersecurity Threats

Co-funded by the European Union

Analytical Bias



Bias in Data

Algorithmic Bias

Co-funded by the European Union

Overcoming Challenges

BETES
Discover Technologies, expand opportunities!

Investing in Quality Data Management

Emphasizing Continuous Learning

Adopting Advanced Technologies

Fostering an Ethical Culture

Strengthening Security Measures

Co-funded by the European Union

Key Ethical Principles in Data Analysis

- **Respect for Privacy:** Emphasize the right to privacy and confidentiality of data subjects. Mention relevant laws like GDPR.
- **Accuracy:** Stress the importance of precision and correctness in data analysis to avoid misinformation.
- **Transparency:** Discuss the need for transparent methodologies and the clear communication of data sources, methods, and limitations.
- **Informed Consent:** Address the necessity of obtaining consent from individuals when collecting and using their data, especially for sensitive information.

Co-funded by the European Union

Data Analysis in a Career Context

BETES
Discover Technologies, expand opportunities!

Role Definition:

Core Responsibilities

- Data Collection
- Data Cleaning
- Data Interpretation
- Data Presentation

Skill Set

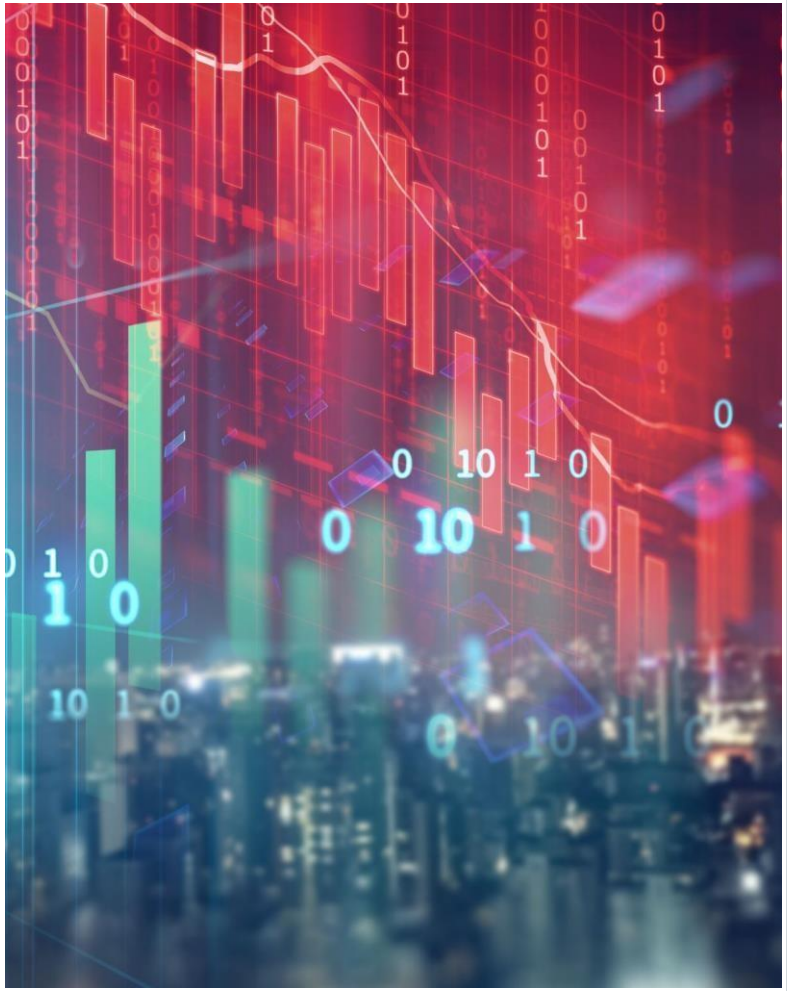
Business Acumen

Ethical Considerations

Career Pathways

Job Market Impact

Continual Learning

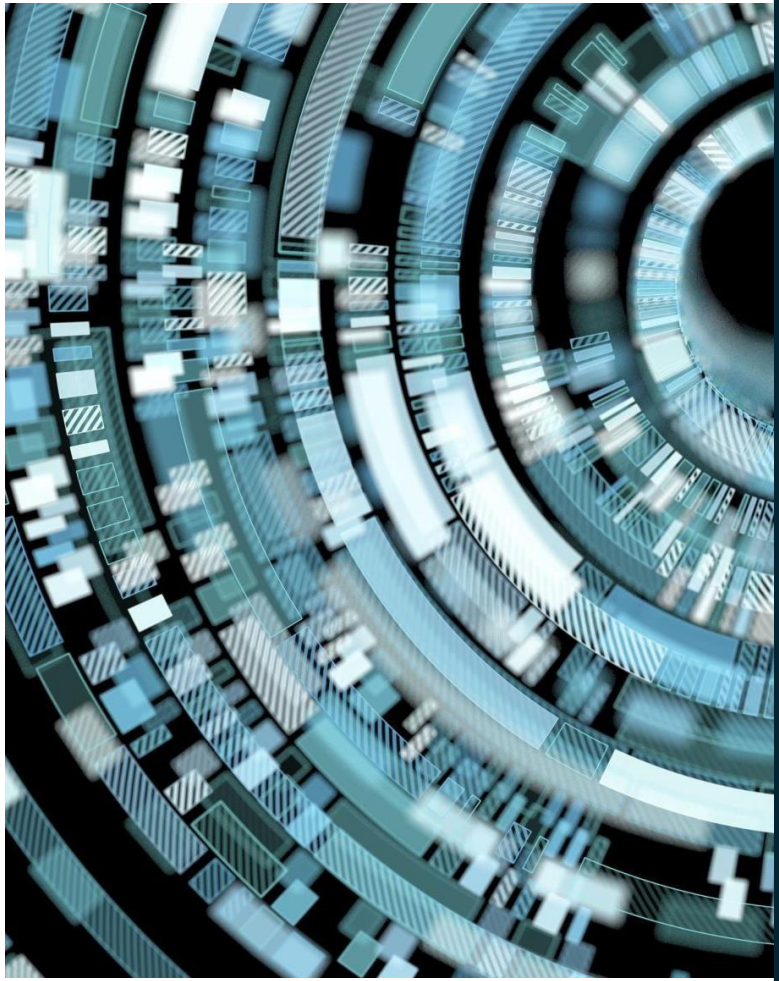


Co-funded by the European Union

Essential Skills for Data Analysts

BETES
Discover Technologies, expand opportunities!

1. **Technical Proficiency:**
   - Programming Languages
   - Analytical Tools
2. **Statistical Knowledge and Machine Learning:**
   - Statistical Methods
   - Machine Learning
3. **Data Visualization**
4. **Soft Skills:**
   - Communication
   - Critical Thinking
   - Collaboration
5. **Business Acumen**



Co-funded by the European Union

Soft Skills and Business Acumen

Communication Skills:	**Clarity and Precision** **Storytelling with Data**
Critical Thinking and Problem-Solving:	**Analytical Mindset** **Problem-Solving**
Collaboration and Teamwork:	**Cross-functional Teams** **Communication Channels**
Business Acumen:	**Industry Knowledge** **Strategic Thinking** **Decision-making**
Adaptability and Continuous Learning:	**Evolving Technologies** **Adaptability**

Co-funded by the European Union

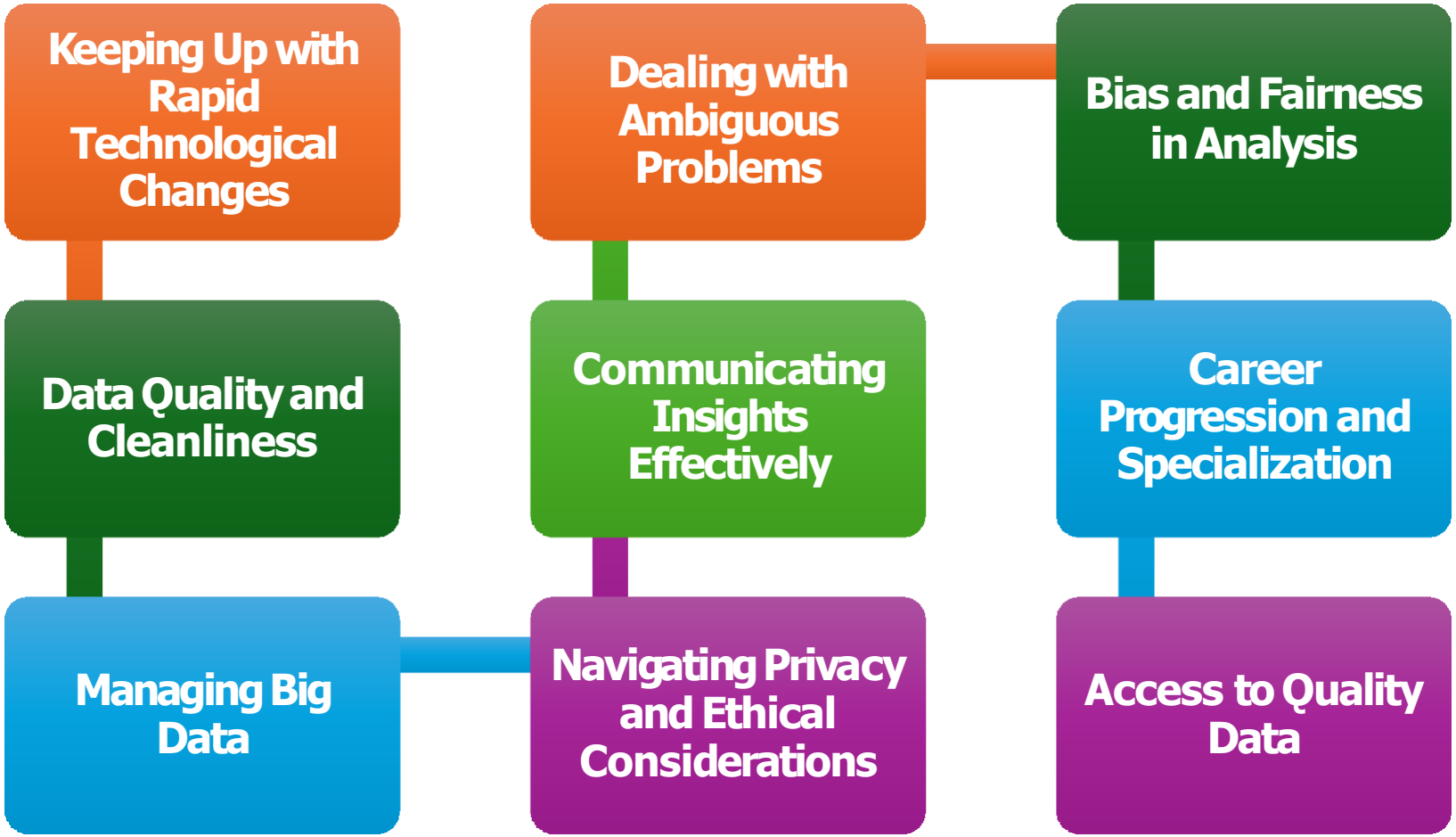
Building a Career in Data Analysis

BETES
Discover Technologies, expand opportunities!

1. **Acquire Fundamental Skills:** Statistical Knowledge, Programming Skills, Data Visualization
2. **Formal Education and Training:** Degree Programs, Online Courses and Certifications
3. **Practical Experience:** Projects, Internships, Freelancing
4. **Networking and Professional Development:** Attend Workshops and Conferences, Professional Organizations, Build an Online Presence
5. **Specialize:** Choose a Specialization, Advanced Certifications
6. **Continuous Learning:** Stay Informed, Adapt to New Technologies
7. **Career Advancement:** Seek Mentoring, Explore Opportunities

Co-funded by the European Union

Challenges



Co-funded by the European Union

Job Opportunities in Data Analysis

- **Current Job Market Trends:**
  - High Demand
  - Salary Prospects
- **Industries Hiring Data Analysts:**
  - Technology and IT
  - Finance and Banking
  - Healthcare
  - Retail and E-commerce
  - Public Sector
- **Roles and Titles in Data Analysis:**
  - Data Analyst
  - Business Intelligence Analyst
  - Data Scientist
  - Data Engineer
- **Emerging Opportunities:**
  - Machine Learning Engineer
  - Data Visualization Specialist
  - Data Governance and Privacy Analyst

Co-funded by the European Union

Advanced-Data Science

Machine learning, Artificial Intelligence

BETES
Discover Technologies, expand opportunities!

Digital Idea
Scientific Association

Funded by the Erasmus+ Programme of the European Union

Machine Learning (ML), Artificial Intelligence (AI), and Data Science

BETES
Discover Technologies, expand opportunities!

Artificial Intelligence (AI):

This is the broadest concept among the three. AI refers to the development of computer systems that can perform tasks typically requiring human intelligence. These tasks include learning, reasoning, problem-solving, perception, and language understanding. AI aims to create machines that can mimic human behavior and thought processes.

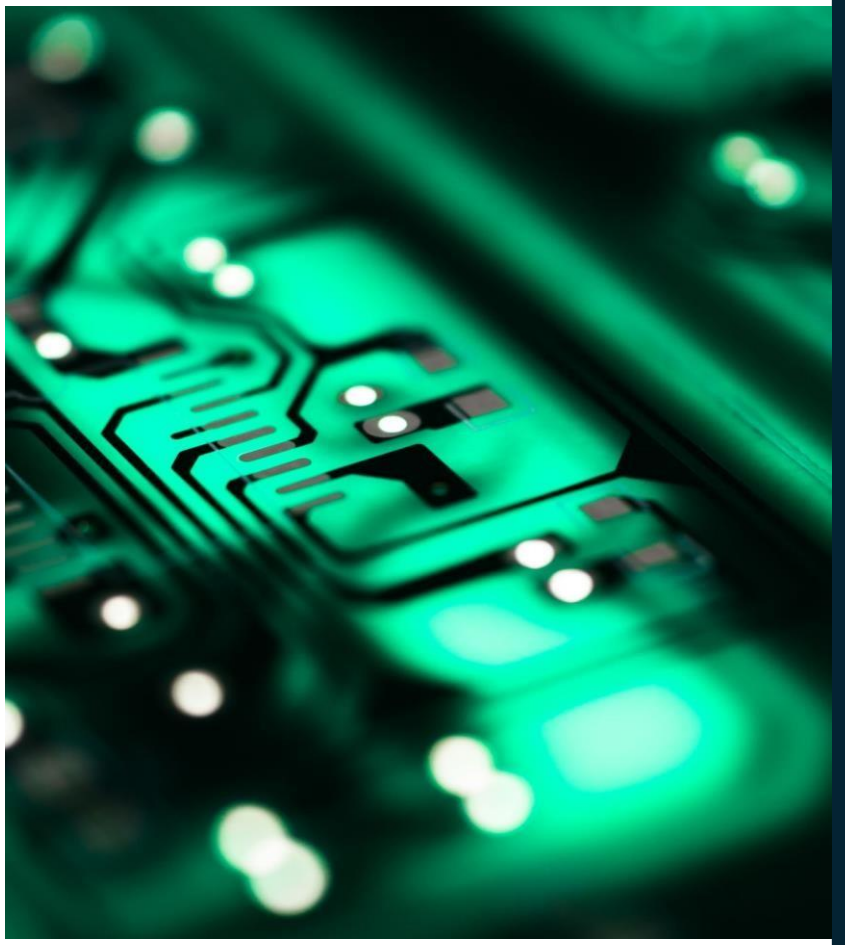
Co-funded by the European Union

Machine Learning (ML), Artificial Intelligence (AI), and Data Science

BETES
Discover Technologies, expand opportunities!

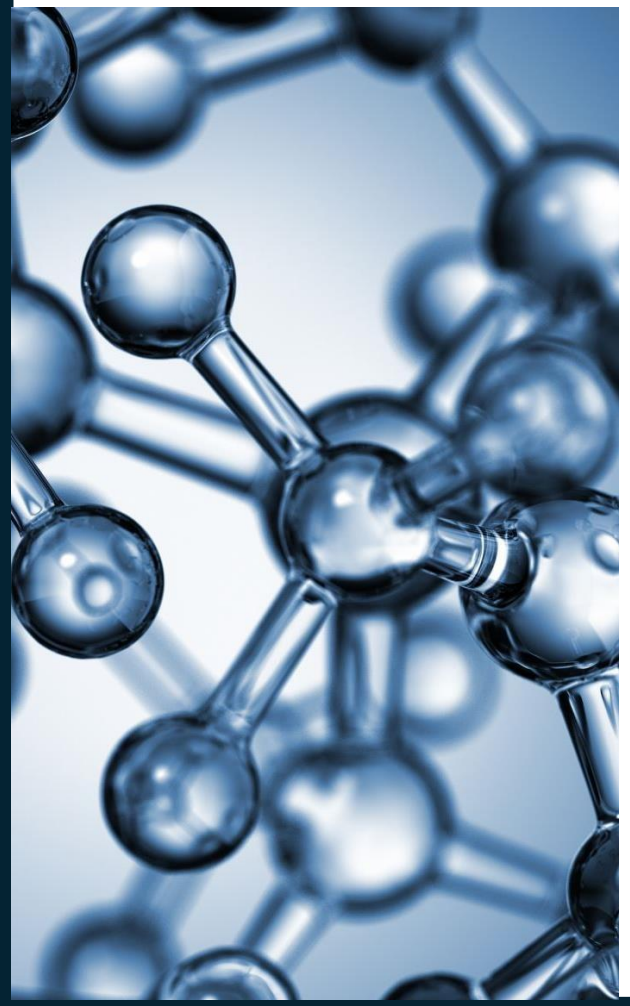
Machine Learning (ML):

Machine Learning is a subset of AI. It involves the use of statistical methods to enable machines to improve at tasks with experience. Essentially, ML is about designing and training algorithms that can learn from and make predictions or decisions based on data. Machine learning algorithms automatically build a mathematical model based on sample data (known as "training data") to make predictions or decisions without being explicitly programmed to perform the task.



Co-funded by the European Union

Machine Learning (ML), Artificial Intelligence (AI), and Data Science



Data Science:

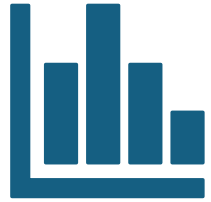
Data Science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. It involves a blend of various tools, algorithms, and machine learning principles to discover hidden patterns in raw data. Data Science is not exclusively about AI or ML but uses various techniques from both fields, along with others from statistics, data analysis, and computer science, to analyze and interpret complex data.

Co-funded by the European Union

Relationship Among AI, ML, and Data Science

Machine Learning as the Goal

Machine Learning as the Means



Data Science as the Foundation

Co-funded by the European Union

The Machine Learning Process

Step 1. **Data preparation**

Step 2a. **Learning data**

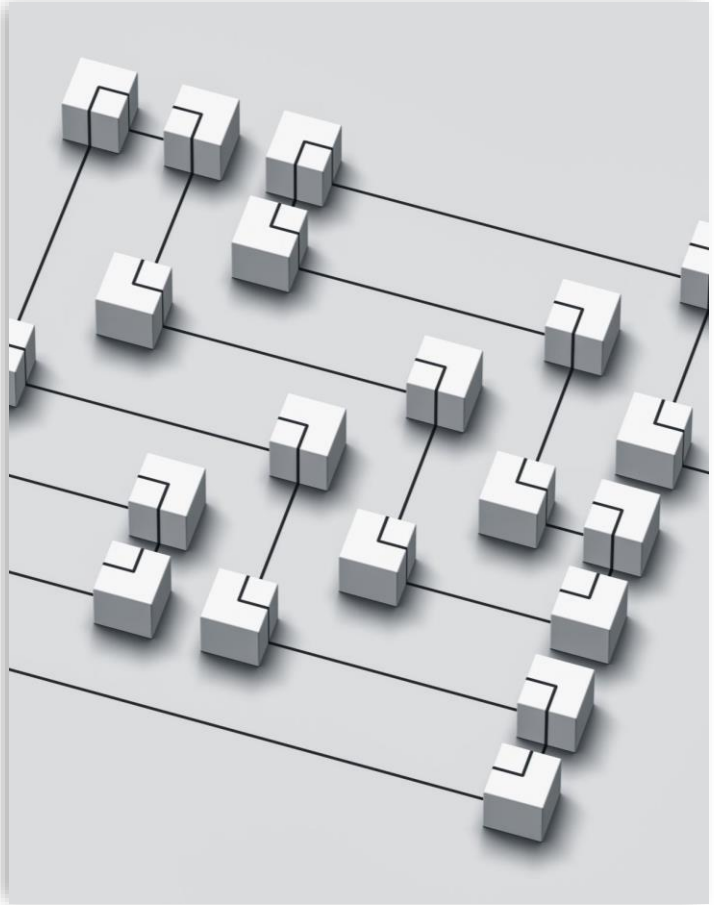
Step 2b. **Testing data**

Step 3. **Learning Process Loop - Selection**

Step 4. **Learning Process Loop - Evaluation**

Step 5. **Model evaluation**

Step 6. **Model implementation**



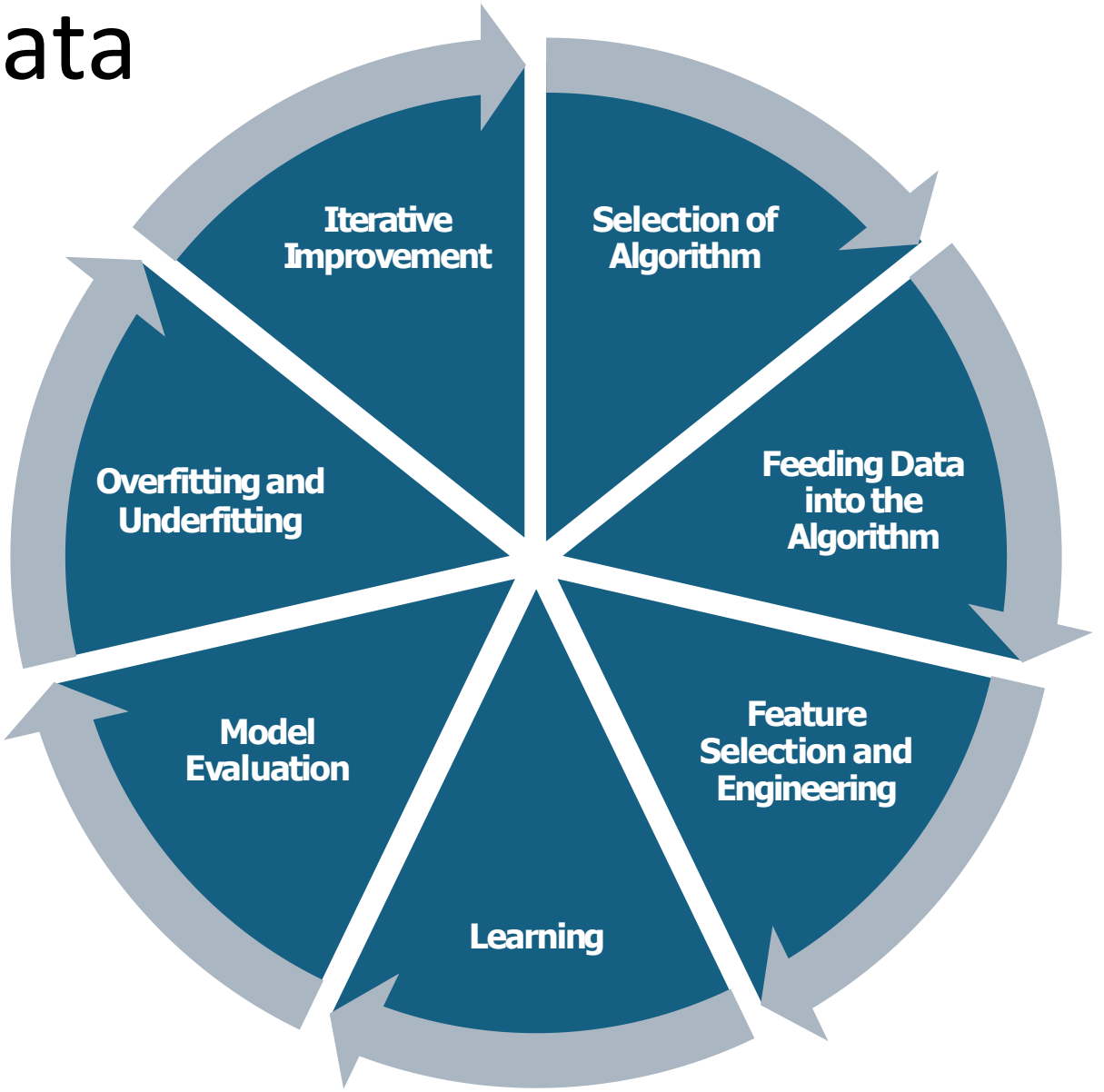
Co-funded by the European Union

Data preparation

1. **Data Collection:** Gathering Data, Data Aggregation
2. **Data Cleaning:** Handling Missing Values, Noise Reduction, Outlier Detection
3. **Data Transformation:** Normalization/Standardization, Encoding Categorical Variables, Feature Engineering
4. **Data Reduction:** Dimensionality Reduction, Feature Selection
5. **Data Splitting:** Training and Testing Split, Validation Set Creation

Co-funded by the European Union

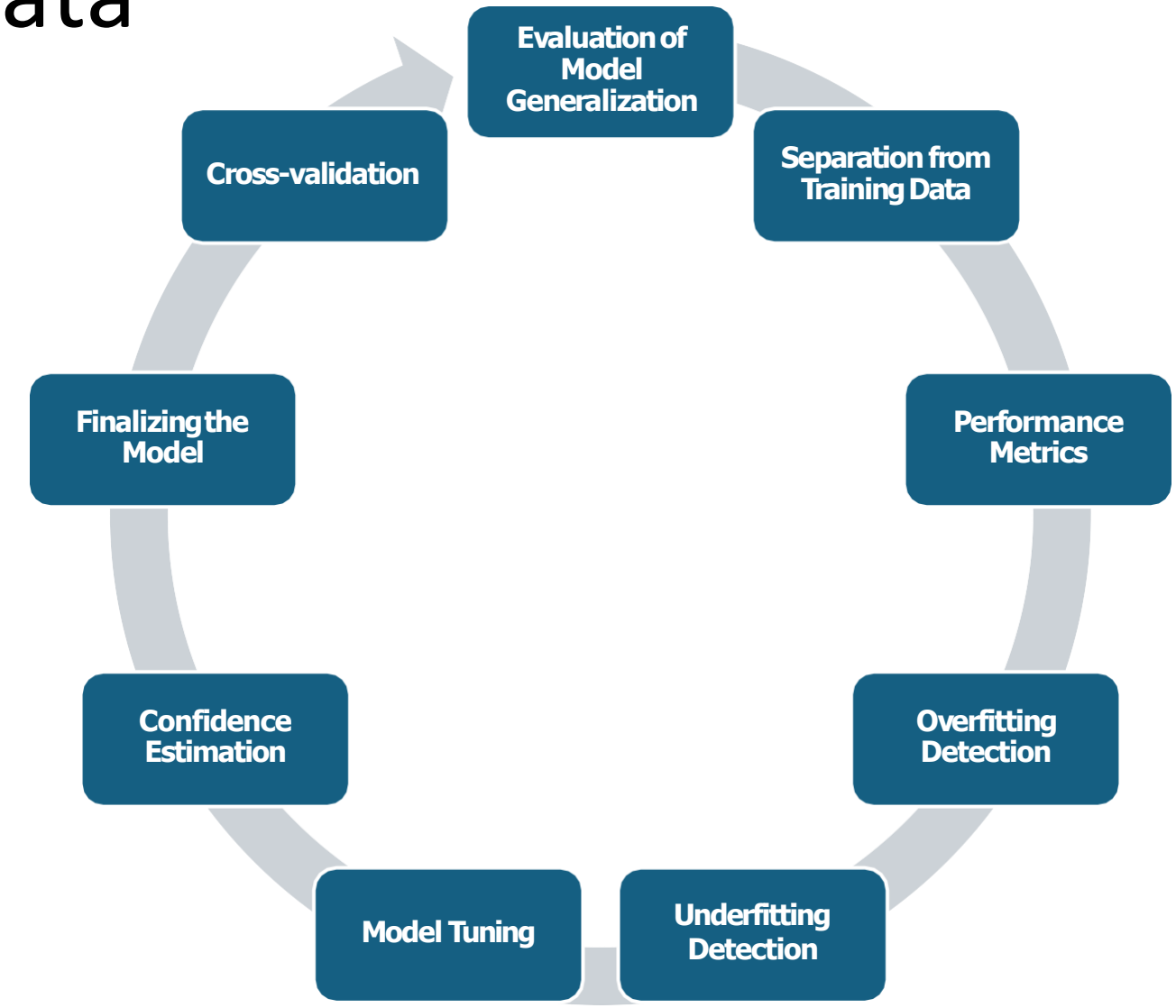
Learning data



BETES
Discover Technologies, expand opportunities!

Co-funded by the European Union

Testing data



Learning Process Loop - Selection

Algorithm Selection

Model Hypothesis

Model Experimentation

Performance Evaluation

Hyperparameter Tuning

Model Comparisons

Feedback Loop

Co-funded by the European Union

Learning Process Loop - Evaluation

BETES
Discover Technologies, expand opportunities!

Performance Metrics

Testing Dataset

Error Analysis

Validation Techniques

Comparison with Baseline

Interpretability

Model Diagnostics

Feedback Integration

Co-funded by the
European Union

Model evaluation

BETES
Discover Technologies, expand opportunities!

Assessment with Test Data

Performance Metrics

Cross-Validation

Analyzing Model Errors

Comparative Evaluation

Understanding Model Behavior

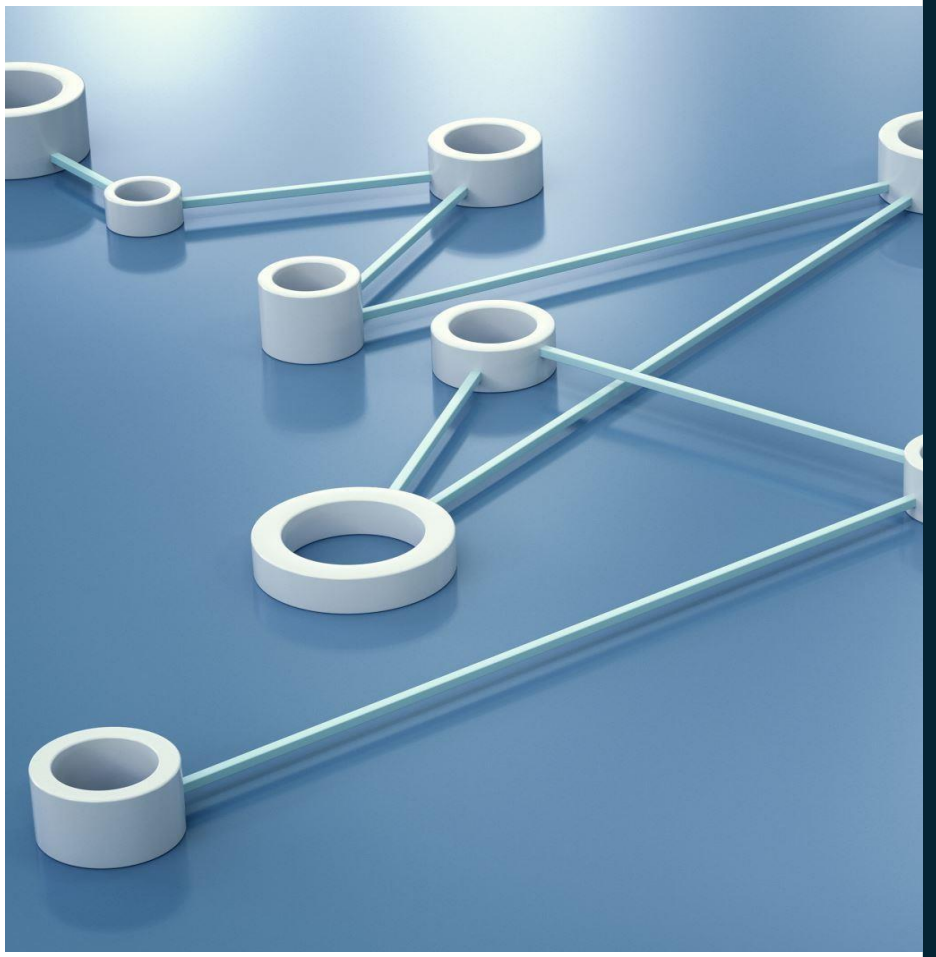
Tuning and Adjustments

Iterative Process

Co-funded by the European Union

Model implementation

- **Integration**
- **Automation**
- **Monitoring**
- **Performance Checks**
- **Maintenance**
- **Feedback Loops**
- **Scalability**
- **Security and Privacy**
- **User Interface**
- **Documentation and Training**



Co-funded by the European Union

Bootcamps for Emerging Technologies and Essential Skills

Thank you!

Developed by

Digital Idea
Scientific Association

For more information:

www.betesproject.eu

www.facebook.com/BETESproject

ecect.projects@gmail.com

☎+ 357 96520112 (Cyprus)

Co-funded by the European Union