

# Unit 1

## Introduction

By- Bhagyashree Udagave(Chougule)

*Bhagyashree*

# Overview

Data analytics is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision making.

Data analytics is a multidisciplinary field that employs a wide range of analysis techniques, including math, statistics, and computer science, to draw insights from data sets.

Data analytics is a broad term that includes everything from simply analyzing data to theorizing ways of collecting data and creating the frameworks needed to store it.

Many industry giants tend to collect data from their audiences. This data is collected in raw form, which often is not in a format that can benefit the business.

A professional data analysis team gives meaning to the data collected.

The term data analytics refers to collecting and transforming raw data into valuable insights that help make actionable business decisions.

It is the entire process that starts from extracting, organizing, analyzing, and ends with transforming the data from numbers to coherent information.

When a data analyst performs the whole process, they then give suggestions to the company regarding what it should do next.

*Bhagyashree*

# Data Analytics in Different Industries

## **Retail and E-commerce**

- Customer Behavior Analysis
- Inventory Management
- Personalized Marketing

## **Healthcare**

- Patient Data Analysis
- Disease Prediction
- Drug Development

## **Marketing**

- Campaign Analysis
- Customer Segmentation
- Social Media Analytics

## **Finance**

- Risk Assessment
- Fraud Detection
- Portfolio Management

## **Manufacturing**

- Quality Control
- Predictive Maintenance
- Supply Chain Optimization

# Data Analytics Applications

Data analytics finds applications across various industries and sectors, transforming the way organizations operate and make decisions.

## Healthcare

Data analytics is revolutionizing the healthcare industry by enabling better patient care, disease prevention, and resource optimization. For example, hospitals can analyze patient data to identify high-risk individuals and provide personalized treatment plans. Data analytics can also help detect disease outbreaks, monitor the effectiveness of treatments, and improve healthcare operations.

# 6 REAL-WORLD APPLICATIONS OF HEALTHCARE DATA ANALYTICS



Sources: Built In, HealthITAnalytics

Pharmasense

## **Finance**

In the financial sector, data analytics plays a crucial role in fraud detection, risk assessment, and investment strategies. Banks and financial institutions analyze large volumes of data to identify suspicious transactions, predict creditworthiness, and optimize investment portfolios. Data analytics also enables personalized financial advice and the development of creative financial products and services.

## **E-commerce**

E-commerce platforms utilize data analytics to understand customer behavior, personalized shopping experiences, and optimize marketing campaigns. By analyzing customer preferences, purchase history, and browsing patterns, e-commerce companies can offer personalized product recommendations, target specific customer segments, and improve customer satisfaction and retention.







### Product Recommendations

- Analyze past behavior and customer profile to recommend the right products, combinations, and bundles.



### Targeted Marketing

- Focus on a group of key customers and leads, and target them at an appropriate moment. Those customers can spread your word.



### Layout Personalization

- Every customer journey can be unique. Tailor the design and onsite experience based on pages that are likely to lead to a conversion.



### Server Load Balance Optimization

- Achieve a business-oriented admission control on a Web server, guaranteeing a higher quality of service to visitors with a higher probability of making a purchase.



### Stock Management

- Forecast and prepare your stock based on your customer's search patterns online.



### Real-Time Customer Service

- Talk to users not registered and temporary visitors, who can be attracted by recommending them the discounted item if their preference is captured in the short-term.



### Purchase Trends Discovery

- Customer journeys online and search patterns can indicate trends in future high demanding products.



### Offers Awareness

- Understand your customer intention online, as a browser, searcher or purchaser, and make them aware of brands and deals matching their profile.

## **Cybersecurity**

Data analytics plays a vital role in cybersecurity by detecting and preventing cyber threats and attacks. Security systems analyze network traffic, user behavior, and system logs to identify anomalies and potential security breaches. By leveraging data analytics, organizations can proactively strengthen their security measures, detect and respond to threats in real-time, and safeguard sensitive information.

## **Supply Chain Management**

Data analytics improves supply chain management by optimizing inventory levels, reducing costs, and enhancing overall operational efficiency. Organizations can identify bottlenecks, forecast demand, and improve logistics and distribution processes by analyzing supply chain data. Data analytics also enables better supplier management and enhances transparency throughout the supply chain.



*Bhagyashree*

## **Banking**

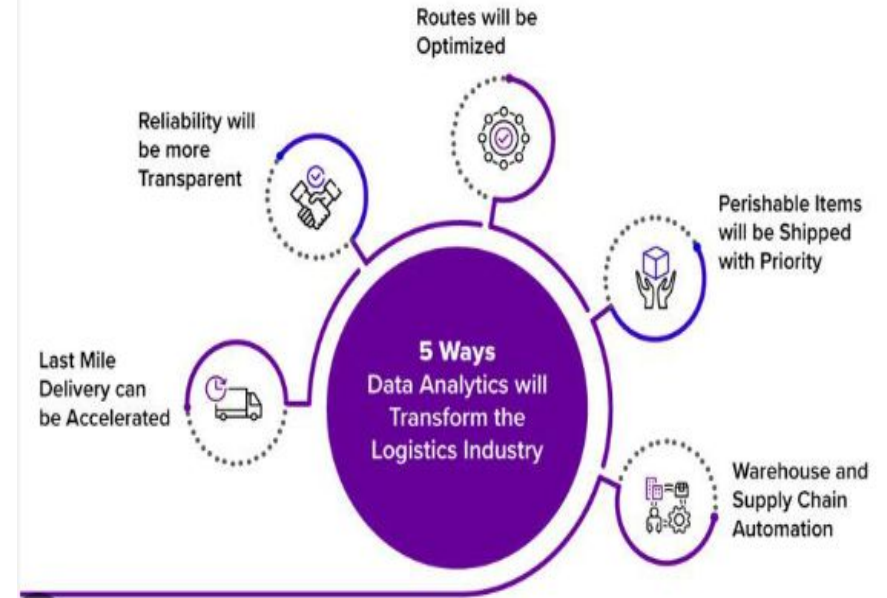
Banks use data analytics to gain insights into customer behavior, manage risks, and personalize financial services. Banks can tailor their offerings, identify potential fraud, and assess creditworthiness by analyzing transaction data, customer demographics, and credit histories. Data analytics also helps banks detect money laundering activities and improve regulatory compliance.

## **Logistics**

In the logistics industry, data analytics plays a crucial role in optimizing transportation routes, managing fleet operations(The processes that go into transporting items, using resources such as vehicles, drivers, fuel, spare parts, and so on), and improving overall supply chain efficiency. Logistics companies can minimize costs, reduce delivery times, and enhance customer satisfaction by analyzing data on routes, delivery times, and vehicle performance. Data analytics also enables better demand forecasting and inventory management.

***Bhagyashree***

## HOW ARE BANKS USING DATA ANALYTICS?

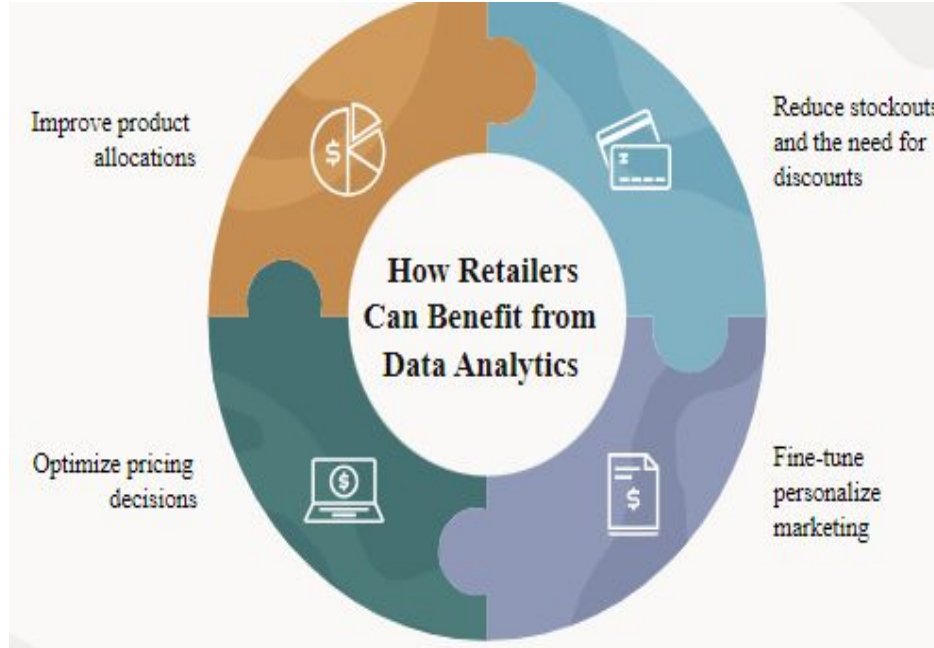


## **Retail**

Data analytics transforms the retail industry by providing insights into customer preferences, optimizing pricing strategies, and improving inventory management. Retailers analyze sales data, customer feedback, and market trends to identify popular products, personalize offers, and forecast demand. Data analytics also helps retailers enhance their marketing efforts, improve customer loyalty, and optimize store layouts.

## **Manufacturing**

Data analytics is revolutionizing the manufacturing sector by enabling predictive maintenance, optimizing production processes, and improving product quality. Manufacturers can predict equipment failures, minimize downtime, and ensure efficient operations by analyzing sensor data, machine performance, and historical maintenance records. Data analytics also enables real-time monitoring of production lines, leading to higher productivity and cost savings.



## **Internet Searching**

Data analytics powers internet search engines, enabling users to find relevant information quickly and accurately. Search engines analyze vast amounts of data, including web pages, user queries, and click-through rates, to deliver the most relevant search results. Data analytics algorithms continuously learn and adapt to user behavior, providing increasingly accurate and personalized search results.

## **Risk Management**

Data analytics plays a crucial role in risk management across various industries, including insurance, finance, and project management. Organizations can assess risks, develop mitigation strategies, and make informed decisions by analyzing historical data, market trends, and external factors. Data analytics helps organizations identify potential risks, quantify their impact, and implement risk mitigation measures.

***Bhagyashree***



## Google Analytics benefits at a glance



Fast web insights and access to custom reports



Tools for reporting and predictive analysis



Customizable dimensions and metrics



Integrates with other tools, platforms



Free, user-friendly

## Applications of data analytics in risk management



Risk assessment and prioritization



Risk modeling and mitigation



Risk identification



Risk monitoring and reporting



*Bhagyashree*

# Need of data analysis process

There is indeed a specific process for data analysis. **Example:** Suppose you are looking to create the best recipe for pizza dough. You could frame your problem as a lack of knowledge—not having a sufficient pizza dough recipe.

What data could help you solve this problem? One way would be to comb through the plethora of online recipes available.

You could then sort this data, filtering out recipes with low reviews or comments noting flaws in the recipe.

Then, once you've collated the best recipes, you can begin to analyze them.

What are the commonalities that emerge? Maybe you find that the best recipe depends on the style of pizza you want to make and that it's best to group certain recipes together.

**The data analysis process won't create the perfect pizza dough recipe for you, but it can get you headed in the right direction.**

*Bhagyashree*

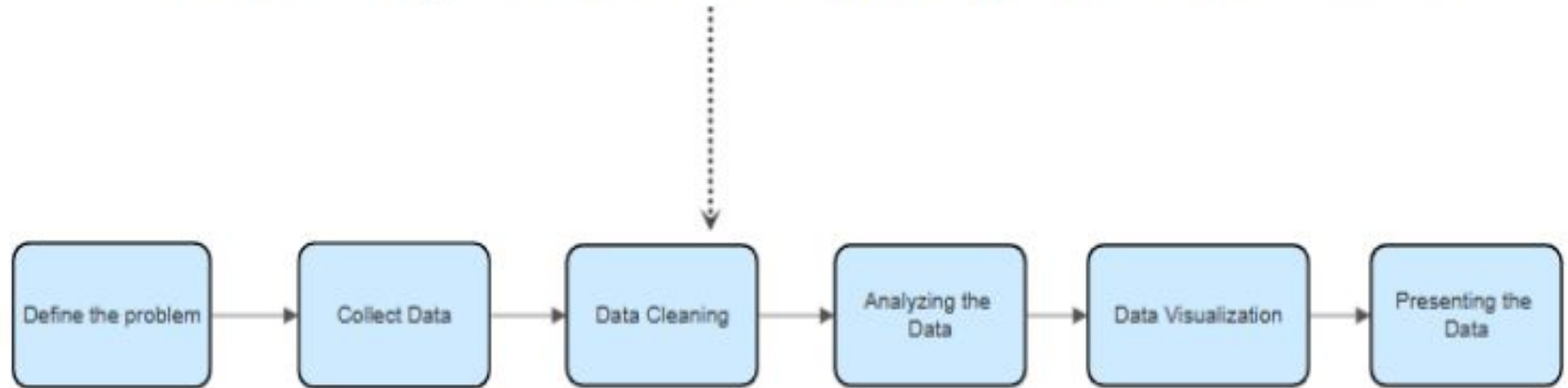
# Data Analytics Process

The process of:

1. Identifying a business problem
2. Finding data that will help solve this problem
3. Cleaning the data
4. Analyzing the data
5. Presenting the data

*Bhagyashree*

# Six Steps of Data Analysis Process



There is indeed a specific process for data analysis. Suppose you are looking to create the best recipe for pizza dough. You could frame your problem as a lack of knowledge—not having a sufficient pizza dough recipe.

What data could help you solve this problem? One way would be to comb through the plethora of online recipes available. You could then sort this data, filtering out recipes with low reviews or comments noting flaws in the recipe. Then, once you've collated the best recipes, you can begin to analyze them. What are the commonalities that emerge? Maybe you find that the best recipe depends on the style of pizza you want to make and that it's best to group certain recipes together. The data analysis process won't create the perfect pizza dough recipe for you, but it can get you headed in the right direction.

*Bhagyashree*

**Let's take a more in-depth look into the data science process:**

## **1. Define the Problem**

In the first step of process the data analyst is given a problem/business task. The analyst has to understand the task and the stakeholders expectations for the solution. (A stakeholder is a person that has invested their money and resources to a project.)

The analyst must be able to ask different questions in order to find the right solution to their problem.

The analyst has to find the root cause of the problem in order to fully understand the problem.

Communicate effectively with the stakeholders and other colleagues to completely understand what the underlying problem is. Questions to ask yourself for the Ask phase are:

- What are the problems that are being mentioned by my stakeholders?
- What are their expectations for the solutions?

## 2. Collect Data

This step includes collecting data and storing it for further analysis. The analyst has to collect the data based on the task given from multiple sources.

The data has to be collected from various sources, internal or external sources. Internal data is the data available in the organization that you work for while external data is the data available in sources other than your organization.

Broadly speaking, there are three different categories of data,

**First-party data** is data that your own organization generates. Often times, this is data about previous customer interactions that can be used to make accurate predictions about your customers' behavior in the future.

You could also use **second-party data**—data that's generated by external sources, but is about your company specifically. This can include what customers are saying on social media platforms or review websites.

**Third-party data** comes from groups like think tanks ( research institutes that seek to play a key role in making and influencing global, regional and national policy.)and government sources and is more concerned with the nature of your customer base, rather than a specific interaction that a customer has had with your company.

# Data Cleaning

Not all the data you collect will be useful or accurate, and you'll need to discard the data points that are irrelevant, duplicated, inconsistent, or outdated.

This is called data cleaning. When combining multiple sources of data, you'll likely wind up with duplicates and outliers. And when you're dealing with millions of data points, as is often the case with data analysis, you can't comb through each piece of data on your own to find the duplicates or outliers. Data analysts estimate that the time spent cleaning data consumes about 70-90% of the data analysis process.

At this stage, you can also do an exploratory analysis, which is an initial and cursory data analysis. Exploratory data analysis will also assist with identifying other data points you may need.

## Data Processing

Once you have all the relevant data, you can begin to process it. This entails organizing the data, sorting the data into relevant categories, and labeling them for easy organization. Now the data is prepared for analysis.



## Analyzing the Data

The fourth step is to **Analyze**.

The cleaned data is used for analyzing and identifying trends.

It also performs calculations and combines data for better results.

The tools used for performing calculations are Excel or SQL. These tools provide in-built functions to perform calculations or sample code is written in SQL to perform calculations.

Using Excel, we can create pivot tables and perform calculations while SQL creates temporary tables to perform calculations.

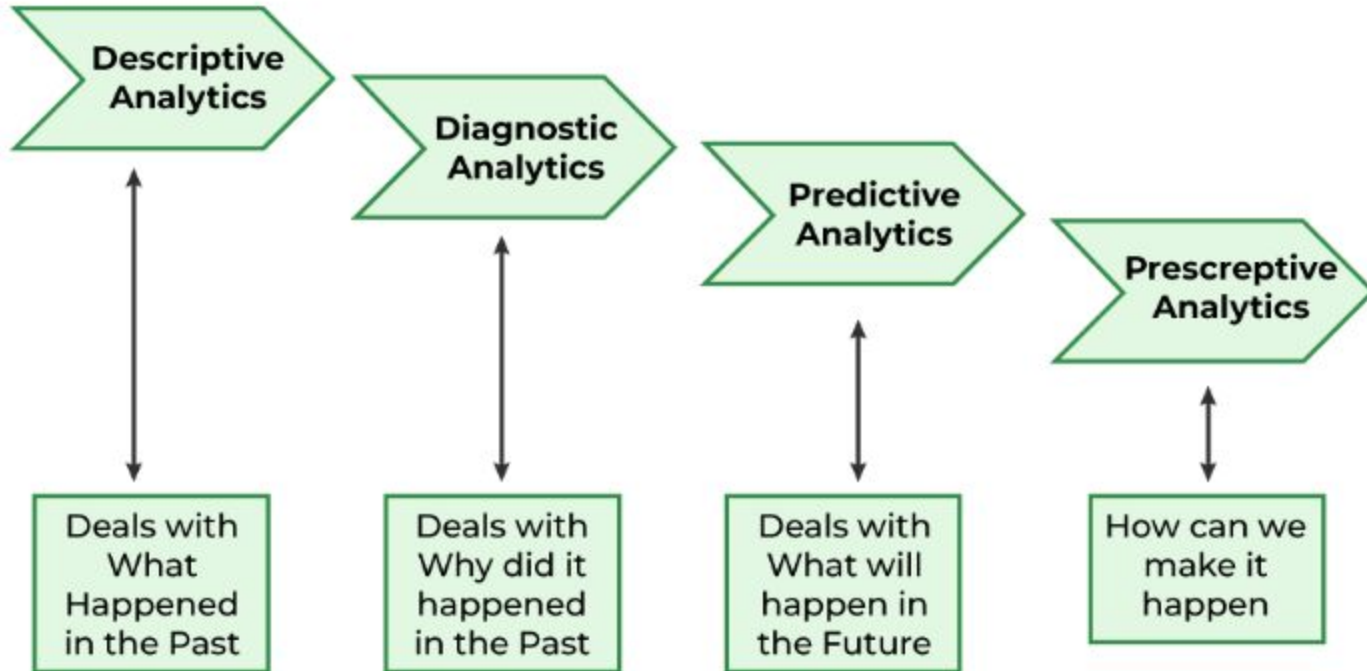
Programming languages are another way of solving problems. They make it much easier to solve problems by providing packages. The most widely used programming languages for data analysis are R and Python.

## Data Visualization

The fifth step is visualizing the data. Nothing is more compelling than a visualization. The data now transformed has to be made into a visual (chart, graph). The reason for making data visualizations is that there might be people, mostly stakeholders that are non-technical. Visualizations are made for a simple understanding of complex data. Tableau and Looker are the two popular tools used for compelling data visualizations.

## Types of Data Analysis

WHAT?	USE IT WHEN YOU WANT TO...
<b>Descriptive analysis</b>	Find out the "what"
<b>Diagnostic analysis</b>	Find out the "why"
<b>Predictive analysis</b>	Gauge what will happen next
<b>Prescriptive analysis</b>	Find out what to do next
<b>Inferential analysis</b>	Derive conclusions from all the above



## **Descriptive Analysis**

As the name suggests, descriptive analysis describes or summarizes the data and its characteristics. It doesn't go beyond explaining what has happened. You use this type of data analysis to deliver a narrative of what has occurred. Descriptive statistics and analysis present scattered data into digestible pointers. You can also do a part of this at the stage of exploratory data analysis.

## **Diagnostic Analysis**

With diagnostic analysis, you begin to focus on the “why,” and diagnose why something is occurring. At this stage, you are not looking for solutions or predictions. The goal is to understand the factors that are contributing to the problem. You use this technique when you want to go into issue identification mode.

## **Predictive Analysis**

Here's where you start generating forecasts based on your data. Data analysts perform predictive analysis when they want to establish a situation in the future. This prediction helps stakeholders gauge business performance.

## **Prescriptive Analysis**

This kind of analysis brings together all of these data analysis techniques to offer recommendations. These form the basis of data-driven decisions.

Prescriptive Analytics not only anticipates what will happen and when to happen but also why it will happen. Further, Prescriptive Analytics can suggest decision options on how to take advantage of a future opportunity or mitigate a future risk and illustrate the implication of each decision option.

## **Inferential Analysis**

With this technique, you derive conclusions based on the data you have collected and analyzed, such as, “lack of employee training is a cause of employee attrition” or “employee attrition affects customer satisfaction.”

BI



*Bhagyashree*





*Bhagyashree*




*Bhagyashree*

## Defining BI

**Business Intelligence refers to the process of gathering, analyzing, and interpreting data to make better business decisions. BI can be used in many different areas of a business, such as sales, marketing, finance, and operations.**

## Applications of BI

**With the help of business intelligence, companies can identify new opportunities, reduce costs, and improve the customer experience.**



## How Business Intelligence Works?

01

### Data Collection

Collect data from customer feedback, sales reports, and website analytics

02

### Data Processing

Perform data cleaning and preprocessing, so that it can be used by BI tools.



03

### Data Exploration

Explore data with descriptive, diagnostic, predictive analysis methodologies.

04

### Data Visualization

Create dashboards and reports using charts, graphs, and maps.

## How Business Intelligence Works?

05

### Actionable Insights

BI tools can provide recommendations for action based on the insights gained from the data.

## How BI Differs From BA?

### Business Intelligence

- Focuses on Descriptive Analytics
- The goal of BI is to provide decision-makers with accurate and timely information, so that they can make informed decisions based on past performance.
- BI tries to answer questions like :
  - What were our sales figures last quarter?
  - How many customers did we have last year?
  - What was our customer satisfaction rating last month?

### Business Analytics

- Focuses on predictive, Prescriptive Analytics
- Involves using statistical and machine learning models to analyze historical data and make predictions about future performance.
- BA tries to answer questions like :
  - What sales figures will look like next year?
  - Which products are most likely to be popular with our customers next year?
  - What actions should we take to improve customer satisfaction?



# BI Tools

## 1. Tableau



Tableau Dashboard

Tableau is one of the most popular BI tools available today, known for its intuitive visualizations and easy-to-use interface. With Tableau, users can easily connect to a wide range of data sources and create interactive dashboards that allow for easy exploration and analysis of data.

## 2. Microsoft Power BI

With Power BI, users can easily create interactive dashboards and reports that provide real-time insights into business operations. Power BI is also tightly integrated with other Microsoft tools such as Excel and Azure, making it easy to connect and automate workflows.



Microsoft Power BI

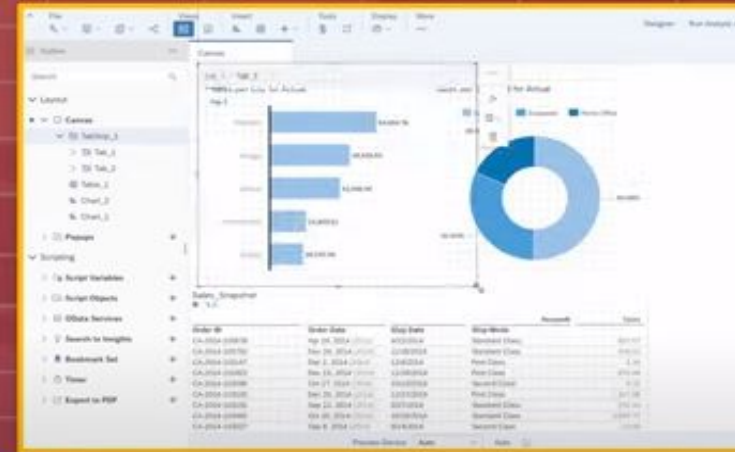


### 3. QlikView



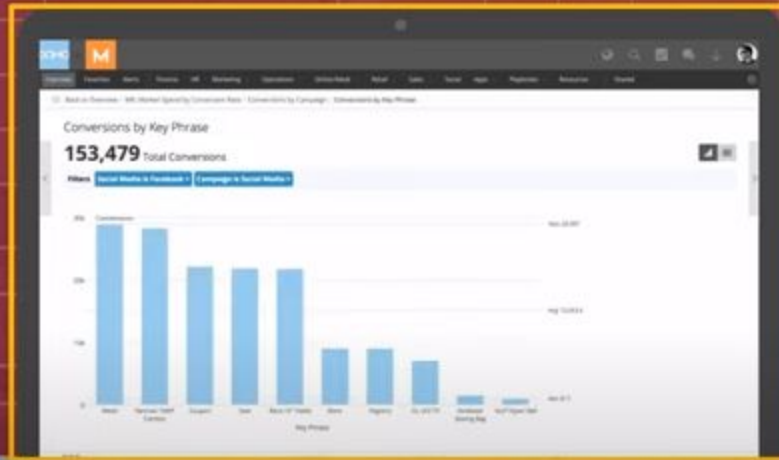
## 4. SAP BusinessObjects

With BusinessObjects, users can create interactive dashboards, reports, and ad hoc queries that provide real-time insights into business operations. BusinessObjects also includes advanced analytics capabilities, such as predictive modeling and data mining, making it a favorite among data scientists.



SAP BusinessObjects

## 5. Domo



Domo Dashboard

With Domo, users can easily connect to a wide range of data sources and create interactive dashboards that provide real-time insights into business operations. Domo also includes advanced collaboration features, such as real-time chat and collaboration tools, making it ideal for teams.

# Decision Support System

## Traditional vs Modern DSS

**Traditional DSS:** Historically, DSS and BI tools relied on preconfigured, historical data with no ability to drive real-time decisions and action. With this approach, decisions are made based on the past.

**Modern DSS:** New tools and processes allow for “active intelligence”, a state of continuous intelligence with an end-to-end analytics data pipeline delivering real-time, up-to-date information designed to trigger immediate insights and actions.

# **Data Analytics and Decision Support Systems: Transforming Businesses in the Digital Age**

Organizations across industries are recognizing the value of data as a strategic asset that can drive innovation, enhance efficiency, and ultimately, boost profitability.

However, the sheer volume and complexity of data generated daily pose a significant challenge.

This is where data analytics and decision support systems step in to provide invaluable insights and support informed decision-making.

## **The Data Explosion**

The digital revolution has led to an exponential growth in data generation.

From social media interactions and online transactions to IoT devices and sensor networks, data is being produced at an unprecedented rate.

This data tsunami offers immense opportunities for businesses to gain a competitive edge by harnessing its potential.

Data analytics is the process of examining large datasets to uncover hidden patterns, correlations, and trends.

It involves using various techniques and tools, such as machine learning, statistical analysis, and data visualization, to extract valuable insights from data.

These insights can inform a wide range of decisions, from operational and tactical to strategic.

***Bhagyashree***

## **Decision Support Systems (DSS)**

A DSS is a software-based solution that aids decision-makers in solving complex problems and making informed choices.

These systems integrate data analytics, business intelligence, and modeling tools to provide a holistic view of the business environment.

## Key Components of a DSS:

- 1. Data Repository:** A centralized database that stores historical and real-time data from various sources.
- 2. Analytics Engines:** The core of the DSS, these engines perform data analysis, predictive modeling, and statistical calculations.
- 3. User Interface:** A user-friendly interface that allows decision-makers to interact with the system, access reports, and run queries.
- 4. Visualization Tools:** Graphs, charts, and dashboards that simplify complex data into easily understandable visuals.
- 5. Reporting Functionality:** Automated report generation and distribution to keep stakeholders informed.
- 6. What-If Analysis:** The ability to simulate different scenarios and assess their impact on outcomes.



## **Applications of DSS:**

**Financial Decision Making:** DSS helps in financial planning, budgeting, and forecasting, aiding organizations in optimizing their resources and investments.

**Supply Chain Management:** It enables efficient inventory management, demand forecasting, and logistics planning.

**Marketing and Sales:** DSS assists in customer segmentation, market analysis, and campaign optimization.

**Healthcare:** In the medical field, DSS helps clinicians make accurate diagnoses and select the most appropriate treatments.

**Risk Management:** Identifying and mitigating risks is crucial in various industries, and DSS can analyze data to assess and manage risks effectively.

# The Synergy of Data Analytics and DSS

While data analytics is focused on processing and extracting insights from data, DSS takes these insights and translates them into actionable decisions.

Here's how the synergy between the two works:

**Data Preparation:** Data analytics processes like cleansing, integration, and transformation prepare raw data for analysis. DSS relies on this clean and well-organized data to generate meaningful reports and insights.

**Insight Generation:** Data analytics identifies patterns and trends in data. DSS uses these insights to provide decision-makers with a clear understanding of the current situation and potential scenarios.

**Decision Making:** Decision-makers can use the information provided by DSS to evaluate options, assess risks, and make informed decisions.

**Monitoring and Feedback:** DSS often includes monitoring features that track the impact of decisions. This data is then fed back into the analytics process to refine future decision-making.

**Challenges :** While data analytics and DSS offer tremendous benefits, they also come with challenges and considerations:

- 1. Data Privacy and Security:** With the increasing emphasis on data, ensuring the privacy and security of sensitive information is paramount.
- 2. Data Quality:** Accurate and reliable data is crucial for meaningful insights. Poor data quality can lead to erroneous conclusions.
- 3. Skill Gap:** Organizations need skilled data analysts and data scientists to extract valuable insights from data effectively.
- 4. Integration:** Integrating DSS into existing IT infrastructure can be complex and require careful planning.
- 5. Ethical Concerns:** The use of data analytics raises ethical questions regarding data usage, bias, and fairness.

# Data Mining

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis.

Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions.

Data mining is a key part of data analytics overall and one of the core disciplines in data science, which uses advanced analytics techniques to find useful information in data sets.

## Why is data mining important?

Effective data mining aids in various aspects of planning business strategies and managing operations. That includes customer-facing functions such as marketing, advertising, sales and customer support, plus manufacturing, supply chain management, finance and HR.

Data mining supports fraud detection, risk management, cybersecurity planning and many other critical business use cases. It also plays an important role in healthcare, government, scientific research, mathematics, sports and more.

Data mining is a crucial component of successful analytics initiatives in organizations.

The information it generates can be used in business intelligence (BI) and advanced analytics applications that involve analysis of historical data, as well as real-time analytics applications that examine streaming data as it's created or collected.

## Data mining process: How does it work?

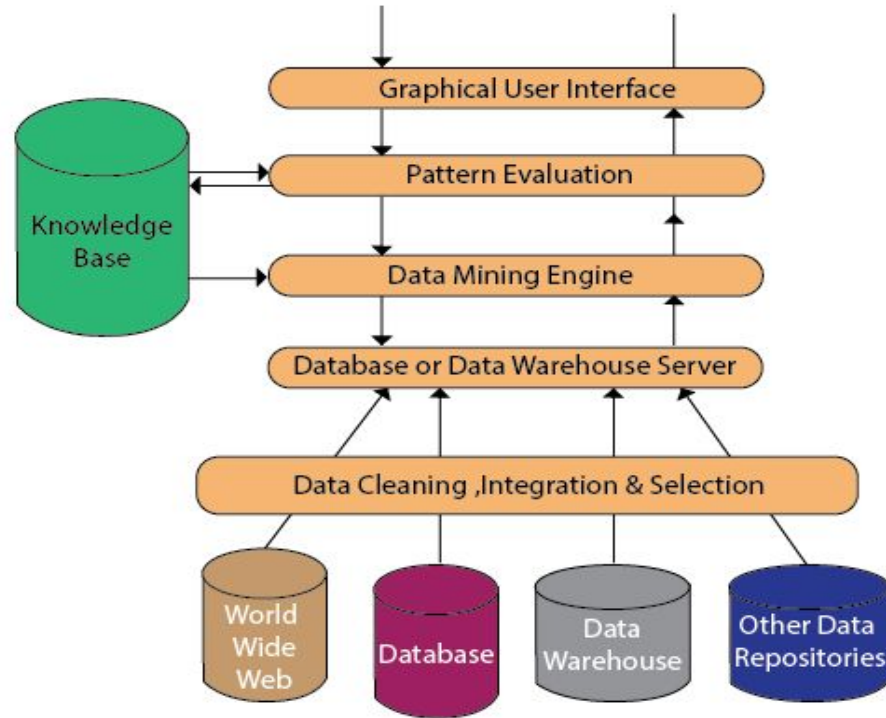


Fig. Data Mining Architecture

## **Data Source:**

The actual source of data is the Database, data warehouse, World Wide Web (WWW), text files, and other documents.

We need a huge amount of historical data for data mining to be successful. Organizations typically store data in databases or data warehouses.

Data warehouses may comprise one or more databases, text files spreadsheets, or other repositories of data. Sometimes, even plain text files or spreadsheets may contain information.

Another primary source of data is the World Wide Web or the internet.

## **Different processes:**

Before passing the data to the database or data warehouse server, the data must be cleaned, integrated, and selected.

As the information comes from various sources and in different formats, it can't be used directly for the data mining procedure because the data may not be complete and accurate. So, the first data requires to be cleaned and unified.

More information than needed will be collected from various data sources, and only the data of interest will have to be selected and passed to the server.

These procedures are not as easy as we think. Several methods may be performed on the data as part of selection, integration, and cleaning.



## **Database or Data Warehouse Server:**

The database or data warehouse server consists of the original data that is ready to be processed. Hence, the server is cause for retrieving the relevant data that is based on data mining as per user request.

## **Data Mining Engine:**

The data mining engine is a major component of any data mining system. It contains several modules for operating data mining tasks, including association, characterization, classification, clustering, prediction, time-series analysis, etc.

In other words, we can say data mining is the root of our data mining architecture. It comprises instruments and software used to obtain insights and knowledge from data collected from various data sources and stored within the data warehouse.

## **Pattern Evaluation Module:**

The Pattern evaluation module is primarily responsible for the measure of investigation of the pattern by using a threshold value. It collaborates with the data mining engine to focus the search on exciting patterns.

This segment commonly employs stake measures that cooperate with the data mining modules to focus the search towards fascinating patterns.

It might utilize a stake threshold to filter out discovered patterns. On the other hand, the pattern evaluation module might be coordinated with the mining module, depending on the implementation of the data mining techniques used.

For efficient data mining, it is abnormally suggested to push the evaluation of pattern stake as much as possible into the mining procedure to confine the search to only fascinating patterns.

## **Graphical User Interface:**

The graphical user interface (GUI) module communicates between the data mining system and the user. This module helps the user to easily and efficiently use the system without knowing the complexity of the process.

This module cooperates with the data mining system when the user specifies a query or a task and displays the results.

## **Knowledge Base:**

The knowledge base is helpful in the entire process of data mining. It might be helpful to guide the search or evaluate the stake of the result patterns.

The knowledge base may even contain user views and data from user experiences that might be helpful in the data mining process.

The data mining engine may receive inputs from the knowledge base to make the result more accurate and reliable. The pattern assessment module regularly interacts with the knowledge base to get inputs, and also update it.