Chapter-1: Introduction to Data Analytics

Dr. Vinod Patidar
Associate Professor
Department of Computer Science and Engineering

Parul® University

NAAC A++

Content

INDEX

1. Data Analytics

- Data analytics is the process of examining raw data to uncover patterns, draw conclusions, and make informed decisions.

- It involves various techniques and tools to convert data into actionable insights.

Fig. 1.1: Data Analytics

Introduction

- This is to introduce conceptual understanding using simple and practical examples.

- Focus on:

  – Why to use a particular technique of procedure

  – how to choose the right technique

  – how to use it correctly

  – how to interpret the result.

2. Data and its importance

- Three terms: Variable, measurement and Data
- What is generating so much data?
- How data add value to the business?
- Why data is important?

Variable, Measurement and Data

- **Variable-** is a characteristic of any entity being studied that is capable of taking on different values.

- **Measurements-** is when a standard process is used to assign numbers to particular attributes or characteristics of a variable.

- **Data-** data are recorded measurements

What is generating so much data?

- Data can be generated by

  ➢ Humans,

  ➢ Machines or

  ➢ Humans-machines combines

- It can be generated anywhere where any information is generated and stored in structured or unstructured formats
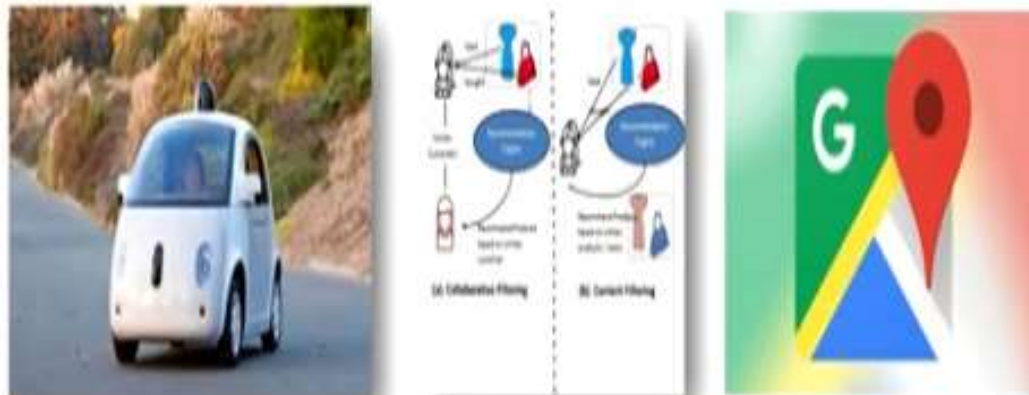
Example of Data Products

Fig. 2.1: Data Products
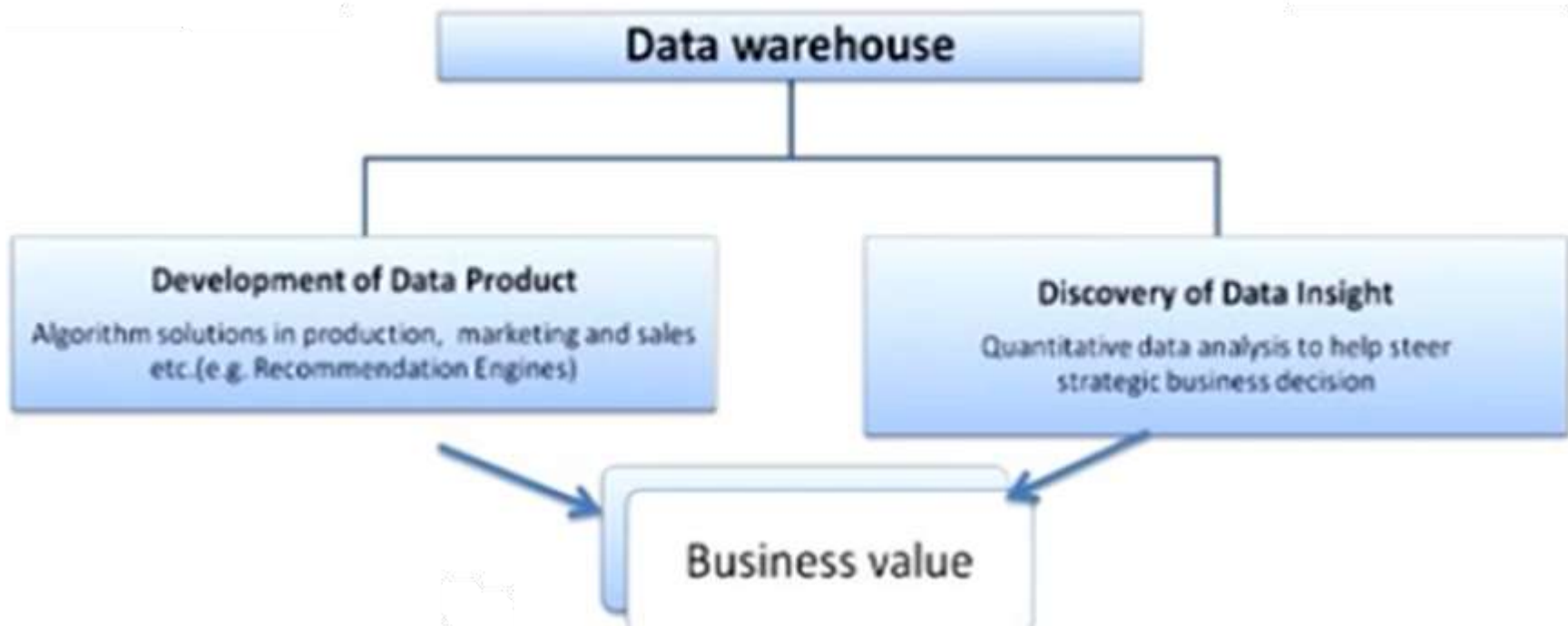
How data add value to business?

Fig. 2.2: Data and data Value

Why data is important?

- Data helps in make better decisions

- Data helps in solve problems by finding the reason for underperformance

- Data helps one to evaluate the performance

- Data helps one improve processes

- Data helps one understand consumers and the market

3. Define Data analytics and its types

- Data analytics is defined as "the scientific process of transforming data into insights for making better decisions."

- Professor James Evans has defined the data analytics that "it is the use of the data information technology, statistical analysis, quantitative methods and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions."

- These insights can be valuable to organizations because they help drive decision-making and strategy formulation.

4. Why Data analytics is important?

Opportunity abounds for the use of analytics and big data such as:

1. Determining credit risk
2. Developing new medicines
3. Finding more efficient ways to deliver products and services
4. Preventing fraud
5. Uncovering cyber threats
6. Retaining the most valuable customers

Data Analysis

- It is the process of examining, transforming, and arranging raw data in a specific way to generate useful information from it.

- It allows for the evaluation of data through analytical and logical reasoning to lead to some sort of outcome or conclusion in some context.

- It is a multi-faceted process that involves a number of steps, approaches, and diverse techniques.

Data Analytics Vs. Data Analysis

Analysis

⟵

Past

Explain

How?   Why?

Data analysis it is something about -
- what has happened in the past.
- why that has happened?
- how it has happened?
- Example: Post mortem analysis
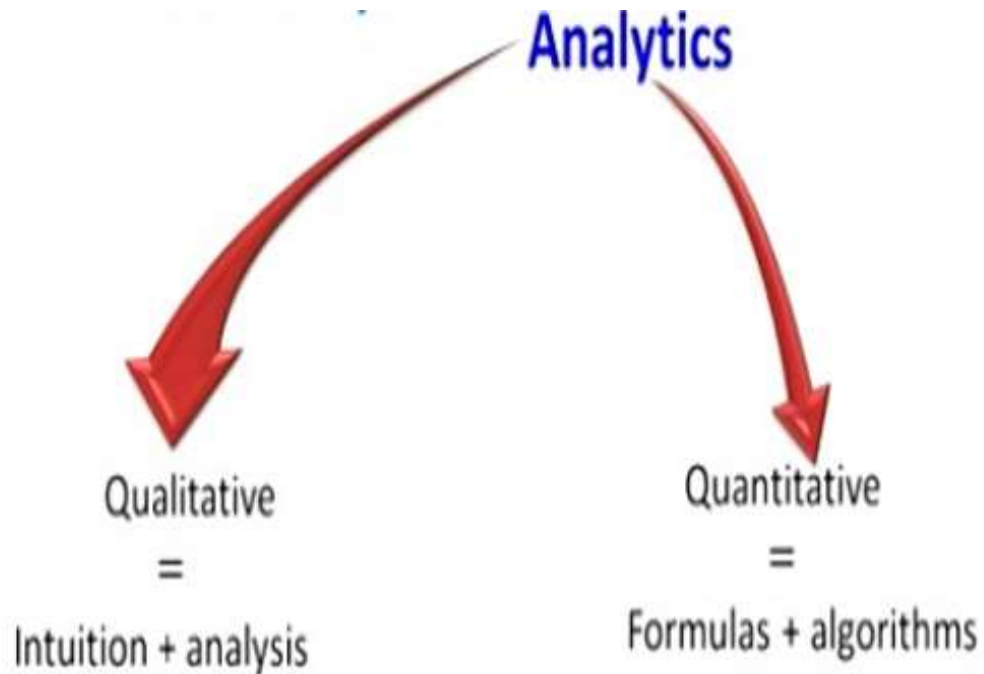
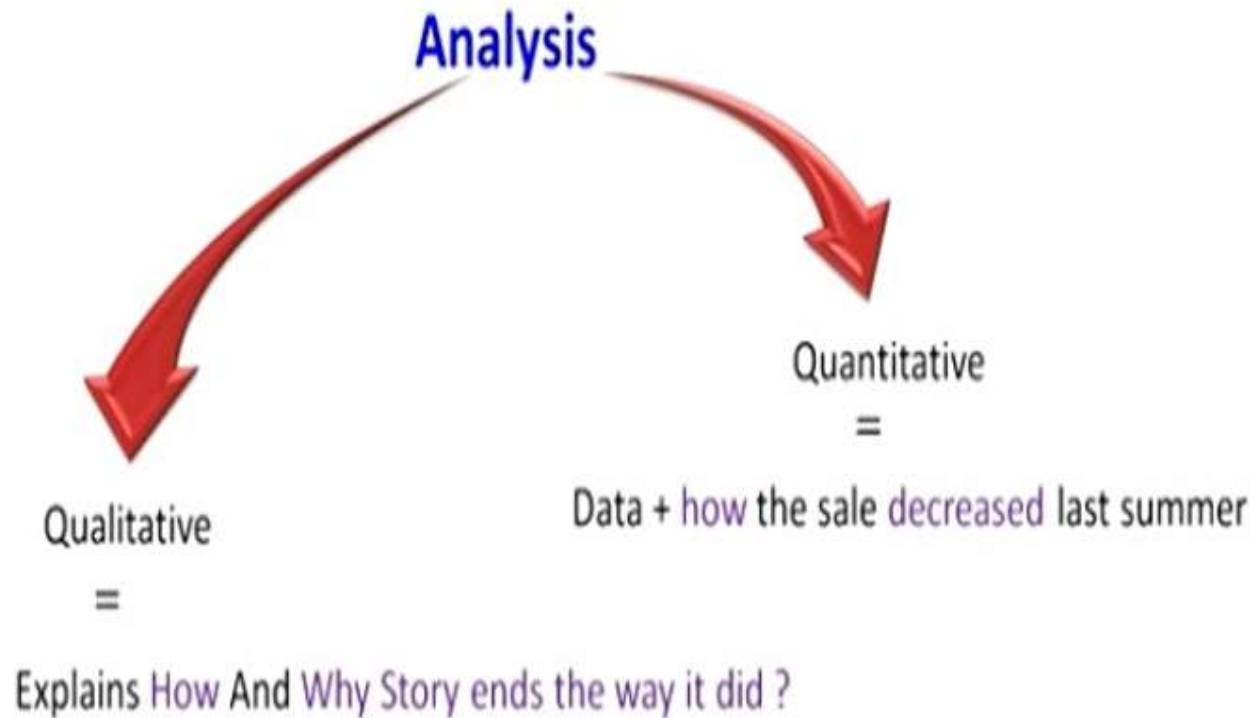## 5. Data Analytics Vs. Data Analysis

Analytics

→

Future

Analytics is studying about –

- what will happen in future.
- predict explore possible potential future events.

Data Analytics Vs. Data Analysis

Data Analytics Vs. Data Analysis

Analysis

Qualitative
=
Explains How And Why Story ends the way it did ?

Quantitative
=
Data + how the sale decreased last summer

6. Classification of Data Analytics…

Based on the phase of workflow and kind of analysis required, there are four major types of data analytics:

1. Descriptive analytics

2. Diagnostic analytics

3. Predictive analytics

4. Prescriptive analytics

Classification of Data Analytics...
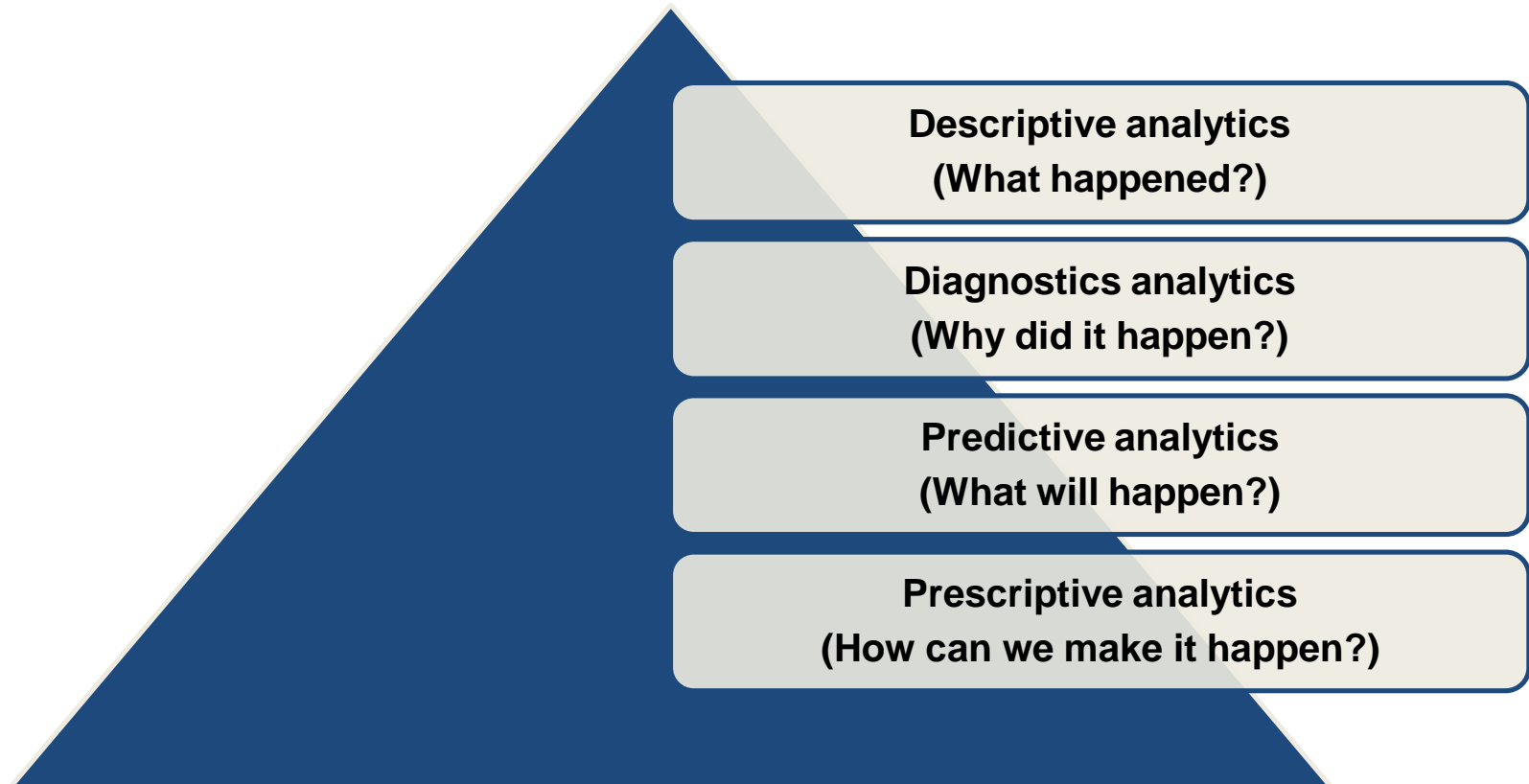
**Descriptive analytics
(What happened?)**

**Diagnostics analytics
(Why did it happen?)**

**Predictive analytics
(What will happen?)**

**Prescriptive analytics
(How can we make it happen?)**

Fig. 6.1: Classification of Data Analytics

Classification of Data analytics...

1. Descriptive analytics(What happened?)-
- Descriptive analytics is the conventional form of business intelligence or data analysis.
- It seeks to provide the depiction or summary view of facts and figures in an understandable format.
- These either inform or prepare data for further analysis. so descriptive analysis or we can say another way in statistics can summarize raw data and convert it into your form that can be easily understood by humans.
- They can describe in detail about an even that has occurred in the past.
- Example: company reports that simply provide the historic review like: data queries, reports, descriptive statistics, data visualization and data dashboard.

Classification of Data analytics...

2. Diagnostic analytics(Why did it happen?)-

- Diagnostic analytics is the process of using data to determine the causes of trends and correlations between variables.

- It can be viewed as a logical next step after using descriptive analytics to identify trends.

- Diagnostic analysis can be done manually, using an algorithm, or with statistical software (such as Microsoft Excel).

Classification of Data analytics…

3. Predictive analytics (What will happen?)

- Example: linear regression, time series analysis and forecasting and data mining.
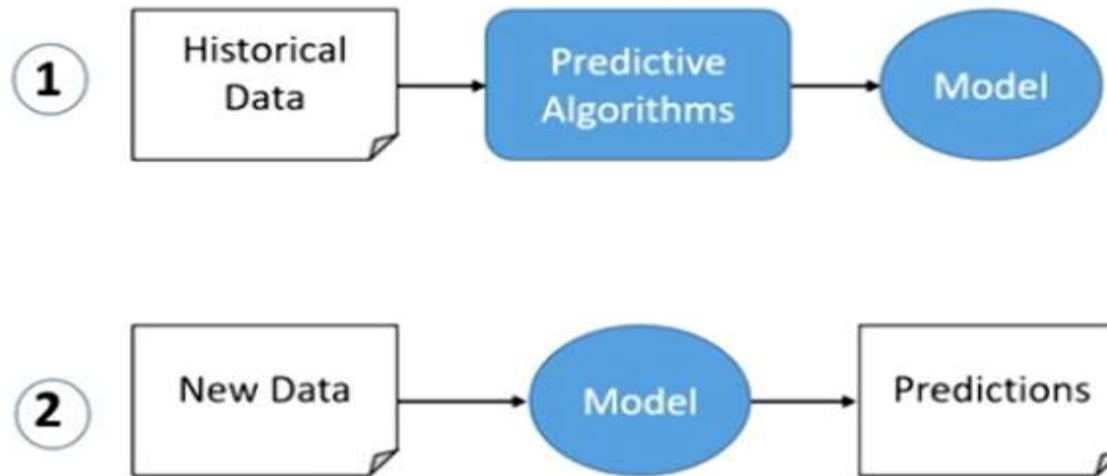- These are the techniques for predictive analytics.

Fig. 6.2: Predictive Analytics

- With the help of historical data by using different algorithm, predictive algorithms you can come with a model.

- Once the model is developed a new data can be fit into this model so we can get some predictions about the past events.

Classification of Data analytics…

4. Prescriptive analytics(How can we make it happen?)-

- A set of techniques to indicate the best course of action.
- It tells what decision to make to optimize the outcome.
- The goal of prescriptive analytics is to enable: quality improvements, service enhancements, cost reductions and increasing productivity.
- Examples: Some of the tools like optimization models, simulation model, and decision analysis.
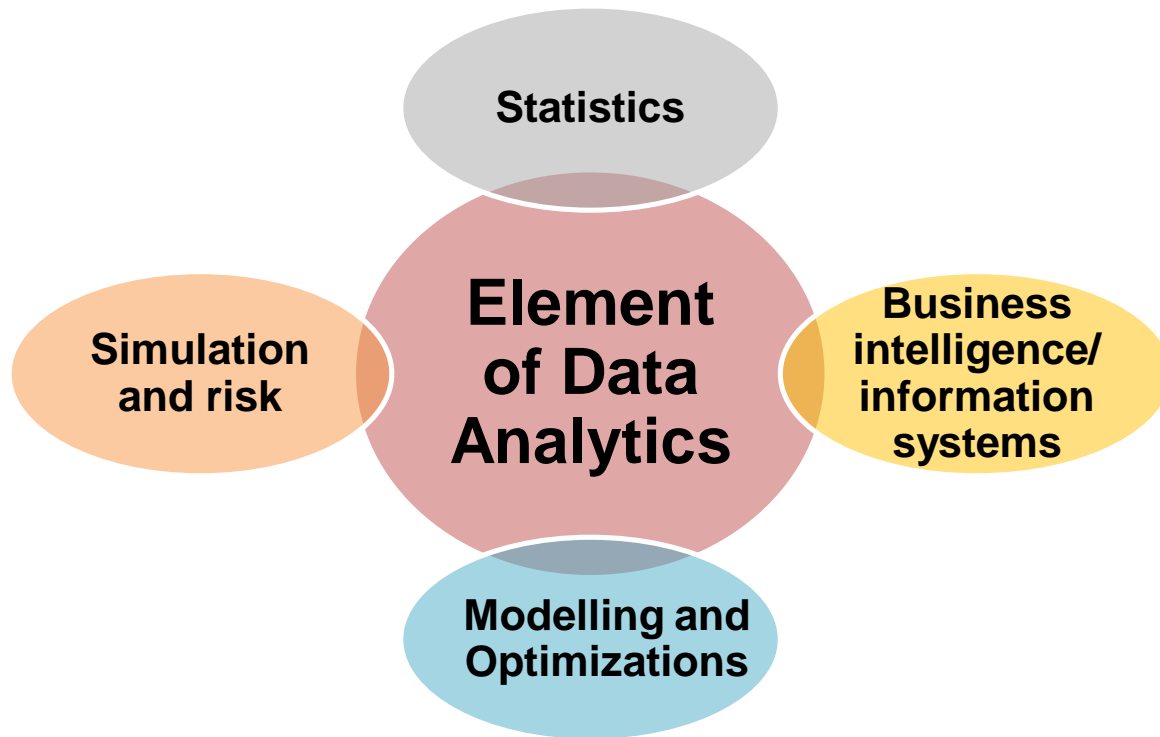
7. Element of Data analytics

Fig. 7.1: Data analytics Elements

8. Data analyst Vs. Data scientist

Aspect	Data Analyst	Data Scientist
Primary Focus	Analyzing data to provide insights for business decisions.	Using advanced statistical and computational methods to solve complex problems.
Skills Required	– Data Cleaning and Preparation – Statistical Analysis – Data Visualization – SQL – Excel Skills – Problem-Solving – Domain Knowledge – Communication Skills – Attention to Detail – Time Management – Continuous Learning	– Machine Learning – Statistical Analysis – Data Cleaning and Preprocessing – Data Visualization – Big Data Technologies (e.g., Hadoop, Spark) – Deep Learning – Natural Language Processing (NLP) – SQL Database Management – Experiment Design and A/B Testing – Cloud Computing Platforms (e.g., AWS, Azure, Google Cloud) – Communication and Presentation Skills – Domain Knowledge – Time Series Analysis – Feature Engineering

Data Analyst Vs. Data Scientist

Aspect	Data Analyst	Data Scientist
Typical Tasks	Cleaning and organizing data, creating reports, generating dashboards, and performing descriptive analytics.	Building predictive models, conducting A/B testing, developing algorithms, and performing exploratory data analysis.
Example Scenario	Analyzing sales data to identify trends and optimize marketing strategies.	Developing a recommendation system for an e-commerce platform based on customer behavior.
Decision Making	Helps businesses make data-driven decisions by providing insights from existing data.	Involves both providing insights and developing solutions to complex problems using data.
Tools Used	Excel, SQL, Tableau, Power BI, Google Analytics.	Python, R, SQL, TensorFlow, PyTorch, Jupyter Notebooks, Big Data Technologies (e.g., Hadoop, Spark).

Applications of Data Analytics:

1. Business: Enhancing marketing strategies, optimizing operations, improving customer service.

2. Healthcare: Predicting disease outbreaks, personalizing patient care, optimizing hospital operations.

3. Finance: Risk management, fraud detection, portfolio management.

4. Retail: Inventory management, customer segmentation, personalized recommendations.

5. Sports: Performance analysis, game strategy optimization, injury prevention.

6. Government: Public policy analysis, resource allocation, public health monitoring.

In summary,

- Data Analytics is important because it transforms raw data into meaningful insights that drive strategic decision-making, operational improvements, and innovation.

- It equips organizations with the knowledge they need to operate more efficiently, respond to market changes swiftly, and meet customer needs effectively.