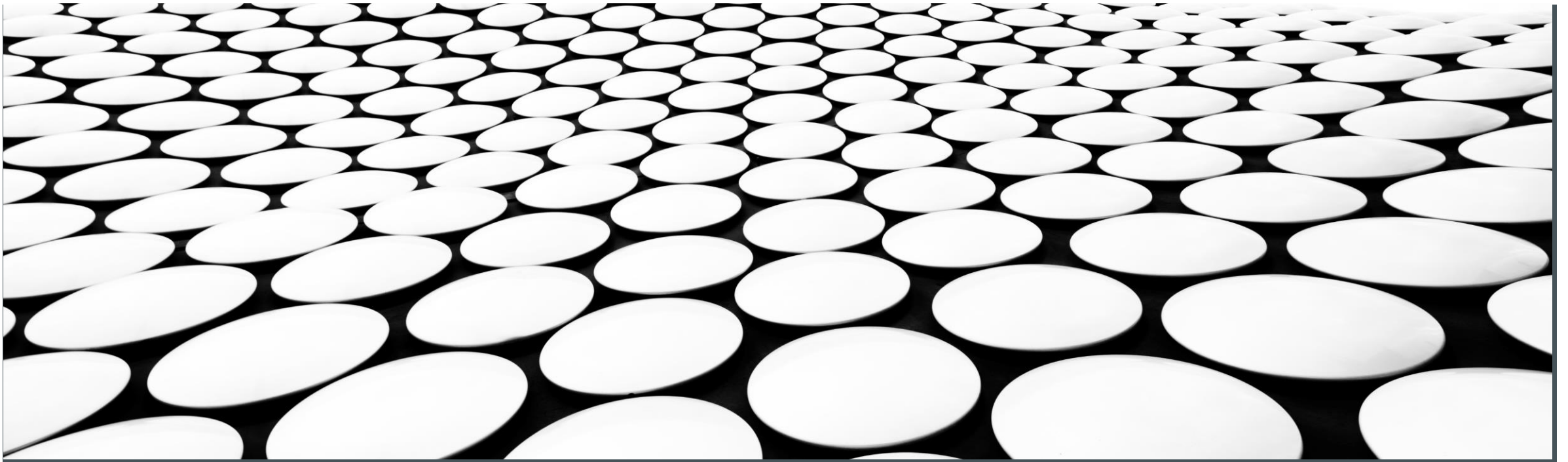


DATA SCIENCE



WHAT IS DATA?

- Data is a collection of information.
- It can be numbers, text, images, or even sounds.
- Data can be structured or unstructured.
- Structured data is organized in a specific format, such as a table.
- Unstructured data is not organized in a specific format, such as a text document or an image.



TYPES OF DATA

- **Quantitative data:** Data that can be counted or measured.
- **Qualitative data:** Data that describes something.
- **Mixed data:** Data that contains both quantitative and qualitative elements.

HOW IS DATA USED

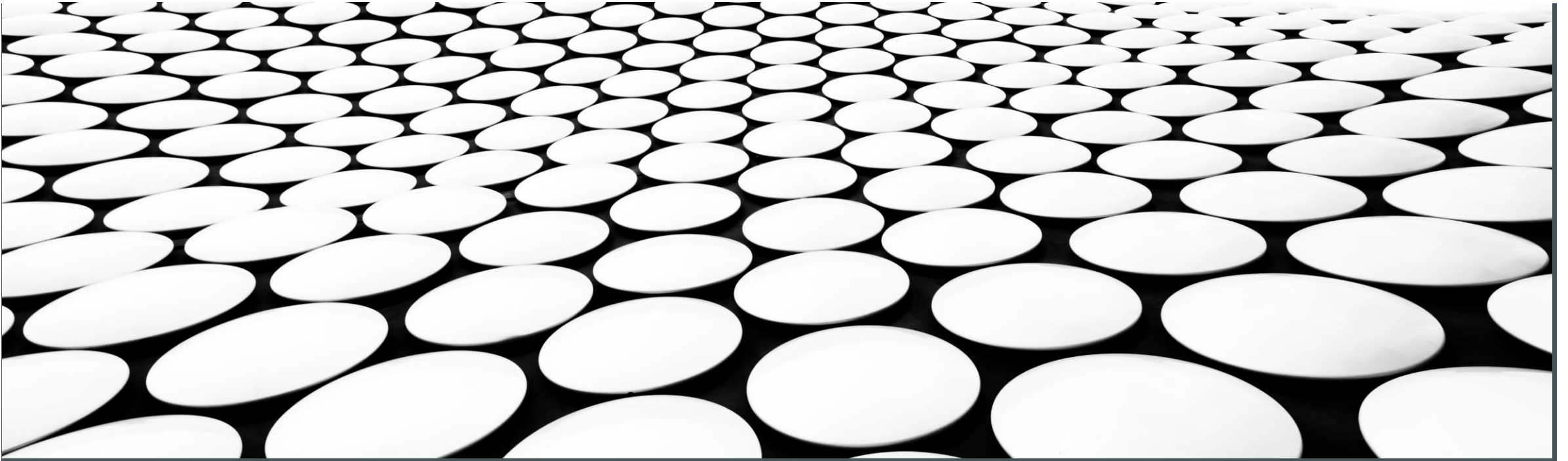
Data is used for a variety of purposes, including:

- Making informed decisions
- Identifying trends
- Solving problems
- Improving products and services
- Understanding customer behavior

THE IMPORTANCE OF DATA

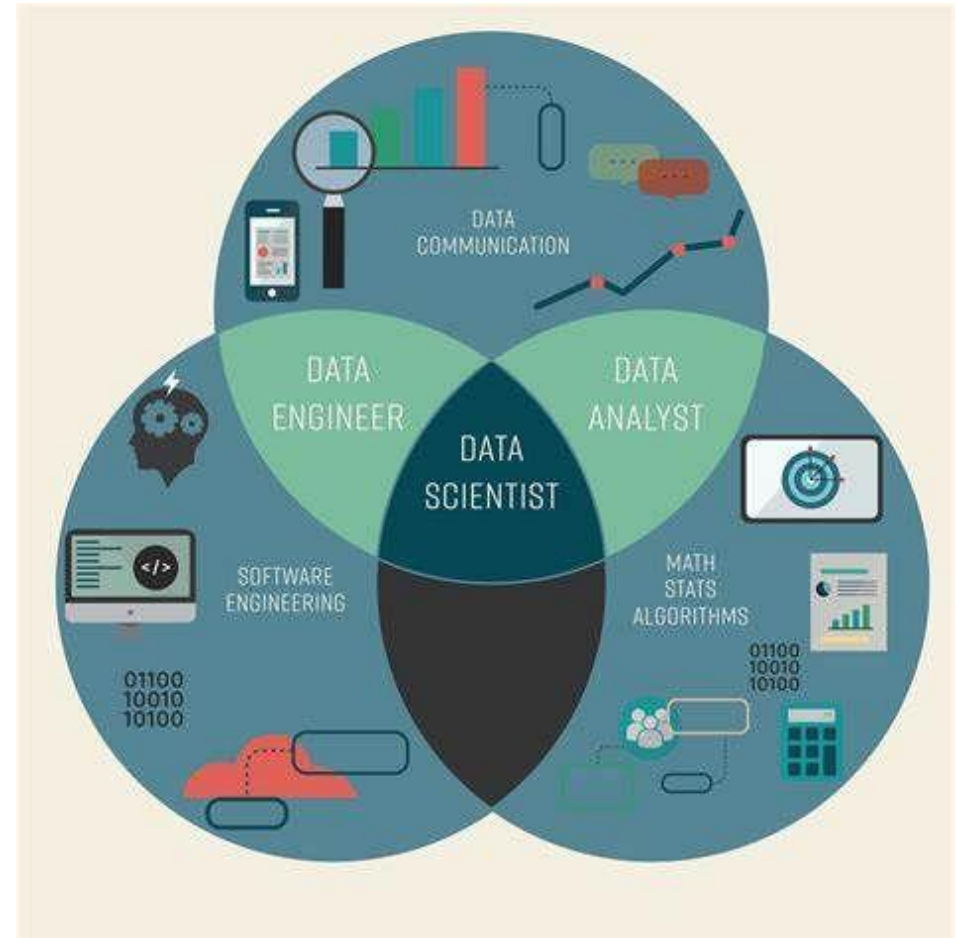
- Data is becoming increasingly important in today's world.
- As the amount of data we collect continues to grow, so does its potential to improve our lives.
- By understanding data, we can make better decisions, solve complex problems, and gain a deeper understanding of the world around us.

DATA SCIENCE VS DATA ANALYTICS



DATA SCIENCE : UNVEILING THE UNKNOWN

- **Broad and Evolving Field:** Data science encompasses a wide range of techniques and tools used to extract knowledge from data.
- **Focus on Unstructured Data:** Data scientists often deal with unstructured data, like social media posts or images, that requires advanced techniques to analyze.
- **Predictive Modeling and Machine Learning:** A core competency of data science is building models that can predict future events or outcomes. This involves machine learning algorithms that can learn from data without explicit programming.



DATA ANALYTICS : MAKING SENSE OF THE PAST

- **Focused on Historical Data:** Data analysts primarily work with structured data, like sales figures or customer information, that is already organized in a specific format.
- **Identifying Trends and Patterns:** Data analysts use statistical methods and data visualization tools to uncover trends and patterns in historical data.
- **Communication and Storytelling:** Data analysts excel at communicating insights from data in a clear and concise way, often through reports, dashboards, and presentations.



DATA ANALYTICS TOOLS

- Python – This object-oriented open-source programming language is used for manipulating, visualizing, and modelling data.
- R – An open-source programming language used in numerical and statistical analysis.
- Tableau – This helps in creating several kinds of visualizations for presenting insights and trends in a better way.
- Power BI – This is a business intelligence tool that supports multiple data sources, helps in asking questions and getting immediate insights.
- SAS – This statistical analysis software helps in performing analytics, visualizing data, writing SQL queries, performing statistical analysis, and building ML models.



PROCESSES IN DATA ANALYTICS

The data analytics practice encompasses many separate processes, which can comprise a data pipeline:

- Collecting and ingesting the data
- Categorizing the data into structured/unstructured forms, which might also define next actions
- Managing the data, usually in databases, data lakes, and/or data warehouses
- Storing the data in hot, warm, or cold storage
- Performing ETL (extract, transform, load)
- Analyzing the data to extract patterns, trends, and insights
- Sharing the data to business users or consumers, often in a dashboard or via specific storage

PRIMARY DATA AND SECONDARY DATA

Primary data



Primary data collection involves the collection of original data directly from the source or through direct interaction with the respondents.

Secondary data



Secondary data collection involves using existing data collected by someone else for a purpose different from the original intent.

1. Primary Data Collection:

- Surveys and Questionnaires
- Interviews
- Observations
- Experiments
- Focus Groups

2. Secondary Data Collection:

- Published Sources
- Online Databases
- Government and Institutional Records
- Publicly Available Data
- Past Research Studies

ETL (EXTRACT TRANSFORM LOAD)

- **Extract:** Retrieve data from various sources, such as databases, files, or APIs.
- **Transform:** Clean, filter, and manipulate data to ensure consistency and prepare it for analysis.
- **Load:** Store the transformed data into a target system or data warehouse for easy access and analysis.



THE FOUR MAIN TYPES OF DATA ANALYSIS

Descriptive

What happened?

Diagnostic

Why did it happen?

Predictive

What is likely to happen in the future?

Prescriptive

What's the best course of action?

DIAGNOSTIC ANALYTICS

- **Definition:** Diagnostic analytics aims to determine the root causes and reasons behind certain events or trends observed in the data.
- **Key Characteristics:** Involves data exploration, drill-down analysis, and correlation identification. Diagnostic analytics answers the question of "why did it happen."
- **Examples:** Data mining techniques, regression analysis, cohort analysis.

DESCRIPTIVE ANALYTICS

- **Definition:** Descriptive analytics focuses on summarizing historical data to gain insights into past events and understand the current state.
- **Key Characteristics:** Involves data aggregation, visualization, and reporting. Descriptive analytics answers the questions of "what happened" and "what is happening."
- **Examples:** Bar charts, line graphs, dashboards displaying key performance indicators (KPIs).

PREDICTIVE ANALYTICS

- **Definition:** Predictive analytics leverages historical data to make predictions about future outcomes or events.
- **Key Characteristics:** Involves statistical modeling, machine learning algorithms, and pattern recognition. Predictive analytics answers the question of "what is likely to happen."
- **Examples:** Forecasting models, time series analysis, classification algorithms.

PRESCRIPTIVE ANALYTICS

- **Definition:** Prescriptive analytics recommends the best course of action based on predictive models, optimization techniques, and business rules.
- **Key Characteristics:** Involves simulation, optimization algorithms, and decision support systems. Prescriptive analytics answers the question of "what should be done."
- **Examples:** Optimization models, simulation tools, decision support systems.

ROLE OF A DATA ANALYST

- A data analyst role is to answer specific questions or address particular challenges that have already been identified and are known to the business.
- To do this, they examine large datasets with the goal of identifying trends and patterns. They then “visualize” their findings in the form of charts, graphs, and dashboards.

CAREER

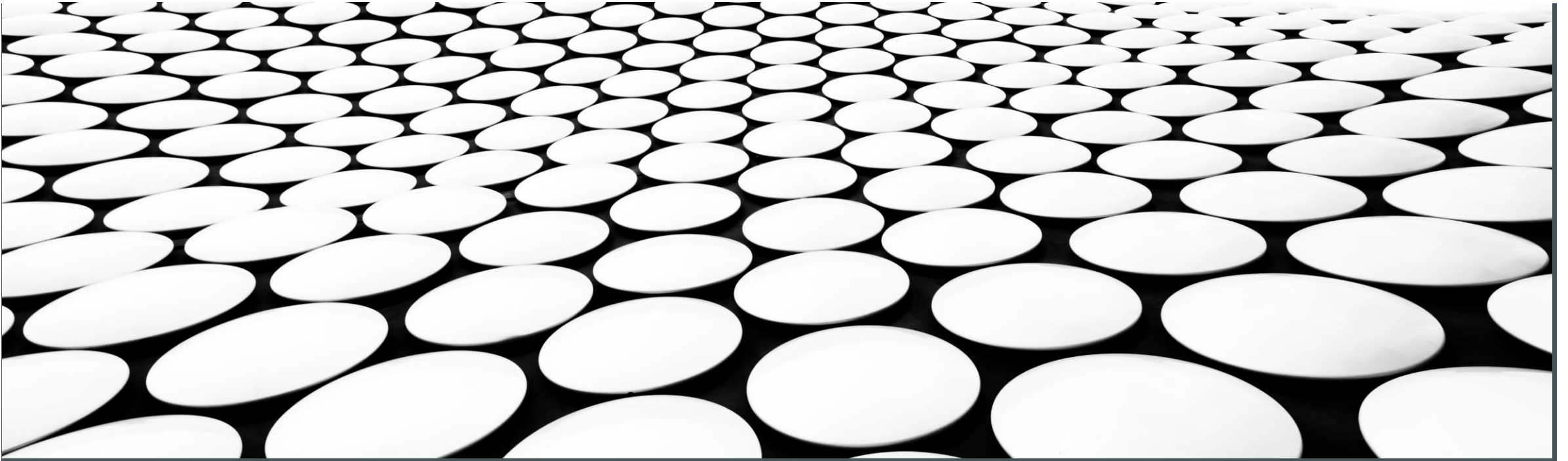
- 1. Data Scientist
- 2. Business Intelligence Analyst
- 3. Data Engineer
- 4. Business Analyst
- 5. Marketing Analytics Manager
- 6. Financial Analyst
- 7. Quantitative Analyst
- 8. Risk Analyst
- 9. Data Governance Analyst
- 10. Data Visualization Engineer



IMPORTANCE OF DATA ANALYTICS IN INDUSTRIES



HOW DATA ANALYTICS TRANSFORMS INDUSTRIES



REVOLUTIONIZING RETAIL: TAILORING THE CUSTOMER JOURNEY

- **Understanding Customer Behavior:** Data analytics helps retailers analyze customer purchase history, browsing patterns, and demographics to understand preferences and predict future behavior.
- **Personalized Marketing:** By leveraging customer insights, retailers can personalize marketing campaigns, product recommendations, and loyalty programs, leading to increased customer engagement and satisfaction.
- **Optimizing Inventory Management:** Data analytics enables retailers to forecast demand and optimize inventory levels, reducing stockouts and overstocking, and improving overall supply chain efficiency.

HEALTHCARE TRANSFORMED: DATA-DRIVEN DIAGNOSIS AND TREATMENT

- **Early Disease Detection:** Data analytics helps analyze patient data, including medical history, lab results, and imaging scans, to identify potential health risks and enable early detection of diseases.
- **Personalized Medicine:** By analyzing patient data, healthcare providers can tailor treatment plans to individual patients, considering their unique genetic makeup and medical history.
- **Improving Treatment Outcomes:** Data analytics enables researchers to analyze clinical trial data and identify more effective treatment options, leading to improved patient outcomes.

FINANCIAL SERVICES: MITIGATING RISK AND MAKING SMARTER INVESTMENTS

- **Fraud Detection:** Data analytics helps financial institutions identify suspicious transactions and patterns that might indicate fraudulent activity, protecting both customers and the institution itself.
- **Credit Risk Assessment:** By analyzing financial data, lenders can make more informed decisions about loan approvals and interest rates, mitigating credit risk and ensuring financial stability.
- **Market Analysis and Investment Strategies:** Data analytics empowers financial institutions to analyze market trends and identify profitable investment opportunities, leading to better returns for their clients.



THANKYOU