INTRODUCTION TO DATA ANALYTICS

Introduction to data analytics
What is data?
 The smallest unit of factual information that can be used as a basis for calculation, reasoning or discussion
Types of data
Unstructured data- information, in many different forms, that doesn't follow conventional data models, making it difficult to store and manage in a mainstream relational database e.g. texts, mails, blogs, images etc.
Structured data-data that has a standardized format for efficient access by software and humans alike. It is typically tabular with rows and columns that clearly define data attributes. Computers can effectively process structured data for insights due to its quantitative nature. Examples

- Excel files.
- SQL databases.
- Point-of-sale data.
- Web form results.
- Search engine optimization (SEO) tags.
- Product directories.
- Inventory control.
- Reservation systems.

Semi-structured Data- Data that does not fit into the formal structure of data models but contains tags or markers to separate data elements.
Example: XML, JSON.

 Big data- refers to extremely large and diverse collections of structured, unstructured, and semi-structured data that continues to grow exponentially over time. These datasets are so huge and complex in volume, velocity, and variety, that traditional data management systems cannot store, process, and analyze them.

5 Vs of big data
1. Volume- amount of data that is generated and collected
2. Velocity- the speed at which data is generated. It deals with real- or near-time data streams that must be analysed in a timely manner.
3. Variety- different types of data that are generated and collected. Big data often deals with structured data (relational database) as well as unstructured data (text, images and videos)
4. Veracity- accuracy and truthfulness of the data
5. Value- potential insights that can be gained from the data.

Data analytics
 the process of examining, cleaning, transforming and modelling data to extract useful information, draw conclusion and support decision making.

The goal of data analytics is to turn data into actionable insights that can help organizations make better decisions, improve their operations and gain competitive advantage.

Importance of data analytics

1.Operational Efficiency

Process Optimization: Data analytics can identify inefficiencies and bottlenecks in processes. By understanding these issues, organizations can optimize operations, reduce costs, and improve productivity.

2.Resource Allocation: Analytics helps in the optimal allocation of resources by identifying areas of high performance and those needing improvement.

3.Decision Making

Informed Decisions: Data analytics provides insights that lead to more informed and evidence-based decision-making. By analyzing data, organizations can identify trends, patterns, and anomalies that inform strategic decisions. Predictive Analysis: Advanced analytics techniques such as machine learning can predict future trends, helping organizations to proactively plan and make better business decisions.

4. Competitive Advantage

Market Insights: Analyzing market data helps organizations understand customer behavior, market trends, and competitive dynamics, which is crucial for staying ahead in the market.

Innovation: Data-driven insights can lead to new product ideas, improvements in existing products, and innovation in services, providing a competitive edge.

5. Personalization

Customer Experience: Data analytics enables organizations to personalize customer interactions and experiences by understanding individual preferences and behaviors.

Targeted Marketing: Marketers can create more effective and targeted campaigns by analyzing customer data, leading to higher engagement and conversion rates.

6. Risk Management

Risk Identification: Analytics helps in identifying potential risks and vulnerabilities, allowing organizations to mitigate them before they escalate.

Fraud Detection: Advanced analytics techniques can detect unusual patterns indicative of fraudulent activities, enhancing security measures.

7. Business Intelligence

Strategic Insights: Data visualization tools transform complex data sets into intuitive visual formats, making it easier for stakeholders to grasp critical insights quickly.

Performance Monitoring: Dashboards and visualizations help in monitoring key performance indicators (KPIs) and tracking progress towards goals in real-time.

8. Enhancing Communication

Clarity: Visualizations make complex data understandable and accessible to a broader audience, including non-technical stakeholders.

Storytelling: Data visualization enables effective storytelling, where data-driven narratives can influence and drive action.

9. Evidence-Based Policy Making

Public Sector: Governments and public institutions use data analytics to develop policies based on empirical evidence, improving public services and resource management.

Healthcare: In healthcare, data analytics improves patient outcomes by identifying effective treatments and operational efficiencies.

Importance of Data Visualization

1. Simplifies Complex Data

Accessibility: Data visualization breaks down complex data into simpler visuals, making it accessible and understandable for a broader audience.

Pattern Recognition: Visual representation of data allows for quick identification of patterns, trends, and outliers that might not be obvious in raw data.

2. Enhances Data Interpretation

Immediate Insight: Visuals like charts, graphs, and maps provide immediate insights, making it easier to grasp the story behind the data without needing to analyze extensive tables.

Comparative Analysis: Visualization tools facilitate the comparison of different data sets, highlighting differences and similarities effectively.

3. Effective Communication

Engagement: Well-designed visuals are more engaging and can capture the audience's attention better than textual data.

Influence: Visuals can be persuasive, influencing decisions and driving actions by presenting data in a compelling manner.

4. Facilitates Exploration

Interactive Exploration: Interactive visualizations allow users to explore data dynamically, drilling down into details, and uncovering insights that static data cannot provide.

User-Friendly: Tools like dashboards make it easy for users to interact with data, customizing their view to meet specific needs and queries.

5. Real-Time Monitoring

Timeliness: Real-time visualizations enable organizations to monitor operations and performance metrics continuously, allowing for prompt responses to emerging issues.

Dashboard Integration: Integrating visualizations into dashboards provides a consolidated view of key metrics, enhancing situational awareness and decision-making.

Types of data analytics

I. **Descriptive analytics**- summarizing and describing the past. It involves analyzing historical data to understand what has happened and identifying patterns and trends. Answers questions what has happened? how has it happened?

II. **Predictive analytics**- deals with forecasting the future. It involves using statistical and machine learning techniques to analyze historical data and make predictions about future events. Answers the question what will happen, how likely is it to happen?

III. **Prescriptive analytics**-goes beyond predicting the future and gives recommendations on what actions to take. It involves using optimization, simulation and decision analysis techniques to find the best course of action. Answers the question what should we do? What will be the outcome of our action. It is the most advanced and rely on machine learning and AI