Big Data Analytics

Difference between Analysis and Analytics

Aspect	Analysis	Analytics
Definition	Examining data in detail to discover patterns, relationships, and insights.	Using tools, technologies, and processes to analyze data and gain insights.
Scope	Focuses on understanding specific questions or problems using historical data.	Includes descriptive, diagnostic, predictive, and prescriptive analytics, covering a wider range of applications.
Methods	Utilizes various statistical and computational methods to interpret and describe data.	Employs advanced techniques like data mining, predictive modelling, and optimization algorithms.
Objective	To explain what has happened and why it happened.	To derive actionable insights, make predictions, and support decision-making processes.
Outcome	Results in reports, summaries, and detailed explanations of the data.	Produces dashboards, visualizations, predictive models, and strategic recommendations.
Examples	Financial statement analysis, customer feedback analysis, market research analysis.	Predictive sales analytics, customer behavior analytics, supply chain optimization, fraud detection.

Big Data:

Big data refers to large and complex datasets that traditional data processing software and methods are unable to handle efficiently. Big data requires advanced analytical methods and technologies, such as distributed computing, machine learning, and data mining, to extract meaningful information and support decision-making processes.

6V's of Big Data

Characteristics of big data (6V's):

1. Volume:

- **Definition:** Volume refers to the vast amount of data generated every second. The sheer quantity of data is immense and continues to grow.
- **Example:** Social media platforms like Facebook and Twitter generate petabytes of data daily through user posts, comments, and interactions. Each day, billions of photos, videos, and text updates are shared across these platforms.

2. Velocity:

- **Definition:** Velocity is the speed at which data is generated, collected, and analyzed. It encompasses the rate of data flow and how quickly it needs to be processed.
- **Example:** Financial markets generate data at high speeds with stock prices fluctuating every millisecond. Realtime trading systems need to process this data instantaneously to make timely investment decisions and execute trades.

3. Variety:

- **Definition:** Variety refers to the different types of data, which can be structured (e.g., databases), semi-structured (e.g., XML files), and unstructured (e.g., text documents, images).
- **Example:** An ecommerce website collects data from various sources, including structured data from databases (customer transactions), semi-structured data from log files (e.g., server logs), and unstructured data from customer reviews and social media posts (e.g., textual reviews, photos).

4. Veracity:

- **Definition:** Veracity involves the accuracy and reliability of the data. It addresses the quality and trustworthiness of the data collected.
- **Example:** In healthcare, patient data needs to be accurate and reliable for effective treatment. Incorrect or incomplete data, such as erroneous patient records or mislabelled test results, can lead to incorrect diagnoses and treatment plans.

5. Value:

- **Definition:** Value refers to the potential insights and benefits that can be derived from analyzing the data. It's not just about having data but extracting meaningful information from it.
- **Example:** In retail, analyzing customer purchase data can reveal trends and preferences, allowing companies to tailor their marketing strategies, optimize inventory management, and improve customer satisfaction based on actionable insights.

6. Variability:

- **Definition:** Variability denotes the inconsistency of the data, which can vary in meaning and context. It captures how data can change over time or have different interpretations.
- **Example:** In social media analytics, the meaning of hashtags and keywords can vary based on current trends, events, or cultural context. For instance, the hashtag #BlackFriday may have different implications during the holiday season compared to other times of the year.

Sources of Big Data:

I. Social Media
II. IOT devices
III. Transactional Data
IV. Healthcare Data
V. Government and Public Data
VI. Multimedia Data

VII. Clickstream Data

Challenges Of Big Data:

A. Storage and Management
B. Data Integration
C. Quality Assurance
D. Privacy and Security
E. Processing and Analysis
F. Scalability
G. Real Time Processing
H. Visualization
I. Cost Management
J. Skill Gaps