**Experiment No 1**

**Aim:** Design Star and Snowflake schema. Compare Business intelligence, Data Science, Machine Learning and Artificial Intelligence.

**LO1**: Identify the sources of data for mining and perform data exploration.

**Theory:**

Business Intelligence (BI), Data Science, Machine Learning (ML), and Artificial Intelligence (AI) are all interconnected fields that involve data processing, analysis, and decision-making. However, they differ in their objectives, methodologies, and applications. Below is a detailed comparison of these four domains.

**1. Business Intelligence (BI)**



**Definition:**  
Business Intelligence refers to technologies, strategies, and practices used to collect, process, and analyze business data to support decision-making. BI primarily focuses on historical and current data to provide insights into business performance.

**Key Features:**

* Collects and processes structured data from various business systems.
* Uses reporting tools, dashboards, and visualization techniques.
* Helps in trend analysis, KPI (Key Performance Indicator) tracking, and business performance improvement.
* Focuses on descriptive analytics (what happened and why).

**Techniques Used:**

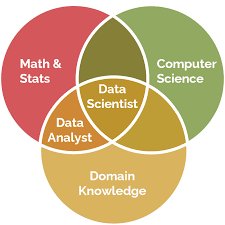
* ETL (Extract, Transform, Load) processes
* Data warehousing
* OLAP (Online Analytical Processing)
* SQL-based reporting
* Dashboards and visualization tools (Tableau, Power BI, QlikView)

**Use Case Example:**  
A retail company uses BI tools to analyze monthly sales trends, compare revenue across regions, and make data-driven marketing decisions.

**Advantages:**  
✔ Helps in real-time business monitoring.  
✔ Improves decision-making and operational efficiency.  
✔ User-friendly with drag-and-drop visualization tools.

**Limitations:**  
✖ Limited to structured data from internal business systems.  
✖ Focuses more on past and current data rather than predictive analytics.

**2. Data Science**



**Definition:**  
Data Science is a multidisciplinary field that extracts meaningful insights from structured and unstructured data using statistical techniques, machine learning models, and domain expertise. It is broader than BI and includes predictive and prescriptive analytics.

**Key Features:**

* Works with structured, semi-structured, and unstructured data.
* Uses statistical analysis, data mining, and machine learning.
* Provides predictive and prescriptive analytics.
* Requires expertise in programming languages like Python, R, and SQL.

**Techniques Used:**

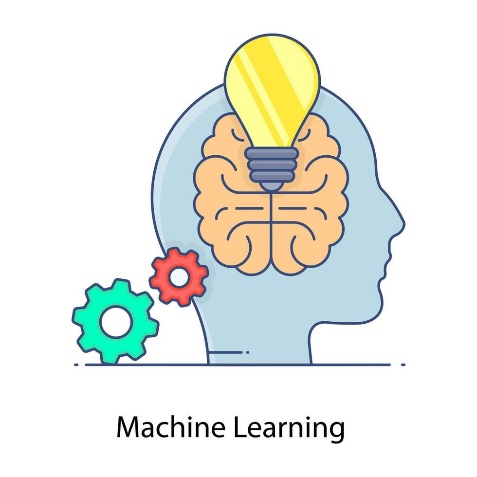
* Data cleaning and preprocessing
* Exploratory Data Analysis (EDA)
* Statistical modeling
* Data visualization
* Machine learning techniques (supervised and unsupervised learning)

**Use Case Example:**  
A telecom company uses data science to analyze customer usage patterns and predict churn rates, allowing them to take preventive action to retain customers.

**Advantages:**  
✔ Provides deeper insights by analyzing large datasets.  
✔ Supports predictive analytics for future trends.  
✔ Can handle both structured and unstructured data.

**Limitations:**  
✖ Requires technical expertise in programming and statistics.  
✖ Data processing and model training can be time-consuming.

**3. Machine Learning (ML)**



**Definition:**  
Machine Learning is a subset of AI that enables systems to learn from data, identify patterns, and make decisions with minimal human intervention. It focuses on improving model performance over time through continuous learning.

**Key Features:**

* Uses algorithms to detect patterns and make predictions.
* Improves automatically with more data and experience.
* Used in recommendation systems, fraud detection, and predictive maintenance.
* Supports supervised, unsupervised, and reinforcement learning methods.

**Techniques Used:**

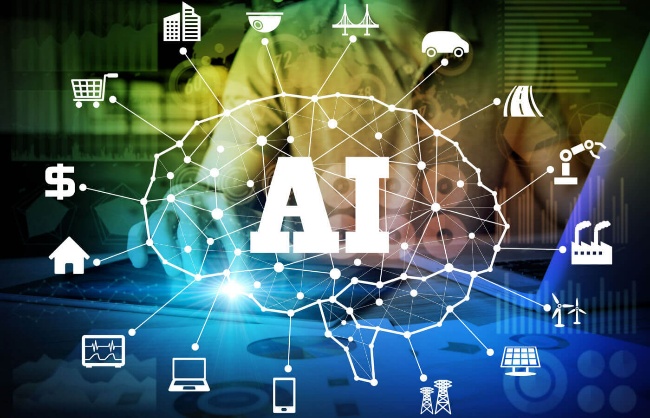
* Supervised learning (Regression, Decision Trees, SVM)
* Unsupervised learning (Clustering, Dimensionality Reduction)
* Reinforcement learning (Q-Learning, Deep Q-Networks)
* Neural networks and deep learning (CNN, RNN, Transformers)

**Use Case Example:**  
A bank uses machine learning models to detect fraudulent transactions by analyzing patterns in user behavior and flagging anomalies.

**Advantages:**  
✔ Automates decision-making without human intervention.  
✔ Continuously improves performance with more data.  
✔ Widely applicable across industries like healthcare, finance, and e-commerce.

**Limitations:**  
✖ Requires large datasets for effective training.  
✖ Can be computationally expensive.  
✖ Some models act as black boxes, making it difficult to interpret decisions.

**4. Artificial Intelligence (AI)**



**Definition:**  
Artificial Intelligence is a broader field that aims to create intelligent systems capable of reasoning, problem-solving, learning, and decision-making—just like humans. AI includes machine learning, natural language processing (NLP), robotics, and cognitive computing.

**Key Features:**

* Simulates human intelligence for automation.
* Includes expert systems, robotics, and NLP.
* Can process and analyze large volumes of unstructured data.
* Used in virtual assistants, self-driving cars, and AI-powered chatbots.

**Techniques Used:**

* Natural Language Processing (NLP) (Speech recognition, Sentiment analysis)
* Computer Vision (Image recognition, Object detection)
* Deep learning (Neural networks, Transformers)

**Use Case Example:**  
An e-commerce website uses AI-driven chatbots to handle customer queries, provide product recommendations, and improve user experience.

**Advantages:**  
✔ Enables automation of complex tasks.  
✔ Enhances accuracy and efficiency in data processing.  
✔ Continuously improves through learning and adaptation.

**Limitations:**  
✖ High computational and implementation costs.  
✖ Ethical concerns, such as data privacy and AI bias.  
✖ Requires extensive training and large datasets for deep learning models.

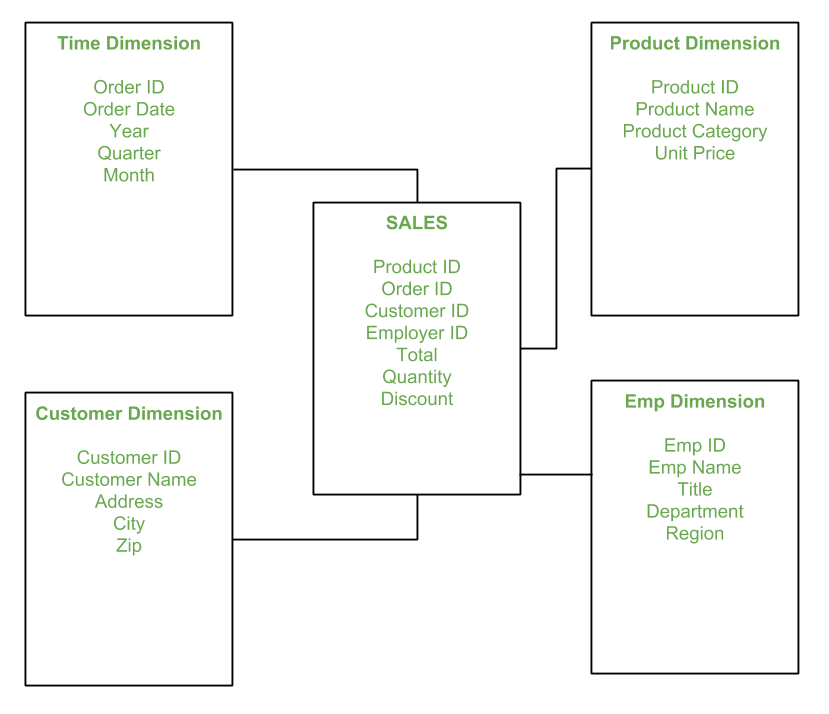
**Comparison Table**

| **Feature** | **Business Intelligence (BI)** | **Data Science** | **Machine Learning (ML)** | **Artificial Intelligence (AI)** |
| --- | --- | --- | --- | --- |
| **Definition** | Analyzes past and current data to improve decision-making. | Extracts insights from structured and unstructured data. | Uses algorithms to learn patterns from data and make predictions. | Simulates human intelligence to perform tasks autonomously. |
| **Focus** | Reporting, dashboards, and data visualization. | Data analysis, statistical modeling, and predictions. | Training models to detect patterns and improve over time. | Cognitive functions like reasoning, learning, and automation. |
| **Techniques Used** | SQL, OLAP, ETL, Data Warehousing. | Statistics, data mining, big data processing. | Supervised & unsupervised learning, deep learning. | NLP, robotics, computer vision, expert systems. |
| **Output** | Reports, dashboards, KPIs. | Predictions, insights, trend analysis. | Predictive models, recommendations. | Autonomous decision-making, automation. |
| **Example Use Case** | A company uses BI to analyze past sales and optimize future strategy. | A retailer analyzes customer behavior to predict demand. | A bank detects fraud by identifying suspicious transactions. | A chatbot provides customer support using AI-driven responses. |

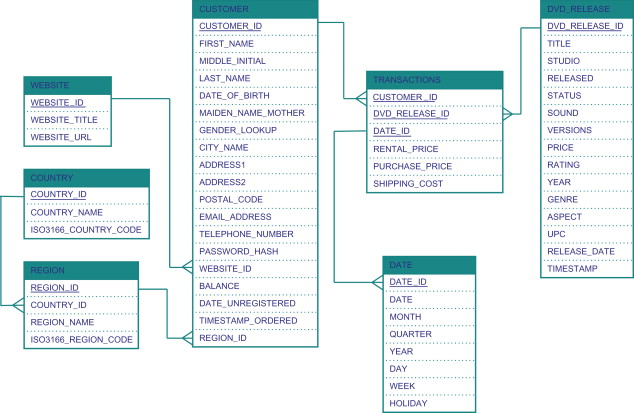
**Key Takeaways:**

* **BI** helps businesses analyze past data for reporting and decision-making.
* **Data Science** extracts insights from structured and unstructured data using statistical techniques and ML.
* **ML** allows systems to learn from data and make predictions without explicit programming.
* **AI** aims to simulate human intelligence, automating complex decision-making tasks.

**Star Schema**



**Snowflake Schema:**



**Conclusion:** A star schema organizes data in a simple, denormalized way for fast queries, while a snowflake schema normalizes data to save space but adds complexity. Business Intelligence (BI) analyzes past and present data to track performance, Data Science uncovers hidden patterns, Machine Learning (ML) predicts future trends, and AI automates complex decisions like a human would. Together, they turn raw data into actionable insights, helping businesses make smarter decisions efficiently.

**LO’s achieved:** LO1

**PO’s achieved:** PO1-PO5, PO8-PO10, PO12