Task 2:

What are the roles of Data analyst , data scientist, data engineer, business intelligence (BI) specialist and machine learning (ML) roles in the data pipeline?

Data Analyst:

- Gathers and cleans data to ensure accuracy and consistency.
- Analyzes data using statistical techniques and tools to derive insights.
- Creates dashboards, charts, and reports to represent data findings clearly.
- Identifies trends, patterns, and correlations to inform business decisions.
- Communicates results to stakeholders and provides actionable recommendations.

Data Scientist:

- Develops predictive and descriptive models using statistical and machine learning methods.
- Cleans, transforms, and structures data for use in modeling and analytics.
- Designs, tests, and optimizes algorithms or machine learning models.
- Works with stakeholders to solve complex problems through data-driven solutions.

Data Engineer:

- Designs and builds scalable data storage and data management systems.
- Develops and manages data ingestion, transformation, and integration pipelines.
- Implements Extract, Transform, Load (ETL) workflows for moving data across different systems.
- Ensures data systems perform efficiently and reliably under heavy workloads.
- Applies best practices for data governance, integrity, and security.

Business Intelligence (BI) Specialist:

- Manages data warehouse solutions for optimized storage and retrieval.
- Builds reports, dashboards, and other visual tools for business intelligence.
- Analyzes key performance indicators (KPIs) to assess business health.
- Prepares and transforms data to provide meaningful insights for business users.
- Works with decision-makers to understand their needs and provide tailored insights.

Machine Learning (ML) Roles in the Data Pipeline:

- Prepares and processes data for training, validation, and testing of models.
- Selects, creates, and transforms data features to improve model accuracy.
- Trains models using various algorithms and techniques on prepared data.
- Integrates and deploys trained models into production systems for real-world use.
- Monitors model performance over time and updates or retrains as needed to maintain accuracy.

Machine Learning Engineer:

- Designs, builds, and deploys ML systems. They optimize the training process, ensure scalability, and integrate models into production environments.