USN:1RVU23CSE325
NAME: PANCHAMI DINESH

Ex No: 1 Date: 13-08-25	Exploring the Data Engineering Lifecycle and Stakeholder Roles

Objective:

This lab provides hands-on experience exploring the data engineering lifecycle and understanding the roles of key stakeholders. Participants will simulate responsibilities of data engineers, data scientists, and business analysts while examining raw data sources and planning a data-driven solution.

Outcomes:

1. Identify and describe each stage of the data engineering lifecycle.
2. Explain the specific responsibilities of stakeholders across the lifecycle.
3. Collaborate to define a business problem using raw data sources.
4. Draft a requirements document based on the business use case.

Materials:

- Raw sales data CSV file (`sales_data_raw.csv`)
- Customer feedback JSON file (`customer_feedback.json`)
- Folder structure representing a mock data warehouse or data lake

Lab Procedure:

Stage 1: Problem Definition and Requirements Gathering (Business Analyst)

1. Review both datasets provided (`sales_data_raw.csv` and `customer_feedback.json`).
2. Formulate a business question, e.g., "What are the top 5 products by revenue, and how does customer sentiment vary for them?"
3. Identify required data points (e.g., `product_id`, `sale_price`, `customer_id`, `sentiment_score`).
4. Create a short requirements document outlining the problem, key metrics, and desired insights.

Stage 2: Role-Based Collaboration Simulation

1. Discuss and map out how the Data Engineer will ingest and clean the data.

USN:1RVU23CSE325
NAME: PANCHAMI DINESH

2. Identify how the Data Scientist will analyze and model insights based on the cleaned data.
3. Define how the Business Analyst will interpret and report results.
4. Define how each stakeholder contributes to the overall data solution.
5. Document the flow of responsibilities and dependencies between roles.

GitHub Link:

https://github.com/panchamidinesh/FDE1