**ETL Pipeline for the Iris Dataset**

**Objective:**

The objective of this task was to build an ETL pipeline for the Iris dataset, involving the extraction, transformation, and loading of the data for further analysis or machine learning purposes.

**Process Overview:**

The ETL pipeline for the Iris dataset was implemented using Python. The process involved the following key steps:

1. **Extract Phase:**
   * I loaded the Iris dataset using the load\_iris function from the sklearn.datasets module.
   * The data was then organized into a DataFrame, making it easier to manipulate and analyze.
2. **Transform Phase:**
   * Normalization: I normalized the feature data to ensure all features had a mean of 0 and a standard deviation of 1. This step is crucial for many machine learning algorithms to perform optimally.
   * Encoding Categorical Variables: The species column, initially in numeric form, was mapped to categorical names ('setosa', 'versicolor', 'virginica'). This mapping made the data more interpretable.
   * One-Hot Encoding: The categorical species data was converted into dummy/indicator variables. This encoding is essential for algorithms that require numerical input.
3. **Load Phase:**
   * I split the transformed dataset into training and testing sets, with 80% of the data used for training and 20% reserved for testing. This split is standard practice in machine learning to evaluate model performance.
   * The training and testing datasets were saved to CSV files. These files serve as a clean, normalized, and well-structured format for future use in analysis or machine learning tasks.

**Conclusion:**

The ETL pipeline successfully processed the Iris dataset, ensuring it was clean, normalized, and appropriately structured. By splitting the data into training and testing sets and saving them into CSV files, I have prepared the data for subsequent analysis or machine learning applications. This exercise demonstrated the practical application of data engineering concepts and the importance of each phase in the ETL process.