Objective: Design and implement an end-to-end data pipeline utilizing Apache Airflow to process and load data from a CSV file into PostgreSQL. The pipeline will follow the Extract, Transform, Load (ETL) methodology while ensuring data cleanliness and dimensional modeling.

Step 1: Extract Data:

- Read the CSV file from the URL: https://raw.githubusercontent.com/plotly/datasets/refs/heads/master/supermarket_Sales.csv.
- Create a Pandas DataFrame from the file.

Step 2: Transform Data:

- Implement data pre-processing (duplicate/missing data handling etc.) methodologies to clean the raw data.
- Load transformed data from Pandas dataframe to PostgreSQL

Step 3: Load Data:

Apply dimensional modeling to create fact and dimension tables from the cleaned dataset. Then load fact & dimension tables from transformed data.

Step 4: Airflow data pipelining Tasks:

For each step of the ETL process, create an individual task in Apache Airflow:

- ✓ Task 1: Extract data
- ✓ Task 2: Transform data
- ✓ Task 3: Load clean data into PostgreSQL
- ✓ Task 4: Create dimension tables & fact table
- ✓ Task 5: Load data into fact table

Step 5: Process scheduling:

Schedule the entire workflow to execute daily at 5:00 AM.