1)findspark-to initializes the spark environment

2)pyspark-to provide the library of pyspark

3) sparksession and functions from pyspark.sql-required for spark sql operations

4)os-for setting environment variables

5)Etl function is defined toperform the etl operation

6)Spark.jars configuration is use for PostgreSQL JDBC driver

7)Spark.read.format is used to load the csv into local path

8)The result\_df dataframe is written to a postgresql database using the jdbc format

9)The spark.stop() method is called to stop the sparksession and release the resources

10)The code imports the PythonOperator from the airflow.operators.python module

11)The PythonOperator is used texecute a Python callable as a task in Airflow DAG

12)The PythonOperator executes the etl\_function from the test1 module as a task in the dag

13)The DAG has a start date of June 29, 2023 (datetime(2023, 6, 29)) and is scheduled to run daily (schedule\_interval="@daily").

14)The DAG has one task named “etl\_task” defined using the PythonOperator. The task executes the etl\_function from the test1 module.