# Summary Explanation of the ETL Script

This Python script is an AWS Glue ETL (Extract, Transform, Load) job script designed to process sales data from an S3 bucket, transform the data using pandas, and load the result into an Amazon Redshift table. Below is a breakdown of its key components:

## 1. Parameters

The script retrieves three parameters using AWS Glue's getResolvedOptions:  
- INPUT\_PATH: S3 path of the input CSV file.  
- OUTPUT\_PATH: S3 path to store the transformed output.  
- REDSHIFT\_CONNECTION: Glue connection name for Redshift.

## 2. S3 Handling

It uses the Boto3 client to interact with S3 and extract the bucket name and object key from the input path.

## 3. Data Transformation

The script performs the following transformations on the data:  
- Converts product IDs to uppercase and strips spaces.  
- Normalizes customer IDs by removing hyphens and zero-padding to 10 digits.  
- Calculates total sales as unit\_price multiplied by quantity.  
- Parses order\_date as a datetime and adds a processed\_at timestamp.  
- Filters out records with null order\_date or non-positive total\_sale.

## 4. Loading to Redshift

Uses SQLAlchemy with connection details extracted from the Glue connection to load the transformed DataFrame into the 'sales\_fact' table in Redshift using batch inserts.

## 5. Output to S3

The transformed data is saved back to S3 with a filename containing a timestamp, under the specified OUTPUT\_PATH.

## 6. Execution

The script's main function ties all steps together: reading from S3, transforming data, writing to Redshift, and optionally saving the processed CSV back to S3.