

Table 1

File name	Description
Lambda_trig.txt	Script defining an AWS Lambda function that triggers AWS Glue jobs in response to S3 event triggers.
normalize.txt	Glue Script to normalize specified numerical columns in a dataset.
outlier.txt	Glue Script for detecting and removing outliers from a dataset based on defined thresholds.
clean.txt	Glue Script to clean datasets by dropping rows that contain any null or missing values.
group.txt	Glue Script to group records in a dataset by a specific key and aggregate counts.
merge.txt	Glue Script for joining two datasets based on a shared key.
clean_input.csv	Input CSV file for the clean job containing raw data with potential null or empty values.
clean_output.csv	Output CSV file from the clean job with null or empty values removed.
group_input.csv	Input CSV file for the group job, containing ungrouped data with common attributes.
group_output.csv	Output CSV file from the group job with data aggregated by specified keys.
merge_input1.csv merge_input2.csv	Initial CSV files for the merge job, containing disjoint datasets to be combined.
merge_output.csv	Output CSV file from the merge job where datasets have been combined.
normalize_input.csv	Input CSV file for the normalize job with data needing scaling to a common range.
normalize_output.csv	Output CSV file from the normalize job with columns scaled to a standard range.
outlier_input.csv	Input CSV file for the outlier job with data that includes potential anomalies.
outlier_output.csv	Output CSV file from the outlier job with anomalies identified and removed.