**Total workflow of the project –**

**Step 1 – Read and Upload Sales Data to AWS S3**

We read the UCI Online Retail Dataset from local Excel and uploaded it to an S3 bucket (raw/).  
This sets up our raw layer in the cloud for future processing.

**Step 2 – Read Data from S3 and Clean/Transform It in Python**

We pulled the file back from S3, removed nulls, fixed datatypes, added TotalPrice, and did minor EDA like time-based trend analysis.

**Step 3 – Save Cleaned Data to S3 and Local**

We saved the cleaned DataFrame as an Excel file both to your local system and back to S3 under the cleaned/ folder.

**Step 4 – Load Cleaned Data to Amazon Redshift**

We will create a Redshift cluster and load the cleaned file from S3 into a staging table using SQL COPY or Python.

**Step 5 – Model Data in Redshift Using dbt**

We’ll use dbt to build dimensional models (dim\_customer, dim\_product, fact\_sales) from the cleaned staging table in Redshift.

**Step 6 – Visualize Insights in Power BI or Tableau**

Connect to Redshift from Power BI and create dashboards for KPIs like monthly sales, top products, and customer trends.

Loading Data into Redshift from S3

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

After that connect the redshift to PowerBI .. for doing visualizations –

A screenshot of a computer

AI-generated content may be incorrect.

Reports generated with PowerBI –

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

Etc...etc…