ETL Project:

import csv

# Read sales data

with open("/content/sample\_data/sales(in)", 'r') as file:

    reader = csv.DictReader(file)

    salesData = list(reader)

# Read product data

with open("/content/sample\_data/Product(in).csv", 'r') as file:

    reader = csv.DictReader(file)

    productData = list(reader)

# Combine sales and product data

combined\_data = []

total\_sales = 0

for sale in salesData:

    for product in productData:

        if sale['product\_id'] == product['product\_id']:

            try:

                # Calculate revenue and accumulate total sales

                revenue = float(product['cost\_price']) \* int(sale['quantity'])

                total\_sales += revenue

                # Create combined data entry

                combined\_entry = {

                    'sales\_id': sale['sale\_id'],

                    'product\_id': sale['product\_id'],

                    'quantity\_sold': sale['quantity'],

                    'sale\_date': sale['date'],

                    'sale\_category': product['category'],

                    'product\_name': product['product\_name'],

                    'price': product['cost\_price']

                }

                combined\_data.append(combined\_entry)

# Write combined data to a new CSV file

with open("/content/sample\_data/combined\_datafile.csv","w",newline='') as file:

    writer = csv.DictWriter(file, ['sales\_id', 'product\_id', 'quantity\_sold', 'sale\_date', 'sale\_category', 'product\_name', 'price'])

    writer.writeheader()

    writer.writerows(combined\_data)

# Output confirmation and total revenue

print(f"Combined data written to {file}")

print(f"Total sales (revenue): ${total\_sales}")