Capstone Project: Customer Insights for a Retail Store

Case Study Overview:
A small retail store has been collecting transaction data in CSV files but wants to organize the data into a database for easier analysis. Your job is to **analyze the data using Pandas and Numpy (optional), normalize it into smaller tables, and load it into PostgreSQL using** psycopg2.

Steps & Deliverables

1. **Data Loading**
   - Read the CSV file(s) into Pandas DataFrames.
2. **Data Cleaning**
   - Handle missing values (e.g., drop or fill where necessary).
   - Ensure numeric columns (like Age, Amount, Zipcode) have the correct data type.
3. **Data Analysis with Pandas (and optional Numpy)**
   - Find the top 5 customers by total amount spent.
   - Calculate the average rating per product type.
   - (Optional, Numpy) Convert the ratings into a Numpy array and calculate the mean.
4. **Database Normalization**
   - Split the big DataFrame into 3 tables:
     - **Customers** (that hold all the data relating to the unique customers)
     - **Orders** (or Transactions. To hold transaction details.)
     - **Payments** (with the payment details. You would have to create an ID for this one)
5. **Load to PostgreSQL**
   - Create the tables in PostgreSQL.
   - Insert the cleaned/normalized data using psycopg2.

Deliverable

- A Jupyter Notebook (or .py file) with:
  - Your Pandas/Numpy analysis
  - Code to normalize the data into the 3 tables
  - SQL loading script using psycopg2
- A short README file (2–3 sentences) describing your approach.