Case study 1:

**Data Preprocessing for a Retail Dataset**

**Objective:** To understand the process of data preprocessing including data integration, data cleaning, and data transformation using a retail dataset. This case study will help students learn the practical aspects of preparing data for analysis, ensuring data quality, and deriving meaningful insights.

**Dataset:** A sample retail dataset containing information about transactions made by customers at a chain of stores. The dataset includes the following columns:

1. TransactionID: Unique identifier for each transaction
2. CustomerID: Unique identifier for each customer
3. ProductID: Unique identifier for each product
4. Quantity: Quantity of the product purchased
5. Price: Price of the product
6. Order Date: Date of the transaction
7. CustomerLocation: location of the customer

**Task:** Integrate the data from two separate tables: Transactions and Products. The Transactions table contains columns: TransactionID, CustomerID, ProductID, Quantity, Price, Date, and CustomerLocation. The Products table contains columns: ProductID, ProductName, and Category.

**Activity:** Merge the Transactions and Products tables on the ProductID column to create a unified dataset.

* + Perform the integration of the Transactions and Products tables and display the first 5 rows of the integrated dataset.
  + Write a Python code snippet to identify and remove duplicate records from the dataset.
  + How would you handle outliers in the Price column using the interquartile range (IQR) method?
  + Perform Min-Max normalization on the Quantity and Price columns and display the transformed dataset.
  + Apply one-hot encoding to the Category column and show the updated dataset.