Case study 2: **Data Preprocessing for a Payroll Dataset**

**Objective:** To understand the process of data preprocessing including data integration, data cleaning, and data transformation using a payroll 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 payroll dataset containing information about employee salaries and department details. The dataset includes the following columns:

1. EmployeeID: Unique identifier for each employee
2. Name: Name of the employee
3. DepartmentID: Unique identifier for each department
4. DepartmentName: Name of the department
5. Salary: Monthly salary of the employee
6. HireDate: Date of hiring
7. Location: Location of the employee

**Case Study Steps:**

**1. Data Integration**

**Task:** Integrate the data from two separate tables: Employees and Departments. The Employees table contains columns: EmployeeID, Name, DepartmentID, Salary, HireDate, and Location. The Departments table contains columns: DepartmentID, DepartmentName.

**Activity:** Merge the Employees and Departments tables on the DepartmentID column to create a unified dataset.

* + Merge the Employees and Departments tables and display the first 5 rows of the merged dataset.
  + Write a Python code snippet to remove duplicate records from the dataset.
  + Explain and apply the interquartile range (IQR) method to detect and handle outliers in the Salary column.
  + Perform Min-Max normalization on the Salary column and display the transformed dataset.
  + Apply one-hot encoding to the Location column and show the updated dataset.