
🧩 Scenario – Customers & Orders

Step 1: Create DataFrames

```
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName("DataFrame-Exercises").getOrCreate()

# Customers Data
customers_data = [
    (1, "Rahul Sharma", "Bangalore", 28),
    (2, "Priya Singh", "Delhi", 32),
    (3, "Aman Kumar", "Hyderabad", 25),
    (4, "Sneha Reddy", "Chennai", 35),
    (5, "Arjun Mehta", "Mumbai", 30),
    (6, "Divya Nair", "Delhi", 29)
]
customers_cols = ["customer_id", "name", "city", "age"]
customers_df = spark.createDataFrame(customers_data, customers_cols)

# Orders Data
orders_data = [
    (101, 1, "Laptop", 55000),
    (102, 2, "Mobile", 25000),
    (103, 1, "Headphones", 3000),
    (104, 3, "Chair", 5000),
    (105, 5, "Book", 700),
    (106, 2, "Tablet", 20000),
    (107, 6, "Shoes", 2500),
    (108, 7, "Camera", 30000)   # Order with non-existent customer
]
orders_cols = ["order_id", "customer_id", "product", "amount"]
orders_df = spark.createDataFrame(orders_data, orders_cols)

customers_df.show()
orders_df.show()
```

Step 2: Exercises (Operations on DataFrames)

Basic Operations

1. Select only `name` and `city` from customers.
2. Filter customers older than 30.
3. Count how many customers are from "Delhi".
4. Find distinct cities in the customer list.

Aggregations

5. Find the average age of customers.
6. Find the maximum and minimum order amount.
7. Count number of orders placed by each customer.
8. Calculate total spending of each customer.

Joins

9. Perform an **inner join** between customers and orders.
10. Perform a **left join** to show all customers (even without orders).
11. Find customers who have **never placed an order**.
12. Find orders that belong to **non-existent customers**.

Sorting & Grouping

13. List customers ordered by age (descending).
14. Show top 3 highest order amounts.
15. Group customers by city and find average age.
16. Group orders by product and find total sales amount.

SQL Operations

17. Register both DataFrames as temp views.
18. Write a SQL query to find total revenue by city.
19. Write a SQL query to list top 2 customers by total spend.
20. Write a SQL query to find all customers who bought products worth more than □20,000.
