



**33\_Hobbies Data Transformation**

Problem Statement:

Transform a dataset with individuals' names and associated hobbies into a new format using PySpark. Convert the comma-separated hobbies into separate rows, creating a DataFrame with individual rows for each person and their respective hobbies.

**Difficult Level :** EASY

**DataFrame:**

**# Sample input data**

**data = [("Alice", "badminton,tennis"),**

**("Bob", "tennis,cricket"),**

**("Julie", "cricket,carroms")]**

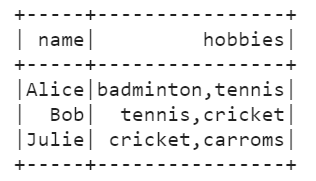
**# Create a DataFrame**

**df = spark.createDataFrame(data, ["name", "hobbies"])**

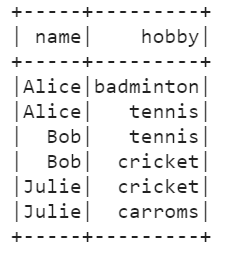
|  |
| --- |



**INPUT**



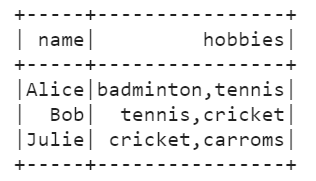
**OUTPUT**

****

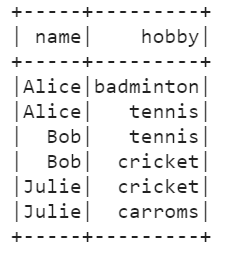


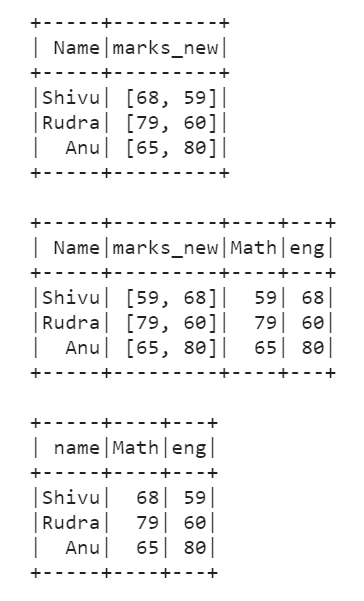
****

****



****

****

****

