



**28\_Join\_in\_pyspark**

Write a pyspark code perform below function

* Write down the query to fetch Project name assign to more than one Employee
* Get employee name, project name order by firstname from "EmployeeDetail" and"ProjectDetail" for those employee which have assigned project already.

**Difficult Level :** EASY

**DataFrame:**

**data = [**

**[1, "Vikas", "Ahlawat", 600000.0, "2013-02-15 11:16:28.290", "IT", "Male"],**

**[2, "nikita", "Jain", 530000.0, "2014-01-09 17:31:07.793", "HR", "Female"],**

**[3, "Ashish", "Kumar", 1000000.0, "2014-01-09 10:05:07.793", "IT", "Male"],**

**[4, "Nikhil", "Sharma", 480000.0, "2014-01-09 09:00:07.793", "HR", "Male"],**

**[5, "anish", "kadian", 500000.0, "2014-01-09 09:31:07.793", "Payroll", "Male"],**

**]**

**# Create a schema for the DataFrame**

**schema = StructType([**

**StructField("EmployeeID", IntegerType(), True),**

**StructField("First\_Name", StringType(), True),**

**StructField("Last\_Name", StringType(), True),**

**StructField("Salary", DoubleType(), True),**

**StructField("Joining\_Date", StringType(), True),**

**StructField("Department", StringType(), True),**

**StructField("Gender", StringType(), True)**

**])**

**pro\_schema = StructType([**

**StructField("Project\_DetailID", IntegerType(), True),**

**StructField("Employee\_DetailID", IntegerType(), True),**

**StructField("Project\_Name", StringType(), True)**

**])**

**# Create the data as a list of tuples**

**pro\_data = [**

**(1, 1, "Task Track"),**

**(2, 1, "CLP"),**

**(3, 1, "Survey Management"),**

**(4, 2, "HR Management"),**

**(5, 3, "Task Track"),**

**(6, 3, "GRS"),**

**(7, 3, "DDS"),**

**(8, 4, "HR Management"),**

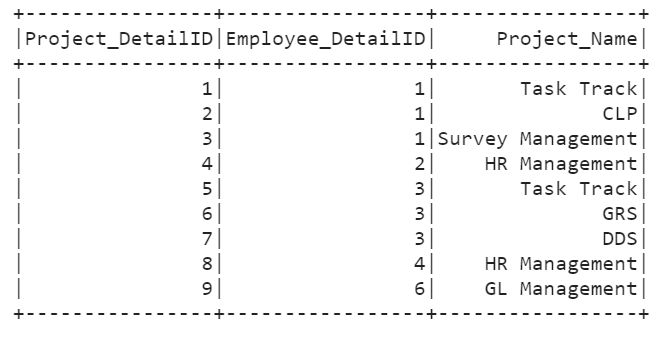
**(9, 6, "GL Management")**

**]**

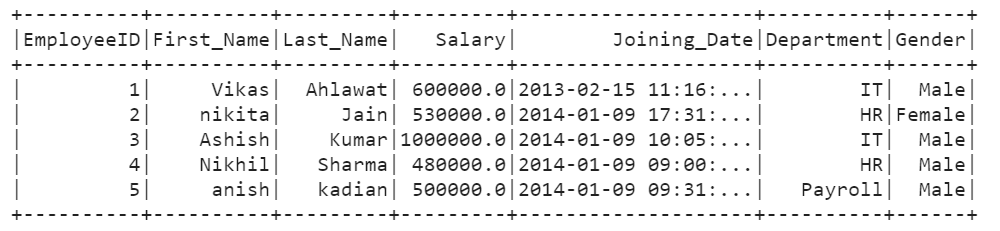


****

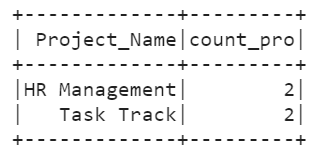
****

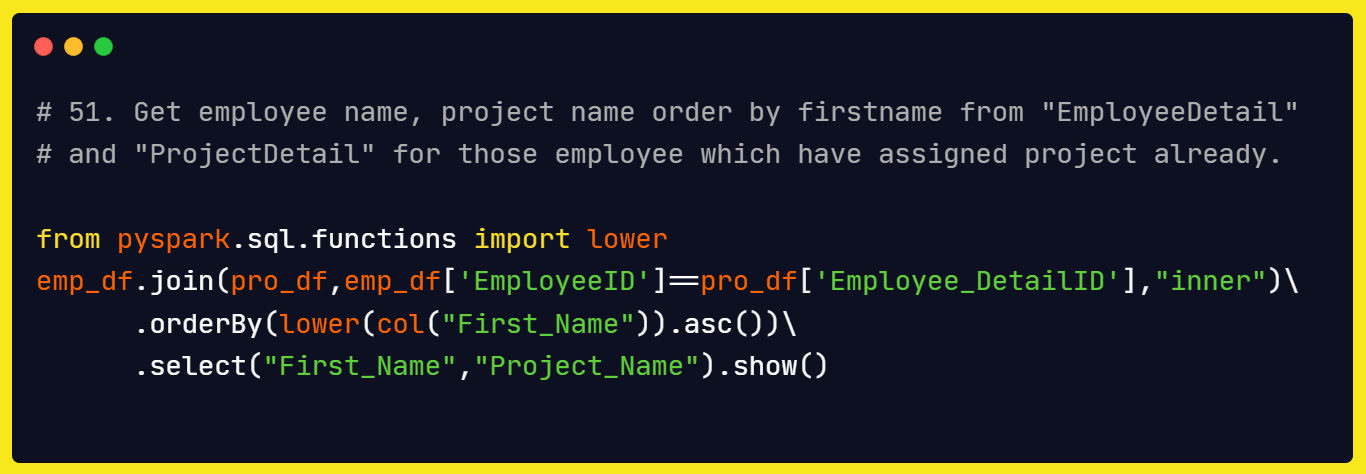
****

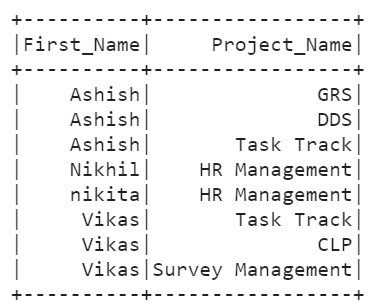
****

****

****

****

****

****

