



**06\_Customers Who Never Order**

Write a Pyspark program to find all customers who never order anything.

**Difficult Level :** EASY

**DataFrame:**

**# Define the schema for the "Customers"**

**customers\_schema = StructType([**

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

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

**])**

**# Define data for the "Customers"**

**customers\_data = [**

**(1, 'Joe'),**

**(2, 'Henry'),**

**(3, 'Sam'),**

**(4, 'Max')**

**]**

**# Define the schema for the "Orders"**

**orders\_schema = StructType([**

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

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

**])**

**# Define data for the "Orders"**

**orders\_data = [**

**(1, 3),**

**(2, 1)**

**]**



**INPUT**

| **INPUT -1 customers** | |
| --- | --- |
| **ID** | **NAME** |
| **1** | **Joe** |
| **2** | **Henry** |
| **3** | **Sam** |
| **4** | **Max** |

| **INPUT - 2 orders** | |
| --- | --- |
| **ID** | **CUSTOMERID** |
| **1** | **3** |
| **2** | **1** |

**OUTPUT**

| **OUTPUT** |
| --- |
| **NAME** |
| **Max** |
| **Henry** |



**# Creating Spark Session**

**from pyspark.sql import SparkSession**

**from pyspark.sql.types import StructType,StructField,IntegerType,StringType**

**#creating spark session**

**spark = SparkSession. \**

**builder. \**

**config('spark.shuffle.useOldFetchProtocol', 'true'). \**

**config('spark.ui.port','0'). \**

**config("spark.sql.warehouse.dir", "/user/itv008042/warehouse"). \**

**enableHiveSupport(). \**

**master('yarn'). \**

**getOrCreate()**

**customers\_schema = StructType([**

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

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

**])**

**# Define data for the "Customers"**

**customers\_data = [**

**(1, 'Joe'),**

**(2, 'Henry'),**

**(3, 'Sam'),**

**(4, 'Max')**

**]**

**# Define the schema for the "Orders"**

**orders\_schema = StructType([**

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

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

**])**

**# Define data for the "Orders"**

**orders\_data = [**

**(1, 3),**

**(2, 1)**

**]**

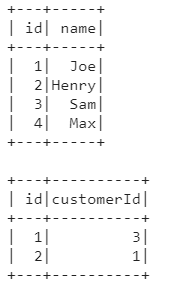
**# Create a PySpark DataFrame**

**cus\_df=spark.createDataFrame(customers\_data,customers\_schema)**

**ord\_df=spark.createDataFrame(orders\_data,orders\_schema)**

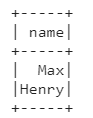
**cus\_df.show()**

**ord\_df.show()**

****

**cus\_df.join(ord\_df,cus\_df.id == ord\_df.customerId,"left\_anti")\**

**.select("name").show()**

****

