001,56,CSE
002,67,DSE
003,78,CCE
004,89,DSE
005,40,CSE
006,51,CCE
007,58,DSE
008,63,CCE
this is 1st dataset and another dataset is: 001,Rajiv,Reddy,21,9848022337,Hyderabad
002,siddarth,Battacharya,22,9848022338,Kolkata
003,Rajesh,Khanna,22,9848022339,Delhi
004,Preethi,Agarwal,21,9848022330,Pune
005,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar
006,Archana,Mishra,23,9848022335,Chennai
007,Komal,Nayak,24,9848022334,trivendram
008,Bharathi,Nambiayar,24,9848022333,Chennai
merge both the dataset and display CSE student name and marks
CODE:

```
// Define a case class for Dataset 1
case class Dataset1(id: String, marks: Int, department: String)

// Define a case class for Dataset 2
case class Dataset2(id: String, name: String, surname: String, age: Int, phone: String, city: String)

// Load Dataset 1 from input.txt
val data1 = spark.read.textFile("input.txt")
.map(_.split(","))
.map(attributes => Dataset1(attributes(0), attributes(1).toInt, attributes(2)))

// Load Dataset 2 from input1.txt
val data2 = spark.read.textFile("input1.txt")
.map(_.split(","))
.map(attributes => Dataset2(attributes(0), attributes(1), attributes(2), attributes(3).toInt,
attributes(4), attributes(5)))

// Convert RDDs to DataFrames
import spark.implicits._
val df1 = data1.toDF()
val df2 = data2.toDF()

// Perform the join operation
val mergedData = df1.join(df2, "id")

// Filter CSE students
val cseStudents = mergedData.filter($"department" === "CSE")
```

```
// Display CSE student names with their marks
cseStudents.select("name", "marks").show()

import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions.col

// Create a SparkSession
val spark = SparkSession.builder
.appName("MergeDatasets")
.getOrCreate()

// Define the data for the first dataset
val data1 = Seq(
("001", 56, "CSE"),
("002", 67, "DSE"),
("003", 78, "CCE"),
("004", 89, "DSE"),
("005", 40, "CSE"),
("006", 51, "CCE"),
("007", 58, "DSE"),
("008", 63, "CCE")
)

// Create DataFrame for the first dataset
val df1 = spark.createDataFrame(data1).toDF("regno", "marks", "department")

// Define the data for the second dataset
val data2 = Seq(
("001", "Rajiv", "Reddy", 21, "9848022337", "Hyderabad"),
("002", "Siddarth", "Battacharya", 22, "9848022338", "Kolkata"),
("003", "Rajesh", "Khanna", 22, "9848022339", "Delhi"),
("004", "Preethi", "Agarwal", 21, "9848022330", "Pune"),
("005", "Trupthi", "Mohanthy", 23, "9848022336", "Bhuwaneshwar"),
("006", "Archana", "Mishra", 23, "9848022335", "Chennai"),
("007", "Komal", "Nayak", 24, "9848022334", "Trivendram"),
("008", "Bharathi", "Nambiayar", 24, "9848022333", "Chennai")
)

// Create DataFrame for the second dataset
val df2 = spark.createDataFrame(data2).toDF("regno", "first_name", "last_name", "age",
"phone", "city")

// Merge the datasets
val merged_df = df1.join(df2, Seq("regno"))

// Filter students who scored more than 70
```

```
val high_score_students = merged_df.filter(col("marks") > 70)

// Display names of high-scoring students
high_score_students.select("first_name", "last_name").show()
```