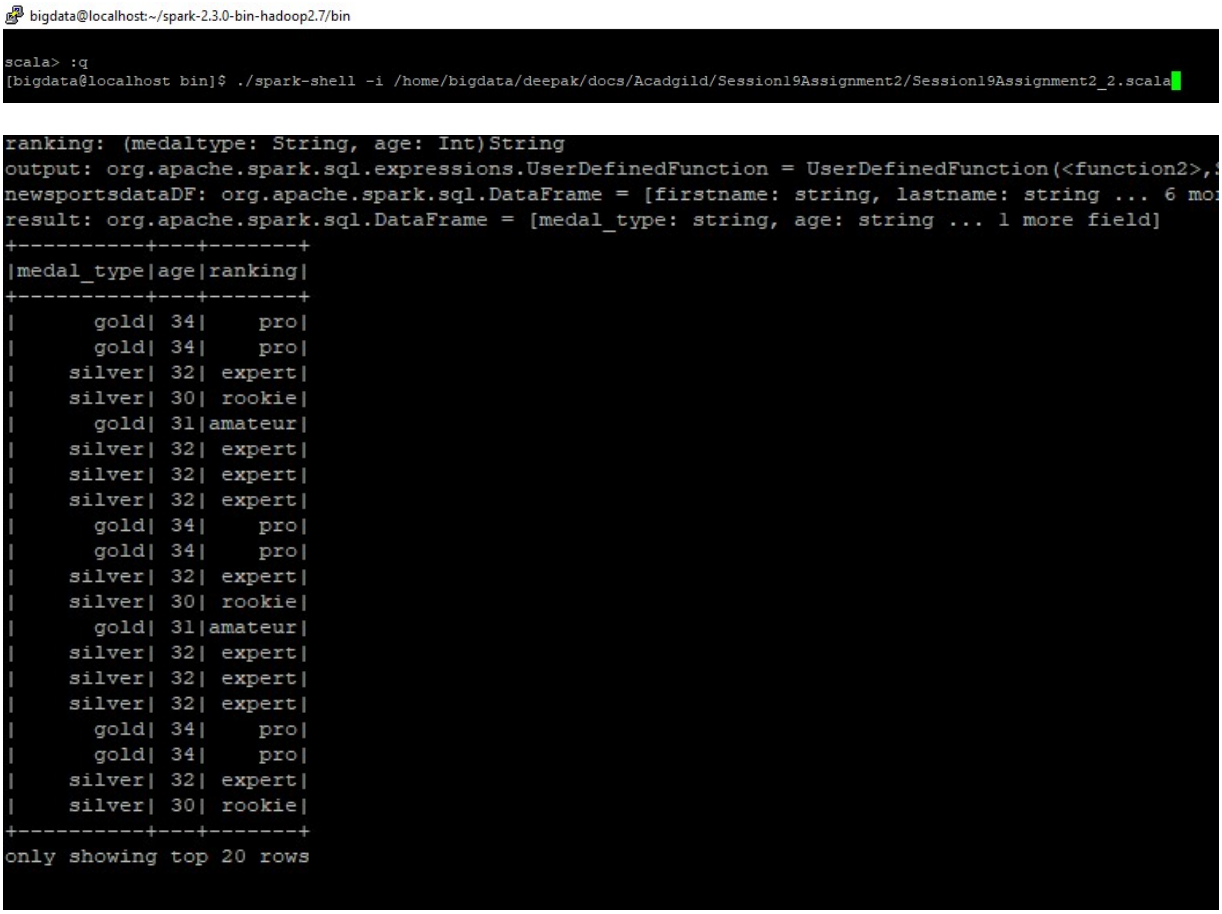
Session 19 Assignment 2

Problem 2 – To add a new column to dataframe using UDF



Code Snapshot

Session19Assignment2_2.scala | Session19Assignment2_1.scala | Session14Assignment1.scala

```
/* Session19 Assignment 2 - UDFs on Dataframe */

//Problem 2 - Adding a new column using UDF
import spark.implicits._

//Read data from input file
val sportsdata=spark.read.csv("/home/bigdata/deepak/docs/Acadgild/Session19Assignment2/Sports_data.txt").toDF
("firstname","lastname","sports","medal_type","age","year","country")

//Get the header
val header = sportsdata.first()

//Skip the header
val sportsdataDF = sportsdata.filter(line => line != header)

//UDF to decide the value of the new column - ranking
def ranking(medaltype: String, age: Int) = {
 if (medaltype=="gold" && age>=32) "pro"
 else if (medaltype=="gold" && age<=31) "amateur"
 else if (medaltype=="silver" && age>=32) "expert"
 else if (medaltype=="silver" && age<=31) "rookie"
 else ""
}

//Register the UDF
val output = udf(ranking(_:String,_:Int))

//Add the new column ranking using UDF
val newsportsdataDF=sportsdataDF.withColumn("ranking",output(sportsdataDF("medal_type"),sportsdataDF("age")))

//Select the new column in the output and display the output
val result=newsportsdataDF.select($"medal_type",$"age",$"ranking")
result.show
```