**TASK1**

1. Spark-shell –packages com.databricks:spark-csv\_2.11:1.3.0
2. Val sqlContext = new org.apache.spark.sql.SQLContext(sc)
3. Val sports\_data = sqlContext.read.format(“com.databricks.spark.csv”).option(“header”, “true”).option(“inferSchema”, “true”).load(“sports\_data.txt”)
4. Sports\_data.registerTempTable(“sportsdata”)
5. Sports\_data.printSchema()
6. Var gold\_medals = sqlContext.sql(“select year, count(\*) from sportsdata where medal\_type = ‘gold’ group by year”)

A screenshot of a social media post

Description generated with very high confidence

How many silver medals have been won by the USA in each sport?

Var silver\_usa = sqlContext.sql(“select sports, count(\*) from sportsdata where medal\_type = ‘silver’ and country = ‘USA’ group by sports)

Print(silver\_usa.show())

A screenshot of a social media post

Description generated with very high confidence

**TASK 2**

1. sqlContext.udf.register(“createnewname”, (colA: String, colB: String) { “Mr.” + colA.substring(0,2) + colB })
2. sqlContext.sql(“select firstname, lastname, createnewname(firstname, lastname) from sportsdata”).show()

A screenshot of a social media post

Description generated with very high confidence

Add new column with category names for athletes based on medals earned and age.

1. Val sqlContext = new org.apache.spark.sql.SQLContext(sc)
2. Val data = sc.textFile(“/home/acadgild/Sports\_data.txt”)
3. Val header = data.first()
4. Data1 = data.filter(row=>row != header)
5. Case class sports\_class(firstname:String, sports:String, medal\_type: String, age: Int, year: Int, country: String)
6. Val sports = data1.map(x=>x.split(“,”)).map(x=>sports\_class(x(0), x(1), x(2), x(3), x(4).toInt, x(5).toInt, x(6))).toDF
7. Sports.registerTempTable(“Sports”)
8. Val sports1 = sqlContext.sql(“select \*, IF((medal\_type = ‘gold’ and age >= 32), ‘pro’, IF((medal\_type = ‘gold’ and age <= 31), ‘amateur’, IF((medal\_type = ‘silver’ and age >= 32), ‘expert’, IF(medal\_type = ‘silver’ and age <= 31), ‘rookie’, ‘none’))))) as category from sports”)
9. Sports1.registerTempTable(“sports1”)
10. sqlContext.sql(“select firstname, category from sports1”).show()A screenshot of a social media post

    Description generated with very high confidence