**ASSIGNMENT-19.2**

Using udfs on dataframe

1. Change firstname, lastname columns into

Mr.first\_two\_letters\_of\_firstname<space>lastname

for example - michael, phelps becomes Mr.mi phelps

2. Add a new column called ranking using udfs on dataframe, where :

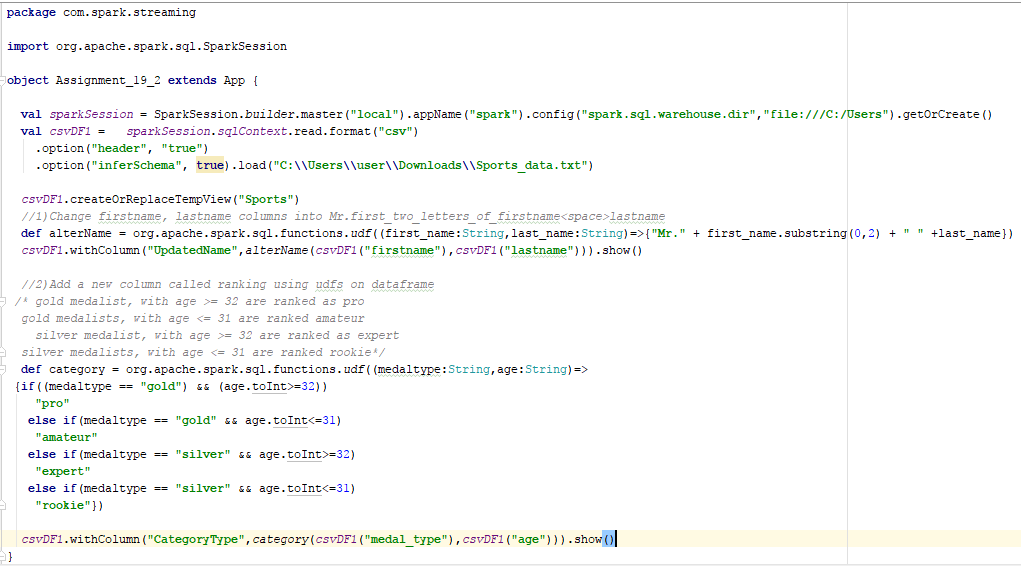
gold medalist, with age >= 32 are ranked as pro

gold medalists, with age <= 31 are ranked amateur

silver medalist, with age >= 32 are ranked as expert

silver medalists, with age <= 31 are ranked rookie

**PROGRAM:**

****

**RESULT:**

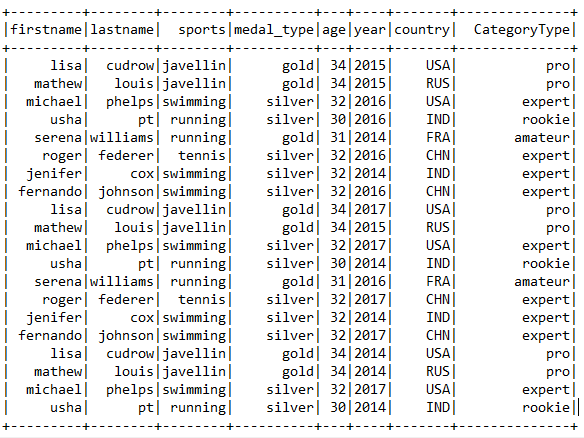
**QUERY:**

**def** alterName = org.apache.spark.sql.functions.*udf*((first\_name:String,last\_name:String)=>{**"Mr."** + first\_name.substring(0,2) + **" "** +last\_name})  
*csvDF1*.withColumn(**"UpdatedName"**,*alterName*(*csvDF1*(**"firstname"**),*csvDF1*(**"lastname"**))).show()

****

**QUERY:**

**def** category = org.apache.spark.sql.functions.*udf*((medaltype:String,age:String)=>  
{**if**((medaltype == **"gold"**) && (age.toInt>=32)) **"pro"  
 else if**(medaltype == **"gold"** && age.toInt<=31) **"amateur"  
 else if**(medaltype == **"silver"** && age.toInt>=32) **"expert"  
 else if**(medaltype == **"silver"** && age.toInt<=31) **"rookie"**})  
  
 *csvDF1*.withColumn(**"CategoryType"**,*category*(*csvDF1*(**"medal\_type"**),*csvDF1*(**"age"**))).show()



*Attached is the file for reference:*

**