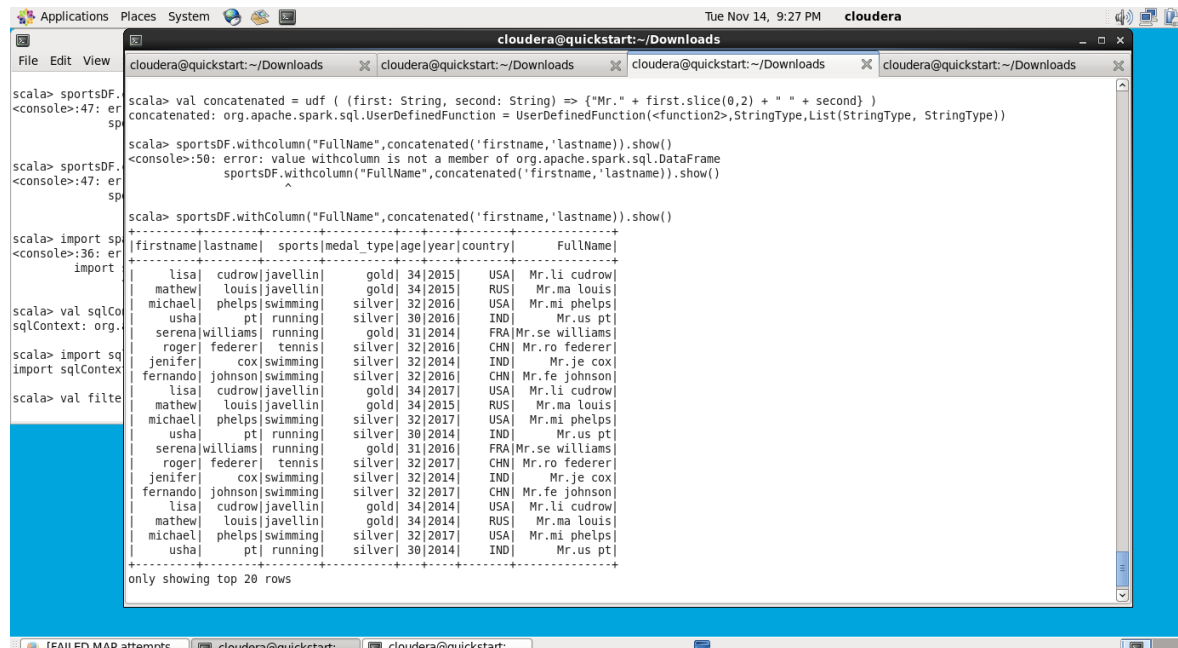


Using udfs on dataframe

1. Change firstname, lastname columns into

Mr.first_two_letters_of_firstname<space>lastname

for example - michael, phelps becomes Mr.mi phelps



```
scala> sportsDF.  
<console>:47: er  
sp  
scala> val concatenated = udf ( (first: String, second: String) => {"Mr." + first.slice(0,2) + " " + second} )  
concatenated: org.apache.spark.sql.UserDefinedFunction = UserDefinedFunction(<function2>,StringType,List(StringType, StringType))  
scala> sportsDF.withColumn("FullName",concatenated('firstname,'lastname)).show()  
<console>:50: error: value withColumn is not a member of org.apache.spark.sql.DataFrame  
sportsDF.withColumn("FullName",concatenated('firstname,'lastname)).show()  
^  
scala> sportsDF.withColumn("FullName",concatenated('firstname,'lastname')).show()  
+-----+-----+-----+-----+-----+-----+-----+-----+  
|firstname|lastname|sports|medal_type|age|year|country|FullName|  
+-----+-----+-----+-----+-----+-----+-----+-----+  
|lisa|cudrow|javelin|gold|34|2015|USA|Mr.li cudrow|  
|matthew|louis|javelin|gold|34|2015|RUS|Mr.ma louis|  
|michael|phelps|swimming|silver|32|2016|USA|Mr.mi phelps|  
|usha|pt|running|silver|30|2016|IND|Mr.us pt|  
|serena|williams|running|gold|31|2014|FRA|Mr.se williams|  
|roger|federer|tennis|silver|32|2016|CHN|Mr.ro federer|  
|jenifer|cox|swimming|silver|32|2014|IND|Mr.je cox|  
|fernando|johnson|swimming|silver|32|2016|CHN|Mr.fe johnson|  
|lisa|cudrow|javelin|gold|34|2017|USA|Mr.li cudrow|  
|matthew|louis|javelin|gold|34|2015|RUS|Mr.ma louis|  
|michael|phelps|swimming|silver|32|2017|USA|Mr.mi phelps|  
|usha|pt|running|silver|30|2014|IND|Mr.us pt|  
|serena|williams|running|gold|31|2016|FRA|Mr.se williams|  
|roger|federer|tennis|silver|32|2017|CHN|Mr.ro federer|  
|jenifer|cox|swimming|silver|32|2014|IND|Mr.je cox|  
|fernando|johnson|swimming|silver|32|2017|CHN|Mr.fe johnson|  
|lisa|cudrow|javelin|gold|34|2014|USA|Mr.li cudrow|  
|matthew|louis|javelin|gold|34|2014|RUS|Mr.ma louis|  
|michael|phelps|swimming|silver|32|2017|USA|Mr.mi phelps|  
|usha|pt|running|silver|30|2014|IND|Mr.us pt|  
+-----+-----+-----+-----+-----+-----+-----+-----+  
only showing top 20 rows
```

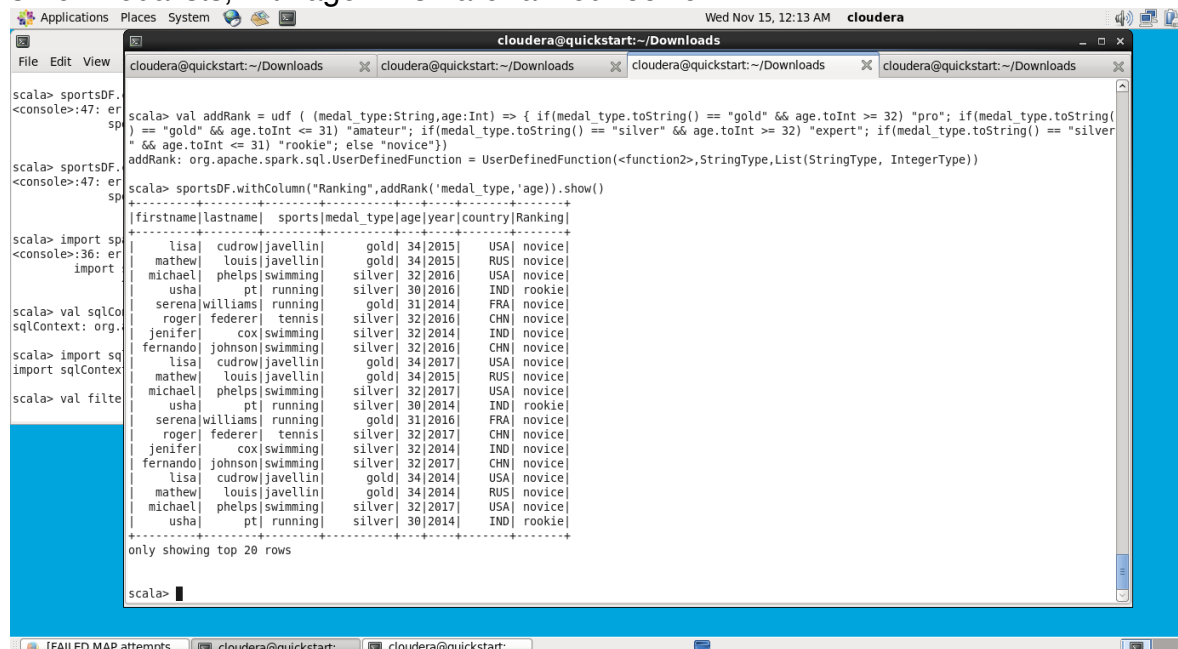
2. Add a new column called ranking using udfs on dataframe, where :

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



```
scala> sportsDF.  
<console>:47: er  
sp  
scala> val addRank = udf ( (medal_type:String,age:Int) => { if(medal_type.toString() == "gold" && age.toInt >= 32) "pro"; if(medal_type.toString()  
) == "gold" && age.toInt <= 31) "amateur"; if(medal_type.toString() == "silver" && age.toInt >= 32) "expert"; if(medal_type.toString() == "silver  
&& age.toInt <= 31) "rookie"; else "novice"} )  
addRank: org.apache.spark.sql.UserDefinedFunction = UserDefinedFunction(<function2>,StringType,List(StringType, IntegerType))  
scala> sportsDF.withColumn("Ranking",addRank('medal_type,'age')).show()  
+-----+-----+-----+-----+-----+-----+-----+-----+  
|firstname|lastname|sports|medal_type|age|year|country|Ranking|  
+-----+-----+-----+-----+-----+-----+-----+-----+  
|lisa|cudrow|javelin|gold|34|2015|USA|novice|  
|matthew|louis|javelin|gold|34|2015|RUS|novice|  
|michael|phelps|swimming|silver|32|2016|USA|novice|  
|usha|pt|running|silver|30|2016|IND|rookie|  
|serena|williams|running|gold|31|2014|FRA|novice|  
|roger|federer|tennis|silver|32|2016|CHN|novice|  
|jenifer|cox|swimming|silver|32|2014|IND|novice|  
|fernando|johnson|swimming|silver|32|2016|CHN|novice|  
|lisa|cudrow|javelin|gold|34|2017|USA|novice|  
|matthew|louis|javelin|gold|34|2015|RUS|novice|  
|michael|phelps|swimming|silver|32|2017|USA|novice|  
|usha|pt|running|silver|30|2014|IND|rookie|  
|serena|williams|running|gold|31|2016|FRA|novice|  
|roger|federer|tennis|silver|32|2017|CHN|novice|  
|jenifer|cox|swimming|silver|32|2014|IND|novice|  
|fernando|johnson|swimming|silver|32|2017|CHN|novice|  
|lisa|cudrow|javelin|gold|34|2014|USA|novice|  
|matthew|louis|javelin|gold|34|2014|RUS|novice|  
|michael|phelps|swimming|silver|32|2017|USA|novice|  
|usha|pt|running|silver|30|2014|IND|rookie|  
+-----+-----+-----+-----+-----+-----+-----+-----+  
only showing top 20 rows  
  
scala>
```