

Ans-1)

```
Hue - File Browser x ritisharmateksystems@ip- x Practice Labs x Webex Training x Practice Labs x +
slbdlh-webconsole.corestack.io
scala> case class Bank(age:Int, job:String, marital:String, education:String, dft:String, balance:Int,
| housing:String, loan:String, contact:String, day:Int, month: String, duration:Int, campaign:Int, pdays:Int,
| previous:Int, poutcome:String, y:String)
defined class Bank

scala> val bankrdd = bfields.map(
| x => Bank(x(0).replaceAll("\\","").toInt, x(1).replaceAll("\\",""), x(2).replaceAll("\\",""), x(3).replaceAll("\\",""), x(4).replaceAll("\\",""),
x(5).toInt, x(6).replaceAll("\\",""),
| x(7).replaceAll("\\",""), x(8).replaceAll("\\",""), x(9).toInt, x(10).replaceAll("\\",""), x(11).toInt, x(12).toInt, x(13).toInt, x(14).toInt, x(15).replaceAll("\\",""),
| x(16).replaceAll("\\","")) )
bankrdd: org.apache.spark.rdd.RDD[Bank] = MapPartitionsRDD[4] at map at <console>:27

scala> val df = bankrdd.toDF()
22/05/27 05:24:27 WARN lineage.LineageWriter: Lineage directory /var/log/spark/lineage doesn't exist or is not writable. Lineage for this application will
be disabled.
df: org.apache.spark.sql.DataFrame = [age: int, job: string ... 15 more fields]

scala> df.registerTempTable("bank")
warning: there was one deprecation warning; re-run with -deprecation for details

scala> val sqlContext = new org.apache.spark.sql.SQLContext(sc)
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@edc900a

scala> spark.conf.set("spark.sql.crossJoin.enabled", "true")

scala> df.show()
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|age|      job| marital|education|dft|balance|housing|loan|contact|day|month|duration|campaign|pdays|previous|poutcome| y|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 58| management| married| tertiary| no| 2143| yes| no| unknown| 5| may| 261| 1| -1| 0| unknown| no|
| 44| technician| single| secondary| no| 29| yes| no| unknown| 5| may| 151| 1| -1| 0| unknown| no|
```

```
Hue - File Browser x ritisharmateksystems@ip- x Practice Labs x Viewing Dileep (Trainer)'s desktop x Practice Labs x +
slbdlh-webconsole.corestack.io
scala> df.show()
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|age|      job| marital|education|dft|balance|housing|loan|contact|day|month|duration|campaign|pdays|previous|poutcome| y|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 58| management| married| tertiary| no| 2143| yes| no| unknown| 5| may| 261| 1| -1| 0| unknown| no|
| 44| technician| single| secondary| no| 29| yes| no| unknown| 5| may| 151| 1| -1| 0| unknown| no|
| 33| entrepreneur| married| secondary| no| 2| yes| yes| unknown| 5| may| 76| 1| -1| 0| unknown| no|
| 47| blue-collar| married| unknown| no| 1506| yes| no| unknown| 5| may| 92| 1| -1| 0| unknown| no|
| 33| unknown| single| unknown| no| 1| no| no| unknown| 5| may| 198| 1| -1| 0| unknown| no|
| 35| management| married| tertiary| no| 231| yes| no| unknown| 5| may| 139| 1| -1| 0| unknown| no|
| 28| management| single| tertiary| no| 447| yes| yes| unknown| 5| may| 217| 1| -1| 0| unknown| no|
| 42| entrepreneur| divorced| tertiary| yes| 2| yes| no| unknown| 5| may| 380| 1| -1| 0| unknown| no|
| 58| retired| married| primary| no| 121| yes| no| unknown| 5| may| 50| 1| -1| 0| unknown| no|
| 43| technician| single| secondary| no| 593| yes| no| unknown| 5| may| 55| 1| -1| 0| unknown| no|
| 41| admin.| divorced| secondary| no| 270| yes| no| unknown| 5| may| 222| 1| -1| 0| unknown| no|
| 29| admin.| single| secondary| no| 390| yes| no| unknown| 5| may| 137| 1| -1| 0| unknown| no|
| 53| technician| married| secondary| no| 6| yes| no| unknown| 5| may| 517| 1| -1| 0| unknown| no|
| 58| technician| married| unknown| no| 71| yes| no| unknown| 5| may| 71| 1| -1| 0| unknown| no|
| 57| services| married| secondary| no| 162| yes| no| unknown| 5| may| 174| 1| -1| 0| unknown| no|
| 51| retired| married| primary| no| 229| yes| no| unknown| 5| may| 353| 1| -1| 0| unknown| no|
| 45| admin.| single| unknown| no| 13| yes| no| unknown| 5| may| 98| 1| -1| 0| unknown| no|
| 57| blue-collar| married| primary| no| 52| yes| no| unknown| 5| may| 38| 1| -1| 0| unknown| no|
| 60| retired| married| primary| no| 60| yes| no| unknown| 5| may| 219| 1| -1| 0| unknown| no|
| 33| services| married| secondary| no| 0| yes| no| unknown| 5| may| 54| 1| -1| 0| unknown| no|
only showing top 20 rows

scala> val sqlContext = new org.apache.spark.sql.SQLContext(sc)
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@14d7d1a8
```

Ans-2a)

```
Hue - File Browser x ritisharmateksystems@ip- x Practice Labs Viewing Deep (Trainer)'s desktop x Practice Labs x +
slbdh-webconsole.corestack.io
58| retired| married| primary| no| 121| yes| no| unknown| 5| may| 50| 1| -1| 0| unknown| no|
43| technician| single| secondary| no| 593| yes| no| unknown| 5| may| 55| 1| -1| 0| unknown| no|
41| admin.| divorced| secondary| no| 270| yes| no| unknown| 5| may| 222| 1| -1| 0| unknown| no|
29| admin.| single| secondary| no| 390| yes| no| unknown| 5| may| 137| 1| -1| 0| unknown| no|
53| technician| married| secondary| no| 6| yes| no| unknown| 5| may| 517| 1| -1| 0| unknown| no|
58| technician| married| unknown| no| 71| yes| no| unknown| 5| may| 71| 1| -1| 0| unknown| no|
57| services| married| secondary| no| 162| yes| no| unknown| 5| may| 174| 1| -1| 0| unknown| no|
51| retired| married| primary| no| 229| yes| no| unknown| 5| may| 353| 1| -1| 0| unknown| no|
45| admin.| single| unknown| no| 13| yes| no| unknown| 5| may| 98| 1| -1| 0| unknown| no|
57| blue-collar| married| primary| no| 52| yes| no| unknown| 5| may| 38| 1| -1| 0| unknown| no|
60| retired| married| primary| no| 60| yes| no| unknown| 5| may| 219| 1| -1| 0| unknown| no|
33| services| married| secondary| no| 0| yes| no| unknown| 5| may| 54| 1| -1| 0| unknown| no|
+-----+
only showing top 20 rows

scala> val sqlContext = new org.apache.spark.sql.SQLContext(sc)
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@14d7d1a8

scala> spark.conf.set("spark.sql.crossJoin.enabled", "true")

scala> val success = sqlContext.sql("select (a.subscribed/b.total)*100 as success_percent from (select count(*) as subscribed from bank where y='yes') a,
(select count(*) as total from bank) b").show()
+-----+
| success_percent|
+-----+
|11.698480458295547|
+-----+

success: Unit = ()

scala>
```

Ans-2b)

```
Hue - File Browser x ritisharmateksystems@ip- x Practice Labs Viewing Deep (Trainer)'s desktop x Practice Labs x +
slbdh-webconsole.corestack.io
60| retired| married| primary| no| 60| yes| no| unknown| 5| may| 219| 1| -1| 0| unknown| no|
33| services| married| secondary| no| 0| yes| no| unknown| 5| may| 54| 1| -1| 0| unknown| no|
+-----+
only showing top 20 rows

scala> val sqlContext = new org.apache.spark.sql.SQLContext(sc)
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@14d7d1a8

scala> spark.conf.set("spark.sql.crossJoin.enabled", "true")

scala> val success = sqlContext.sql("select (a.subscribed/b.total)*100 as success_percent from (select count(*) as subscribed from bank where y='yes') a,
(select count(*) as total from bank) b").show()
+-----+
| success_percent|
+-----+
|11.698480458295547|
+-----+

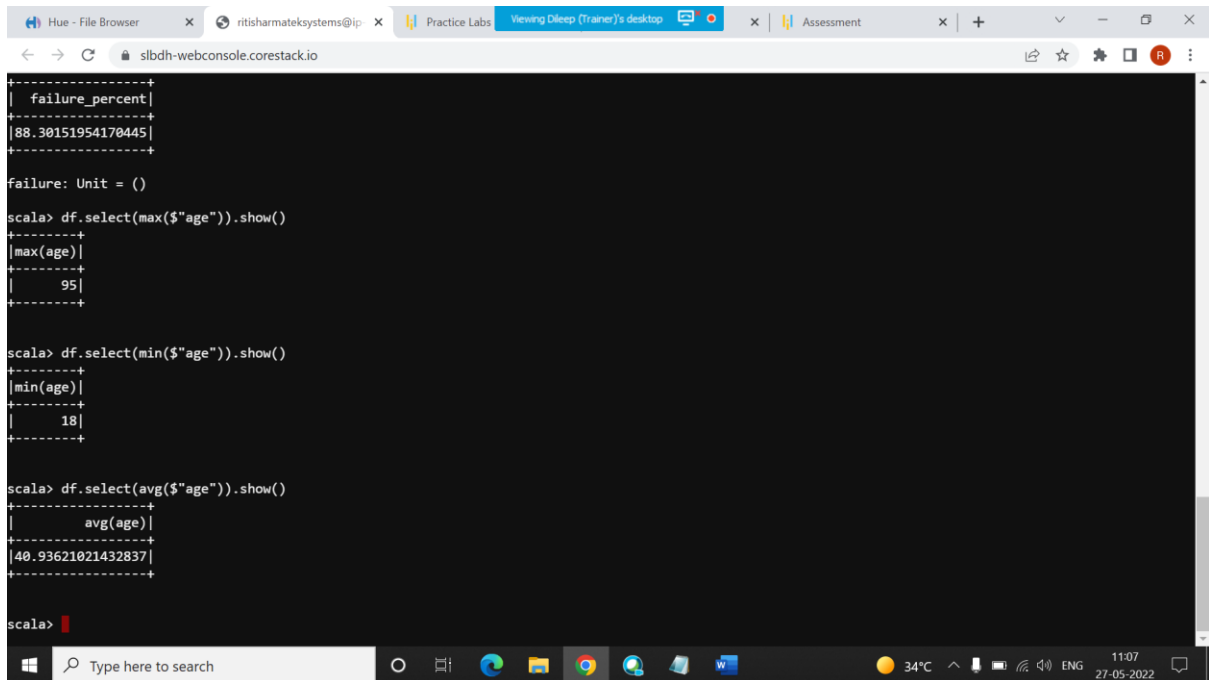
success: Unit = ()

scala> val failure = sqlContext.sql("select (a.not_subscribed/b.total)*100 as failure_percent from (select count(*) as not_subscribed from bank where y='
no') a,(select count(*) as total from bank) b").show()
+-----+
| failure_percent|
+-----+
|88.30151954170445|
+-----+

failure: Unit = ()

scala>
```

Ans:1)



The screenshot shows a web browser window with multiple tabs. The active tab is titled "Viewing Deep (Trainer)'s desktop" and displays a Scala REPL interface. The code executed in the console includes:

```
failure_percent|
+-----+
|88.30151954170445|
+-----+

failure: Unit = ()

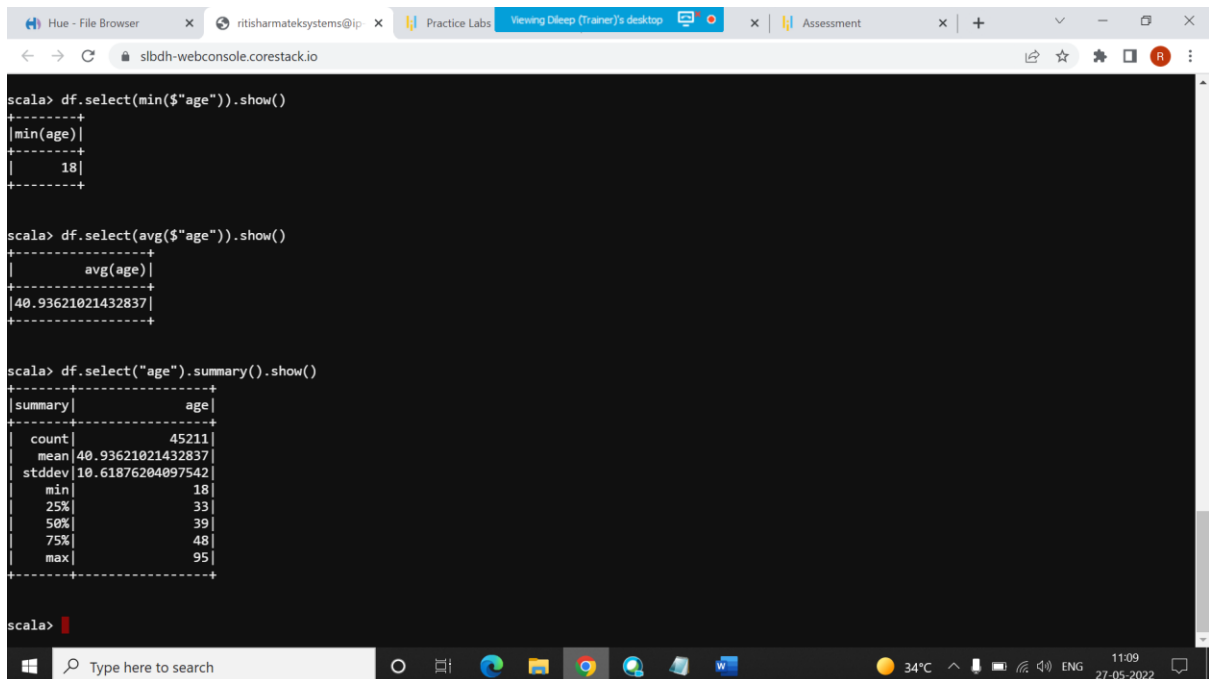
scala> df.select(max($"age")).show()
+-----+
|max(age)|
+-----+
|      95|
+-----+

scala> df.select(min($"age")).show()
+-----+
|min(age)|
+-----+
|      18|
+-----+

scala> df.select(avg($"age")).show()
+-----+
| avg(age) |
+-----+
|40.93621021432837|
+-----+

scala>
```

The bottom of the image shows a Windows taskbar with a search bar and system tray icons.



The screenshot shows the same web browser window as above, but with additional code executed in the Scala REPL:

```
scala> df.select(min($"age")).show()
+-----+
|min(age)|
+-----+
|      18|
+-----+

scala> df.select(avg($"age")).show()
+-----+
| avg(age) |
+-----+
|40.93621021432837|
+-----+

scala> df.select("age").summary().show()
+-----+
|summary|      age|
+-----+
|count|      45211|
|mean|40.93621021432837|
|stddev|10.61876204097542|
|min|      18|
|25%|      33|
|50%|      39|
|75%|      48|
|max|      95|
+-----+

scala>
```

The bottom of the image shows the same Windows taskbar as above.

Ans:2)

```
Hue - File Browser x ritisharmateksystems@ip- Practice Labs Viewing Dileep (Trainer)'s desktop x Assessment x + - x
slbdh-webconsole.corestack.io
scala> df.select(avg($"age")).show()
+-----+
|      avg(age)      |
+-----+
|40.93621021432837|
+-----+

scala> df.select("age").summary().show()
+-----+
|summary|      age|
+-----+
|count|      45211|
|mean|40.93621021432837|
|stddev|10.61876204097542|
|min|      18|
|25%|      33|
|50%|      39|
|75%|      48|
|max|      95|
+-----+

scala> df.select(avg($"balance")).show()
+-----+
|      avg(balance)  |
+-----+
|1362.2720576850766|
+-----+

scala>
```

```
Hue - File Browser x ritisharmateksystems@ip- Practice Labs Viewing Dileep (Trainer)'s desktop x Assessment x + - x
slbdh-webconsole.corestack.io
+-----+
|summary|      age|
+-----+
|count|      45211|
|mean|40.93621021432837|
|stddev|10.61876204097542|
|min|      18|
|25%|      33|
|50%|      39|
|75%|      48|
|max|      95|
+-----+

scala> df.select(avg($"balance")).show()
+-----+
|      avg(balance)  |
+-----+
|1362.2720576850766|
+-----+

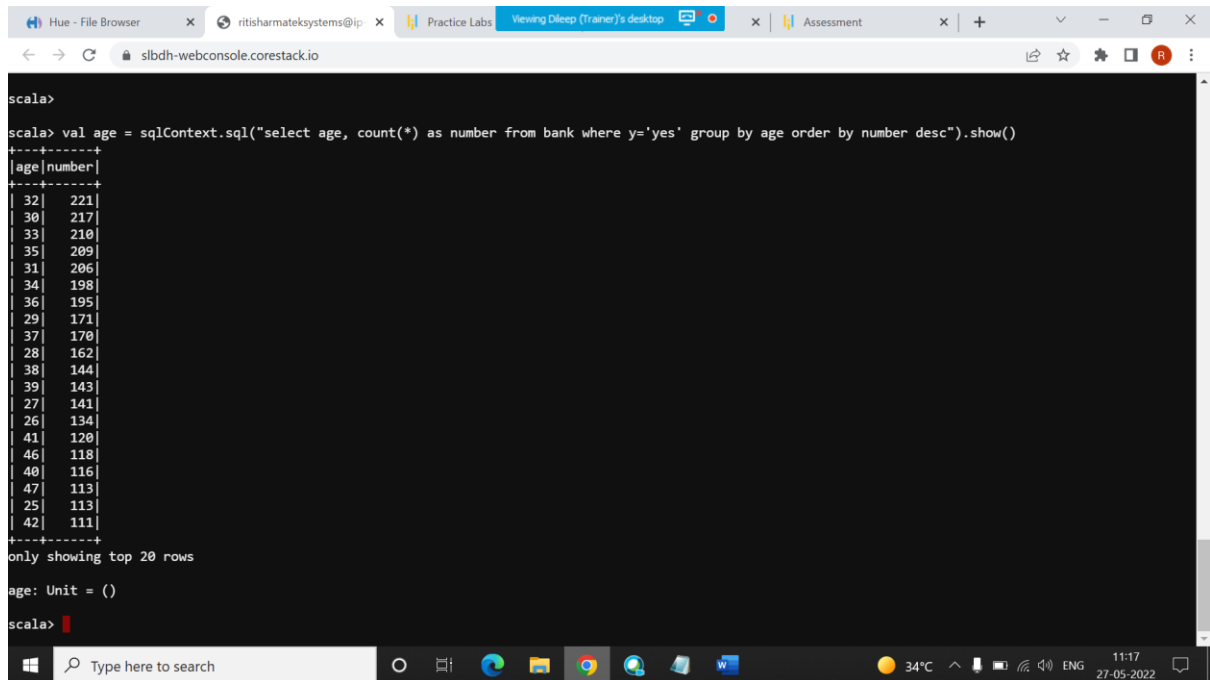
scala> val median = sqlContext.sql("SELECT percentile_approx(balance, 0.5) FROM bank").show()
+-----+
|percentile_approx(balance, CAST(0.5 AS DOUBLE), 10000)|
+-----+
|                        448|
+-----+

median: Unit = ()

scala>

scala>
```

Ans:3)



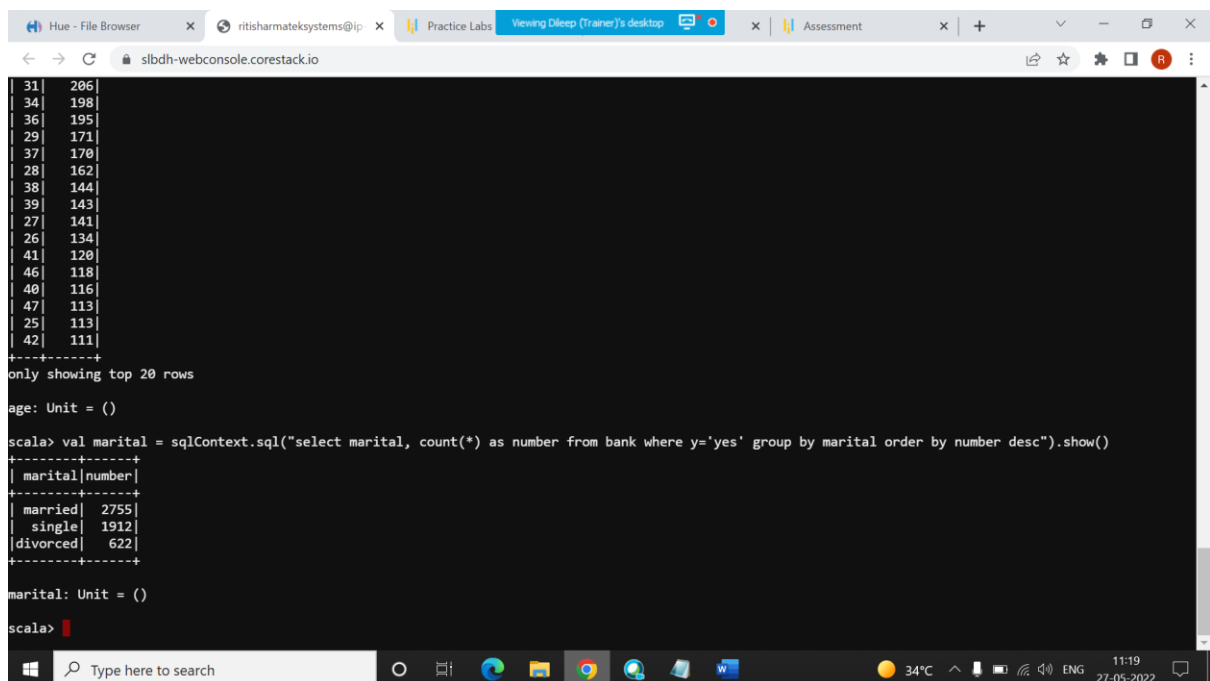
The screenshot shows a web browser window with multiple tabs. The active tab is 'Viewing Deep (Trainer)'s desktop'. The address bar shows 'slbdh-webconsole.corestack.io'. The main content area displays a Scala REPL session. The user has entered a query to select age and count(*) as number from a table named 'bank' where y='yes', grouped by age and ordered by number descending. The result is shown as a table with 20 rows.

```
scala> val age = sqlContext.sql("select age, count(*) as number from bank where y='yes' group by age order by number desc").show()
+-----+
|age|number|
+-----+
|32|221|
|30|217|
|33|210|
|35|209|
|31|206|
|34|198|
|36|195|
|29|171|
|37|170|
|28|162|
|38|144|
|39|143|
|27|141|
|26|134|
|41|120|
|46|118|
|40|116|
|47|113|
|25|113|
|42|111|
+-----+
only showing top 20 rows

age: Unit = ()

scala>
```

Ans:4)



The screenshot shows the same web browser window as the previous one. The user has entered a new Scala query to select marital status and count(*) as number from the 'bank' table where y='yes', grouped by marital status and ordered by number descending. The result is shown as a table with 3 rows.

```
scala> val marital = sqlContext.sql("select marital, count(*) as number from bank where y='yes' group by marital order by number desc").show()
+-----+
|marital|number|
+-----+
|married|2755|
|single|1912|
|divorced|622|
+-----+

marital: Unit = ()

scala>
```

```
Hue - File Browser x ritisharmateksystems@ip- Practice Labs Viewing Dileep (Trainer)'s desktop x Assessment x + - x
slbdh-webconsole.corestack.io

26| 134|
41| 120|
46| 118|
40| 116|
47| 113|
25| 113|
42| 111|
+-----+
only showing top 20 rows

age: Unit = ()

scala> val marital = sqlContext.sql("select marital, count(*) as number from bank where y='yes' group by marital order by number desc").show()
+-----+
| marital|number|
+-----+
| married| 2755|
| single | 1912|
| divorced| 622|
+-----+

marital: Unit = ()

scala> df.groupBy($"y".alias("Did the customer subscribed")).agg(count($"marital").alias("marital count")).show()
+-----+
|Did the customer subscribed|marital count|
+-----+
| no                         | 39922|
| yes                        | 5289|
+-----+

scala>
```

Ans:5)

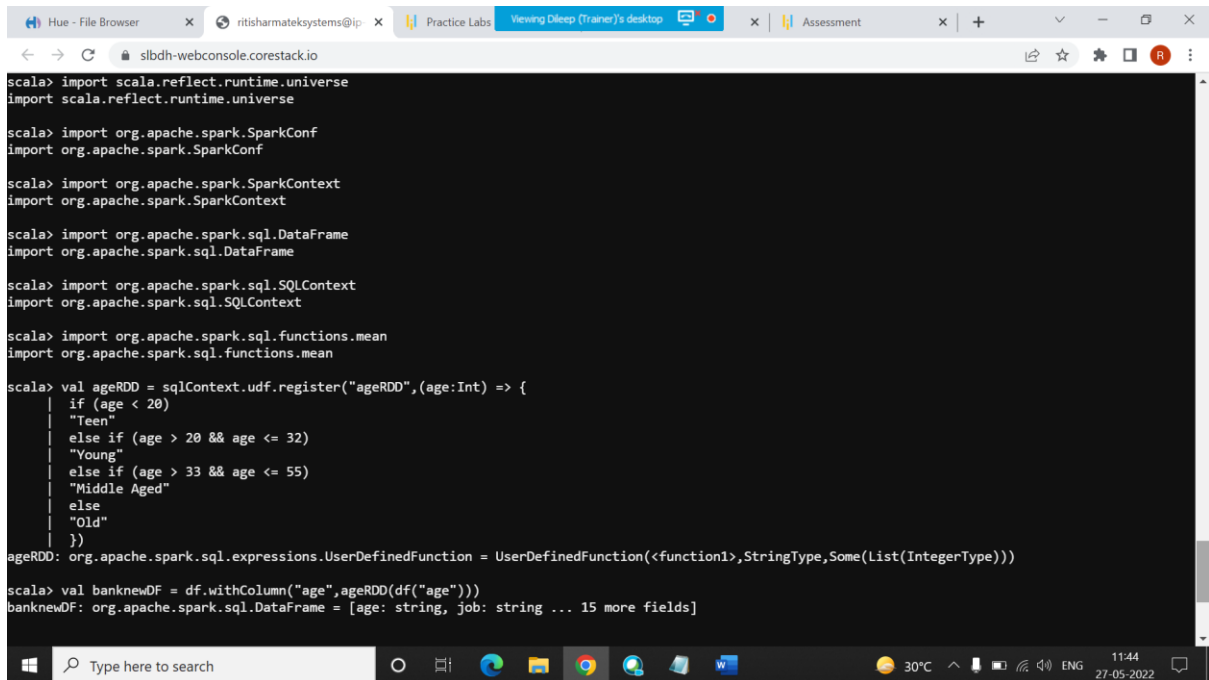
```
Hue - File Browser x ritisharmateksystems@ip- Practice Labs Viewing Dileep (Trainer)'s desktop x Assessment x + - x
slbdh-webconsole.corestack.io

scala> val age_marital = sqlContext.sql("select age, marital, count(*) as number from bank where y='yes' group by age,marital order by number desc").show()
+-----+
|age|marital|number|
+-----+
| 30| single | 151|
| 28| single | 138|
| 29| single | 133|
| 32| single | 124|
| 26| single | 121|
| 34| married| 118|
| 31| single | 111|
| 27| single | 110|
| 35| married| 101|
| 36| married| 100|
| 25| single | 99|
| 37| married| 98|
| 33| single | 97|
| 33| married| 97|
| 32| married| 87|
| 39| married| 87|
| 38| married| 86|
| 35| single | 84|
| 47| married| 83|
| 31| married| 80|
+-----+
only showing top 20 rows

age_marital: Unit = ()

scala>
```

Ans:6)



The screenshot shows a Hue web console interface with a terminal window. The terminal displays Scala code for registering a User Defined Function (UDF) and creating a DataFrame. The code defines a UDF named 'ageRDD' that categorizes ages into 'Teen', 'Young', 'Middle Aged', and 'Old'. It then creates a DataFrame 'banknewDF' with a new column 'age' based on the UDF.

```
scala> import scala.reflect.runtime.universe
import scala.reflect.runtime.universe

scala> import org.apache.spark.SparkConf
import org.apache.spark.SparkConf

scala> import org.apache.spark.SparkContext
import org.apache.spark.SparkContext

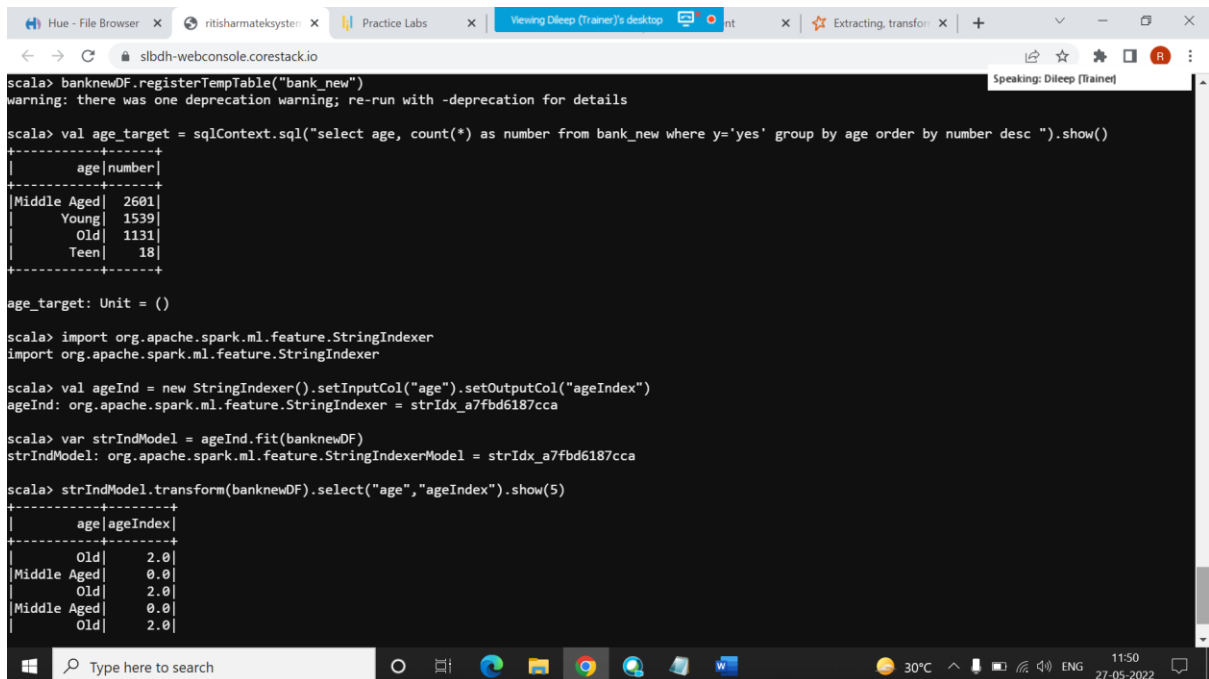
scala> import org.apache.spark.sql.DataFrame
import org.apache.spark.sql.DataFrame

scala> import org.apache.spark.sql.SQLContext
import org.apache.spark.sql.SQLContext

scala> import org.apache.spark.sql.functions.mean
import org.apache.spark.sql.functions.mean

scala> val ageRDD = sqlContext.udf.register("ageRDD", (age: Int) => {
  | if (age < 20)
  |   "Teen"
  | else if (age > 20 && age <= 32)
  |   "Young"
  | else if (age > 33 && age <= 55)
  |   "Middle Aged"
  | else
  |   "Old"
  | })
ageRDD: org.apache.spark.sql.expressions.UserDefinedFunction = UserDefinedFunction(<function1>, StringType, Some(List(IntegerType)))

scala> val banknewDF = df.withColumn("age", ageRDD(df("age")))
banknewDF: org.apache.spark.sql.DataFrame = [age: string, job: string ... 15 more fields]
```



The screenshot shows a Hue web console interface with a terminal window. The terminal displays Scala code for creating a temporary table, querying it, and using a StringIndexer. The code creates a temp table 'bank_new', queries it to get the count of rows for each age group, and then uses a StringIndexer to create an 'ageIndex' column from the 'age' column.

```
scala> banknewDF.registerTempTable("bank_new")
warning: there was one deprecation warning; re-run with -deprecation for details

scala> val age_target = sqlContext.sql("select age, count(*) as number from bank_new where y='yes' group by age order by number desc ").show()
+-----+
| age|number|
+-----+
|Middle Aged| 2601|
| Young| 1539|
| Old| 1131|
| Teen| 18|
+-----+

age_target: Unit = ()

scala> import org.apache.spark.ml.feature.StringIndexer
import org.apache.spark.ml.feature.StringIndexer

scala> val ageInd = new StringIndexer().setInputCol("age").setOutputCol("ageIndex")
ageInd: org.apache.spark.ml.feature.StringIndexer = strIdx_a7fbd6187cca

scala> var strIndModel = ageInd.fit(banknewDF)
strIndModel: org.apache.spark.ml.feature.StringIndexerModel = strIdx_a7fbd6187cca

scala> strIndModel.transform(banknewDF).select("age", "ageIndex").show(5)
+-----+
| age|ageIndex|
+-----+
| Old| 2.0|
|Middle Aged| 0.0|
| Old| 2.0|
|Middle Aged| 0.0|
| Old| 2.0|
+-----+
```