scala> val bank_people_data = spark.read.option("multiline","true").json("/user/charangowda845_gmail/bank_edited.json"); bank_people_data: org.apache.spark.sql.DataFrame = [age: bigint, balance: bigint . . . 15 more fields]

scala> bank_people_data.show()

+	++		+	++			+	+	+	+	+	+	++		+	++
age	balance	campaign	contact	day	default	duration	education	housing	job	loan	marital	month	pdays	poutcome	previous	∣ y∣
+	++		+	++			+	+	+	+	+	+	+			+
58	2143	1	unknown	5	no	261	tertiary	yes	management	no	married	may	-1	unknown	0	no
44	29	1	unknown	5	no	151	secondary	yes	technician	no	single	may	-1	unknown	0	no
33	2	1	unknown	5	no	76	secondary	yes	entrepreneur	yes	married	may	-1	unknown	0	no
47	1506	1	unknown	5	no	92	unknown	yes	blue-collar	no	married	may	-1	unknown	0	no
33	1	1	unknown	5	no	198	unknown	no	unknown	no	single	may	-1	unknown	0	no
35	231	1	unknown	5	no	139	tertiary	yes	management	no	married	may	-1	unknown	0	no
28	447	1	unknown	5	no	217	tertiary	yes	management	yes	single	may	-1	unknown	0	no
42	2	1	unknown	5	yes	380	tertiary	yes	entrepreneur	no	divorced	may	-1	unknown	0	no
58	121	1	unknown	5	no	50	primary	yes	retired	no	married	may	-1	unknown	0	no
43	593	1	unknown	5	no	55	secondary	yes	technician	no	single	may	-1	unknown	0	no
41	270	1	unknown	5	no	222	secondary	yes	admin.	no	divorced	may	-1	unknown	0	no
29	390	1	unknown	5	no	137	secondary	yes	admin.	no	single	may	-1	unknown	0	no
53	6	1	unknown	5	no	517	secondary	yes	technician	no	married	may	-1	unknown	0	no
58	71	1	unknown	5	no	71	unknown	yes	technician	no	married	may	-1	unknown	0	no
57	162	1	unknown	5	no	174	secondary	yes	services	no	married	may	-1	unknown	0	no
51	229	1	unknown	5	no	353	primary	yes	retired	no	married	may	-1	unknown	0	no
45	13	1	unknown	5	no	98	unknown	yes	admin.	no	single	may	-1	unknown	0	no
57	52	1	unknown	5	no	38	primary	yes	blue-collar	no	married	may	-1	unknown	0	no
60	60	1	unknown	5	no	219	primary	yes	retired	no	married	may	-1	unknown	0	no
33	0	1	unknown	5	no	54	secondary	yes	services	no	married	may	-1	unknown	0	no
+	++		+	++		+	+	+	+	+	+	+	++			++

only showing top 20 rows

scala>

```
scala> bank_people_data.select(max($"age")).show()
max(age)
+----+
95
+-----
scala> bank_people_data.select(min($"age")).show()
min(age)
+----+
scala> bank_people_data.select(avg($"age")).show()
avg(age)
40.93621021432837
scala> bank_people_data.select(avg($"balance")).show()
avg(balance)
+-----
1362.2720576850766
+----+
```

```
scala> val median = spark.sql("SELECT percentile_approx(balance, 0.5) FROM datanewtable").show()
|percentile_approx(balance, CAST(0.5 AS DOUBLE), 10000)|
median: Unit = ()
scala> val agedata = spark.sql("select age, count(*) as number from datanewtable where y='yes' group by age order by number desc") agedata: org.apache.spark.sql.DataFrame = [age: bigint, number: bigint]
scala>
scala> agedata.show()
|age|number|
1 321
           221
  30
           217
  33 İ
           210
  35
           209
  31
           206
           198
  34
  36
           195
  29
           171
  37
           170
  28
           162
  38
           144
  39
           143
  27
           141
  26
           134
  41
           120
  46
           118
  40
           116
  25
           113
  47
           113
42
          111
only showing top 20 rows
```

scala> val agedata = spark.sql("select age, count(*) as number from datanewtable where y='yes' group by age order by number desc") agedata: org.apache.spark.sql.DataFrame = [age: bigint, number: bigint]

scala> val maritaldata = spark.sql("select marital, count(*) as number from datanewtable where y='yes' group by marital order by number desc") maritaldata: org.apache.spark.sql.DataFrame = [marital: string, number: bigint]

```
scala> val ageandmaritaldata = spark.sql("select age, marital, count(*) as number from datanewtable where y='yes' group by age,marital order by numbe
r desc")
ageandmaritaldata: org.apache.spark.sql.DataFrame = [age: bigint, marital: string ... 1 more field]
scala> ageandmaritaldata.show()
|age|marital|number|
  30 single
28 single
29 single
32 single
34 married
31 single
37 single
35 married
36 married
36 married
37 married
33 married
33 married
                        138
133
124
121
                        118
                        111
                        110
                         101
                        99
98
97
97
87
   33 single
32 married
39 married
38 married
                          87
                          86
  35 single
47 married
31 married
                          84
                          83
                          80
only showing top 20 rows
```

```
| ))
agedata: org.apache.spark.sql.expressions.UserDefinedFunction = UserDefinedFunction(<function1>,StringType,Some(List(IntegerType)))
\label{eq:continuous} $$ scala> val banknewOF = bank\_people\_data.withColumn("age",agedata(bank\_people\_data("age"))) $$ banknewOF: org.apache.spark.sql.DataFrame = [age: string, balance: bigint ... 15 more fields] $$ balance: bigint ... 15 more fields $$ bala
  scala> banknewDF.show()
               age|balance|campaign|contact|day|default|duration|education|housing|
                                          e|balance|campaign|contact|day|default|duration|education|housing| job|loan| marital|month|pdays|poutcome|previous| y|
                                                                                                                                                                                                                                                                    yes management no married
yes technician no single
yes entrepreneur yes married
yes blue-collar no married
                                old
                                                        2143
                                                                                                   1 unknown l
                                                                                                                                                                                                    261 | tertiary|
                                                                                                                                                                                                                                                                                                                                                                                                                    -1 unknown
                                                                                                                                               51
                                                                                                                                                                                                                                                                                                                                                                                          mav
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Middle Aged
| old
| Middle Aged
                                                         29
2
1506
                                                                                                   1 unknown
1 unknown
1 unknown
1 unknown
                                                                                                                                                                                                    151 secondary
76 secondary
92 unknown
198 unknown
                                                                                                                                                                                                                                                                                                                                                                                           may|
may|
may|
                                                                                                                                                                                                                                                                                                                                                                                                                    -1 unknown
-1 unknown
-1 unknown
-1 unknown
                                                                                                                                               no
no
no
no
no
    old
|Middle Aged
                                                                                                                                                                                                                                                                         no
                                                                                                                                                                                                                                                                                          unknown
management
                                                                                                                                                                                                                                                                                                                                        no
                                                                                                                                                                                                                                                                                                                                                        single
                                                                                                                                                                                                                                                                                                                                                                                           may
                                                                                                                                                                                                    198 unknown
139 tertiary
217 tertiary
380 tertiary
50 primary
55 secondary
222 secondary
                                                             231
                                                                                                    1 unknown
                                                                                                                                                                                                                                                                      ves
                                                                                                                                                                                                                                                                                                                                        noİ
                                                                                                                                                                                                                                                                                                                                                     married
                                                                                                                                                                                                                                                                                                                                                                                           mavİ
                                                                                                                                                                                                                                                                                                                                                                                                                      -1 unknown
  Young
Young
Middle Aged
old
Middle Aged
Middle Aged
                                                                                                   1 unknown
1 unknown
1 unknown
1 unknown
1 unknown
                                                                                                                                                                                                                                                                     yes management
yes management
yes entrepreneur
yes retired
yes technician
yes admin.
yes admin.
                                                                                                                                                                                                                                                                                                                                    yes single
no divorced
no married
no single
no divorced
                                                            447
                                                                                                                                                                                                                                                                                                                                                                                                                      -1 unknown
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      2
121
593
                                                                                                                                                                                                                                                                                                                                                                                                                      -1 unknown

-1 unknown

-1 unknown

-1 unknown

-1 unknown
                                                                                                                                                                    yes
no
no
no
no
no
no
no
no
                                                                                                                                                                                                                                                                                                                                                                                           may
                                                             270
                                                                                                                                                                                                                                                                                                                                      no divorced
no single
no married
no married
no married
no single
no married
no married
no married
no married
                                                                                                                                                                                                                                                                                                                                                                                           may
                          Young
                                                             390
                                                                                                     1 unknown
                                                                                                                                                                                                     137 secondary
                                                                                                                                                                                                                                                                                                                                                                                           mav
                                                                                                                                                                                                                                                                                                                                                                                                                      -1 unknown
   Young
|Middle Aged
| old
| old
                                                                                                   1 unknown
1 unknown
1 unknown
1 unknown
1 unknown
1 unknown
                                                                                                                                                                                                    517 secondary
71 unknown
174 secondary
                                                                                                                                                                                                                                                                      yes
yes
yes
                                                                                                                                                                                                                                                                                                                                                                                           may
may
may
                                                                                                                                                                                                                                                                                                                                                                                                                                 unknown
unknown
unknown
unknown
                                                                                                                                                                                                                                                                                          technician
                                                                                                                                                                                                 51/|-
| 71| unkm
| 174|secondary|
| 353| primary|
| 98| unknown|
| primary|
                                                             71
162
229
                                                                                                                                                                                                                                                                                           technician
technician
services
retired
                                                                                                                                                                                                                                                                     yes services
yes retired
yes admin.
yes blue-collar
yes retired
yes services
     Middle Aged
                                                                                                                                                                                                                                                                                                                                                                                           may|
may|
    Middle Aged
                                                                13
52
                                                                                                                                                                                                                                                                                                                                                                                                                      -1
                                                                                                                                                                                                                                                                                                                                                                                                                                 unknown
                                                                                                                                                                                                         38 primary
219 primary
54 secondary
                                                                                                                                                                                                                                                                                                                                                                                          may|
may|
may|
                                 oldi
                                                                                                     1 unknown
                                                                                                                                                                                                                                                                                                                                                                                                                                 unknown
                                                                                                    1 unknown
1 unknown
                                 oldi
                                                                                                                                                                                                     219
                                                                                                                                                                                                                                                                                                                                                                                                                                  unknown
```

only showing top 20 rows

scala>	banknewDF	.show()

+				+	+	+		+						+			++
	age	balance	campaign	contact	day	default	duration	education	housing	job	loan	marital	month	pdays	poutcome	previous	y l
+	+		+	+	++	+		+	+		4	++			+	+	·+
1	old	2143	1	unknown	5	no	261	tertiary	yes	management	no	married	may	-1	unknown	0	no
Middle	Aged	29	1	unknown	5	no	151	secondary	yes	technician	no	single	may	-1	unknown	0	no
İ	old	2	1	unknown	5	no	76	secondary	yes	entrepreneur	yes	married	may	-1	unknown	0	no
Middle	Aged	1506	1	unknown	5	no	92	unknown	yes	blue-collar	no	married	may	-1	unknown	0	no
İ	old	1	1	unknown	5	no	198	unknown	no	unknown	no	single	may	-1	unknown	0	no
Middle	Aged	231	1	unknown	5	no	139	tertiary	yes	management	no	married	may	-1	unknown	0	no
į v	oung	447	1	unknown	5	no	217	tertiary	yes	management	yes	single	may	-1	unknown	0	no
Middle	Aged	2	1	unknown	5	yes	380	tertiary	yes	entrepreneur	no	divorced	may	-1	unknown	0	no
i	old	121	1	unknown	5	no	50	primary	ves	retired	no	married	may	-1	unknown	9	no
Middle	Aged	593	1	unknown	5	no	55	secondary	ves	technician	no	single	may	-1	unknown	9	nol
Middle	Aged	270	1	unknown	5	no	222	secondary	ves	admin.	no	divorced	may	-1	unknown	9	nol
	oung/	390	1	unknown	5	no	137	secondary	yes	admin.	no	single		-1	unknown	0	no

```
scala> banknewDF.registerTempTable("banknewtable") warning: there was one deprecation warning; re-run with -deprecation for details
scala> val targetage = spark.sql("select age, count(*) as number from banknewtable where y='yes' group by age order by number desc") targetage: org.apache.spark.sql.DataFrame = [age: string, number: bigint]
scala> targetage.show()
 age|number|
|Middle Aged| 2601
| Young| 1539
| old| 1131
| Teen| 18
scala>
scala> import org.apache.spark.ml.feature.StringIndexer
import org.apache.spark.ml.feature.StringIndexer
scala> val agedata2 = new StringIndexer().setInputCol("age").setOutputCol("ageindex")
agedata2: org.apache.spark.ml.feature.StringIndexer = strIdx_d413a7d2b256
scala> var strindModel = agedata2.fit(banknewDF)
strindModel: org.apache.spark.ml.feature.StringIndexerModel = strIdx_d413a7d2b256
scala> strindModel.transform(banknewDF).select("age","ageIndex").show(5)
          age ageIndex
          old
                       2.0
|Middle Aged
                     0.0
                     2.0
0.0
           old
|Middle Aged|
old
                     2.0
only showing top 5 rows
```

scala>