Welcome to Zeppelin.

This is a live tutorial, you can run the code yourself. (Shift-Enter to Run)

```
import org.apache.commons.io.IOUtils
import java.net.URL
import java.nio.charset.Charset

// Zeppelin creates and injects sc (SparkContext) and sqlContext (HiveContext or SqlContext)
// So you don't need create them manually

// load bank data
val bankText = sc.parallelize(
    IOUtils.toString(
        new URL("https://s3.amazonaws.com/apache-zeppelin/tutorial/bank/bank.csv"),
        Charset.forName("utf8")).split("\n"))

case class Bank(age: Integer, job: String, marital: String, education: String, balance: Integer)

val bank = bankText.map(s => s.split(";")).filter(s => s(0) != "\"age\"").map(
    s => Bank(s(0).toInt,
            s(1).replaceAll("\"", ""),
            s(2).replaceAll("\"", ""),
            s(3).replaceAll("\"", ""),
            s(5).replaceAll("\"", "").toInt
        )
).toDF()
bank.registerTempTable("bank")
```

```
import org.apache.commons.io.IOUtils
import java.net.URL
import java.nio.charset.Charset
bankText: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[36] at parallelize at <console>:43
defined class Bank
bank: org.apache.spark.sql.DataFrame = [age: int, job: string ... 3 more fields]
warning: there were 1 deprecation warning(s); re-run with -deprecation for details
```

```
%sql
select age, count(1) value
from bank
where age < 30
group by age
order by age
```

age	value
19	4
20	3
21	7
22	9
23	20
24	24
25	44
26	77
27	94
28	103
29	97

```
%sql
select age, count(1) value
from bank
where age < ${maxAge=30}
group by age
order by age
```

age	value
19	4
20	3
21	7
22	9
23	20
24	24
25	44
26	77
27	94
28	103
29	97
30	150
31	199
32	224
33	186
34	231

```
%sql
select age, count(1) value
from bank
where marital="${marital=single,single|divorced|married}"
group by age
order by age
```

age	value
19	4
20	3
21	7
22	9
23	17
24	13
25	33
26	56
27	64
28	78
29	56
30	92
31	86
32	105
33	61
34	75
35	46
36	50

```
37      43
38      44
39      30
40      25
41      19
42      23
43      21
44      20
45      15
46      14
47      12
48      12
49      11
50      8
51      6
52      9
53      4
55      3
56      3
57      2
58      7
59      2
60      5
66      2
69      1
```

Congratulations, it’s done.

You can create your own notebook in ‘Notebook’ menu. Good luck!

About bank data

```
Citation Request:
  This dataset is public available for research. The details are described in [Moro et al., 2011].
  Please include this citation if you plan to use this database:

  [Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology.
  In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, GuimarÃ£es, Portugal, October, 2011. EUROSIS.

  Available at: [pdf] http://hdl.handle.net/1822/14838
                [bib] http://www3.dsi.uminho.pt/pcortez/bib/2011-esm-1.txt
```