Project 1 - Apache Spark—Real Time Project—Marketing Analysis

# Pre-requisites:

The csv data file was cleaned and loaded as comma separated text file P1\_Bank\_DataSet.txt.

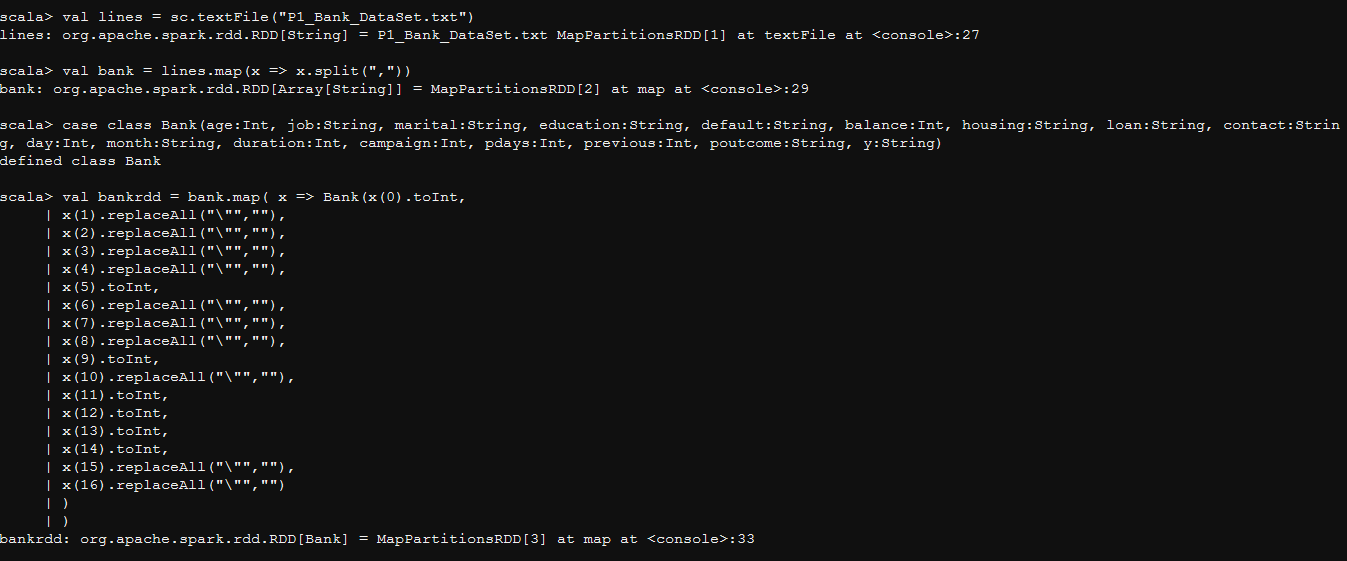
Then the file was uploaded to cloudlab using FTP service.

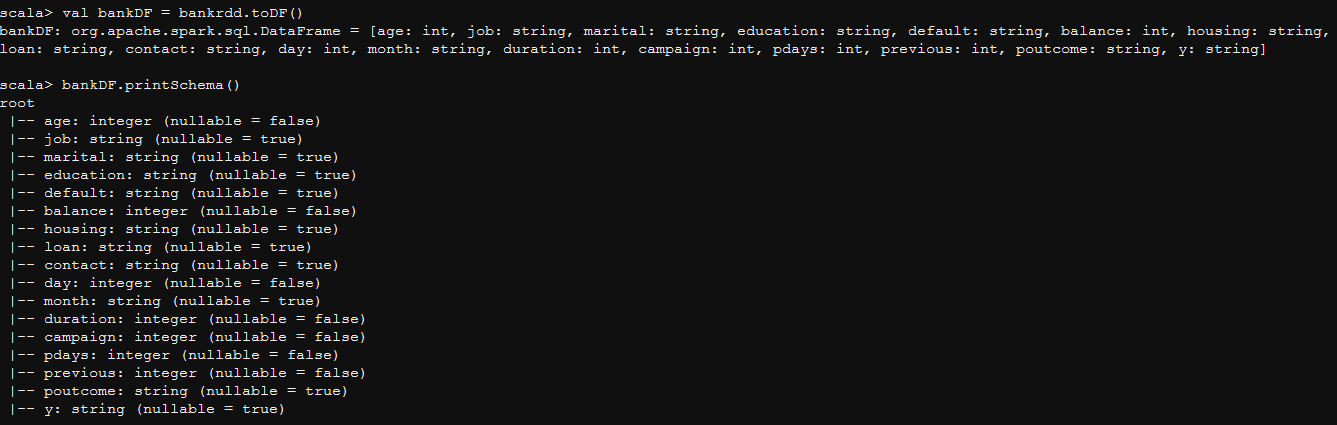
Then they were uploaded to Hadoop FS using the command:

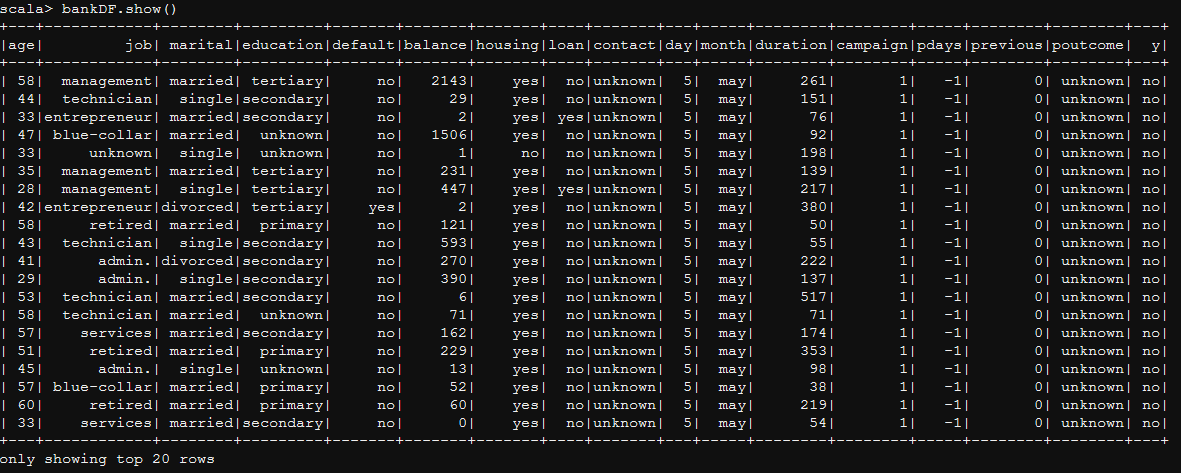
***hadoop fs –put P1\_Bank\_DataSet.txt .***

The spark shell is then launched and the data processing starts.

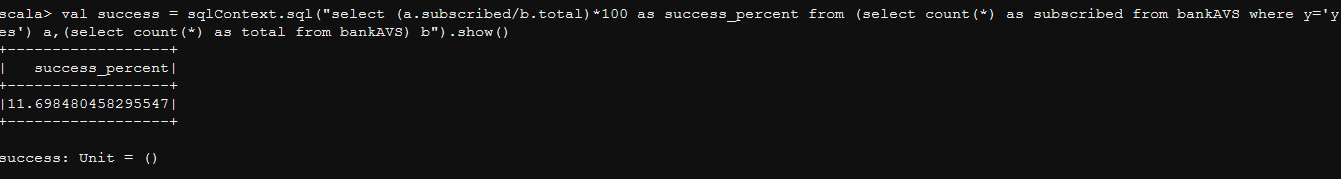
# Load data and create Spark data frame





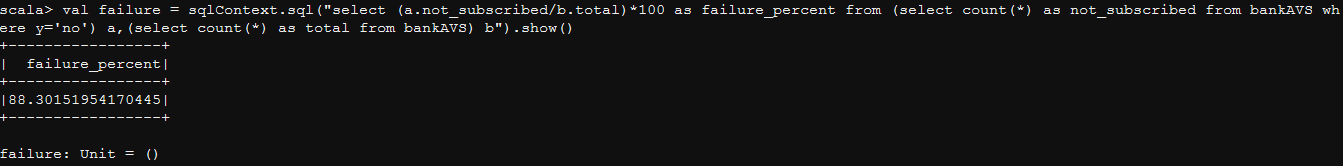


# Give marketing success rate. (No. of people subscribed / total no. of entries)



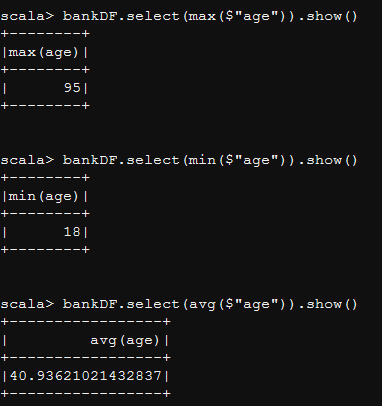
**Marketing Success Rate : 11.698 %**

# 2a.Give marketing failure rate



**Marketing Failure Rate : 88.301 %**

# Maximum, Mean, and Minimum age of average targeted customer

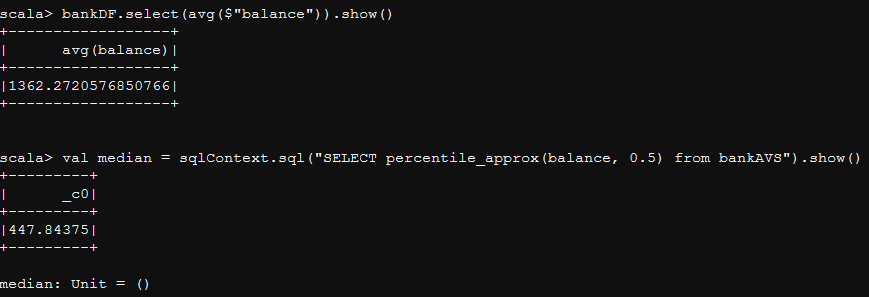


**Max Age of Targeted Customer: 95**

**Min Age of Targeted Customer: 18**

**Average Age of Targeted Customer: 40.936**

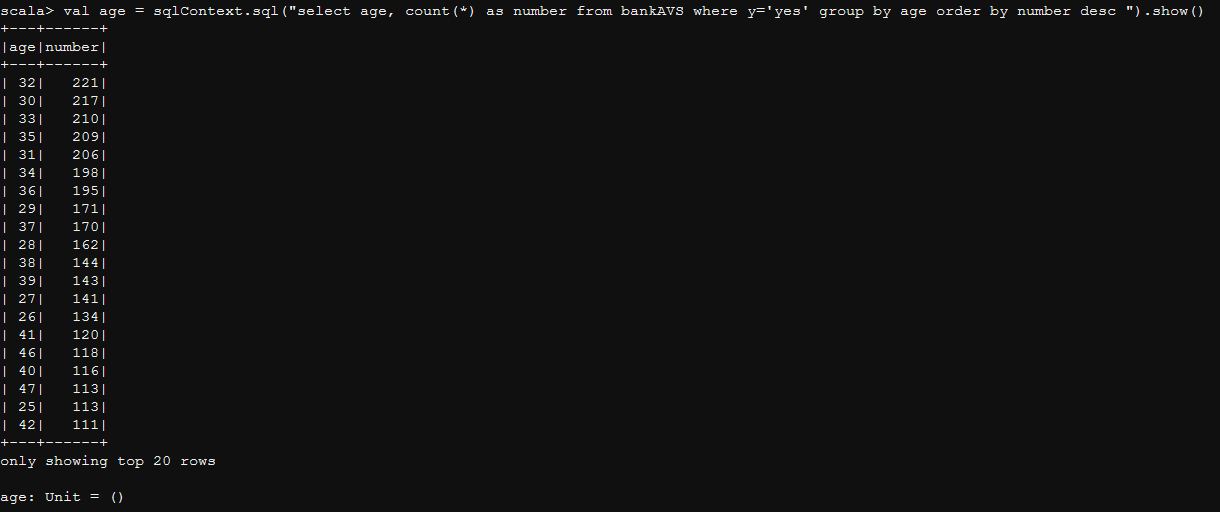
# Check quality of customers by checking average balance, median balance of customers



**Average Balance of customers: 1362.272**

**Median Balance of customers: 447.843**

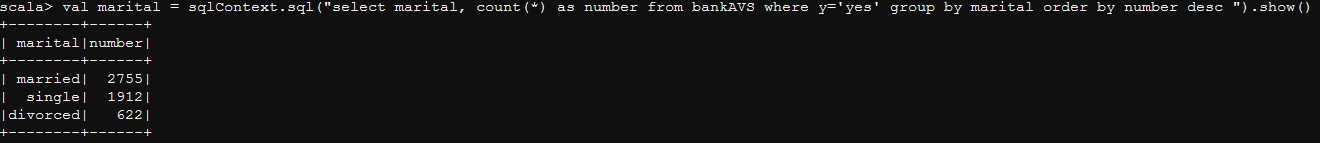
# Check if age matters in marketing subscription for deposit



**Conclusion:**

**Age matters. The age range of 30-36 is quite strong here.**

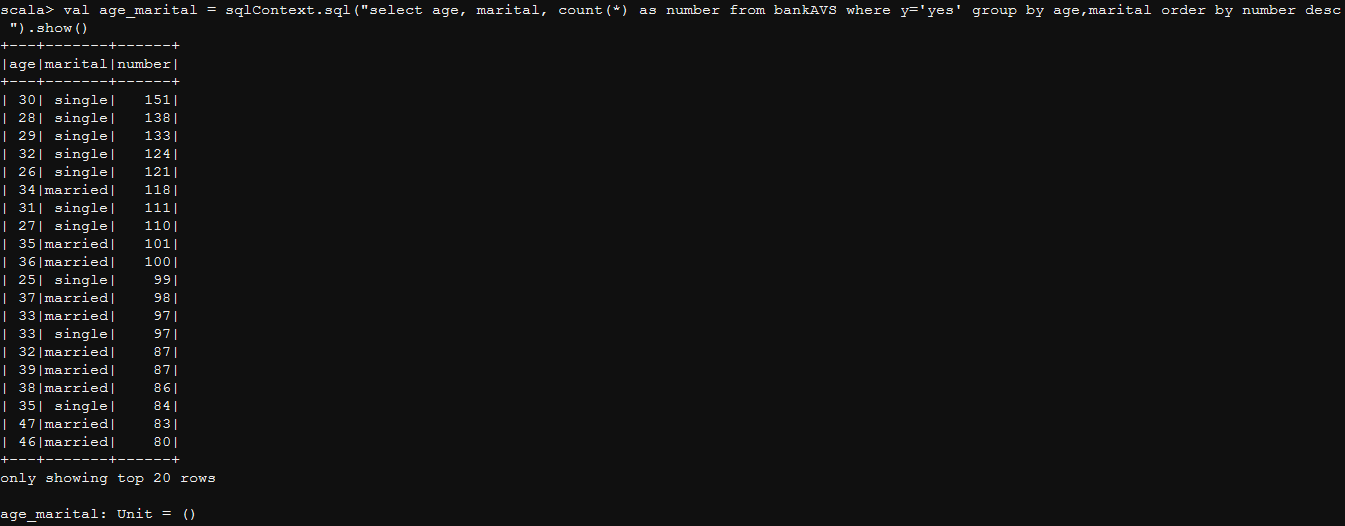
# Check if marital status mattered for subscription to deposit.



**Conclusion:**

**Marital Status also matters. Married people tend to do it more.**

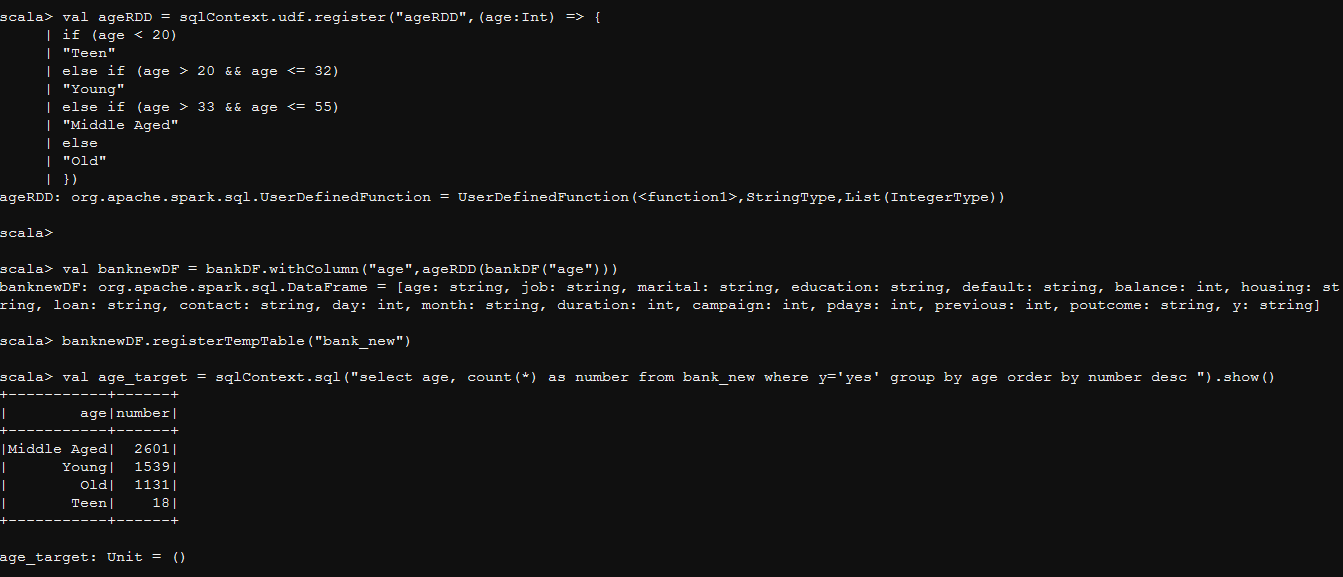
# Check if age and marital status together mattered for subscription to deposit scheme



**Conclusion:**

**Single people in the age 30-35 dominate the subscriptions.**

# Do feature engineering for column—age and find right age effect on campaign



**Conclusion:**

**Age < 20 🡪 Teen**

**Age in between 21-32 🡪 Young**

**Age in between 33-55 🡪 Middle Aged**

**Age > 56 🡪 Old**

**We can conclude here that the ‘Middle Aged’ people between age 33 and 55 is the right age for the campaign.**