

Market analysis in banking domain

Thobelani Mkhize

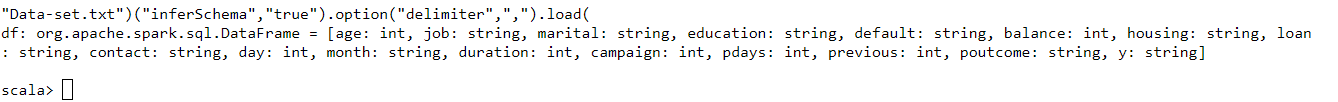


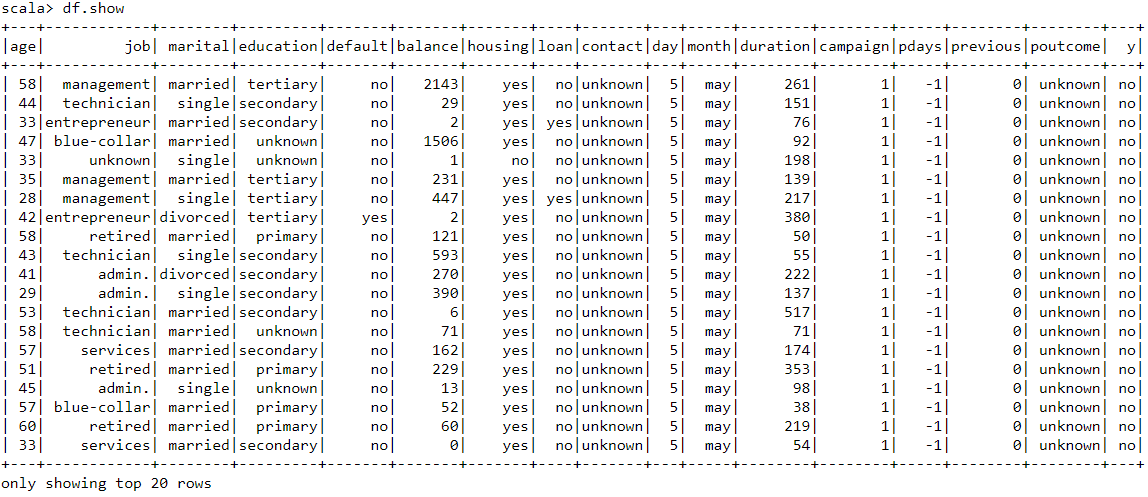
February 20, 2020

**Question One**

* Load data and create a spark data frame.
* val df = sqlContext.read.format("com.databricks.spark.csv").option("header","true").option("inferSchema","true").option("delimiter",",").load("Data-set.txt")

**Out Put:**





**Analysis:**

* The data has been loaded and the data frame also created.

As we can see on our data frame that it shows different ages of different people and their marital status and their personal information.

**Question Two**

* Give marketing success rate (Number of People subscribed, Total number of entries)
* val totalcount = df.count().toDouble
* val subscription\_count= df.filter($"y" === "yes").count().toDouble
* val success\_rate = subscription\_count/totalcount

**Out Put:**

****

****

****

**Analysis:**

* As the calculation shows the total number of entries of 45211.0, and their subscriptions of 5289.0 to the bank term deposit, It shows that the marketing campaign was not successful.
* Even the bank success rate of 0.11, shows clearly that the campaign was not successful.
* Marketing failure rate

**Out Put:**

****

**Analysis:**

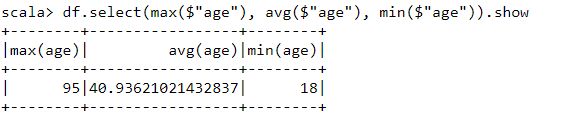
* The calculation of failure rate shows that the bank campaign failed with 0.88.

Which means that the bank failed to make success into their campaign using phone calls with 90% failure rate.

That shows poor marketing campaign.

* Give the maximum, mean, and minimum age of the average targeted customer

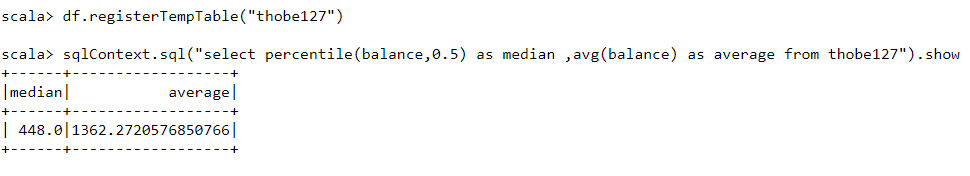
**Out Put:**



**Analysis:**

* The table above shows the age of average targeted customers, which displays the Maximum age of 95 years, average age of 41 years and the minimum age of 18 years.
* Check the quality of customers by checking average balance, median balance of customers

**Out Put:**



**Analysis:**

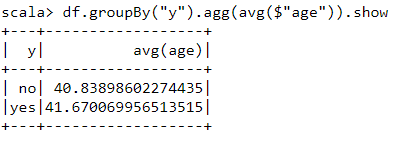
* The table above shows the quality of customers through their balances on their accounts.

The median is 448.0

The average is 1362.272

* Check if age matters in marketing subscription for deposit

**Out Put:**



**Analysis:**

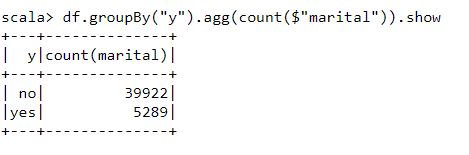
* The analysis table above shows clearly that the age do matters when it comes to the subscriptions of the customers.

People with age of 40 years and below, are not allowed to do subscriptions.

People with age of 41 years and above, do make subscriptions.

* Check if marital status mattered for a subscription to deposit

**Out Put:**



**Analysis:**

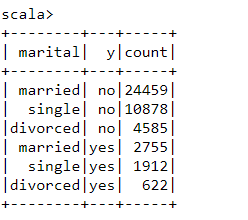
* The analysis table above shows clearly that the marital status do not matter when it comes to the subscriptions of the customers.

The count of 39922 says No, The marital status does not matter when it comes to subscriptions.

While the 5289 says Yes, do matters.

* Check if age and marital status together mattered for a subscription to deposit scheme
* df.groupBy("marital","y").count().sort($"count".desc).show

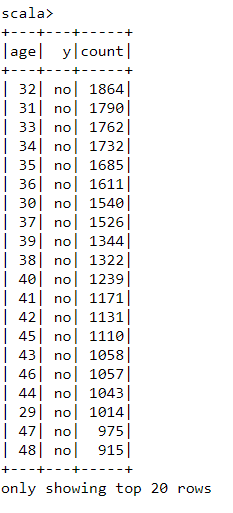
**Out Put:**



**Analysis:**

* The analysis table above shows clearly that the age and marital status together do not matter when it comes to the subscriptions to deposit of the customers.
* Do feature engineering for the bank and find the right age effect on the campaign
* df.groupBy("age","y").count().sort($"count".desc).show

**Out Put:**



**Analysis:**

* The table of analysis shows that there’s no age effect on the campaign