**Market Analysis in Banking Domain Output Screenshots**

DESCRIPTION

**Background and Objective:**

Your client, a Portuguese banking institution, ran a marketing campaign to convince potential customers to invest in a bank term deposit scheme.   
The marketing campaigns were based on phone calls. Often, the same customer was contacted more than once through phone, in order to assess if they would want to subscribe to the bank term deposit or not. You have to perform the marketing analysis of the data generated by this campaign.

**Domain**: Banking (Market Analysis)

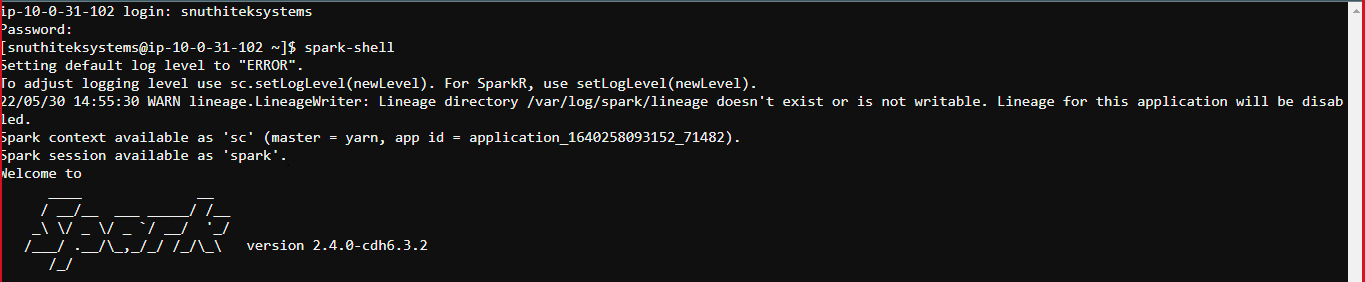
**Analysis tasks to be done-:**

The data size is huge, and the marketing team has asked you to perform the below analysis-

1. Load data and create a Spark data frame
2. Give marketing success rate (No. of people subscribed / total no. of entries)
3. Give marketing failure rate
4. Give the maximum, mean, and minimum age of the average targeted customer
5. Check the quality of customers by checking average balance, median balance of customers
6. Check if age matters in marketing subscription for deposit
7. Check if marital status mattered for a subscription to deposit
8. Check if age and marital status together mattered for a subscription to deposit scheme
9. Do feature engineering for the bank and find the right age effect on the campaign.

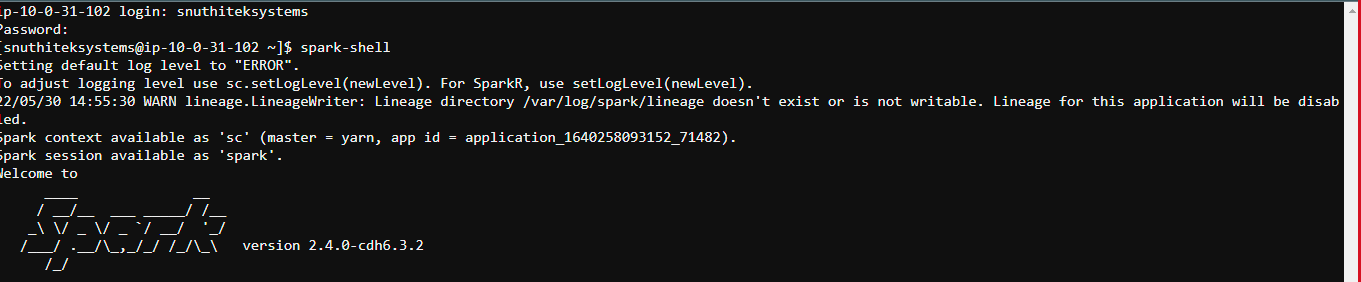
Step 1:

Login with your credentials and run the spark shell to create scala environment and to start the project



Step 2:

Loading data from the path it is uploaded to in hue.



Step 3:

Once the file is loaded, we need to clean the data to make data informative. We have raw data that is separated by ‘;’ so to we make a split based on ‘;’.



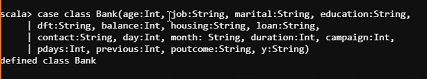
Step 4:

Then, extract the fields using map partitions (map ';')



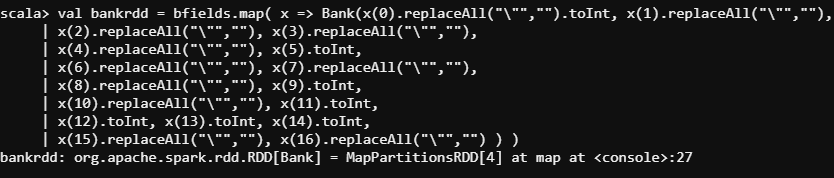
Step 5:

Next, define a class "bank" for schema and assign appropriate the datatypes to the required columns



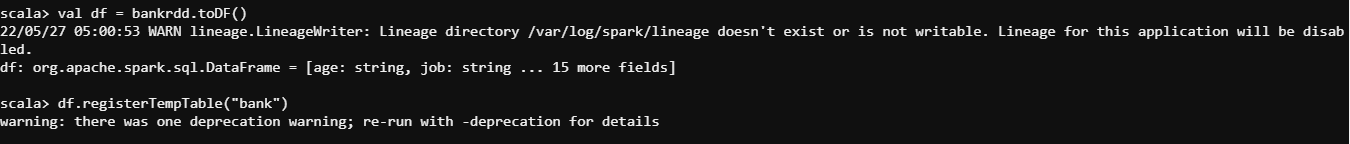
Step 6:

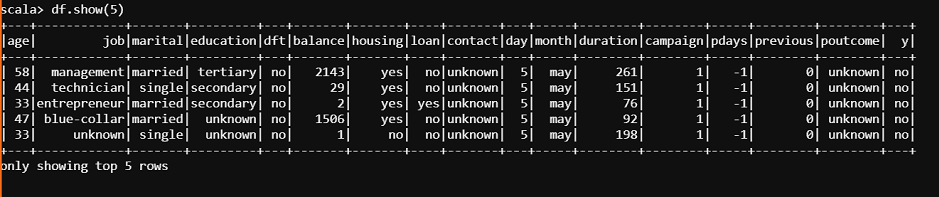
Next, specify the datatypes and remove quotations (" ") from all the fields by replacing them with empty values



Step 7:

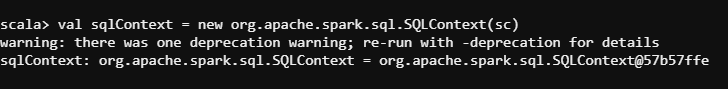
‘toDf’ method is used to covert a data into dataFrame so here we are making the given data as spark dataframe



Displaying top 5 row of dataframe

Step 8:

Define sqlContext by assigning org.apache.spark.sql.SQLContext(sc)



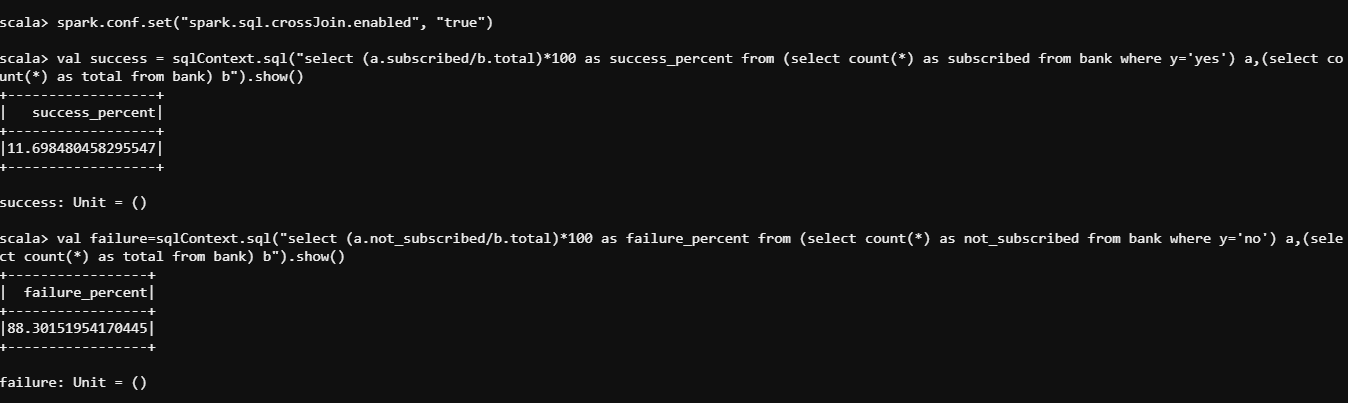
Step 9:

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

By selecting yes values of y and total values of y can give the result

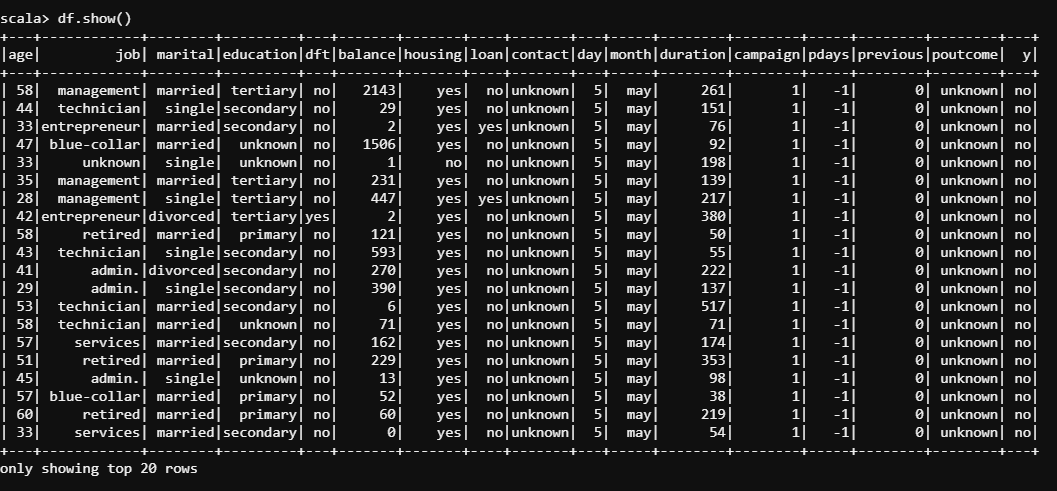
And

Calculate marketing failure rate



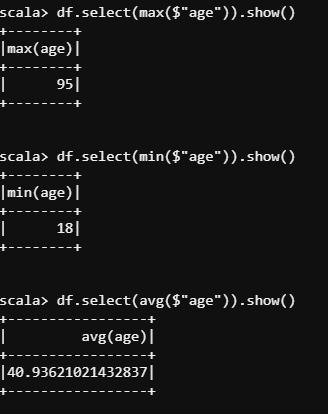
step 10:

check if the data is loaded and cleaned as required



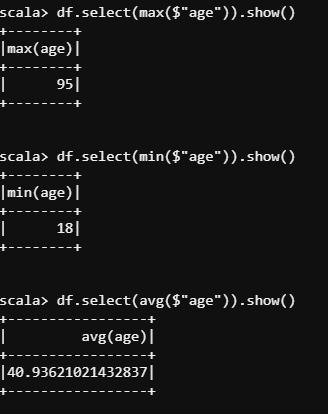
Step 11:

Calculation of maximum age of target customer



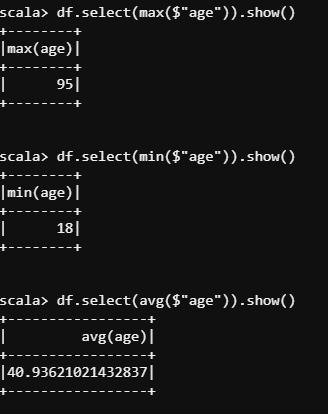
Step 12:

Calculation of minimum age of target customer



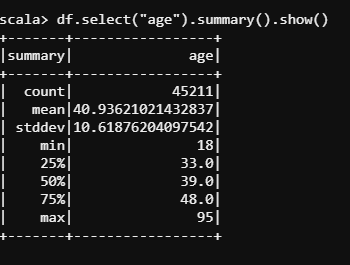
Step 13:

Calculation of average age of target customer



Step 14:

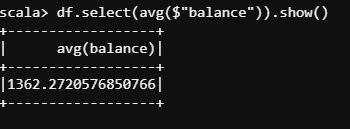
Calculation of mean age of target customer

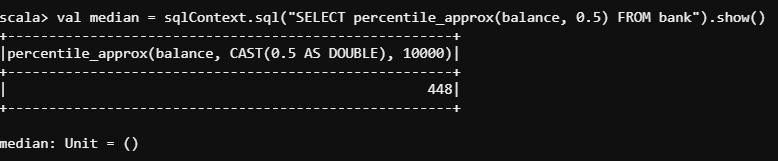


Step 15:

Check the quality of customers by checking average balance and median balance of customers

Mean balance of customers



Median balance of customers

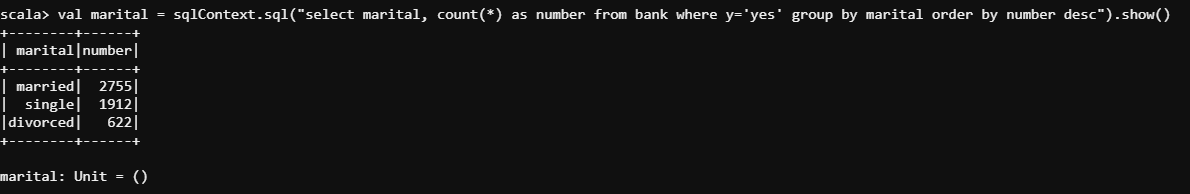
Step 16:

Check if age matters in marketing subscription for deposit



Step 17:

Check if marital status mattered for a subscription to deposit



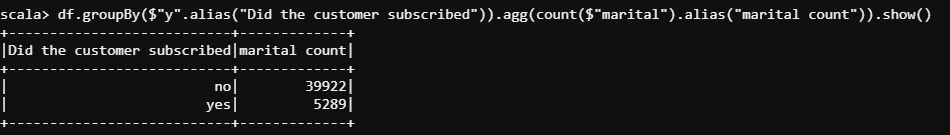
Step 18:

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



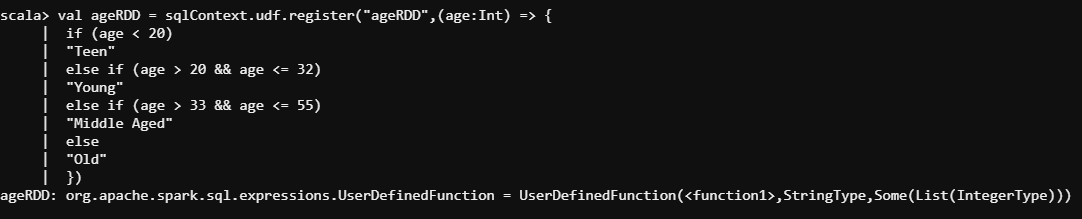
Step 19:

Feature engineering for the bank and find the right age effect on the campaign.

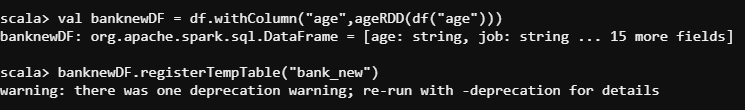


Importing required packages

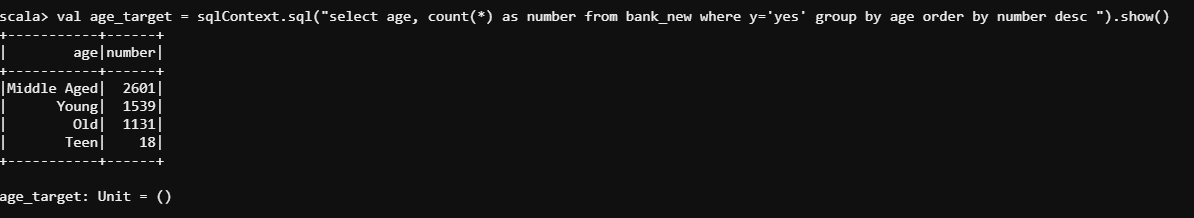




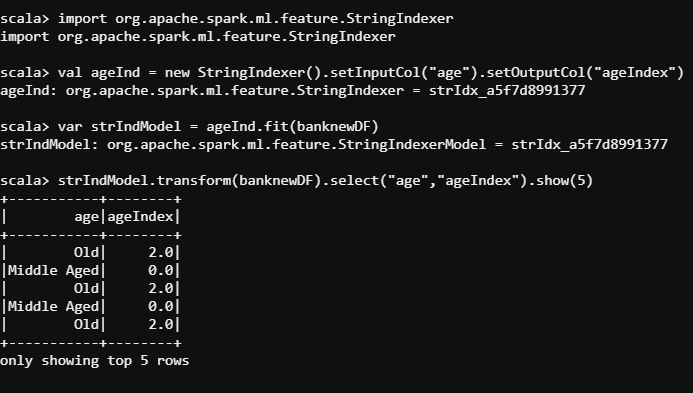
Replacing old “age” column with new “age” column



Running a query to see the age group which subscribed the most. It’s ‘Middle-Aged’



Pipeline:



Fitting the model:

