**Market Analysis in Banking Domain**

PROJECT SUBMISSION AS PART OF **BIGDATA HADOOP AND SPARK PROGRAMMER** BY SIMPLILEARN

TRAINER: PRADNYA PATRIKE

TRAINEE: KRISHNAMURTHY LAKSHMINARAYANA

[KRISHL@ZOHO.COM](mailto:KRISHL@ZOHO.COM)

SUBMISSION DATE: 04.July.2021

# Overview

## Project Background and Description

|  |  |
| --- | --- |
|  | 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) |

## Dataset Description

|  |  |
| --- | --- |
|  | 1 - age (numeric)  2 - job : type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')  3 - marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)  4 - education (categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')  5 - default: has credit in default? (categorical: 'no', 'yes', 'unknown')  6 - housing: has housing loan? (categorical: 'no', 'yes', 'unknown')  7 - loan: has personal loan? (categorical: 'no', 'yes', 'unknown')  # related to the last contact of the current campaign:  8 - contact: contact communication type (categorical: 'cellular', 'telephone')  9 - month: Month of last contact (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')  10 - day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri') © Copyright 2016, Simplilearn. All rights reserved. P a g e | **2**  11 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (example, if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call “y” is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.  # other attributes:  12 - campaign: number of times a customer was contacted during the campaign (numeric, includes last contact)  13 - pdays: number of days passed after the customer was last contacted from a previous campaign (numeric; 999 means customer was not previously contacted)  14 - previous: number of times the customer was contacted prior to (or before) this campaign (numeric)  15 - poutcome: outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')  # social and economic context attributes  16 - emp.var.rate: employment variation rate―quarterly indicator (numeric)  17 - cons.price.idx: consumer price index―monthly indicator (numeric)  18 - cons.conf.idx: consumer confidence index―monthly indicator (numeric)  19 - euribor3m: euribor 3 month rate―daily indicator (numeric)  20 - nr.employed: number of employees―quarterly indicator (numeric)  Output variable (desired target):  21 - y - has the customer subscribed a term deposit? (binary: 'yes', 'no') |

SPARK SQL CodE

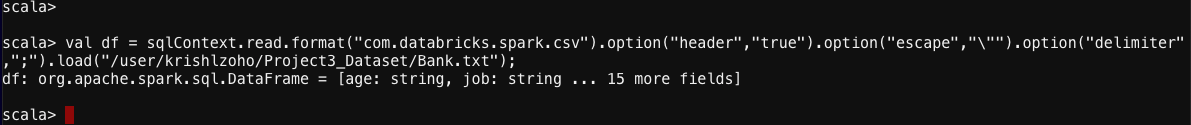
## Load data and create a Spark data frame

A picture containing text

Description automatically generated

Text

Description automatically generated



A picture containing calendar

Description automatically generated

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

Text

Description automatically generated

## Give marketing failure rate

Text

Description automatically generated

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

Text

Description automatically generated

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

Text

Description automatically generated with medium confidence

Text

Description automatically generated

## Check if age matters in marketing subscription for deposit

Text

Description automatically generated

**## Conclusion -**

From the results, we could conclude that AGE really matters; Age between from **30** to **36** shows more Promising Subscriptions for Deposits.

## Check if marital status mattered for a subscription to deposit

Text

Description automatically generated

**## Conclusion -** From the results, we could conclude that **Married** Customers are topping the Subscription.

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

Text

Description automatically generated

## **Conclusion** -

From the results, we could conclude that **Single** Customers with an Age group from 26 to 32 are showing more interest for Subscriptions.

## Do feature engineering for the bank and find the right age effect on the campaign.

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

## **Conclusion** - From the results, we could conclude that **Middle Aged** Customers (between age-group of 26 to 59) should be the Targeted customers as they are making the Most subscriptions.

\*\* In particular, ‘**Married and Middle-age**’ customers are Topping the Subscriptions.

## Feature Engineering with effect of Campaigns on Age Groups (with some more Deep-Dive).

\*\* Middle-age is further divided into sub-groups of **10** years.

Text

Description automatically generated

Text

Description automatically generated

**Most affected with Campaigns::**

* Married: 30 to 60+ [ ag3040, ag4050, ag5060 & ag60plus ]
* Single: **<** 40 [ ag30minus & ag3040 ]

**Less affected with Campaigns::**

* Divorced: -30 to 60+ [ ag30minus, ag3040, ag4050, ag5060 & ag60plus ]

<End of Document>