1. Load data and create a Spark data frame.

Following script is used to load the data.

val bank\_df = spark.read.option("multiline",true).json("/user/ribishnugmail/bank\_edited.json")

A picture containing calendar

Description automatically generatedbank\_df.show()

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

Following codes are used to generate maximum, minimum and mean of the targeted customer.

The result shows that maximum age of the customer is 95 and minimum is 18 and average is 40.9

Text

Description automatically generated

1. Check the quality of customers by checking average balance, median balance of customers.

Below are provided the codes to show median of the target aged group and the median balance is 448 and average balance is 1362.27. This shows the data is right skewed.

Text

Description automatically generated

1. Check if age matters in marketing subscription for deposit.

Text

Description automatically generated

Data show that age group does not matter in marketing subscription for deposit as there is no defined relationship between subscription and age group.

1. Check if marital status mattered for a subscription to deposit.

Text

Description automatically generated

The above data show that subscription depends upon marital status as married subscribe significantly greater by almost 4-fold compared to divorced group and single group subscribe nearly 3-fold compared to divorced group.

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

Text

Description automatically generated

Clearly, the result shows the subscription depends upon age and marital status.

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

A picture containing text

Description automatically generated

Text

Description automatically generated