Q1 . Load data and create a Spark data frame ?

Ans1 : val df = spark.read.format(“csv").option("delimiter",";").option("header","true").load("/user/santoshsaxenaofficialgmai/banking.csv")

Q2 . Give marketing success rate (No. of people subscribed / total no. of entries) ?

Ans2 : df.filter('y === “yes").count.toFloat/df.count.toFloat

Q2 B . Give marketing failure rate ?

Ans3. df.filter('y === "no").count.toFloat / df.count.toFloat

Q1 . Give the maximum, mean, and minimum age of the average targeted customer ?

Ans1 : df.agg(min("age") , max("age") , round(avg(“age"),2)).show

Q2. Check the quality of customers by checking average balance, median balance of customers ?

Ans2 : sql("select avg(balance) as mean , percentile\_approx(balance,0.5) as median from banking”).show

Q3 . Check if age matters in marketing subscription for deposit

Ans3. sql("select age as age , count(\*) as count from banking where y = 'yes' group by age order by count desc limit 10”).show

Q4. Check if marital status mattered for a subscription to deposit

Ans 4 . sql("select marital as marital , count(\*) as count from banking where y = 'yes' group by marital order by count desc limit 10”).show

Q5. Check if age and marital status together mattered for a subscription to deposit scheme

Ans5. sql("select age as age , marital as marital , count(\*) as count from banking where y = 'yes' group by age,marital order by count desc limit 10”).show

Q6. Do feature engineering for the bank and find the right age effect on the campaign. ?

Ans 6. sql("select case when age < 25 then 'Young' when age between 25 and 60 then 'Middle' when age > 60 then 'older' end as age\_range ,y as y ,count(\*) as count from banking group by age\_range, y “).show