**1. Load data and create Spark data frame**

hdfs dfs -put /mnt/home/prabhatvarma1985gmail/BDH\_Project1/data.csv Proj1

var rdd = sc.textFile("Proj1/data.csv")

**On Spark Console – Cleaning and Creating Data Frame**

rdd.take(3).foreach(println)

formatted\_data.take(3).foreach(println)

formatted\_data.coalesce(1).saveAsTextFile("Proj1/formatted/")

var df = spark.read.format("csv").option("header", "true").option("delimiter", ";").option("inferSchema", "true").load("Proj1/formatted/")

**2. Give marketing success rate. (No. of people subscribed / total no. of entries**

var suc\_rate = df.filter($"y" === "yes").count.toFloat / df.count.toFloat \*100

**2a Give marketing failure rate**

val fail = df.filter($"y" === "no").count.toFloat / df.count.toFloat \*100

**3. Maximum, Mean, and Minimum age of average targeted customer**

import org.apache.spark.sql.functions.{min, max, avg}

df.agg(max($"age"),min($"age"), avg($"age")).show()

**4. Check quality of customers by checking average balance, median balance of customers**

import org.apache.commons.math3.stat.descriptive

df.createOrReplaceTempView("sample")

val minmaxavgBal = sql("SELECT max(balance) as max, min(balance) as min, avg(balance) as average FROM sample");

minmaxavgBal.show()

**5. Check if age matters in marketing subscription for deposit**

df.groupBy($"y".alias("Did the customer Subscribed")).agg(count($"age").alias("Age Count")).show

**6. Check if marital status mattered for subscription to deposit.**

df.groupBy($"y".alias("Did the customer Subscribed")).agg(count($"marital").alias("Marital Count")).show

**7. Check if age and marital status together mattered for subscription to deposit scheme**

df.groupBy("marital","age").count.sort($"count").show

**8. Do feature engineering for column—age and find right age effect on campaign**

import org.apache.spark.sql.functions.udf

var bank\_f = df.filter($"y" === "yes")

var age\_cat = bank\_f.withColumn("age\_cat", when($"age" < 25, "young").otherwise(when($"age" > 60, "old").otherwise("mid")))

var result = age\_cat.groupBy("age\_cat").count()

result.show()