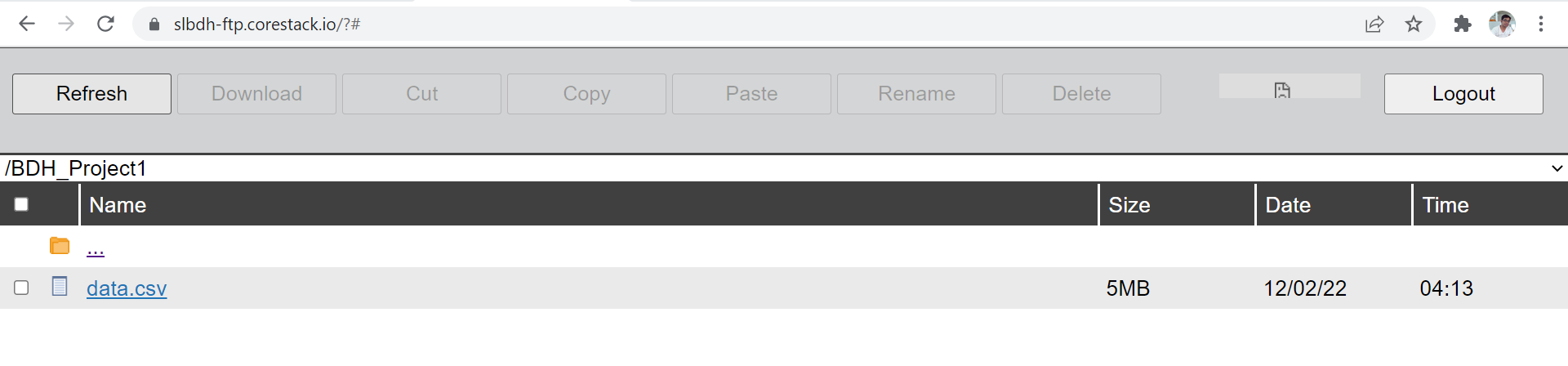
Project 1 –

**Move File from Local To WebConsole**



**Move File To HDFS**

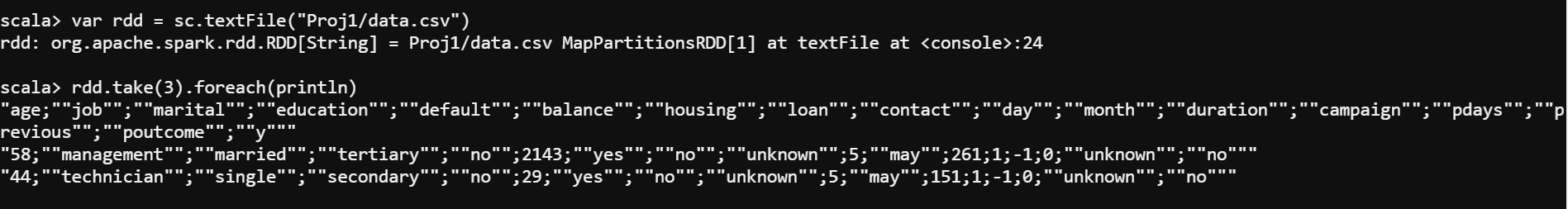
hdfs dfs -mkdir Proj1

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

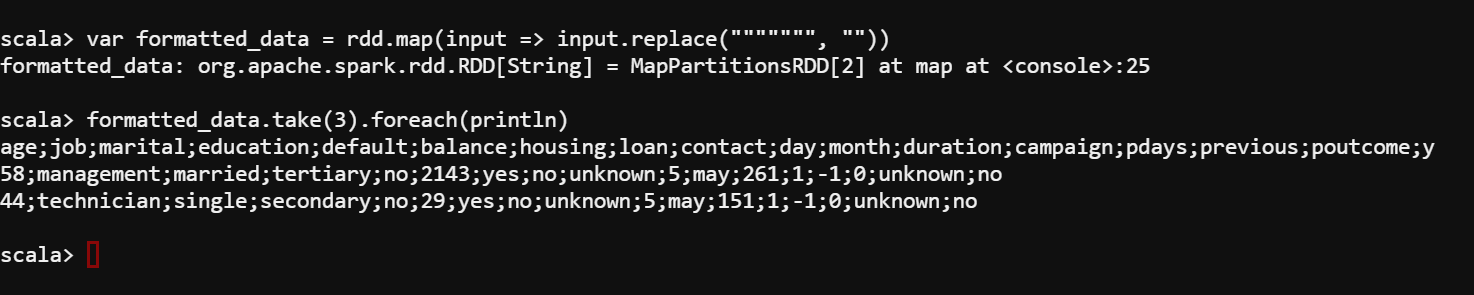
**Spark Shell**

**Question 1 :** Load data and create Spark data frame

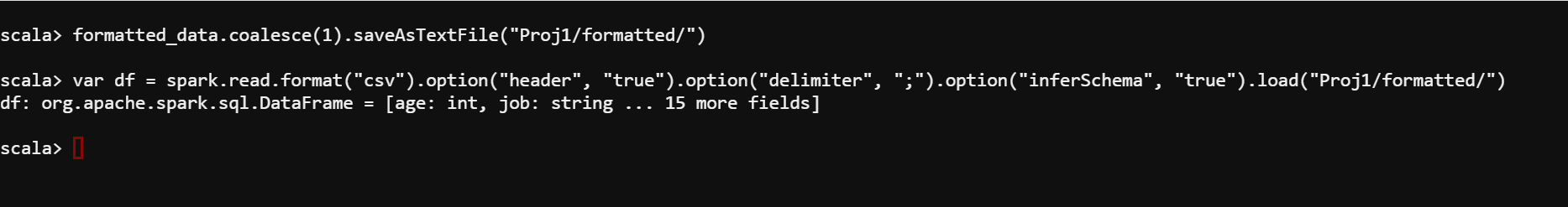
Loading Data



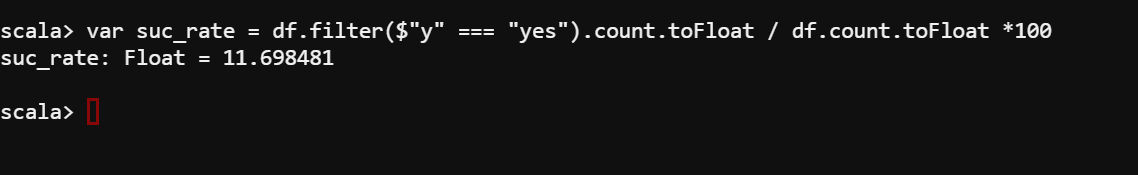
Cleaning Data (Replacing double quote from Data Header and Data Value)



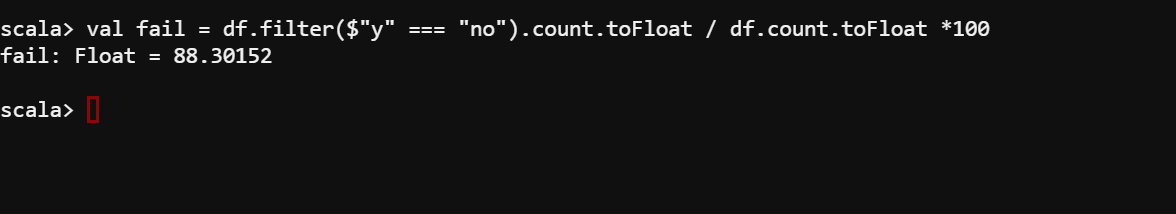
Creating DataFrame as **df**



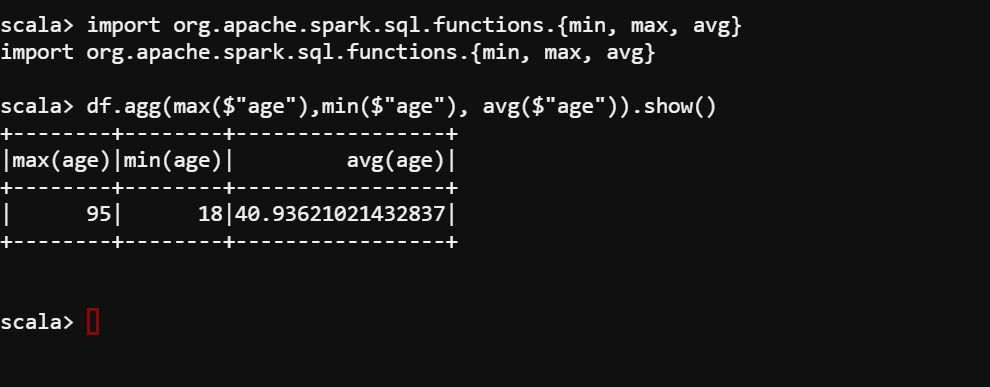
**Question 2 :** Give marketing success rate. (No. of people subscribed / total no. of entries)



**Question 2a :**Give marketing failure rate

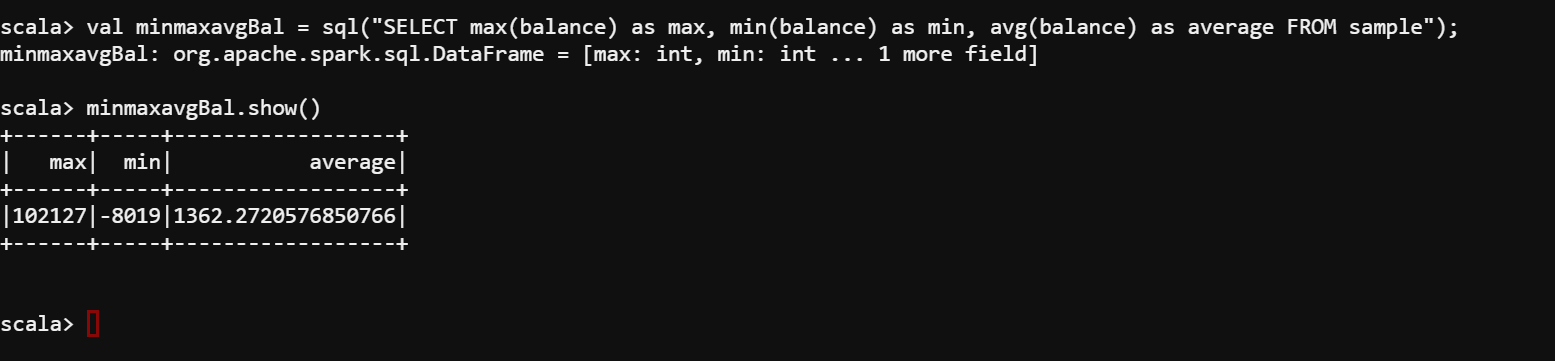


**Question 3** : Maximum, Mean, and Minimum age of average targeted customer



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

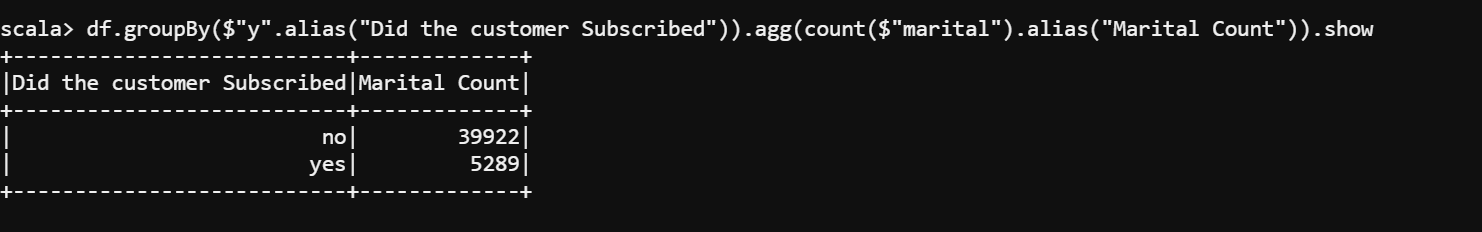
***Note – Not considered Median in this below query***



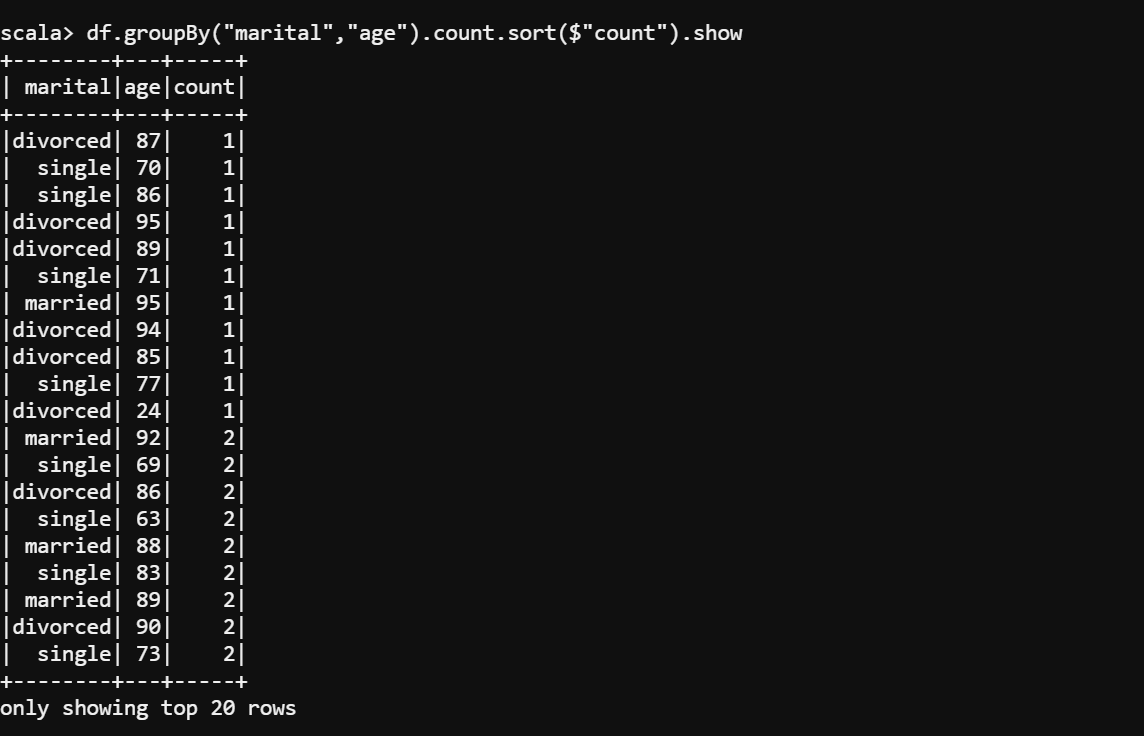
**Question 5 :** Check if age matters in marketing subscription for deposit



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



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



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

