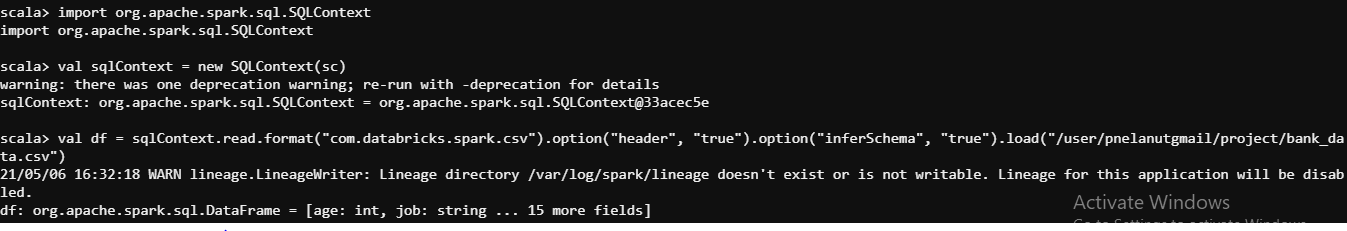
# PROJECT

**Copying Dataset to HDFS**



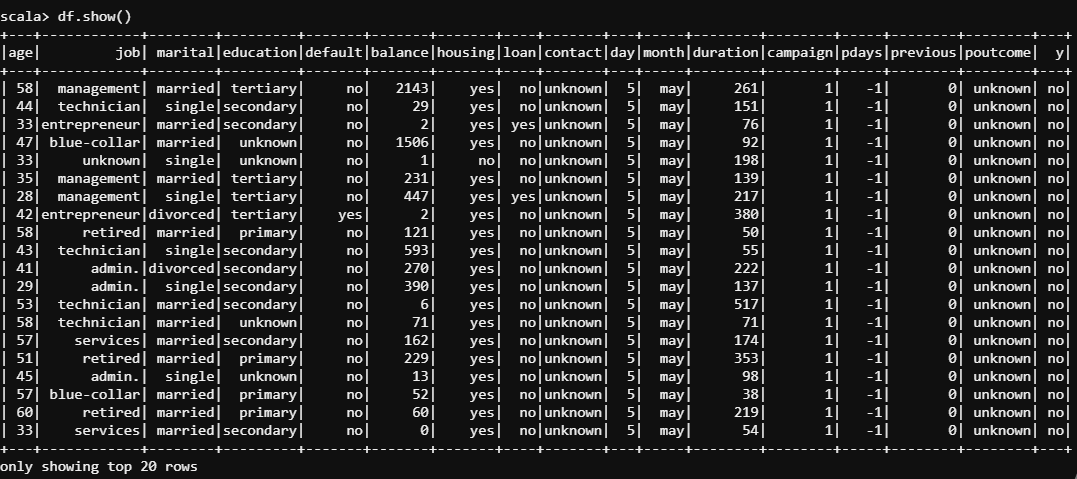
1. **1. Loading Data into Data frame**



**Print Schema**



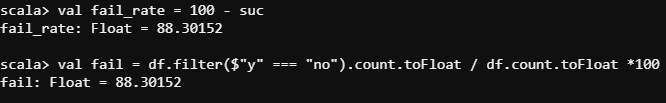
**Checking Data in Data frame**



1. **2. Marketing Success rate (in percentage)**

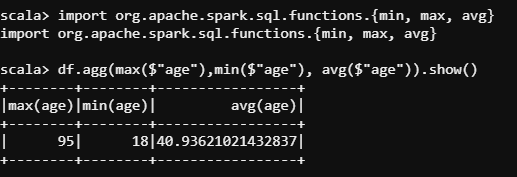


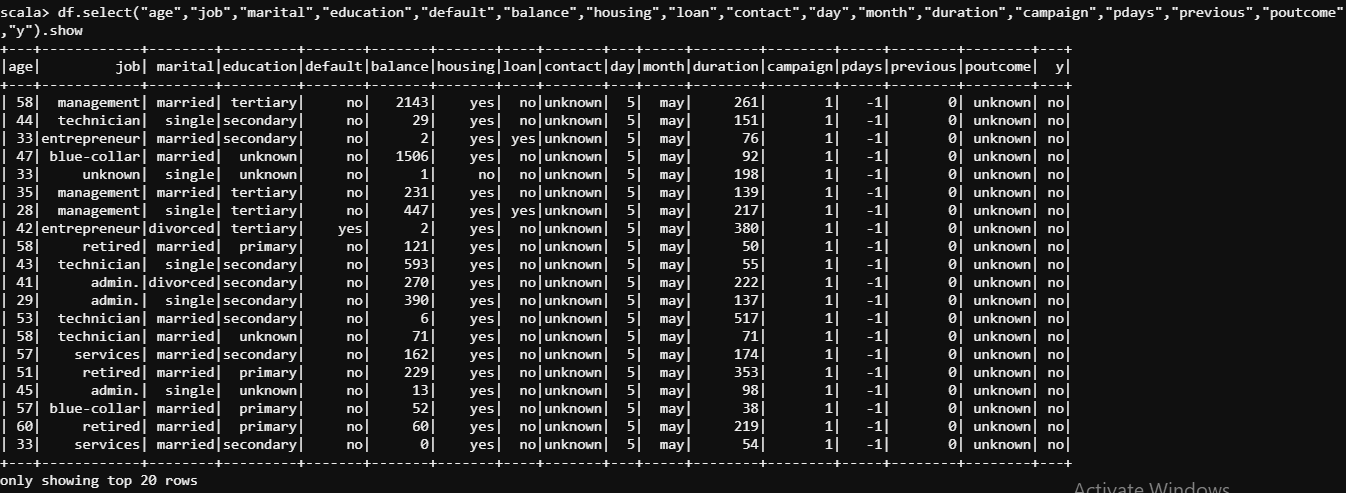
**Marketing Fail rate (in percentage)**



Since the calculation by count of no and 100 – success is same we can conclude there is no null value under ‘y’

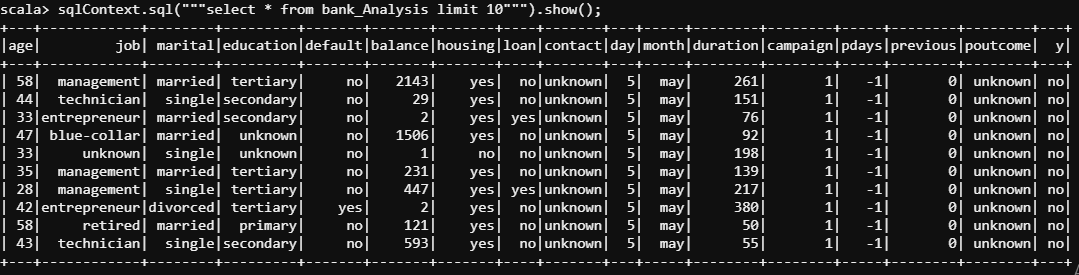
1. **3. Maximum, Mean, and Minimum age of the average targeted customer**

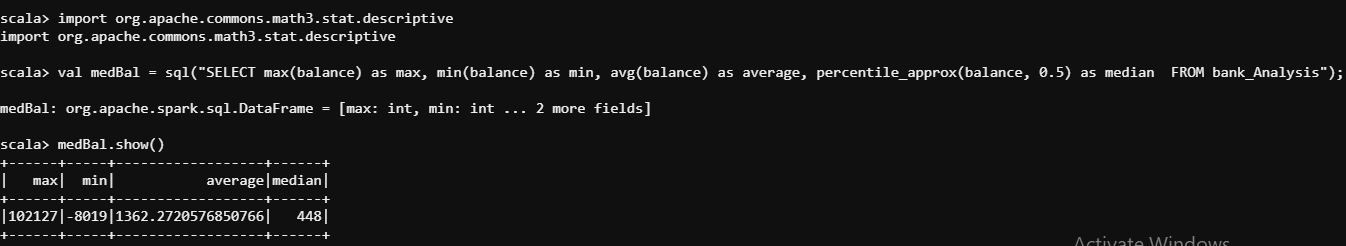
****

1. **4. Quality of customers by checking Average balance, Median balance of customers**
2. **creating table for Spark SQL**
3. ****

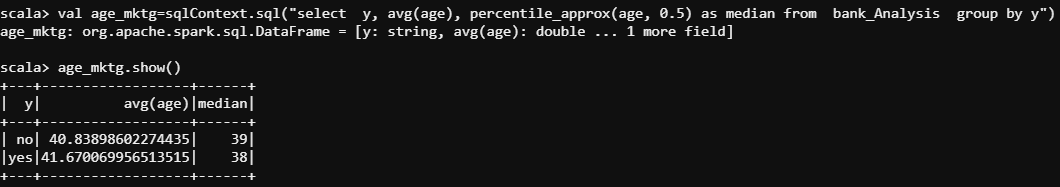
****

**Use sqlContext to calculate data**

****

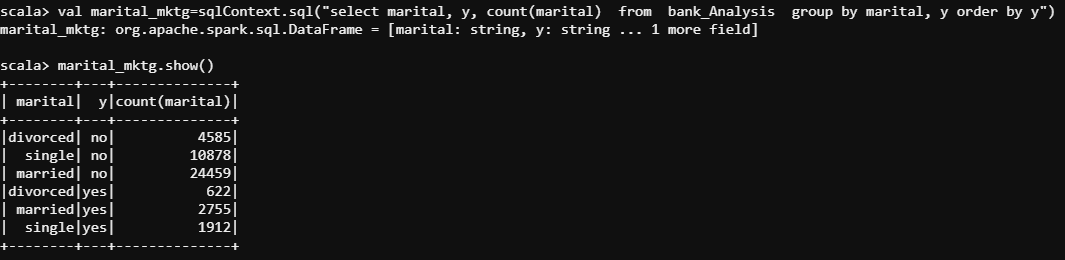
****

1. **5. Age matters in marketing subscription for deposit?**

****

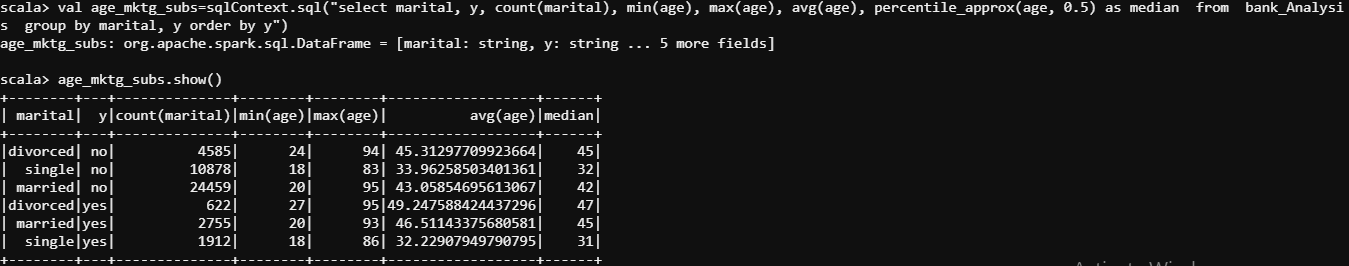
As per the above data we can conclude age doesn’t matter for subscription.

1. **6. Marital status mattered for a subscription to deposit?**

****

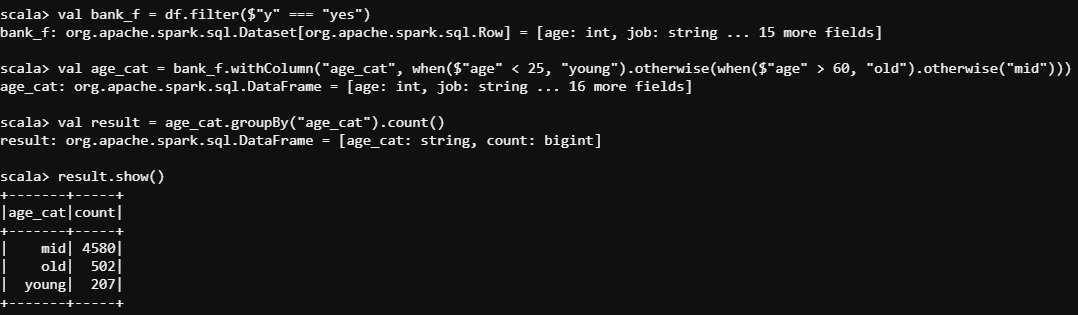
The data is ambiguous, conclusion about marital status a matter for subscription cannot be driven purely by this information.

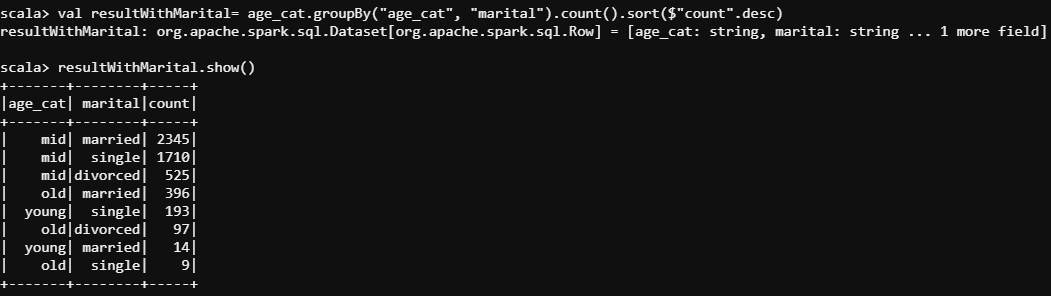
1. **7. Age and marital status together mattered for a subscription to deposit scheme?**

****

Though parameters are closely indicating for subscription and no subscription, a minor observation can be made such as Married and divorced people who are elderly and single people who are younger are more inclined for subscription.

1. **8. Feature engineering for the bank and find the right age effect on the campaign**

****

****