**PROJECT 3**

**Market Analysis in Banking Domain (snapshots)**

**DESCRIPTION**

**Background and Objective:**

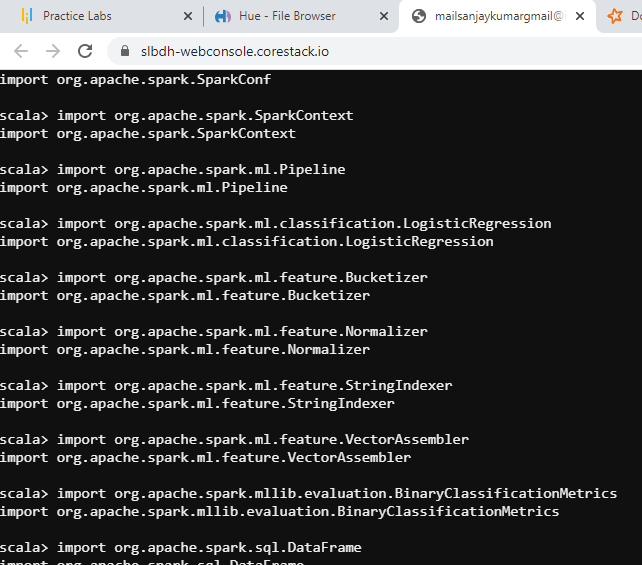
**Your client, a Portuguese banking institution, ran a marketing campaign to convince potential customers to invest in a bank term deposit scheme.   
The marketing campaigns were based on phone calls. Often, the same customer was contacted more than once through phone, in order to assess if they would want to subscribe to the bank term deposit or not. You have to perform the marketing analysis of the data generated by this campaign.**

**Domain: Banking (Market Analysis)**

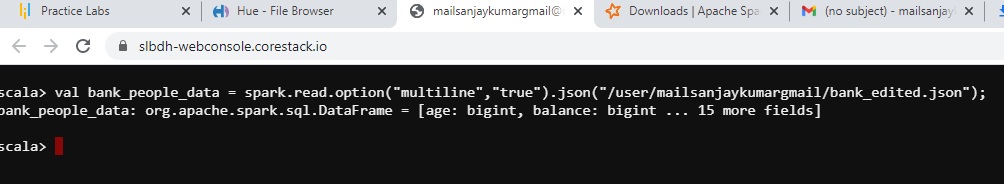
**Following are the steps:**

**Open spark-shell**

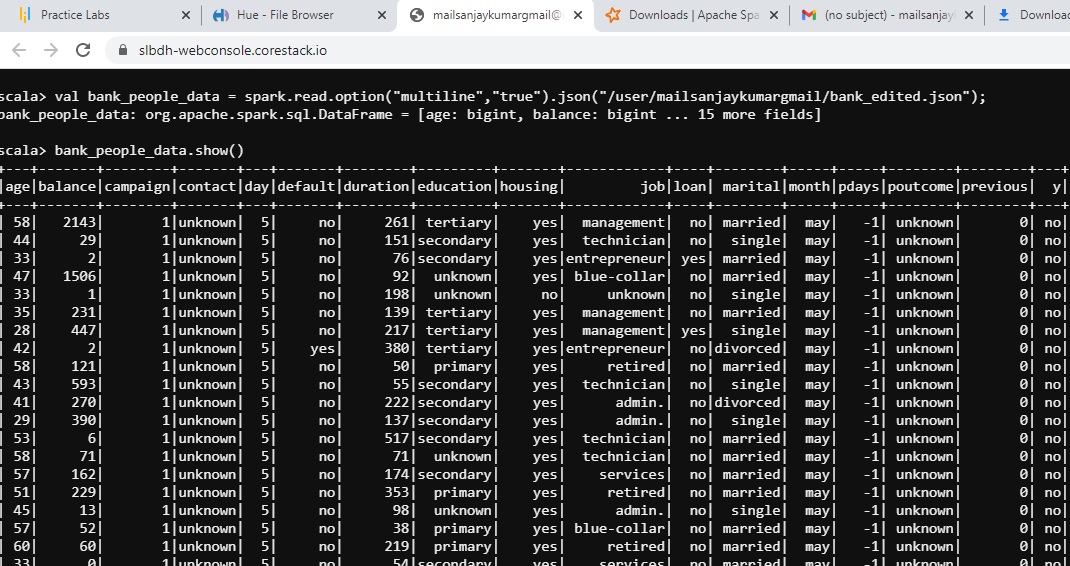
**Importing required spark libraries**



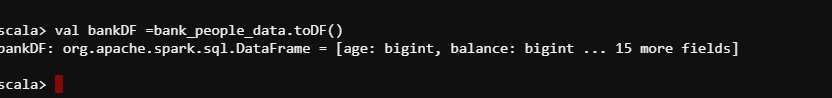
1. **Load data and create a Spark data frame**

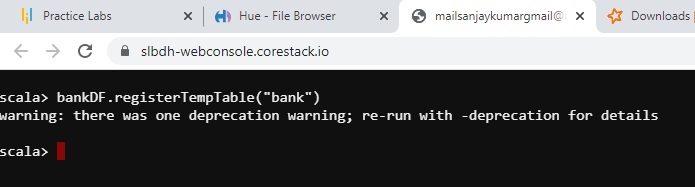


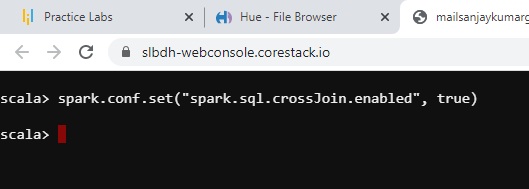
* **show the loaded data**



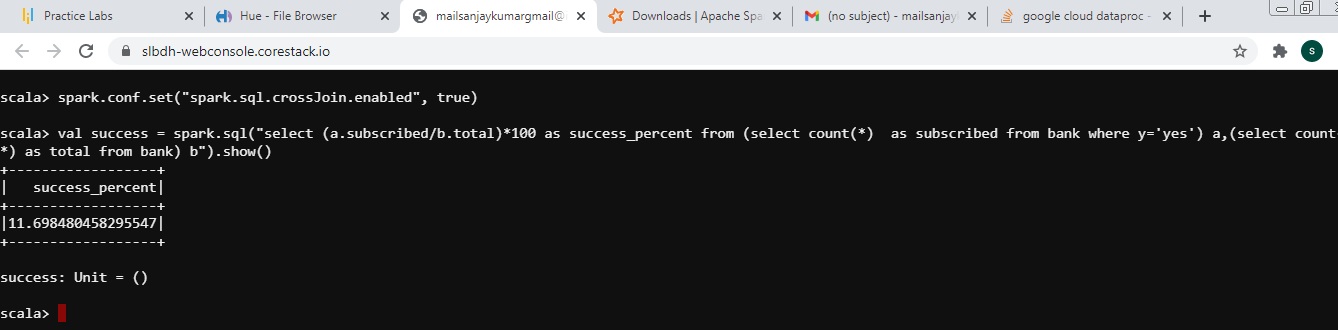
* **convert to dataframe**



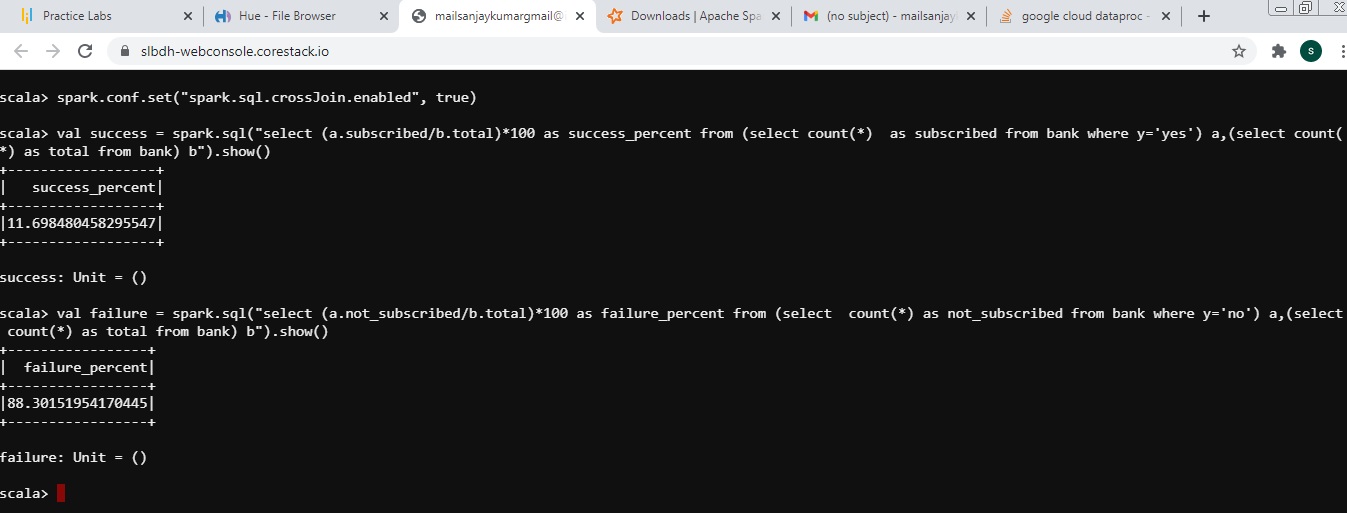
****



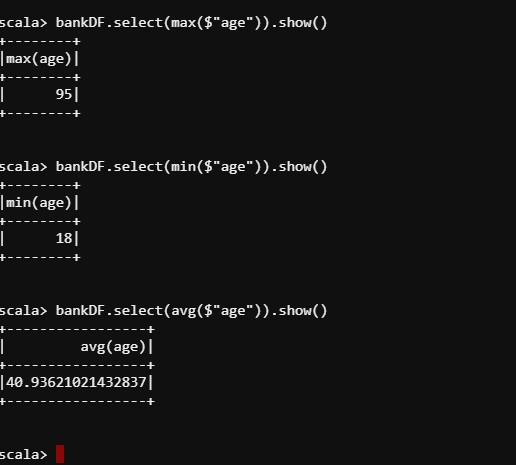
1. **Give marketing success rate (No. of people subscribed / total no. of entries)**



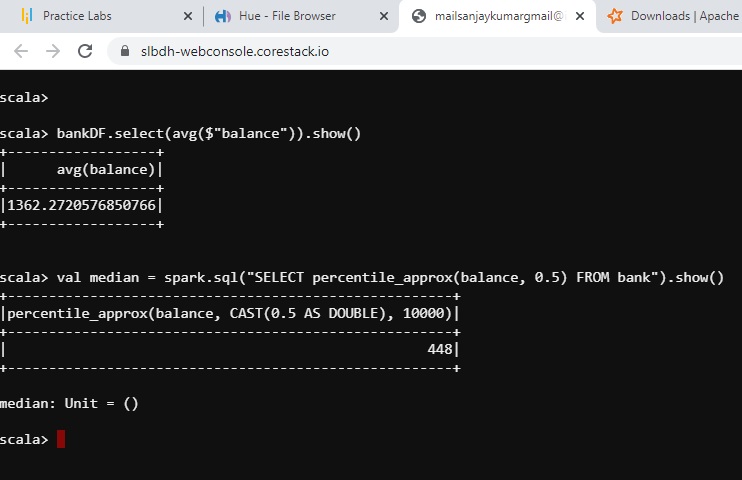
* **Give marketing failure rate**



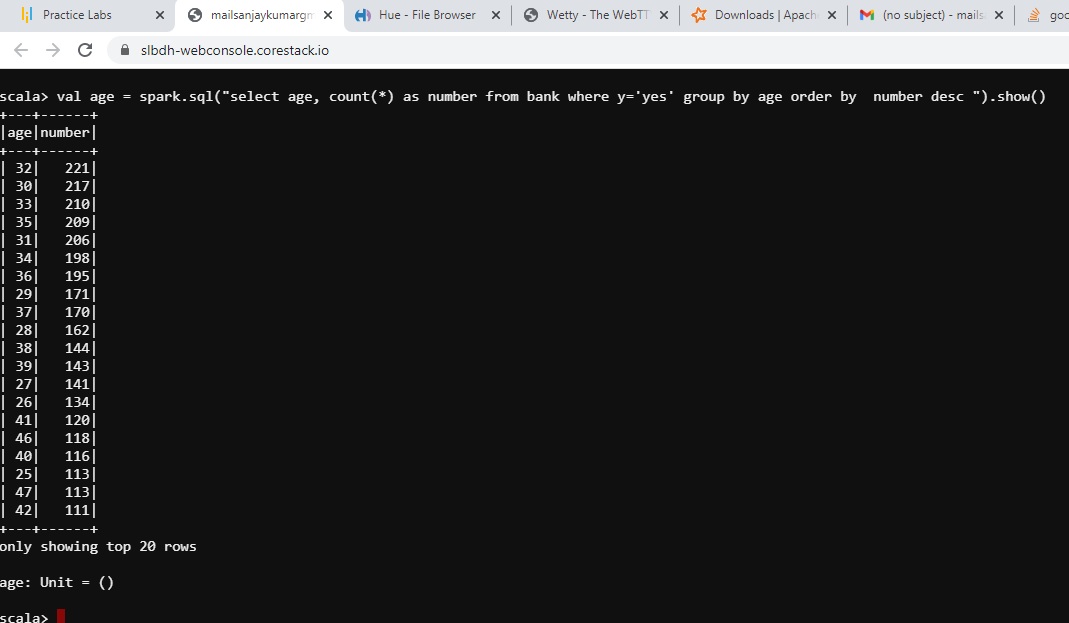
1. **Give the maximum, mean, and minimum age of the average targeted customer**



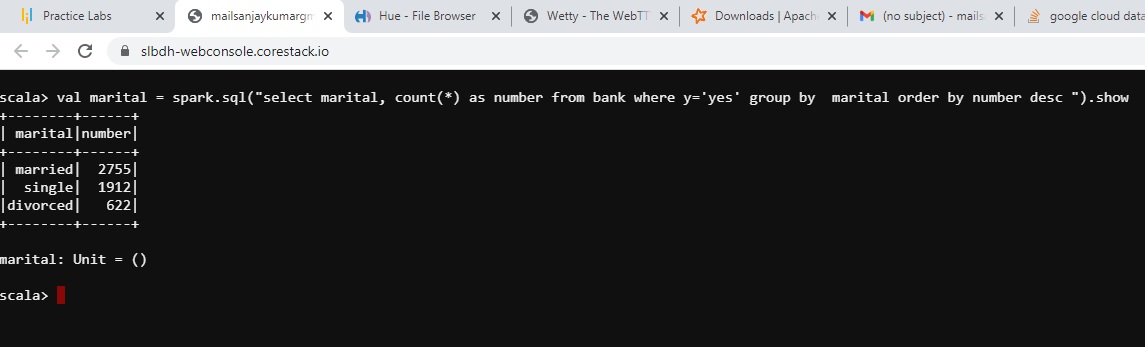
1. **Check the quality of customers by checking average balance, median balance of customers**



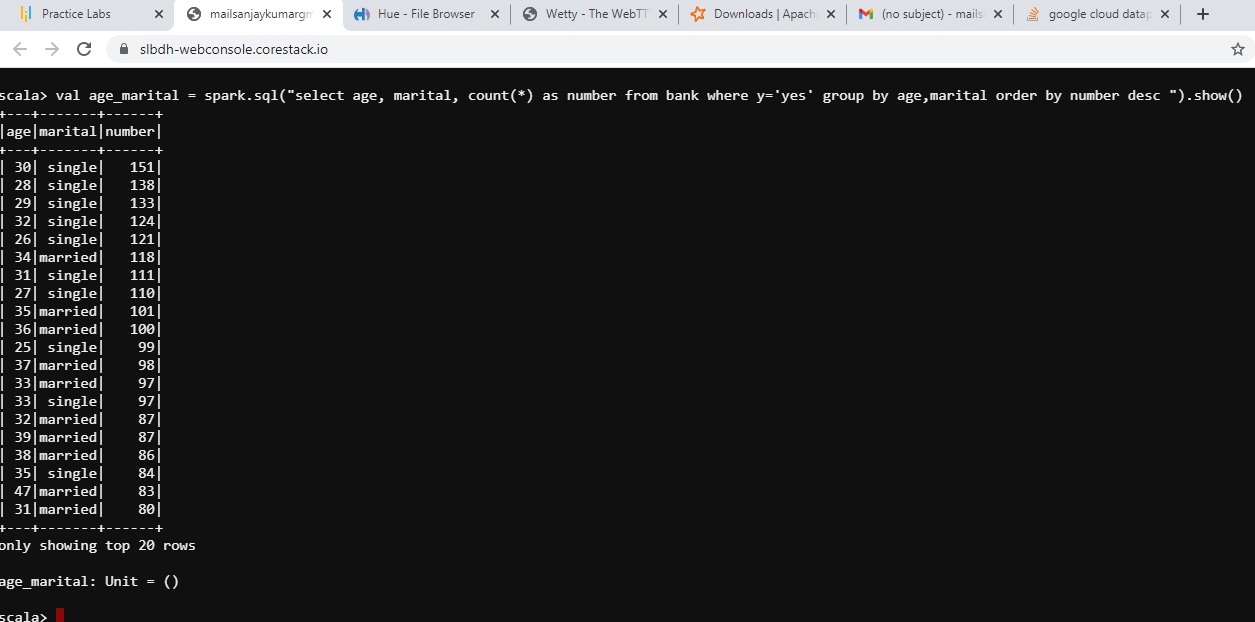
1. **Check if age matters in marketing subscription for deposit**



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

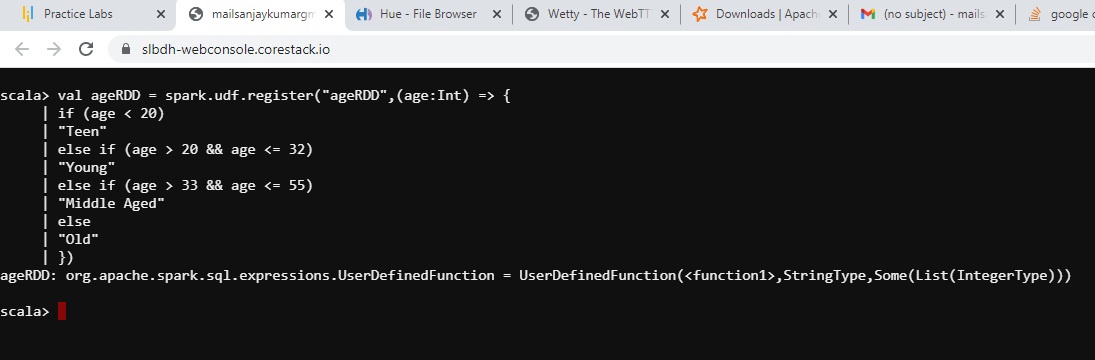


1. **Check if age and marital status together mattered for a subscription to deposit scheme**

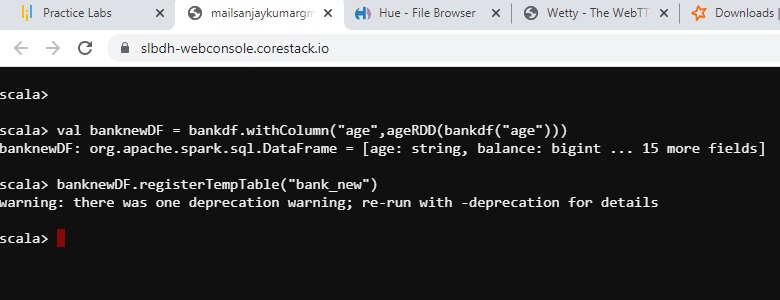


1. **Do feature engineering for the bank and find the right age effect on the campaign**.

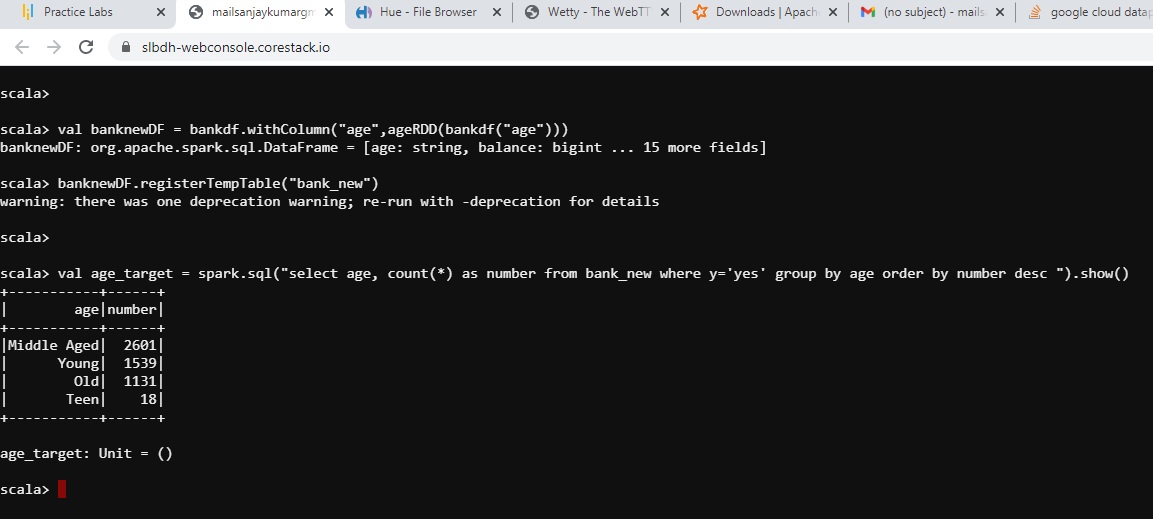
--Defining a new UDF with which we will generate new features.We divide the age groups into 4 categories.



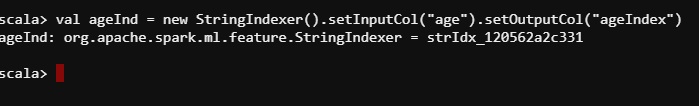
--Replacing old “age” column with new “age” column



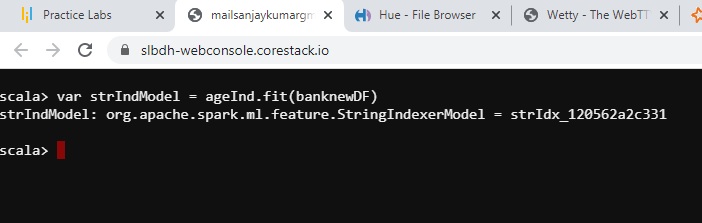
--Running a query to see the age group which subscribed the most. We see it’s ‘Middle-Aged’



--Pipeline



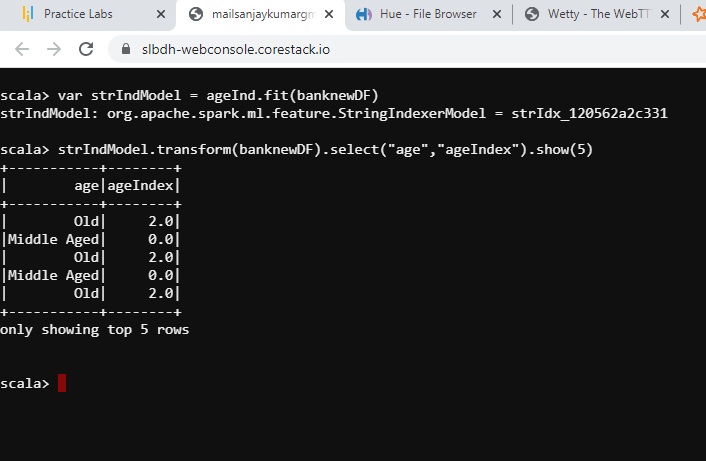
--Fitting the model



--StringIndexerModel.transform() assigns the generated index to each value of the column in the

given DataFrame.

--Middle aged is the most frequent word in this data, so it is given index 0



**So we can conclude from the Feature Engineering that It is the ‘Middle Aged’ people between age 33 and 55 who should be the targeted customers as they subscribe the most**