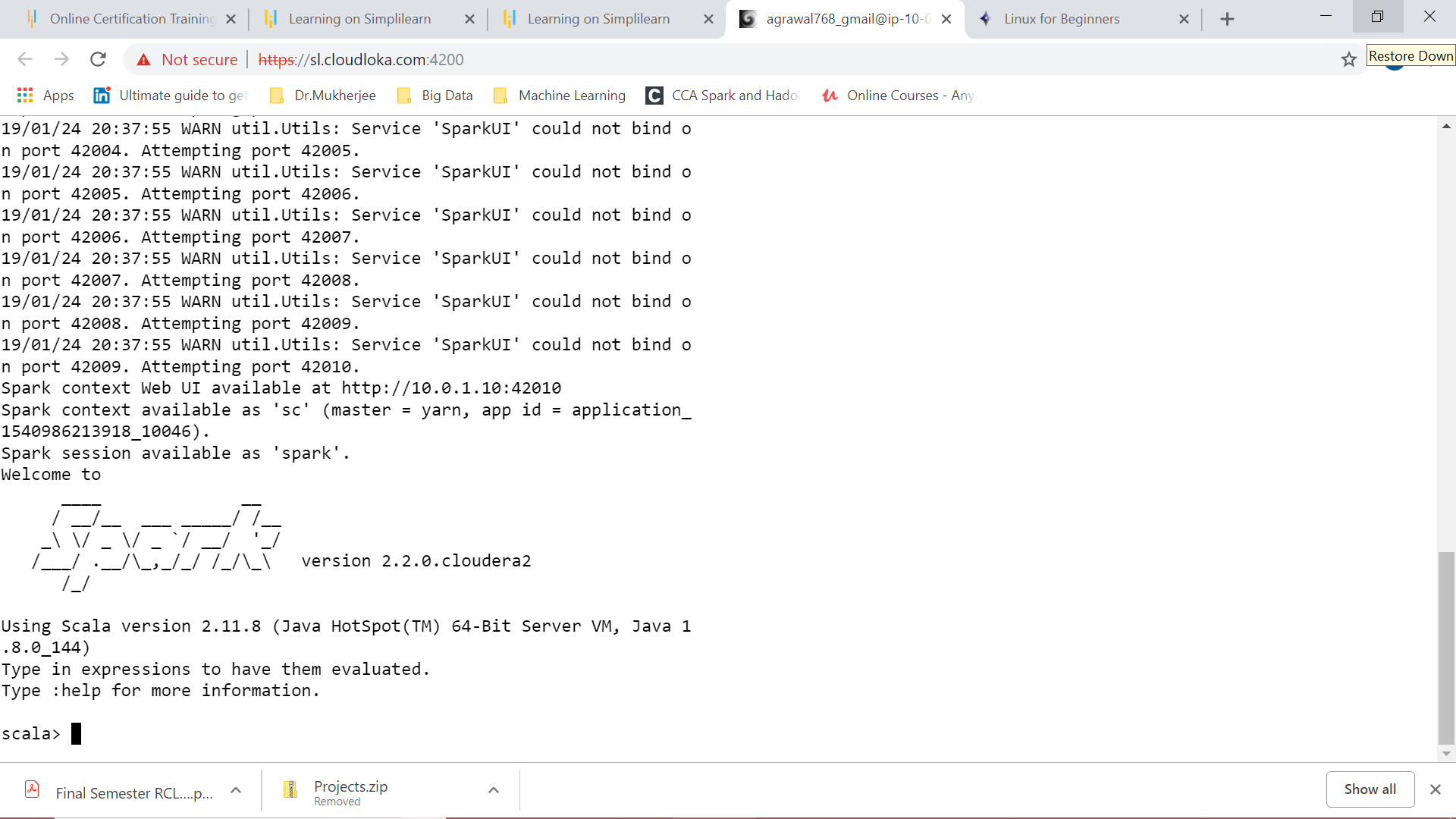
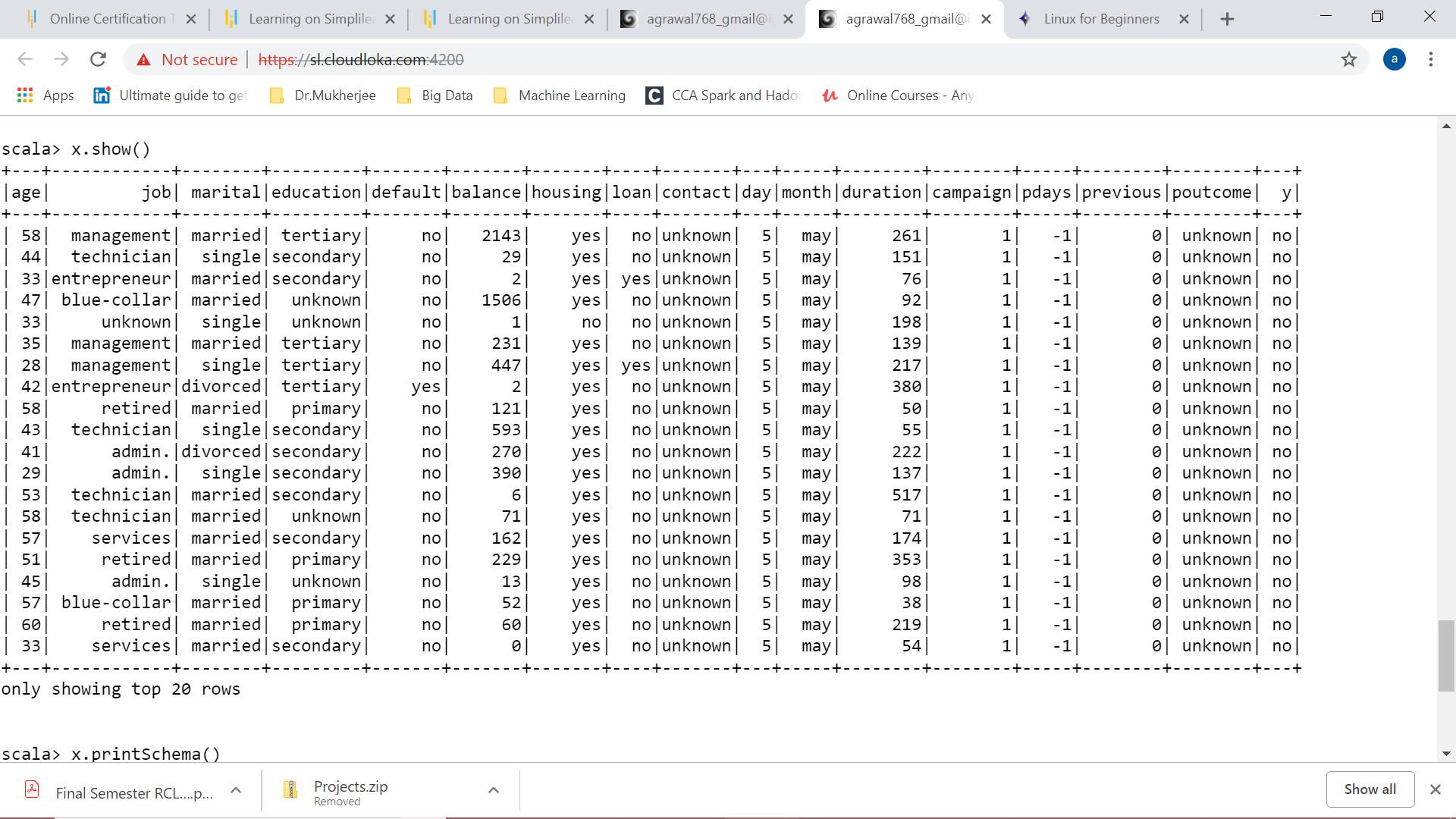
Project 1- Spark Data Analysis using Dataframes

1. Starting spark

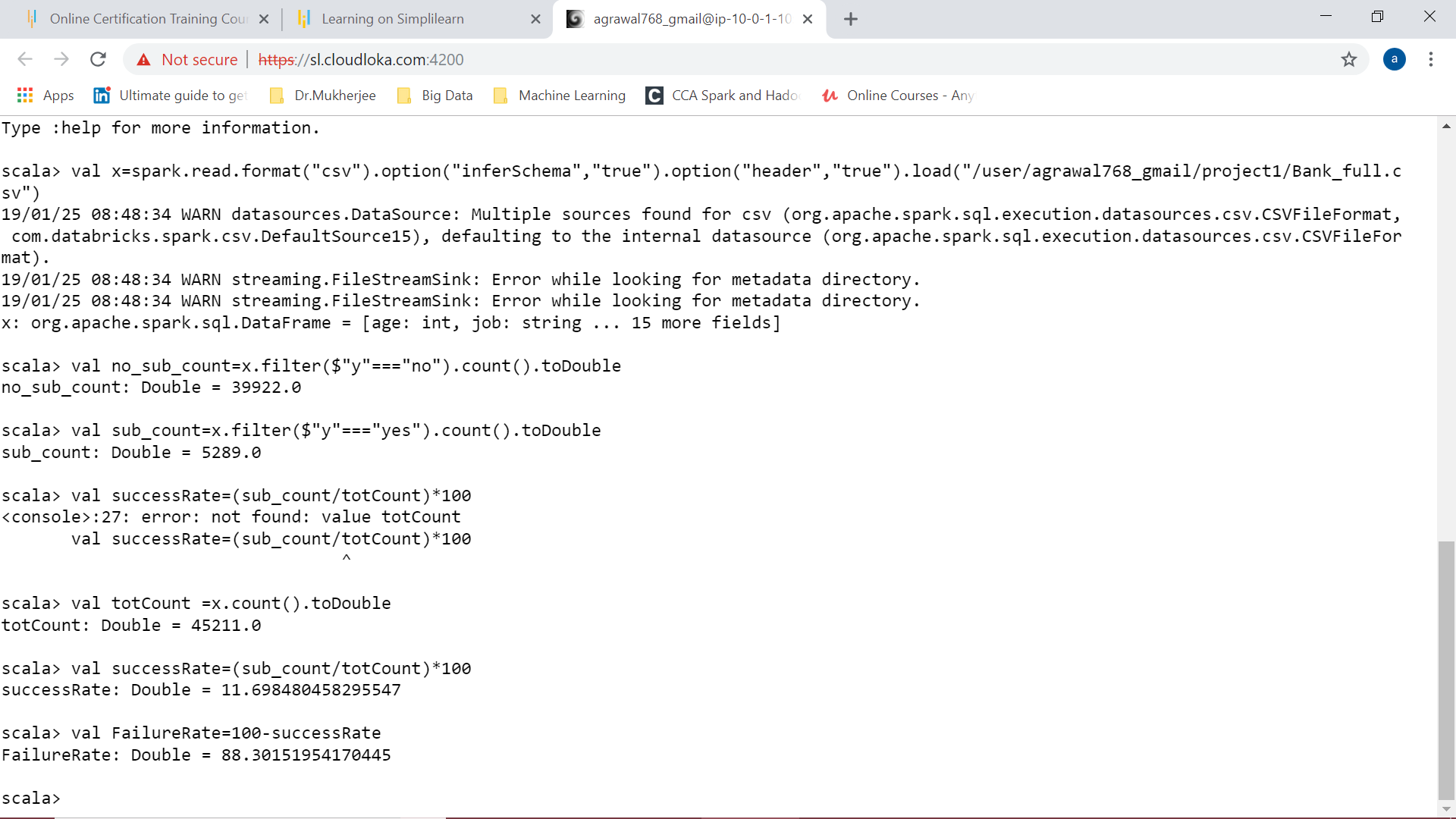


Printing the data frame

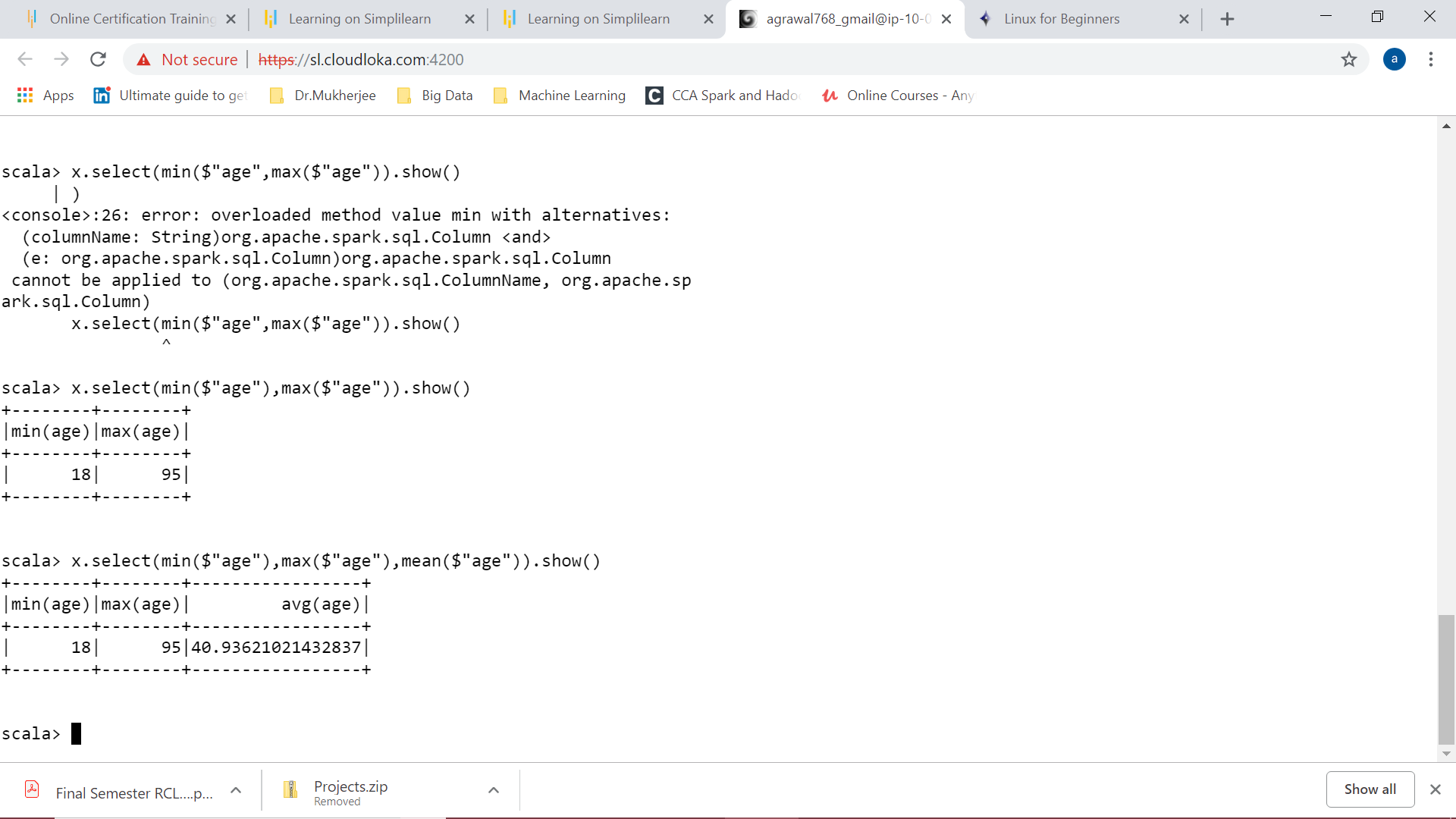


1. Total number of customers, Subscribed Customer count, Non Subscribed Customer count

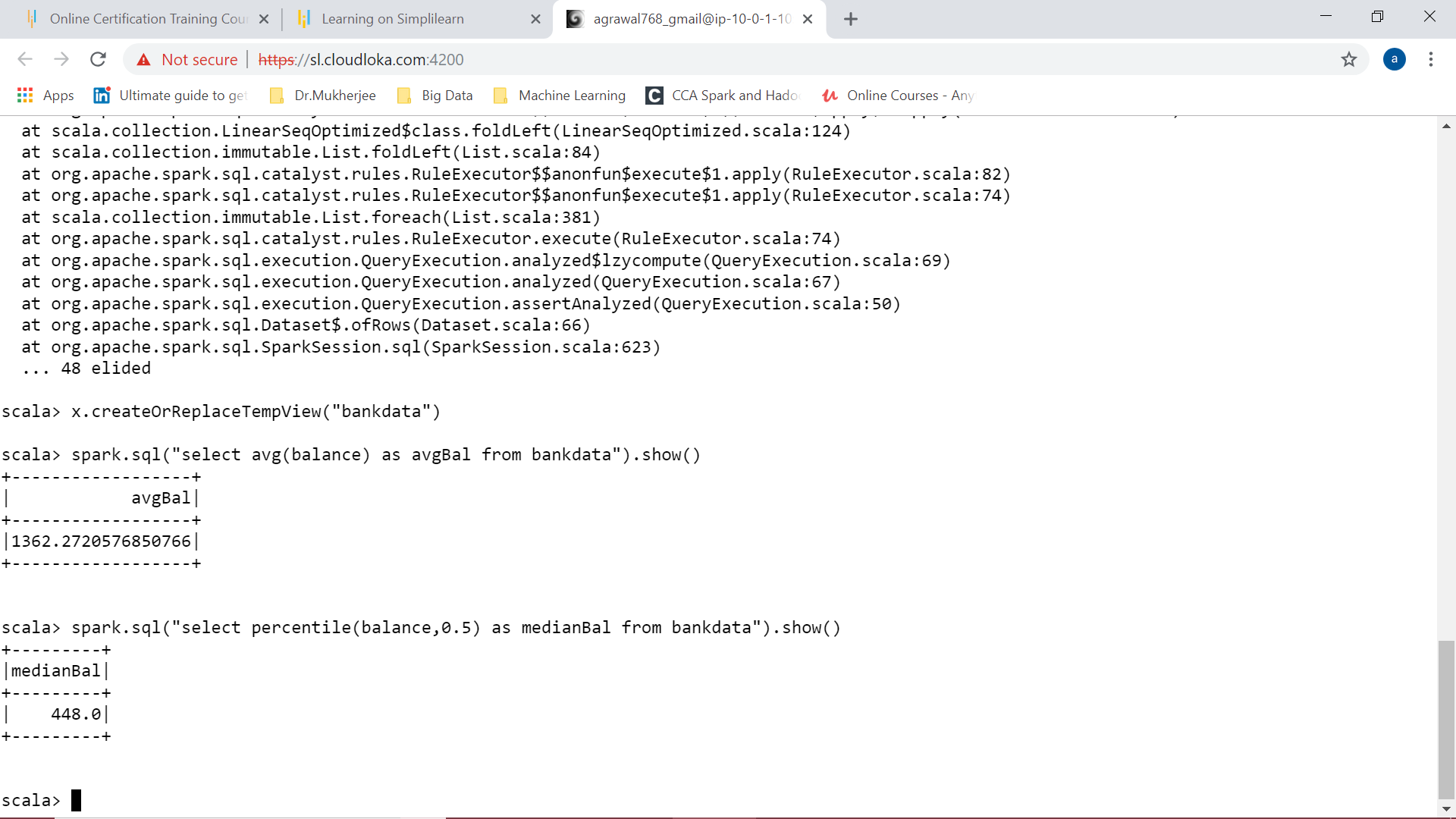
Success ratio and Failure ratio



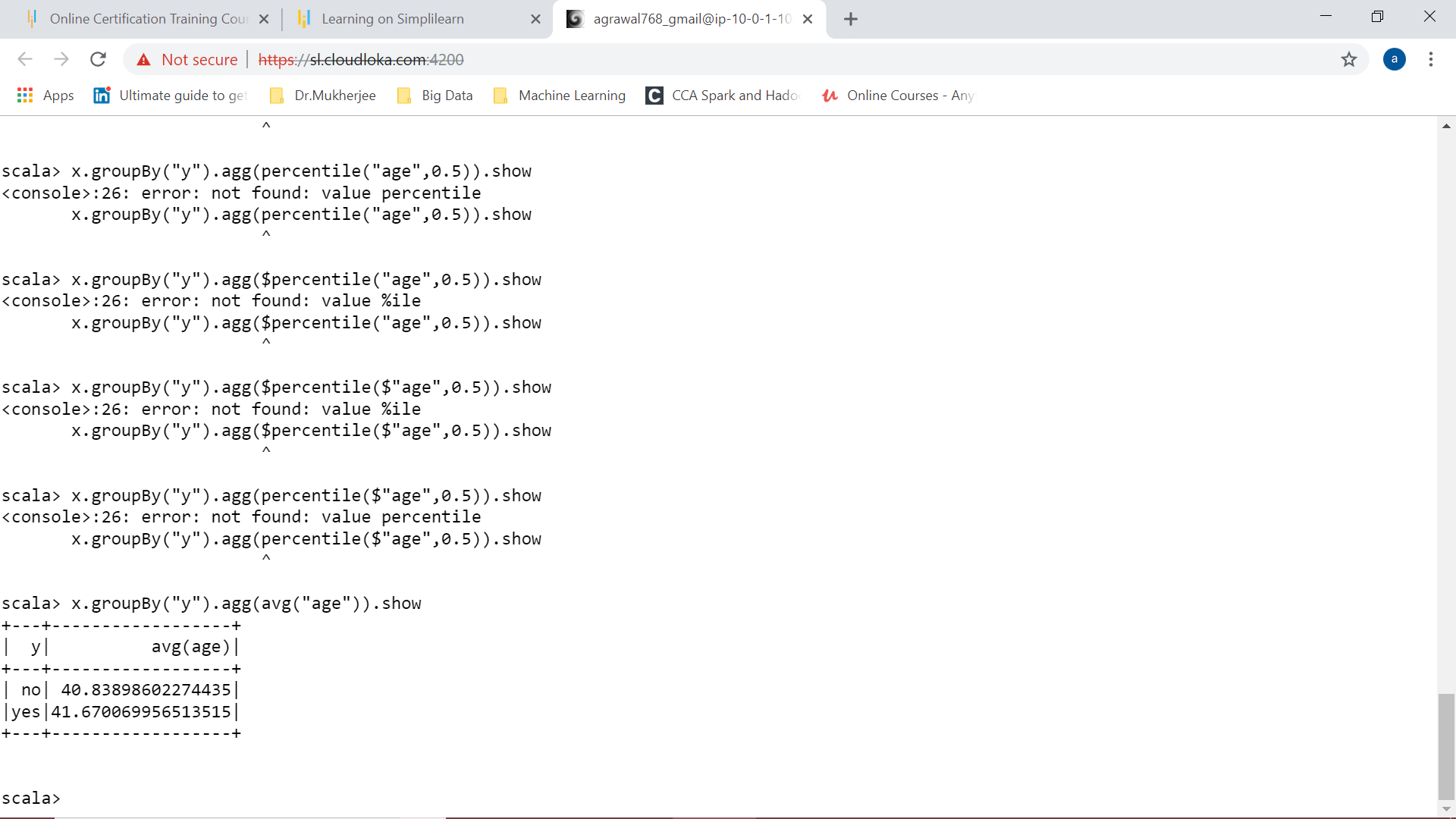
1. Finding mean, min and max of the age



1. Finding the avg and median balance

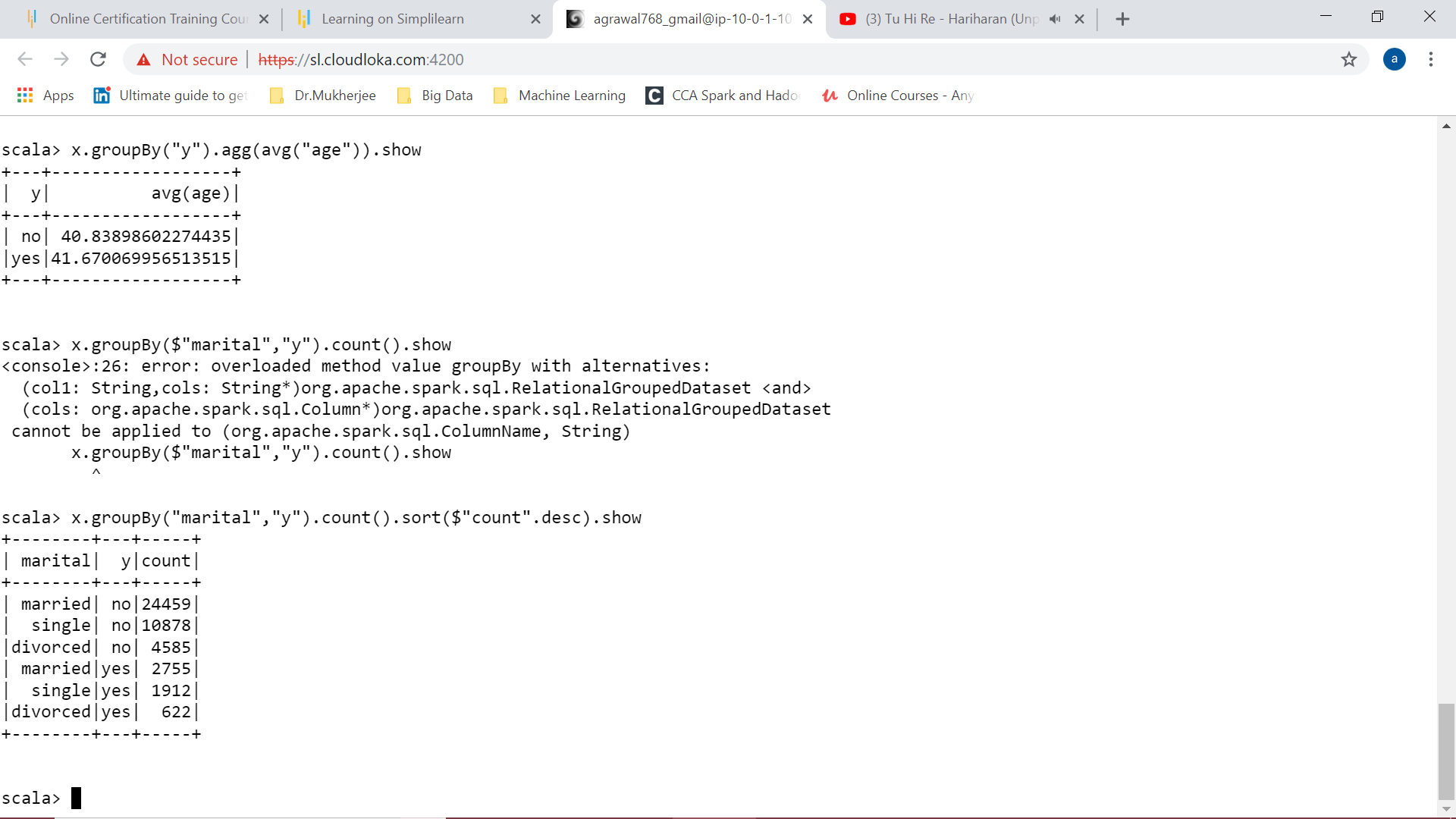


1. Age does not matter for subscription because the mean remains the same for both:

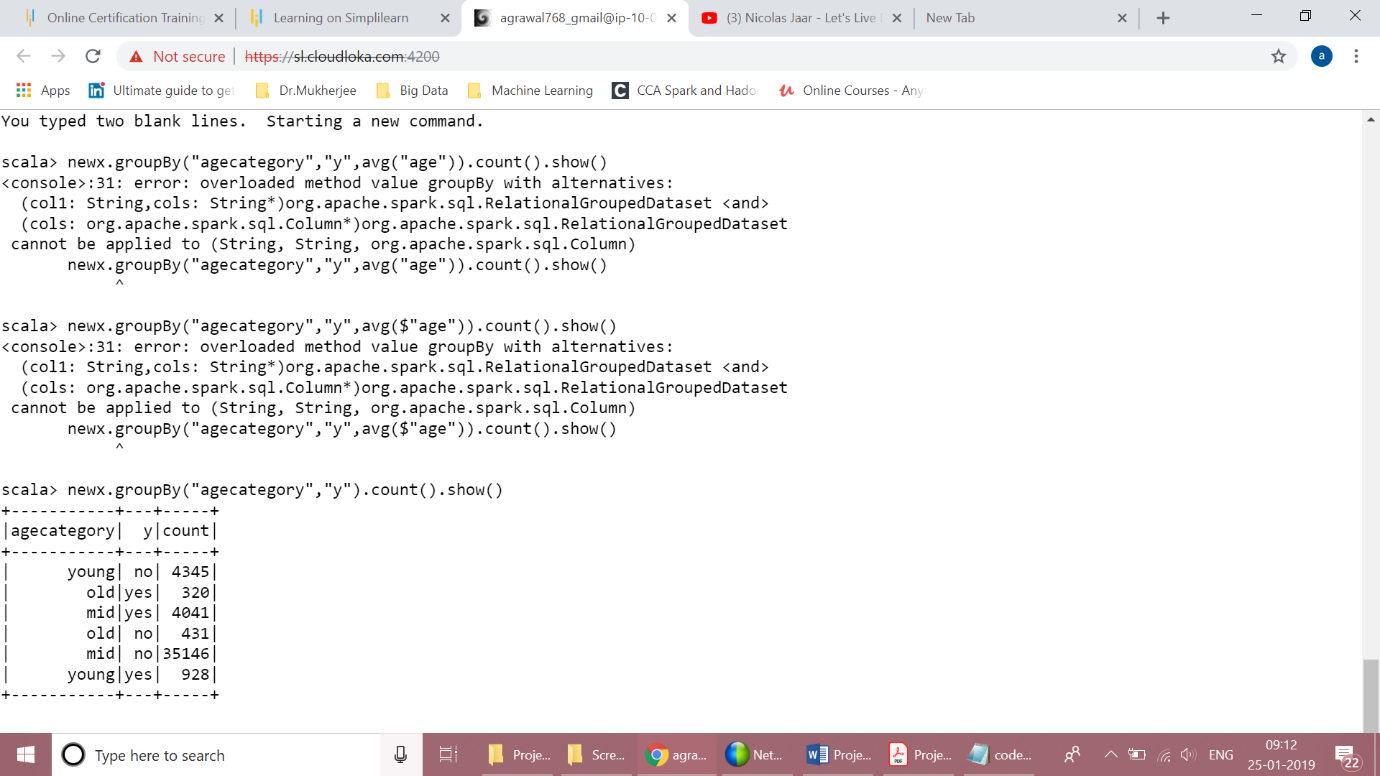


1. Importance of marriage on subscription:

If we think of subscriptions married people barely subscribe. So, it’s an important variable.



1. From ans to 6, Yes both the variables age and marital status affect the subscriptions
2. Feature Engineering:



Use User defined functions to categorise the age into young, mid and old group. On that basis, mid age people are found to be not much interested on term deposits or subscriptions.