

PS4

Rafael Alfena Zago

February 2019

1 Question 5

d) It's a "data.frame".

2 Question 6

4) df's class is a "SparkDataFrame", while df1's class is a "data frame".

5b) There is an error message: "Error in (function (classes, fdef, mtable) :
unable to find an inherited method for function 'select' for signature '"data."".
This is probably because of the difference in the classes of each data frame.

6b) The same thing happens as in the previous exercise.

3 Question 7

I would like to automatize downloading several links available at the same page. I don't know if this is actually considered scrapping, but I have heard of it as such and I needed it when I was downloading the data for my research. So, for example, I had to download the .csv file for all the crimes registered in a certain month of a certain year. However, I wanted it for every month of several years, which made me click on every single file. The same thing happened for some other files I had to download for another research project. It took me a week to download all the files I wanted for the crimes, not to mention the tedious job of waiting and clicking. That's why I wanted to learn how to do it.