

PS4

Colton Steele

February 23, 2023

1 Introduction

The data sources that I would like to scrape from are ones that house sports data. A website that I have scraped before is Pro Football Reference, and I could do a similar scraping for Pro Baseball Reference. I think doing sentiment analysis with Twitter data would be fun too. Additionally, as you know, I've done a lot of work with `nflfastr` and `cfbfastr` which scrape from various pages such as `sports-reference.com`, `espn.com`, `nfl/ncaa.com` etc.

2 Answers

Problem 5

5d) The class of the object is character

Problem 6

6.7) `df` provides the output of `tbl_spark`, `tbl_sql`, `tbl_lazy`, and `tbl` which means it's a spark table

`df1` outputs `tbl_df`, `tbl`, and `data.frame` which shows it is a data frame.

6.8) My column names in `df` have an underscore between the words while `df1` has periods. My assumption is that in spark tables, you can't have periods between the words as periods can't be used in variable names while in R data frames, they can be.