

PS4 Dunkleberger

Davis Dunkleberger

February 15 2023

1 Scraping

I am interested in scraping data from many sports related websites. My program is sports data analytics and there are all sorts of ways to access data. Many of these sites do have APIs that I know of or recommend R packages that provide play by play data similar to nflfastR. With the wide variety of packages ready for me to use, scraping might be a last resource for me and others in my field.

2 mydf\$date

1. mydf\$date is of the type character

3 class(df1)

1. tbl_df
2. tbl
3. data.frame

4 class(df)

1. tbl_spark
2. tbl_sql
3. tbl_lazy
4. tbl

5 Column Names

The names are different between the first two objects. SQL does not like a . in the names of things and instead prefers to use underscores in names. Spark has traits similar to SQL and allows us to use functions in R and other RDD options as well. Spark recreated the object in a format that is easier to read for other types of processes.