

PS4

Luke Denton

February 25, 2021

Exercise 1

I would be interested in scraping from Jane Eyre on project gutenber. I really disliked reading that book in high school and never actually finished it, so I think I could have some fun quantifying how terrible of a book it is (in my opinion).

I am also interested in scraping twitter's trending hashtags. I would like to investigate how twitter bots play a role in creating trending hashtags. I think trending hashtags are a particularly vulnerable area of twitter that is easily manipulated. I understand that they may have practices in place to alleviate bots from spamming hashtags to promote things that real people aren't tweeting/talking about, but I would still like to look into it.

I would also be interested in scraping sports data, specifically about the March Madness Men's NCAA Basketball Tournament, given the time we are in. I don't know what specifically I would like to scrape, but I think it would be a fun exercise.

Exercise 2

Question 5.D

"mydf" has the types: "tbl_df", "tbl", "data.frame"

"mydf\$date" has the type "character"

Question 6.7

"df1" has the types: "tbl_df", "tbl", "data.frame"

"df" has the types: "tbl_spark", "tbl_sql", "tbl_lazy", "tbl".

Question 6.8

The column names are different, perhaps because Spark shows the data in a different way.