

# Problem Set 4

Mason Ross Hayes

February 2020

## 1 Possible Web Scraping

- I would like to look at financial data such as cryptocurrency markets or the stock market. Lately I have been messing around with gekko trading bot as well as learning how to get stock data through the tiingo API. I think I could use one of these sources for the project.

## 2 PS4 Answers

1.
  - df1 is a data frame
  - df is a SparkDataFrame
2.
  - At first I was using tidyverse to create df1, which I later realized overrided the native RDD functions of sparkR. Neither the select or filter function worked on df until I reloaded sparkR without the tidyverse library.
  - Once doing it correctly (without tidyverse), the select and filter functions do not work on df1 because it is a data frame, not a spark data frame, and those functions only work on RDDs.
  - With tidyverse, of course, the select and filter functions work on df1 because sparkR is calling the tidyverse functions, not the RDD functions.