

PS4 Holt

richard.n.holt

February 2019

1 Data Source Interests

I would be interested in scraping data off of different news websites in order to analyze word connotations and other potential biases in articles. I would cross-reference these articles either by author or by corporate source to see where these biases lie.

I would also be interested in scraping data from websites which track economic movements and make predictions about economic trends. I would take that web data from the past several years to see which sources track those movements most accurately.

2 Question 5

5d. mydf is a list; mydf student is a data.frame.

3 Question 6

6.4 `class(df1)` is a dataframe. `class(df)` is a `SparkDataFrame`.

6.5 After trying to run the `select` function on `df1`, there was an error message related to the inability to find the inherited method `'select'`.

6.6 This same error occurred for the `filter` method. No inherited method was found. I believe this happens because R inherits the method from Spark, and the `df1` is a normal dataframe, not a `SparkDataFrame`.