

PS4\_Maddy

ethanmaddy

February 2021

## 1 Part 1

I think that it would be interesting to scrape data from Reddit. With so many forums and topics, it would be interesting to see how posts and engagement relate to measurable, quantifiable outcomes. For example, scraping popular trading forums (r/wallstreetbets, r/pennystocks, etc.) could lead to a predictive model for personal investments. Additionally, it would be interesting to conduct sentiment analysis on r/news and compare it with other news platforms (WSJ, NYT, etc.)

## 2 Part 2

### 2.1 Question 5.D

mydf is a tbl\_df, tbl, data.frame. mydf\$date is a character type.

### 2.2 Question 6.7

df1 is a tbl\_df, tbl, and data.frame. df is a tbl\_spark, tbl\_sql, tbl\_lazy, tbl.

### 2.3 Question 6.8

Yes, the column names are different across the objects. The columns might be named differently since they are copied into spark and spark the total data displayed.