

ECON 5253: PS4

William Lorton

February 23, 2021

1 Data Sources for Scraping

- Financial news websites; perhaps search for mentions of EUR/USD, observe sentiment towards the cross, and see how this relates to its movement.
- Search Twitter for mentions of Daft Punk and see how people are reacting to the news of their breakup after 28 years.
- Search for statistics related to the balance (or lack thereof in some seasons) between OU football's defensive and offensive performances and see how this relates to each season's outcome.

2 Answers for Question 6

- **Verify that the two dataframe are different types: type class(df1) and class(df). What is the class of each?**
 - df1 is a regular tibble (tbl_df, tbl, data.frame)
 - df is a spark tibble (tbl_spark, tbl_sql, tbl_lazy, tbl)
- **Are the column names any different across the two objects? If so, why might that be?**
 - Yes. df1 uses the period as a separator while df uses the underscore as a separator. Spark may not be able to use the period (or other special characters other than the underscore) when naming variables.