

Problem Set 4

Jack Vincent

February 19, 2025

1 Data Sources for Scraping

- **Athletic Performance Data:** APIs such as World Athletics or NCAA track and field databases provide access to historical and real-time race results, which could be useful for performance analysis.
- **Federal Reserve Economic Data:** The FRED database, as it is commonly referred to, is a useful place to find historical data dating back to the 1940's regarding metrics such as the inflation rate, GDP, unemployment, among various economic statistics. These could be used to find trends in the data or relationships between separate data points.

2 Working with JSON Data

: The type of object that "mydf date" is classifies as a character.

3 Spark Class Type & Column Names

: The class type of each dataframe is as follows: df1 is "tbl underscore df, tbl, data.frame," while df is "data.frame." The column names are different, yes, and may be so because Spark may be enforcing certain naming rules.