PS4_Wardwell

lwardwell

February 24th, 2025

PS4 Question 7

1. What data sources would you be interested in scraping from the web?

   - I am interested in bank failures and information around them, which is sometimes very sanitized from official sources. I might like to scrape tweets with the #bankfailure hashtag to see if anything interesting comes up.

2. API sources I am interested in:

   - The FDIC has a "Statistics on Depository Institutions" (SDI) website, which has an API option to get data.

PS4 Question 6 Answers

(6.7) What class is each dataframe?

   - df = "tbl_spark" "tbl_sql" "tbl_lazy" "tbl"
   - df1 = "tbl_df", "tbl", "data.frame"

(6.8) Are the column names any different across the two objects? If so, why?

   - The column names in the spark dataframe have "_" separators instead of "." separators. For example, "Sepal_Length" instead of "Sepal.Length". According to claude.ai, this is because Spark uses SQL naming conventions, and because "." has a special meaning in SWL, Spark is converting the "." to underscores.