Problem Set Five

Devin Williams

March 2, 2025

- Part One: Web scraping without API

  ⬦ What is the data?

    ⬦ The data interests me as additional information related to my class project. I am looking to analyze transfer evaluations of professional soccer players. Although this data does not apply directly to current players, it could be used to analyze player value to teams in the past. It is interesting to understand the history of the league that is currently the biggest in the world and that will be a big part of research these next coming months and hopefully in the future.

  ⬦ Will it be useful?

    ⬦ As mentioned, the data does relate to my project for this class. I also have an interest in continuing to analyze player transfer values and how they relate to country of origin, goal contributions, league played in, etc. I am looking to create a regression looking at what affects the evaluation of a player, and this information may help give context to previous player evaluations.

  ⬦ Did you use any online tutorials to help? Which ones?

    ⬦ Yes, I used two online tutorials to help. I used the 'Data Science for Economists' slides by Grant R. McDermott that we discussed in class. And, a little additional information from "https://r4ds.hadley.nz/webscraping". I also used Claude to help format the code correctly, as selectorgadget did not give me an accurate path to download the data.

- Part Two: Using an API

  ⬦ What was found in the data?

    ⬦ I discovered that for my project, I will have to do quite a bit of data cleaning based on this API. It will take a bit of effort to be able to properly express players names, countries, and clubs based on this API. Currently, these variables have placeholder numbers that will have to be changed for better analysis.

  ⬦ What packages were used?

    ⬦ I used httr and jsonlite for this API. Httr was required to work with HTTP requests and APIs. The GET(), and content() functions were used from this package. Jsonlite was used to handle the JSON data that came from the API. fromJSON() function was used within this assignment.