PS5

Bradley Rann

March 2023

1 Introduction

1.1

The first data set I pulled are all the Wimbledon winners in the open era of tennis. I pulled this data table from Wikipedia then converted it to a data frame. Then I specifically pulled the winners from the data that had the year, score, opponents name etc. This project isn't important for my research but I'm a tennis fan and thought it'd be a good source of data to pull from. I did ultimately use a tutorial as I don't have any clue how python packages work and I needed to see how BS4 worked. It was this youtube video "Python Tutorial: Web Scraping with BeautifulSoup and Requests."

1.2

The second data set is from an API from healthdata.gov and it contained thousands of rows and a massive amount of columns, After pulling the data I converted it from a CSV to a dataframe with pandas and looked specifically at how many hospital beds each state had. This data might actually be important for my research because I'm particularly interested in health data and might use it for my project down the line. I almost used a package called Pyhealth but it wasn't compatible with my version of python unfortunately.