

Project 1 Approach

Benjamin Bravo

2026-02-19

Approach

In this project, I will transform a semi-structured chess tournament text file into a clean dataset that can be exported as a CSV file. Since the file is not formatted like a normal table, I will read it line by line and use pattern matching to extract the needed information for each player.

First, I will identify the repeating two-line structure for each player. From these lines, I will extract the player's name, state, total points, pre-tournament rating, and opponent pair numbers. Then, I will match each opponent's pair number to their pre-rating and calculate the average opponent rating for each player.

Anticipated Data Challenges

The main challenge is that the file contains irregular spacing, special characters, and match result codes (W, L, D), which makes parsing more complex and difficult than reading a standard CSV. Extracting only the correct opponent numbers without capturing unrelated values will require careful pattern matching. Additionally, some players may have byes or incomplete rounds, which must be handled properly when calculating averages.

Overall, the focus of this project is converting messy text data into a structured dataset while ensuring accurate matching and calculations.