Kingdom

Andy Vo

Description / Objectives

Kingdom is a data analysis on the top Chess players in history to show how many pieces players have won or lost in their career. This data metric is supposed to symbolize their “winnings” in terms of chess pieces.

- Collect PGN data on top Chess masters throughout history
- Clean and Parse the Data
- Analyze their chess matches and collect data on how many pieces they won, lost, and sacrificed
- Compare them amongst each other based on their average games played

Background

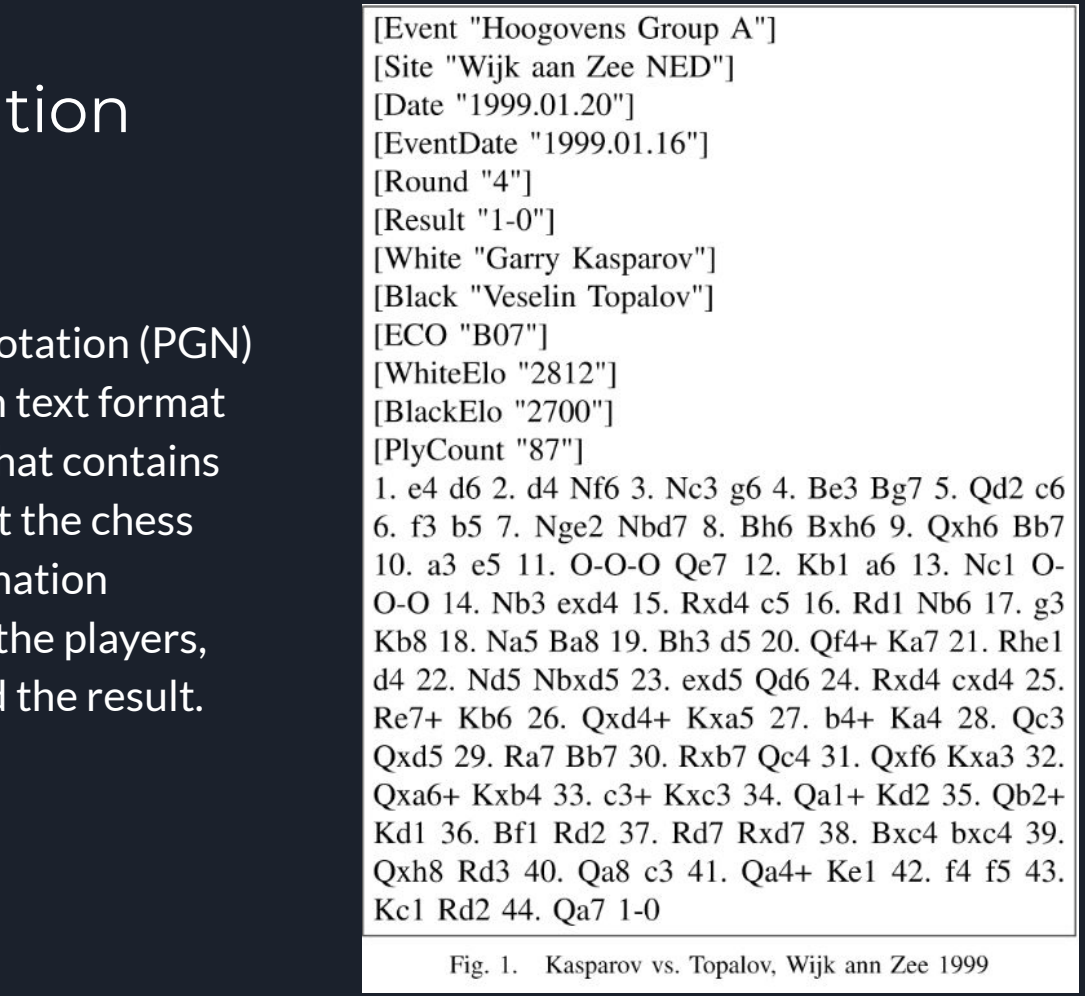
- Chess is a popular strategy-based board game between two players and is derived from the ancient Indian game, Chaturanga. It consists of 32 pieces, 16 pieces for each player: one king, one queen, two rooks, two knights, two bishops, and eight pawns. The board is a square chessboard with an eight-by-eight grid of 64 squares.
- The main motivation for this project was to collect a different dataset that can symbolize the path to becoming a grandmaster. As such, I was curious as to how many pieces were won or lost throughout the player's entire match history.

Data Acquisition

- Originally, the project was supposed to use LiChess's official API for data acquisition. However, LiChess's API required a specific profile created on their website, and since the matches I were looking for were done in person, I had to search elsewhere.
- Fortunately, there is a website called PGNmentor that contains PGN files with most if not all of the games played for the top chess players in history. I was able to collect around 47, 347 games/PGN files.

Data Explanation

- Portable Game Notation (PGN) is a standard plain text format for chess games that contains information about the chess match. The information provided include the players, moves plated, and the result.



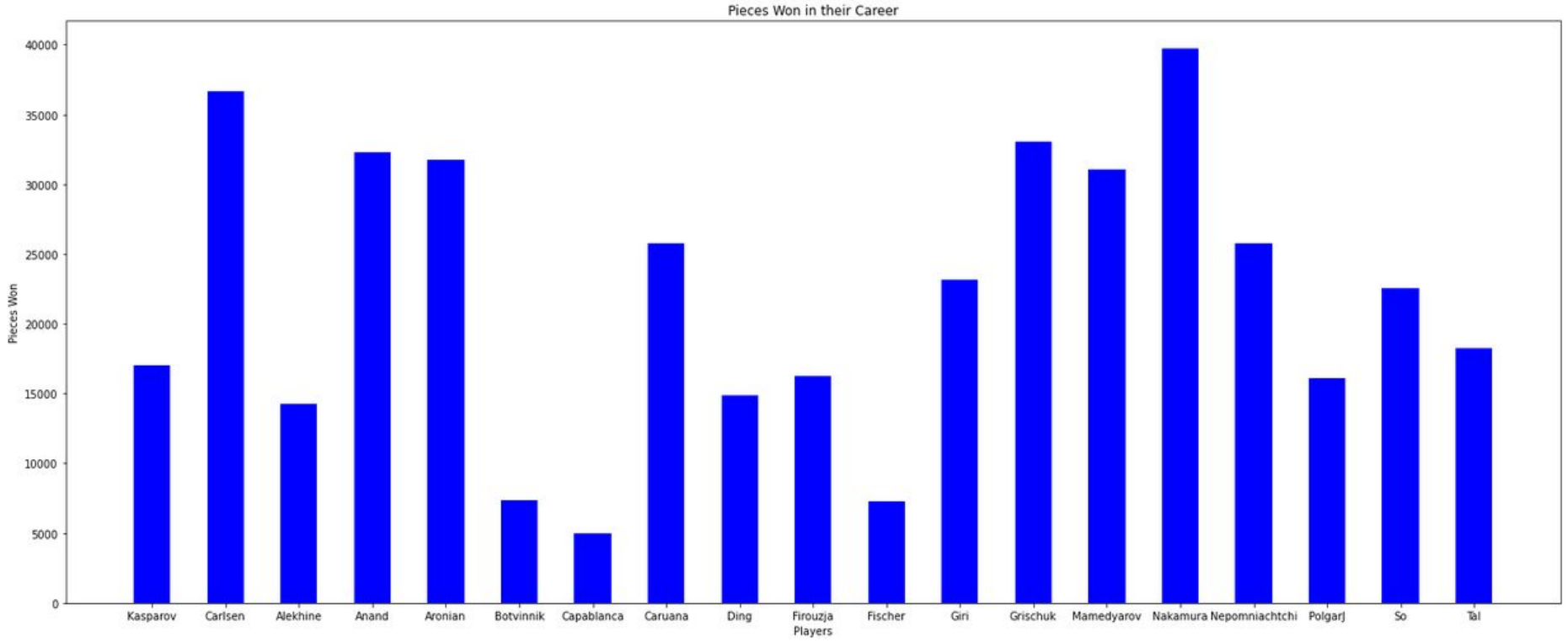
[Event "Hoogovens Group A"]
[Site "Wijk aan Zee NED"]
[Date "1999.01.20"]
[EventDate "1999.01.16"]
[Round "4"]
[Result "1-0"]
[White "Garry Kasparov"]
[Black "Veselin Topalov"]
[ECO "B07"]
[WhiteElo "2812"]
[BlackElo "2700"]
[PlyCount "87"]
1. e4 d6 2. d4 Nf6 3. Nc3 g6 4. Be3 Bg7 5. Qd2 c6 6. f3 b5 7. Nge2 Nbd7 8. Bh6 Bxh6 9. Qxh6 Bb7 10. a3 e5 11. O-O-O Qe7 12. Kb1 a6 13. Nc1 O-O-O 14. Nb3 exd4 15. Rxd4 c5 16. Rd1 Nb6 17. g3 Kb8 18. Na5 Ba8 19. Bh3 d5 20. Qf4+ Ka7 21. Rhe1 d4 22. Nd5 Nbxd5 23. exd5 Qd6 24. Rxd4 cxd4 25. Re7+ Kb6 26. Qxd4+ Kxa5 27. b4+ Ka4 28. Qc3 Qxd5 29. Ra7 Bb7 30. Rxb7 Qc4 31. Qxf6 Kxa3 32. Qxa6+ Kxb4 33. c3+ Kxc3 34. Qa1+ Kd2 35. Qb2+ Kd1 36. Bf1 Rd2 37. Rd7 Rxd7 38. Bxc4 bxc4 39. Qxh8 Rd3 40. Qa8 c3 41. Qa4+ Ke1 42. f4 f5 43. Kc1 Rd2 44. Qa7 1-0

Fig. 1. Kasparov vs. Topalov, Wijk ann Zee 1999

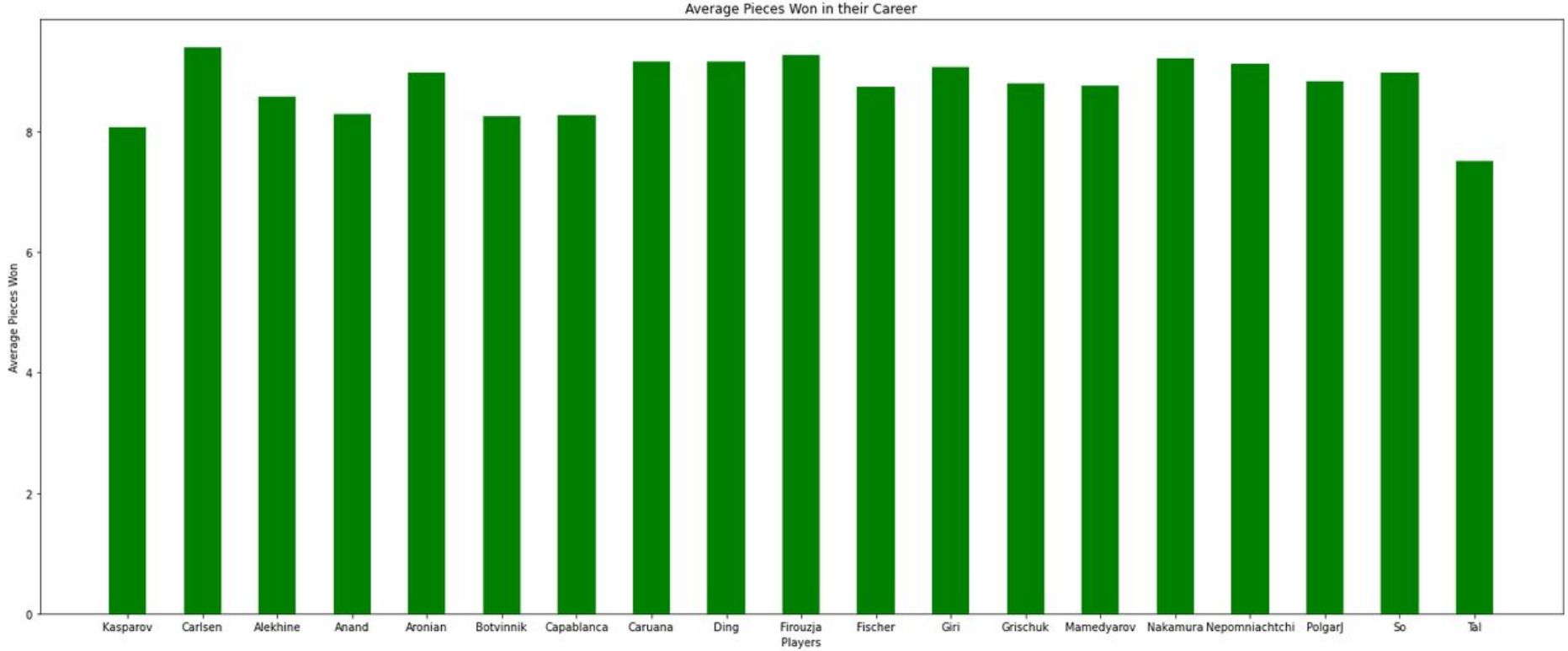
Parsing and Cleaning

- To parse the data, I used pgnparser, a simple open-sourced PGN parser in Python.
- It follows the PGN format and collects data from the file and sorts them through their associated tags.
- Then, I used the moves list generated for my algorithm to analyze the data.

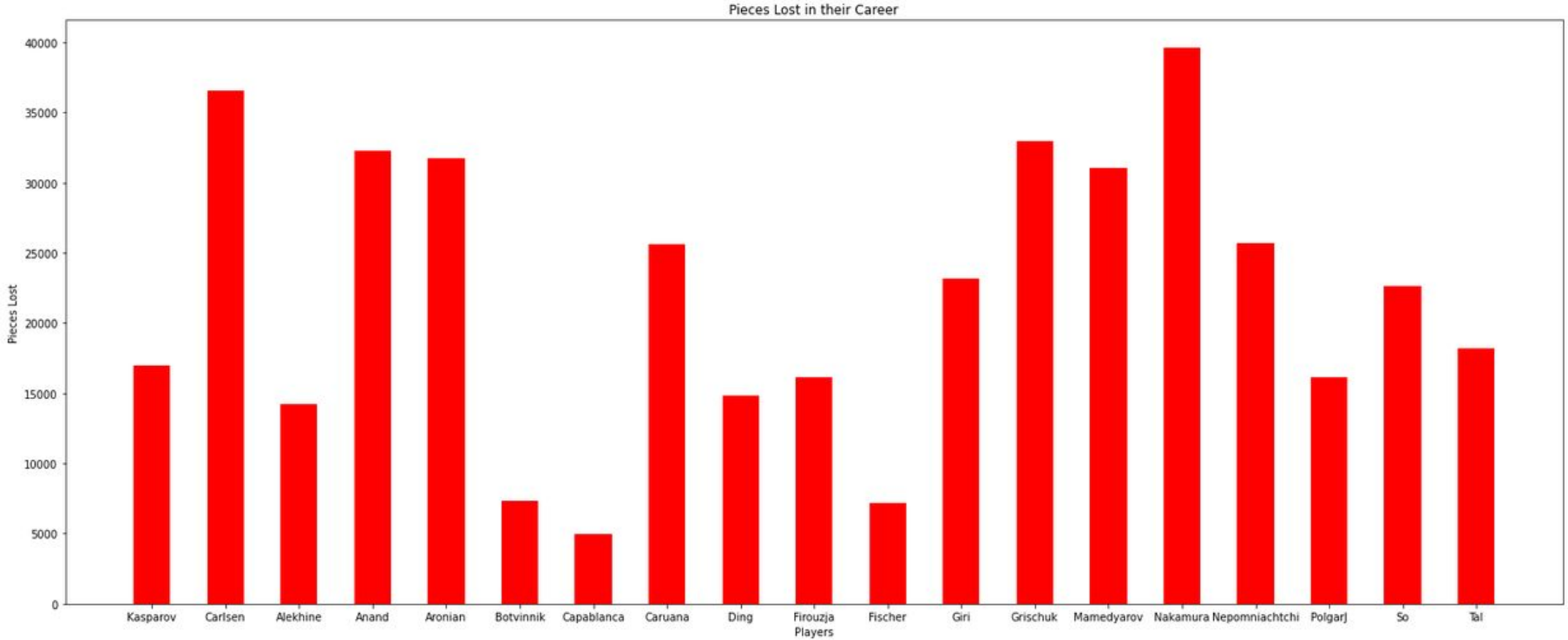
Data Analysis



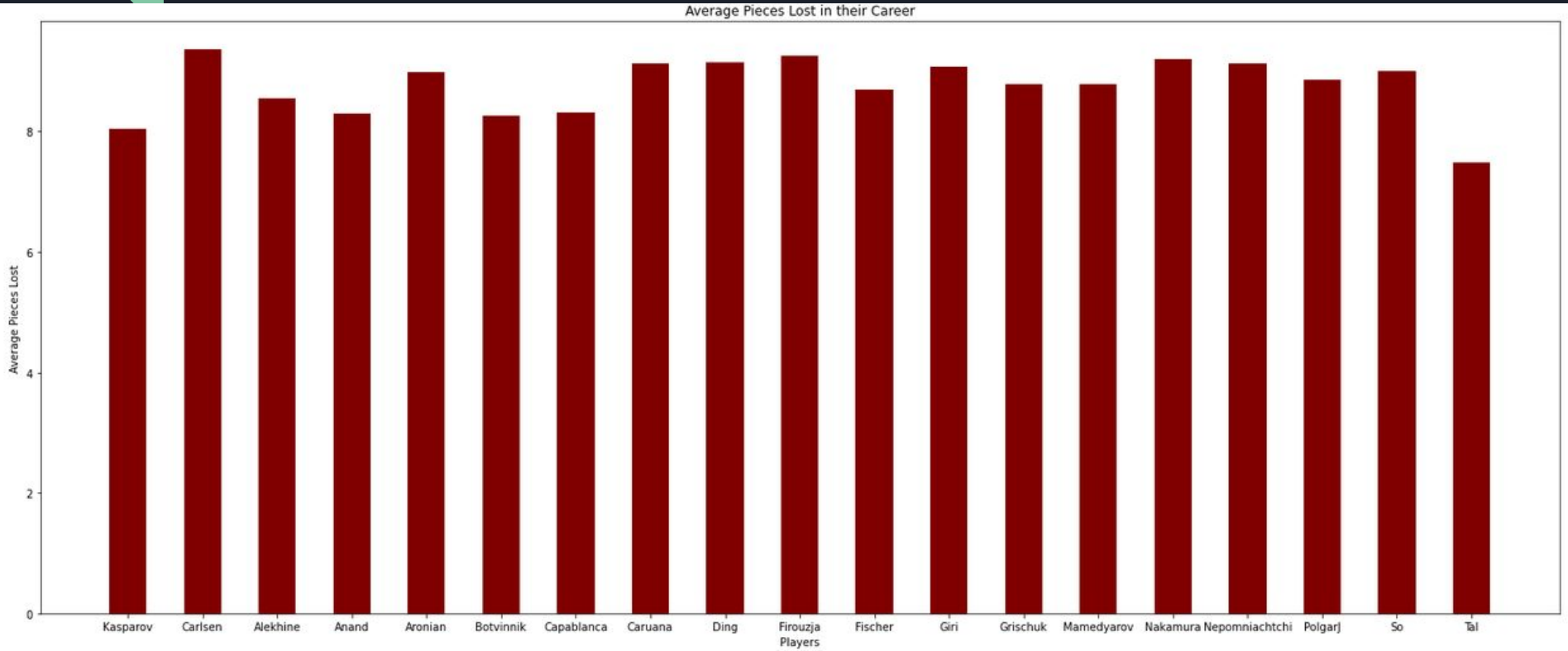
Data Analysis



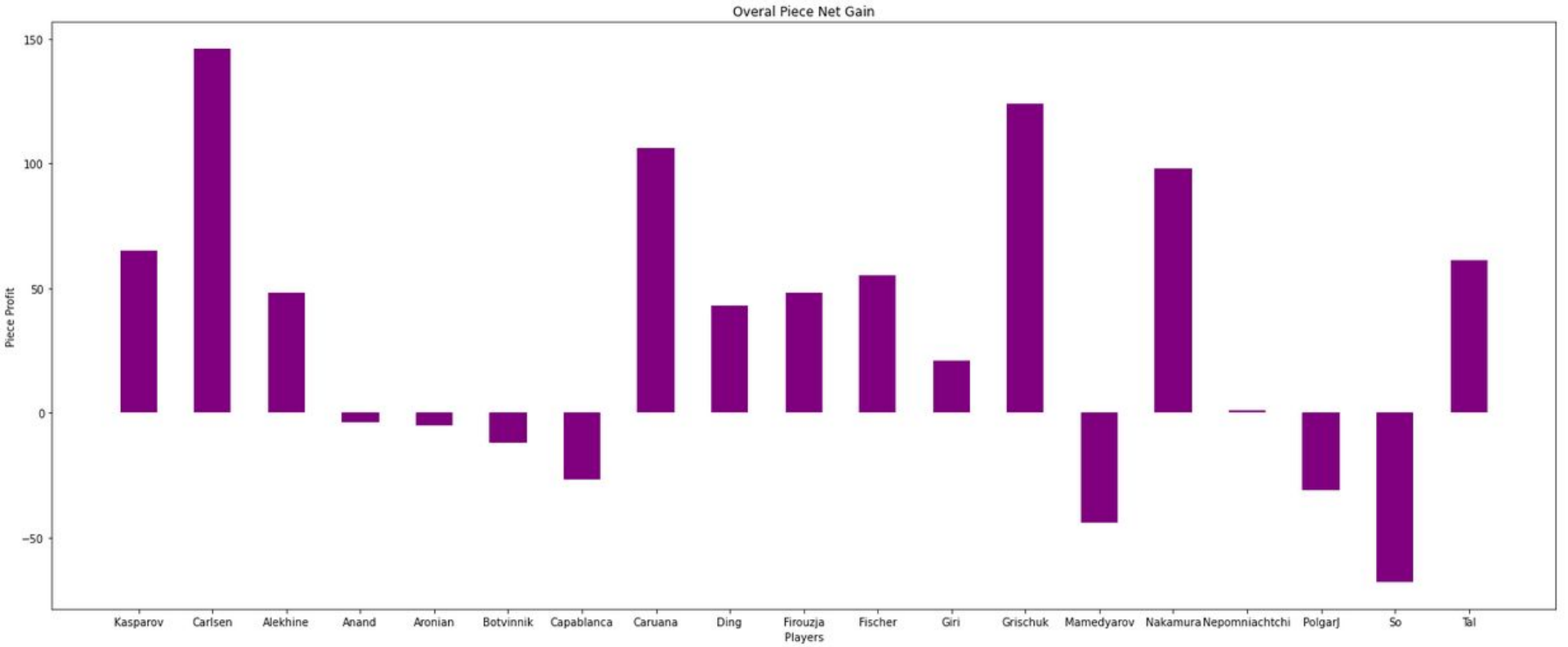
Data Analysis



Data Analysis



Data Analysis



Project Timeline

Week	Milestones
1	Research Chess Database
2	Clean and Parse the Data
3	Develop Data Structures and Algorithms to Analyze the Data
4	Optional Statistical Research
5	Analyze and Visualization the Data
6	Final Report and Presentation

Conclusion / Future Plans

- Overall, this project was a valuable learning experience with data and analysis. I gain insight and knowledge on using Python and PGN files.

- In the future, I would like to work with various Chess bots such as Stockfish to collect data and analyze more specific moves such as sacrifices. I would also like to see if I can timeline their rating history.

Questions?