

Online Chess Games

Salokya

2023-09-25

Contents

1 Abstract	1
2 Introduction	1
3 Dataset Description	2
4 Exploratory analysis with plots	2
4.1 How win probability varies by color?	2
4.2 Which is the most frequent mode of victory?	3
4.3 Which is the most preferred opening move?	4
4.4 If a player has higher rating than his opponent, how likely is that the higher rated player wins?	4
4.5 How many moves are required to determine game outcome?	5
4.6 What is the average player rating? Is the average player rating of black and white similar?	5
4.7 Do people prefer rated game or casual play?	6
4.8 How rating difference influences win probability of higher rated player?	6
5 Results	7
6 Conclusion	7

1 Abstract

In this project, I have done a thorough data analysis for 20,000+ chess games played on Lichess, an online chess application. Analysis of player parameters such as moves, victor, rating, opening details etc. helps to better understand the nuances of online match-making, the diverse or most preferred game modes and overall gives a broader perspective of the game itself and also about the gaming platform. Understanding the working mechanism facilitates understanding of player expectations, common player tendencies which further influences app development. But most importantly, it is a necessary for fair play and enjoyable user experience.

2 Introduction

To comprehend the game and the gaming application, we first need to specify what we want to know and define the questions accordingly. As a chess player of intermediate calibre and regular user of Lichess, I have found the following questions seemingly obvious/fundamental:

- How win probability varies by color?
- Which is the most preferred game mode?

- Which is the most frequent mode of victory?
- Which is the most preferred opening move?
- What is the average player rating? Is the average player rating of black and white similar?
- How variation of rating difference influences win probability of higher rated player?

3 Dataset Description

The dataset for this project is provided by Mysar Ahmad Bhat on Kaggle. It has 20,058 records/rows(of which 19112 are unique) and 17 variables/columns.

The different variables are described below:

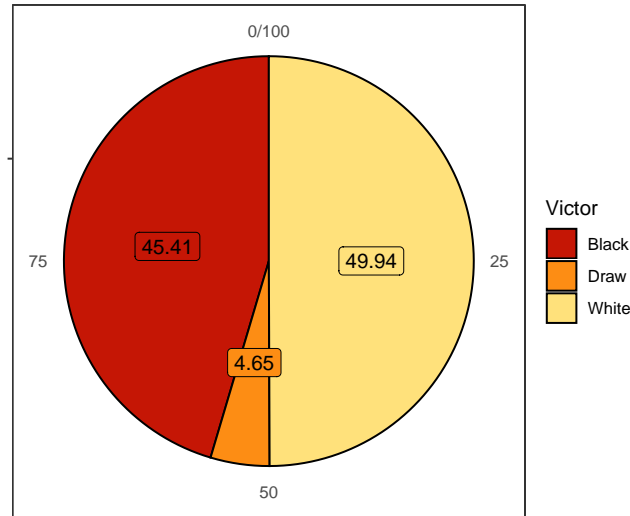
1. **game-id**: Every chess game has a unique identification number to ensure distinction from other games and for future access
2. **rated**: Unrated game is for casual play i.e. match outcome will not influence player rating.
3. **turns**: It is a natural number denoting the total number of moves made in the match for a definite outcome.
4. **victory_status**: It defines the nature of victory.
5. **winner**: It describes which player won or if the match ended in a draw
6. **time_increment**: The value before the '+' indicates the initial allocated time per player in minutes. The value after the '+' indicates the increase in allocated time in seconds per move.
7. **white_id**: Profile id of the player playing as white. This is imperative to uniquely identify the player.
8. **white_rating**: A natural number denoting the skill level of the player. Higher value indicates greater competency. Rating increases on winning and decreases on losing.
9. **black_id**: similar to white id
10. **black_rating**: similar to white rating
11. **moves**: denotes the sequence of moves played by each player.
12. **opening_code**: denotes the code of the opening move of the game.
13. **opening_moves**: The opening stage of the game is where a player activates the chess pieces and creates a strategic formation for further play. Opening moves is a natural number denoting the number of moves consumed in the opening stage.
14. **opening_fullname**: Based on the type of opening, each opening has a universally accepted nomenclature, which is indicated by opening fullname.
15. **opening_shortname**: shorter version of the opening fullname.
16. **opening_response**: Depending on white's opening move, black has a typical/standard response move. opening response denotes if player black followed the standard response(also known as book move) or played a different move.
17. **opening_variation**: Just as there are numerous opening moves and opening responses, every opening move/response has further various possibilities of subsequent moves termed as opening variation. The opening variation plays a crucial role in determining the intermediate gameplay.

4 Exploratory analysis with plots

4.1 How win probability varies by color?

From **Fig 1**, we observe that of all the games played, almost 50% resulted in victory of white, while of the remaining half, 45.41% was won by black and the rest i.e. 4.65% games resulted in a draw. If we ignore the draw, then white wins $(49.94/(49.94+45.41))\%$ i.e. 52.37% of all the games. This slightly skewed probability of win in favor of white is actually not due to randomness, but due to the "*first move advantage*" in chess which states that for similar rated players, white's win probability lies between 52% and 56%, which supports our observation of 52.37%.

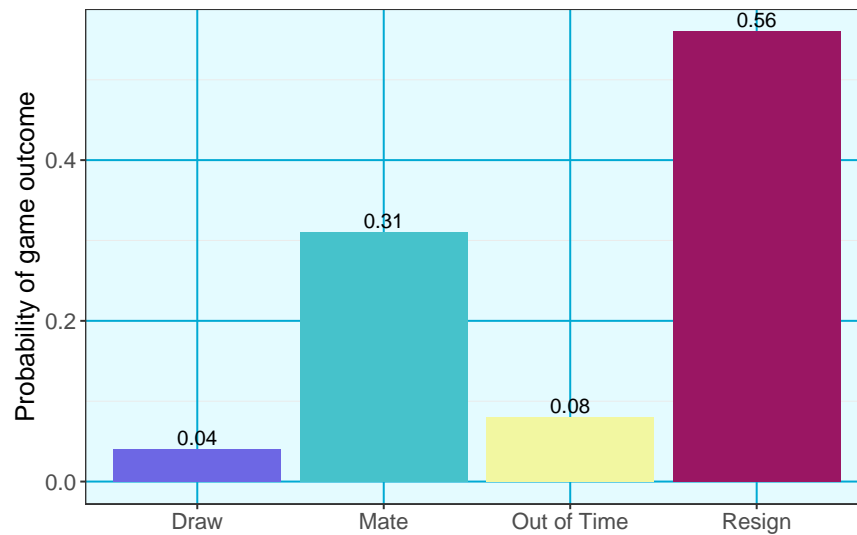
Fig 1: Variation of win probability by color



4.2 Which is the most frequent mode of victory?

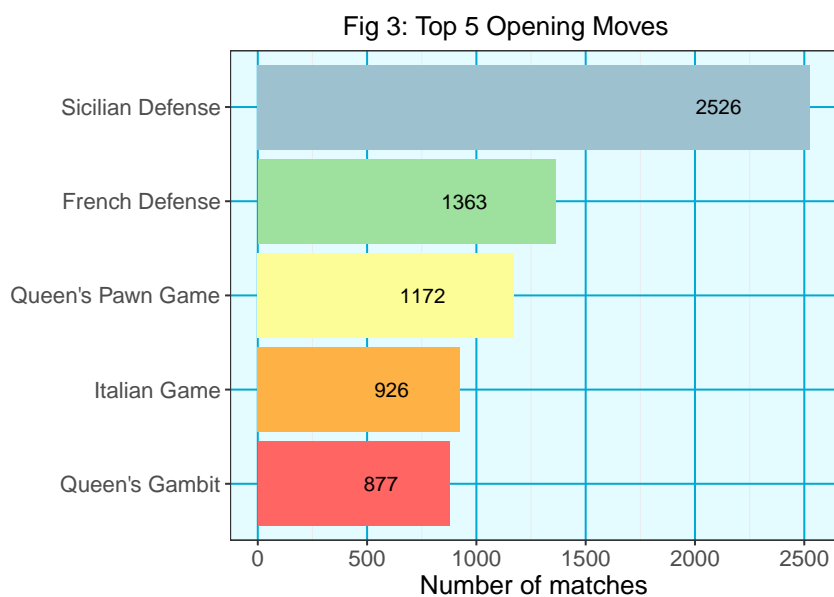
From **Fig 2**, we observe that probability of draw(including stale mate) is only 4%. Win by resignation is 56% whereas win by checkmate is 31%. It can be concluded that most players prefer to resign early on when they feel game is out of hand rather than face the embarrassment of checkmate.

Fig 2: Probability of different victory modes



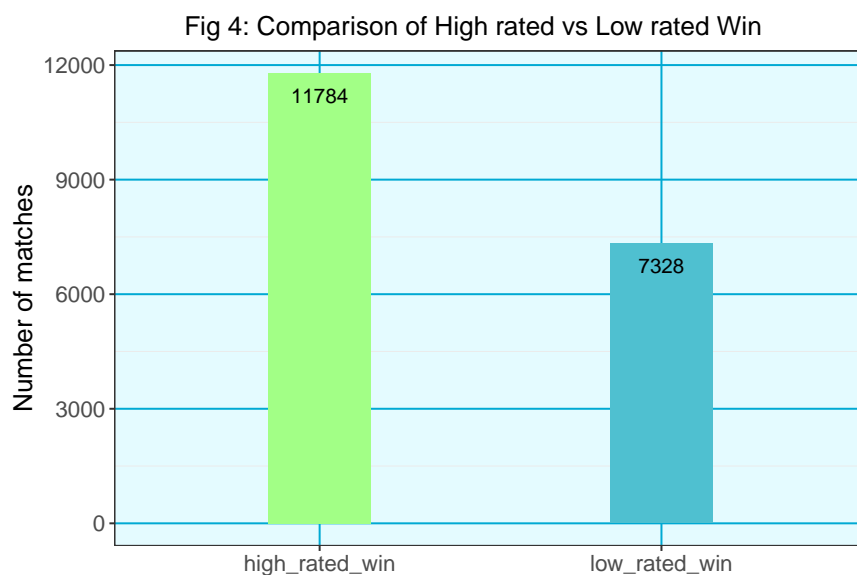
4.3 Which is the most preferred opening move?

From **Fig 3** below, we see that the most common opening move is Sicilian Defense. The popularity of this opening is even more than the next 2 openings combined.



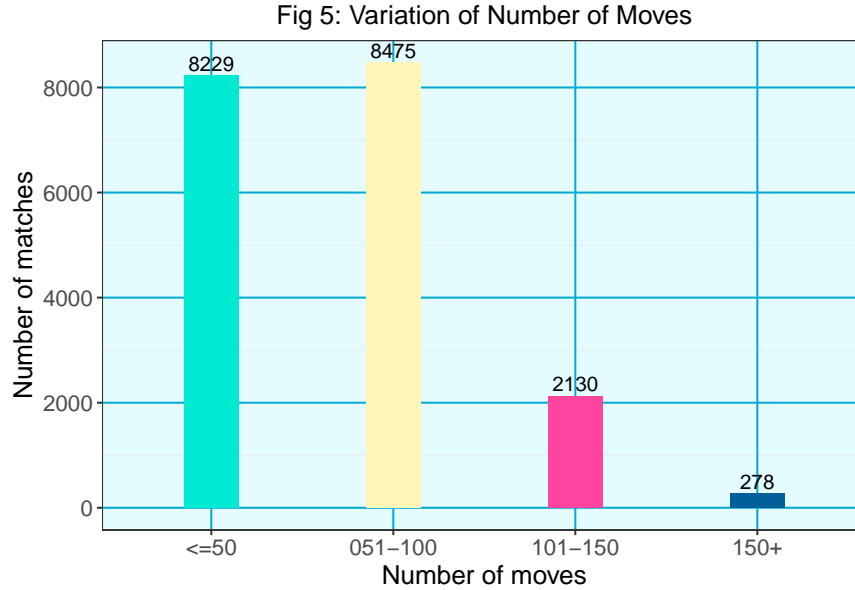
4.4 If a player has higher rating than his opponent, how likely is that the higher rated player wins?

From **Fig 4** below, we see that of the 19112 matches, there are 11784 higher rated wins and 7328 lower rated wins. So, it can be inferred that a higher rated player is about 50% more likely to win than his opponent. In Fig 8, we will see how higher rated win probability varies with the rating difference.



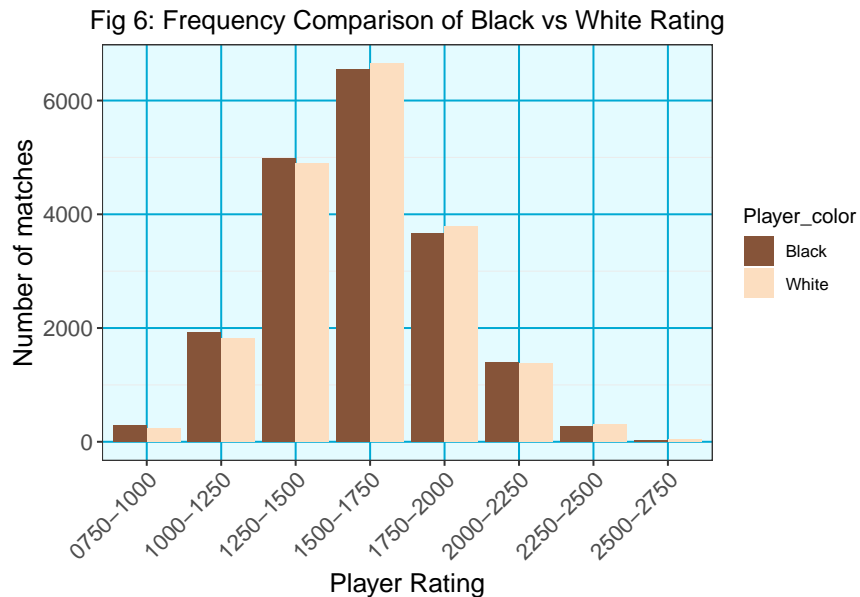
4.5 How many moves are required to determine game outcome?

From **Fig 5**, we observe that of 19112 games, more than 16000 games were over in less than or equal to 100 moves. The highest number of moves in a single match is 349. But there are just about 3000 matches where more than 100 moves were played. So, it can be inferred that the number of matches where more than 100 moves were played drops drastically with increase in number of moves.



4.6 What is the average player rating? Is the average player rating of black and white similar?

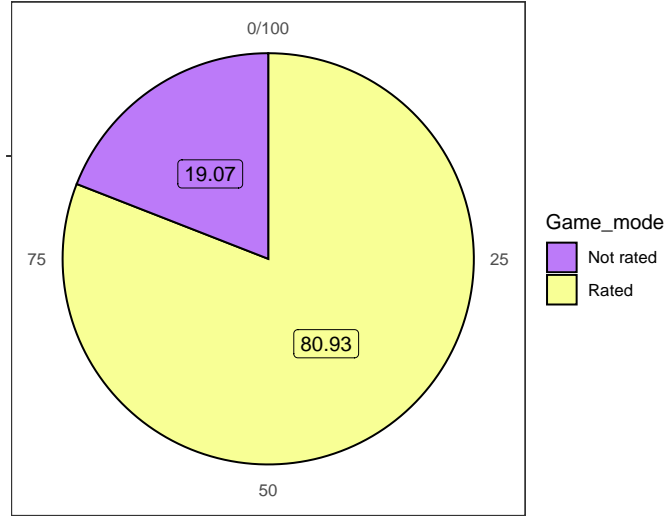
From the data we find that the lowest rating is that of white viz 784 and the highest rating is that of black viz 2723. From **Fig 6**, we see that most players have rating between 1500 to 1750. Also another important observation that can be drawn from the graph is that in each rating band, both black and white have similar number of players. Assuming a player's rating remains in a particular band throughout, this implies he has equal chance to play as white or black, which is important for fair play (First Move Advantage!!)



4.7 Do people prefer rated game or casual play?

From **Fig 7**, we see that more than 80% of the games are rated in nature. Users have distinctly higher tendency to play rated games.

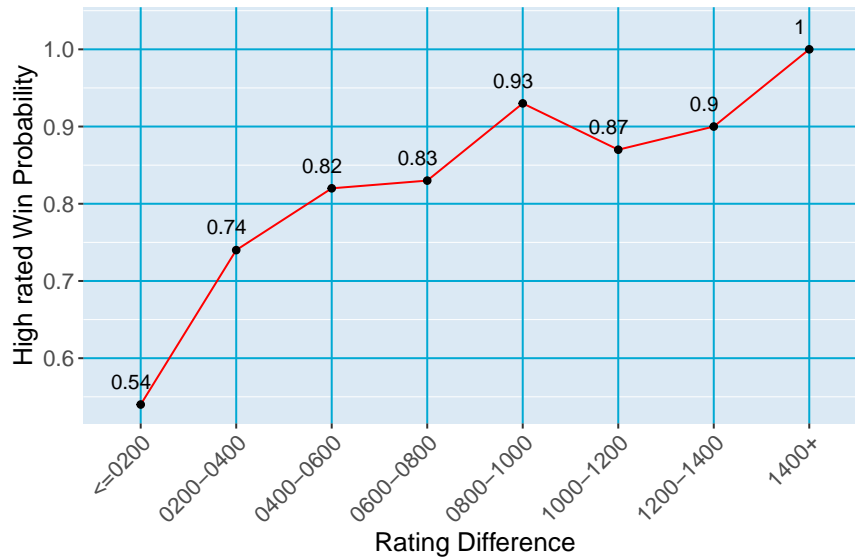
Fig 7: Percentage of rated vs unrated games



4.8 How rating difference influences win probability of higher rated player?

High rated win probability is the probability of victory of the higher rated player. From **Fig 8** we see that when rating difference is less than 200, win probability of higher rated player is about 50%; and this high rated win probability increases when rating difference increases. When the rating difference becomes 1400 or more, this probability becomes 1 and victory of the higher rated player is predetermined. Hence it is essential during match-making that the rating difference be as less as possible or at least less than 200 to ensure fair play.

Fig 8: Variation of high rated win probability with rating difference



5 Results

Summarizing the observations so far, we learn that white is slightly more likely to win than black, given they have equal/similar ratings. Also most played games are rated and are likely to finish within 100 moves. The ratings of black and white are more or less identically distributed with similar minimum,maximum and average ratings. This ensures fair play, but most importantly the rating difference should be less. If the game doesn't result in a draw, the losing player is more likely to resign by themselves than be checkmated.

6 Conclusion

The above results confirm that thorough analysis is necessary to better understand the procedure of an online gaming platform. A beginner in chess will understand the game better from the recorded observations and results. We can also infer that our results are essential for proper match making, app development and better user experience.