

Chess Games Analysis

Disha Saha

Summer 2021

https://github.com/dishasaha/Chess_Games_Analysis

Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data.

1. https://en.wikipedia.org/wiki/Shannon_number#Number_of_sensible_chess_games
The shannon number is the number of permutations that can be played during a given match of 10^{120}
2. <https://www.chess.com/article/view/history-of-chess>
This article covers the history of chess, which began around the 16th century.
3. <https://www.365chess.com/eco.php>
This webpage covers ECO openings and their movement of the pieces.
4. <https://chessprizes.com/>
This webpage shows how much each player has received in prize money since 2017. Magnus Carlsen has won an estimated 2.7 million since 2017.
5. <https://www.chess.com/terms/chess-strategy>
This webpage displays the top 5 concepts to learn in chess strategy.
6. <https://www.cnn.com/2020/12/06/us/queens-gambit-chess-popularity-trnd/index.html>
Article covering the increase in popularity of chess due to Netflix show 'The Queen's Gambit'
7. https://en.wikipedia.org/wiki/Chess_opening
Wikipedia page explaining chess openings.
8. <https://www.chessable.com/blog/chess-openings-white-surprise-attacks/>
This article has the top 5 best chess openings for white, which we will use to compare against our data analysis.
9. <https://www.ichess.net/blog/5-best-chess-openings-black/>
This article has the best openings for black depending on white's opening moves.
10. https://en.wikipedia.org/wiki/Chess_rating_system
Wikipedia page for the chess rating systems. We will use the chart to determine the groups in which we will analyze opening successes.

Which Data?

What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

The datasets I will be examining are from kaggle and chess.com. The kaggle dataset contains 20,000 chess games from Lichess and the chess.com data will come from the chess.com API.

<https://www.kaggle.com/datasnaek/chess>

<https://www.chess.com/news/view/published-data-api>

Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.

1. What opening has the best win rate for white?
2. What openings are most played by FIDE ratings?
3. What openings have the best win rate for blitz games?
4. What openings have the best win rate for rapid games?
5. Who has won the most games? What was their distribution in victory status?

What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

I will be performing exploratory data analysis using the datasets provided. My plan is to try to use the API from chess.com and apply some cleaning and preparation to the API data. I will be falling back on the kaggle dataset if I am unsuccessful. I would like to create some nice visuals for my analysis so that I can see which are the best openings to use to win at chess.

Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

The challenges and issues I could face while working on this project is getting the API to work properly to pull data from the chess.com website. If I continue to have issues I am going to fall back on the kaggle dataset for analysis.

Concluding Remarks

Tie it all together. Think of this section as your final report's abstract.

Chess is a game that has existed for a long time and is played all over the world. It is a very complex and complicated game, as a single game of chess can have as many permutations as 10^{120} (Wikipedia). While a game that is as complex as it is there are very strategic openings that can be played to better your chances of winning. During my analysis we will analyze a variety of games played by different players of

FIDE level and titles and different time constraints of rapid or blitz. We will come up with a list of top openings played by these games with the best winning rate for players playing with white or black pieces.

https://en.wikipedia.org/wiki/Shannon_number#Number_of_sensible_chess_games