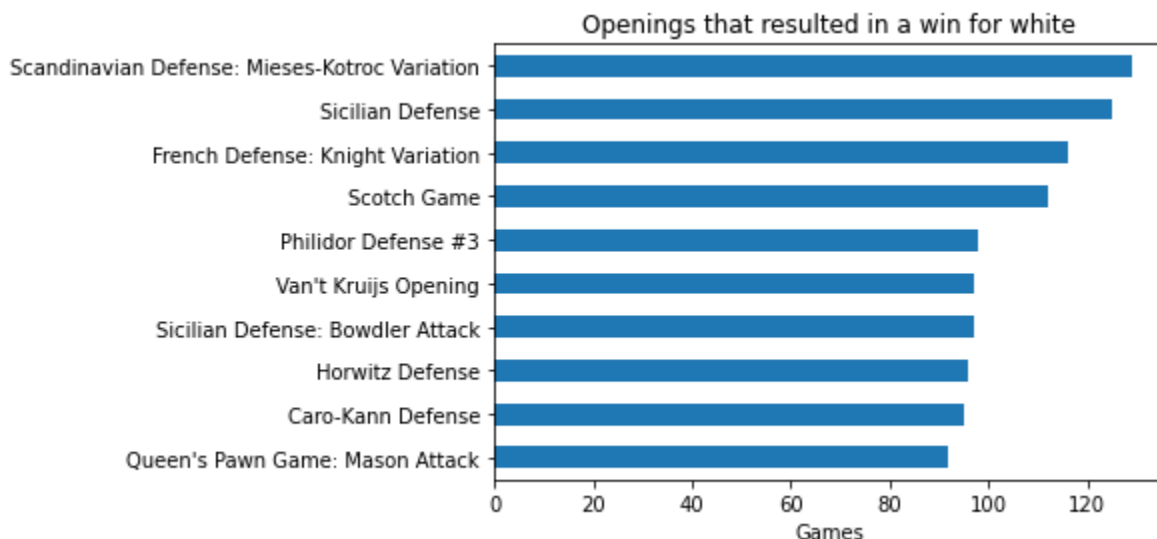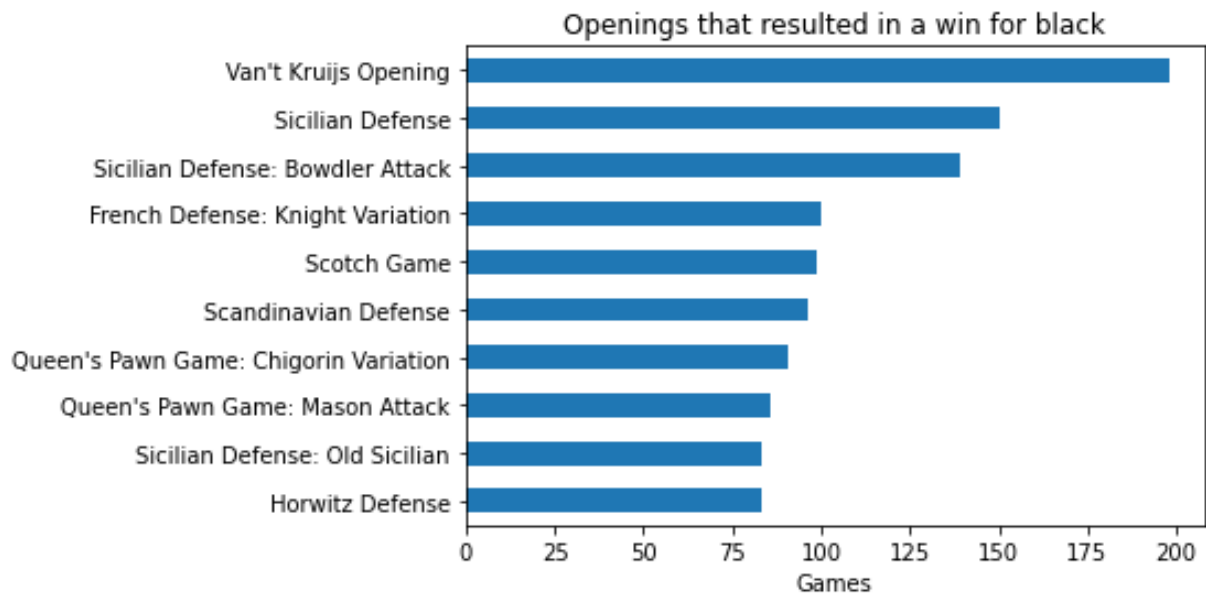Disha Saha
Project 3 Paper
DSC 680

Chess is a strategy game that has been around for a long time, as early as the 16th century (Team,

C). It consists of a 64 square board with 16 pieces where the pieces can have different ways to maneuver

around the board. Due to this it is a very complex game and can have almost 10^ 120 permutations on the

board (Shannon Number). Despite all of this, there have been certain strategic opening moves that have

been played to better position pieces on the board to win the game. Through my analysis of over 20,000

chess games we will explore the openings that have resulted in a win for either side, which openings have

been played the most by novices through grandmasters, and out of the 20,000 games who has won the

most games.

To start with we began to analyze which openings have the best win rates for either the white or

black side. In order to do this we needed to load our dataset of over 20,000 games in pandas, and filter on

only rated games. I am filtering to just rated games as non rated games could skew numbers as players

might be quickly resigning or not trying as much since it has no impact on their rating. I then filtered the

games where the winner was white or black and plotted the opening name for the top 10 openings.

Openings that resulted in a win for black

It is interesting to note that the Sicilian defense resulted in a top opening for winning with white or black and is one of the most popular openings in the game. The Sicilian defense starts with white playing a pawn to E4 and black playing the move pawn to C5. After that there are different variations of the Sicilian depending on what the next moves are for both sides, which can be seen as the next most effective opening for black as the Sicilian defense: Bowdler attack.

The next step in my analysis was to investigate the most popular openings played by different ELO scales. The ELO system is used by FIDE which is the international chess federation and depending on your rating there are different tiers ranging from novices to grandmasters (Chess Rating System). With my dataset I had applied the different ELO tiers based on white's rating and followed the ranges based on the chess rating system wiki (Chess Rating System). The table below shows the breakdown with the most popular openings being played by each of the tiers.
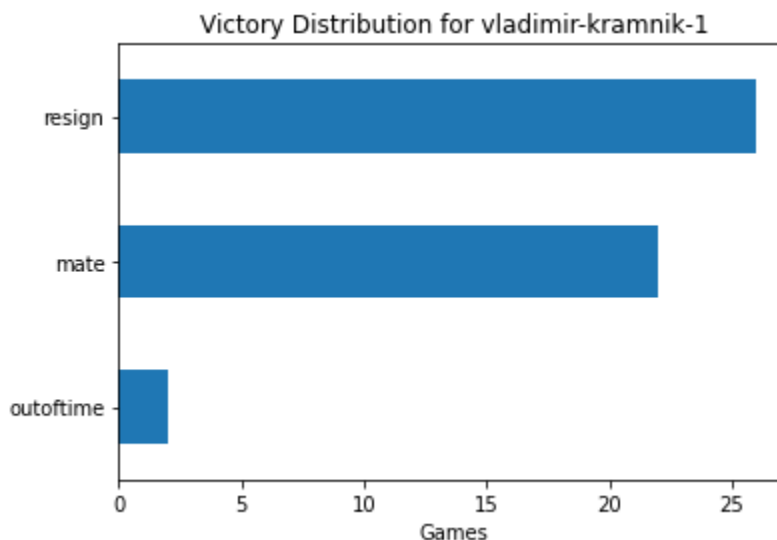
tier	opening_name	opening_counts
CM	Sicilian Defense	12
FM	Giuoco Piano	2
FM	Latvian Gambit Accepted | Main Line	2
FM	Nimzo-Indian Defense: Panov Attack | Main Line	2
FM	Queen's Pawn Game: Zukertort Variation	2
FM	Saragossa Opening	2
FM	Sicilian Defense: Closed #2	2
FM	Sicilian Defense: Delayed Alapin	2
GM	King's Indian Attack: Symmetrical Defense	1
GM	Nimzowitsch Defense | Declined Variation	1
GM	Pirc Defense #4	1
GM	Zukertort Opening: Sicilian Invitation	1
IM	Italian Game: Giuoco Pianissimo | Italian Fou...	2
IM	Nimzo-Indian Defense: Normal Variation | Rago...	2
NM	Queen's Pawn Game: Mason Attack	7
class A	Queen's Pawn Game: Mason Attack	40
class B	French Defense: Knight Variation	68
class C	Sicilian Defense: Bowdler Attack	109
class D	Van't Kruijs Opening	137
novices	Van't Kruijs Opening	74

It is interesting to note that for novices through national masters (NM) most just have one popular opening being played, however above that there are multiple openings per tier. This is partly due to the fact that there are less players reaching those tiers and therefore less games but also higher level players generally know and master more than one opening so they can react to better position themselves during a game.

The next step in my analysis was to investigate which openings were most played by rapid and blitz games. In chess, there are different game types determined by what the time controls are. The platform Chess.com defines a bullet game by less than 3 minutes, a blitz game from 3-5 minutes and a rapid game 10 minutes or more. Using these time constraints I applied labels to each of the games in the dataset. Then by counting each of the games in those categories and taking the top openings I came up with the following.

	Game Type	opening_name	Game_Openings_Count
0	blitz	Van't Kruijs Opening	86
717	bullet	Scandinavian Defense	7
792	rapid	Sicilian Defense	237

The final step in my analysis was to see who has won the most games on the platform and what their distribution was in the way they had won those games. The way I had tackled this issue was to create two dataframes in which we had a count of white wins and black wins grouped by each of the player IDs. I then needed to join these two dataframes and create a new column called combined wins which summed the wins for each player with each of the pieces. The result is shown below where the user vladimir kramnik 1 had won the most games with 50 and his victories mainly consisted of resignations from the other players and checkmates.

To conclude my analysis, while it will be helpful to learn the Sicilian Defense as a novice player, moving up the ranks in chess will require more discipline to learn a variety of openings and strategies to win the games. It is also important to look at how the strategies can change

based on the game type as we have seen before the time constraints can affect which openings may be more effective and better position your pieces to come out victorious.

References:

1) Chess rating system. (2021, August 02). Retrieved from
   https://en.wikipedia.org/wiki/Chess_rating_system
2) Shannon number. (2021, August 08). Retrieved from
   https://en.wikipedia.org/wiki/Shannon_number#Number_of_sensible_chess_games
3) Team, C. (2019, September 30). History of Chess: From Early Stages to Magnus.
   Retrieved from https://www.chess.com/article/view/history-of-chess

10 Audience Questions

1) What is the importance of a chess opening?
   Chess openings are important as they can set up the board strategically for either player. There is a lot of theory around opening moves and how each player can better position themselves to win the game based on the opponents moves.

2) How many chess openings are there?
   There are numerous openings in chess but the most popular ones that you can learn from chess.com include 34 different openings.
   https://www.chess.com/openings

3) What does each of the chess tiers mean?
   The different tiers of players is based on the ELO rating system for which more information can be found here (https://en.wikipedia.org/wiki/Elo_rating_system)
   For each of the titles starting with candidate master through grandmaster there are specific requirements that have to be met in order to achieve that title (https://en.wikipedia.org/wiki/FIDE_titles)

4) Who are the top chess players?
   A list of the top chess players can be found here https://www.chess.com/players, which includes Magnus Carlsen, Fabiano Caruana, and Ding Liren.

5) Where did this data come from?
   The dataset came from kaggle and was derived from LiChess.org games
   https://www.kaggle.com/datasnaek/chess

6) What is LiChess?
   LiChess is an online platform where individuals can register and play chess for free. They allow individuals to play against each other online and also have chess puzzles that can be solved.

7) What would be the best opening to learn for a beginner?
   Based on my analysis the best opening to learn for a beginner would be the sicilian defense.

8) What other strategies would you recommend as a beginner?
   Other strategies I would recommend is to try to anticipate your opponent's moves and look out for areas where you could be in check or lose pieces. Always guard your pieces and look for opportunities where your opponent doesn't guard a piece.

9) What other chess platforms are there?
   Another platform you can play on is chess.com (https://www.chess.com/home)

10) What is the difference between rapid and blitz?
    The difference between rapid and blitz games are the time constraints on the players. Blitz games are anywhere between 3-10 minutes per player and rapid are games greater than 10 minutes per player.