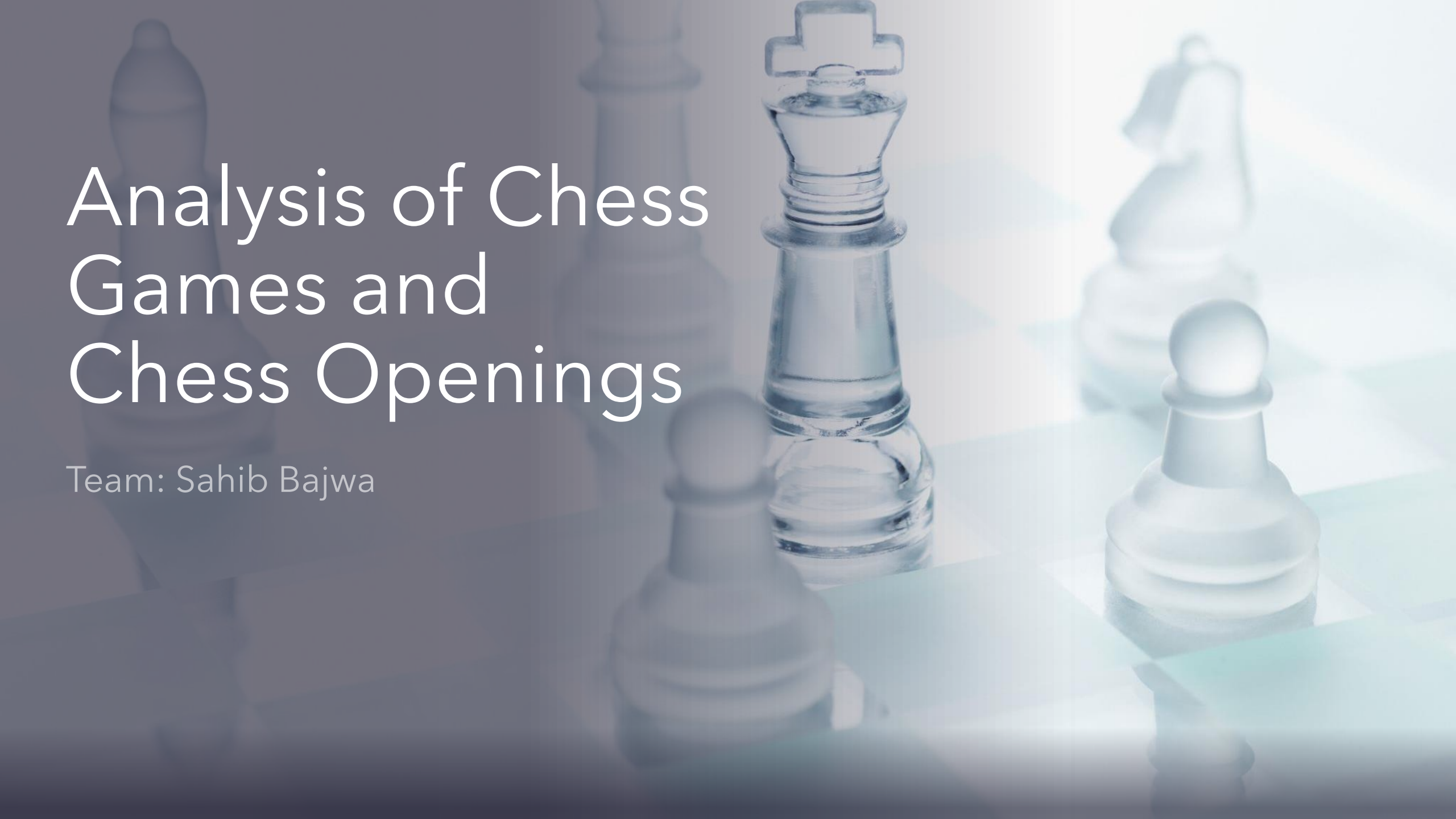

Analysis of Chess Games and Chess Openings

Team: Sahib Bajwa

Introduction

The Shannon Number tells us that the conservative lower bound of the game-tree complexity of chess is 10^{120}. To give this number some perspective, there are an estimated 10^{80} atoms in the universe. It is very difficult to determine the best way to play the game, even given the use of AI (Stockfish, AlphaZero, Leela Chess Zero, etc.). Having a strong opening is one of the most important moments of a chess game as it sets the pace and structure for the rest of the game. Given that humans do not have the computational ability of a strong AI, finding out what the best opening is for a player would help them increase their chances of winning a large amount.

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

Related Work

- Chess AI:
  - There are Chess AI that are built around specific types of playstyles that prefer specific types of game states. Thus, they start a specific way in order to push the game towards that game state.
- Chess Theory:
  - There has been a lot of theory crafting about how to play in specific game states. One great example of this is the book *FCO: Fundamental Chess Openings* by Paul van der Sterren.
  - There is also lots of chess theory that is done on the individual level by high level chess players before specific games. This is usually not seen in/by lower level players.

Completed Work

To determine the best openings for players to use we take in multiple factors. Some of these factors include:

- Player(s) Rating
- Player Color
- Opponent
- Opening Move
- Game Type

This was done by analyzing previously played chess games and then taking in the stated variables. There are many factors going into this analysis, so the output for the best opening will vary greatly dependent on the player.

Initial Evaluation vs New Evaluation

- Human winning percentage when given an opening
- Game state variation when given an opening.
  - Ease of transitioning from opening to later game state.
- Overall player thought when given an opening they should practice/implement.
- AI winning percentages given different ratings and specific openings.

Completed Milestones

- Determine if datasets should be merged.
    - No datasets were merged, but two different datasets were considered. The main dataset used for calculation used was the Lichess data set
- Clean data (determine what data should be used)
    - Data called by specified cells, even if there is missing data, we can learn a lot from the remaining data (example of missing player ids).
- Specified analysis of openings for all players.
- Specified analysis of openings for specific players.
    - Specified analysis of openings by player rating.
        - Specified analysis of openings by color.
- Setting games with openings vs each other to determine viability of openings.

Specified analysis of openings for all players

- The most popular opening for all players when considering double variation was Van’t Kruijs Opening
  - Second option: Sicilian Defense. Third option: Sicilian Defense – Bowdler Attack
- White has a much higher chance of winning at 49.86%!
  - Black: 45.40% and Ties: 4.736

Specified analysis of openings for specific (white starting) players

- White Starting Player Data
  - At rating 0 - 1000
    - First: Queen’s Pawn Game, Second: Scandinavian Defense
  - At rating 1000 – 2000
    - First: Van’t Kruijs Opening, Second: Sicilian Defense
  - At rating 2000 or above
    - First: Indian Game, Second: Queen’s Pawn Game

Specified analysis of openings for specific (black starting) players

- Black Starting Player Data
  - At rating 0 - 1000
    - First: Van’t Kruijs Opening, Second: Scandinavian Defense
  - At rating 1000 – 2000
    - First: Van’t Kruijs Opening, Second: Sicilian Defense
  - At rating 2000 or above
    - First: Indian Game, Second: French Defense: Knight Variation

Preferred Openings based on starting color and removing double variation

- Interestingly enough, if we remove rating, the most popular openings when removing double variation for both colors is similar
  - Queen’s Pawn Game #2
  - Van’t Kruijs Opening
  - French Defense: Knight Variation

Findings when using real players

- 7 players of 1000-1500 rating given openings
  - Overall majority said that Van't Kruijs Opening was their favorite to use
  - Player(s) who had a leaning higher rating (near 1500) said that they would prefer to change their opening based on the current game
- Conclusion
  - Lower rated players prefer being given a specific opening to use, and Van't Kruijs is their preferred one
  - The higher the player rating gets, the more they would like to play on their own and not stick to a specific opening

AI Findings

- It turns out that trying to program chess ai to act in a specific way is difficult
  - For example, it has taken multiple years for top chess AI company Stockfish to release a Chess AI that plays like specific veteran players
- My AI
  - Simple, made it so that the AI can analyze slightly, taking high valued pieces from the opponent player
  - Winning percentages fall in line with Lichess data set, where the white starting player has an advantage
  - *Other dataset