Chess: Opening Moves

Elena Adame

20 November 2021

For this project, I sought to determine whether the opening moves made in a chess game affected the outcome of the game. I discovered through my analysis that there is a statistically significant relationship between these two variables. This significance does not necessarily mean that one opening move over the other guarantees a win or a loss, but it does show that it can impact the game. For this project, I chose to perform analysis of the top five most used chess openings. These were the Sicilian Defense, French Defense, Queen’s Pawn Game, Italian Game, and King’s Pawn Game. I found through my scatterplot and pmf, that these openings did not have a high correlation to games ending in a draw, or games that ended due to running out of time. Three openings, Sicilian Defense, French Defense, and Queen’s Pawn Game, related to a large number of games ending in a player resigning.

I would need to perform further analysis to determine whether there are other openings that impact the game’s outcome. I feel that I missed out on analyzing the openings in total given how many openings were included within the dataset. Given more time, I would determine a method to capture a portion of openings whose usage is in the middle section and analyze those in conjunction with the top 5 least used openings. Additionally, I would like to further explore the impact of the players rating with regard to the openings used and the outcome of the game. I would segment the data into highly skilled, average skilled, and low skilled players. Then, I would explore whether the openings, used for each of those skillsets, impacted the game’s outcome.

Additionally, one other variable I feel could have been utilized better and helped with the analysis was the Victory Status variable. I feel that while the Winner variable provided information about which player won, the Victory Status provided more in-depth information about the real outcome of the game.

When beginning my EDA, I assumed that the opening of the game would not have any impact on the games outcome and that I would find that there was no statistical significance. However, I was proven wrong.

One challenge I experienced during this project was the handling of the different object types within my data. My dataset was clean when I began my project, meaning that I found no empty cells for the variables I had chosen. However, when I went to perform various tasks in my EDA of the data, I was often faced with the ValueError from python. My code could not handle one variable being an ‘int’ object type and one variable being a ‘str’ object type. To get around this, especially when calculating Spearman and Pearson correlation, I changed two of my variables over from ‘str’ to ‘int’. Moving forward, I would like to find a less clunky way than the method I found. I found this project both rewarding and difficult.