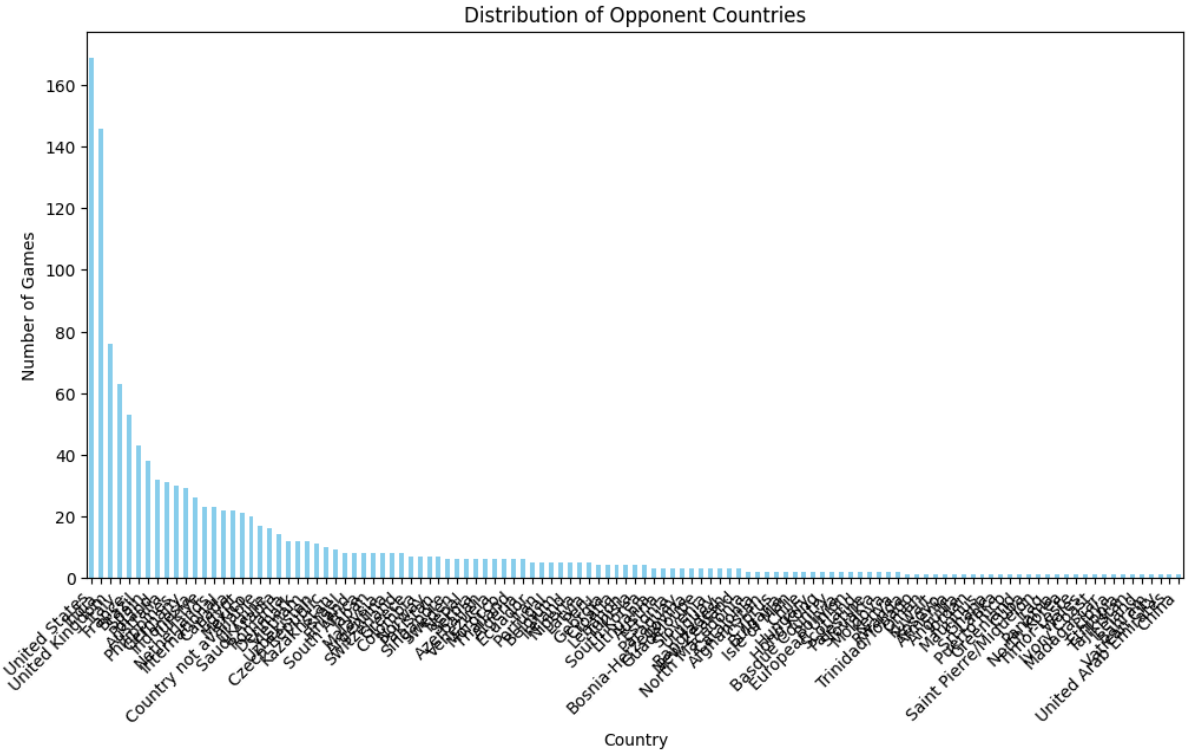
Buizza Federico 33656

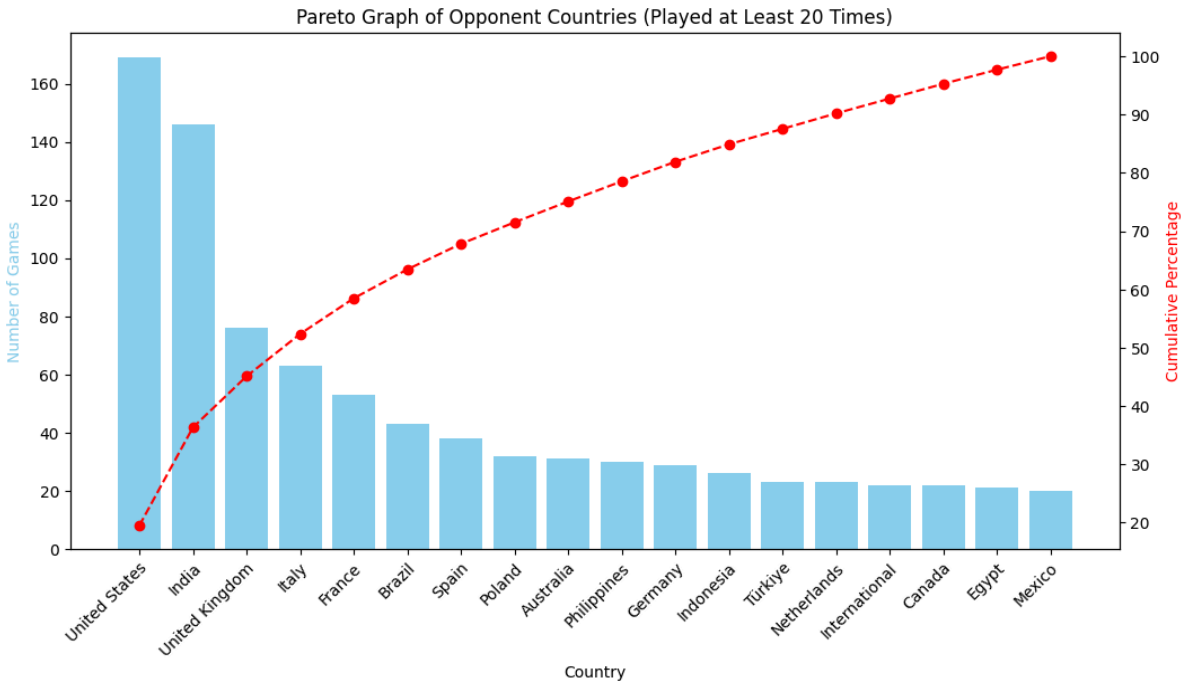
FINAL PROJECT CS210

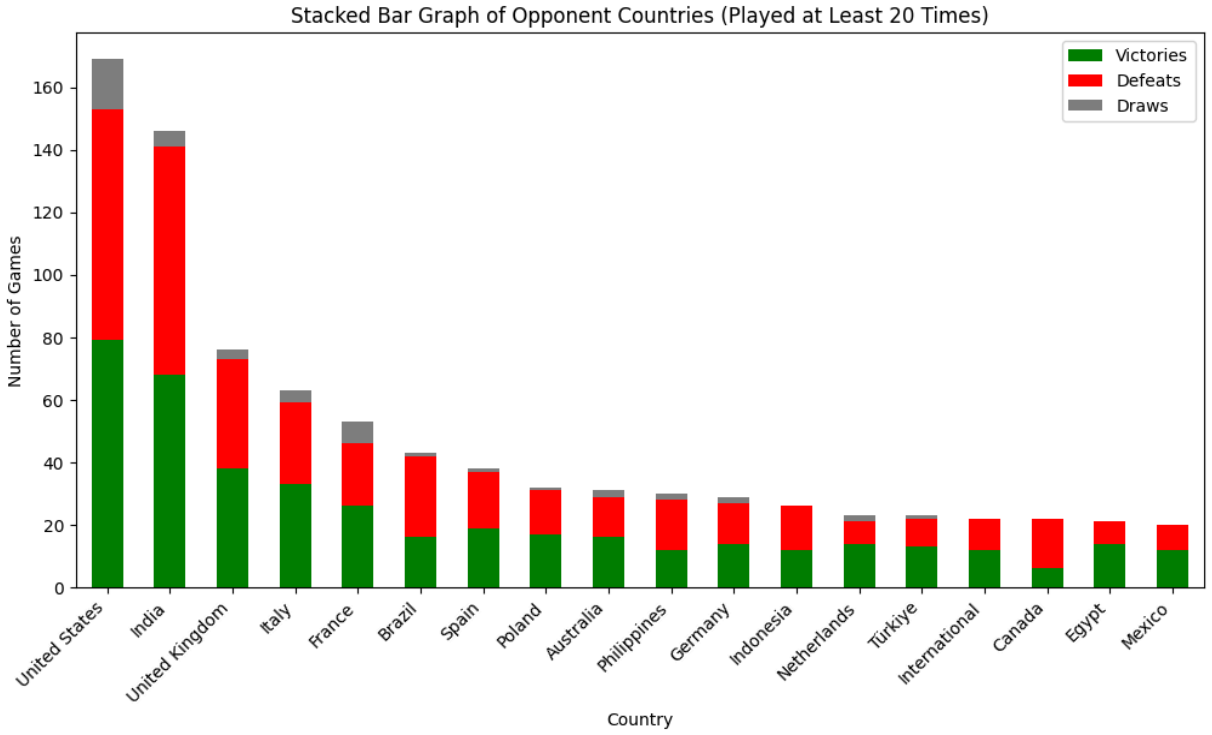
For this project, I am going to analyze my chess games. Specifically, I will test the long-debated theory of whether being the first mover provides an advantage. Additionally, I will analyze whether purchasing a premium subscription has helped me improve my game.

Before delving into this analysis, I would first like to explore something related to one of my favorite topics: geography. Below, you can see a graph of all the countries from where my opponents were.

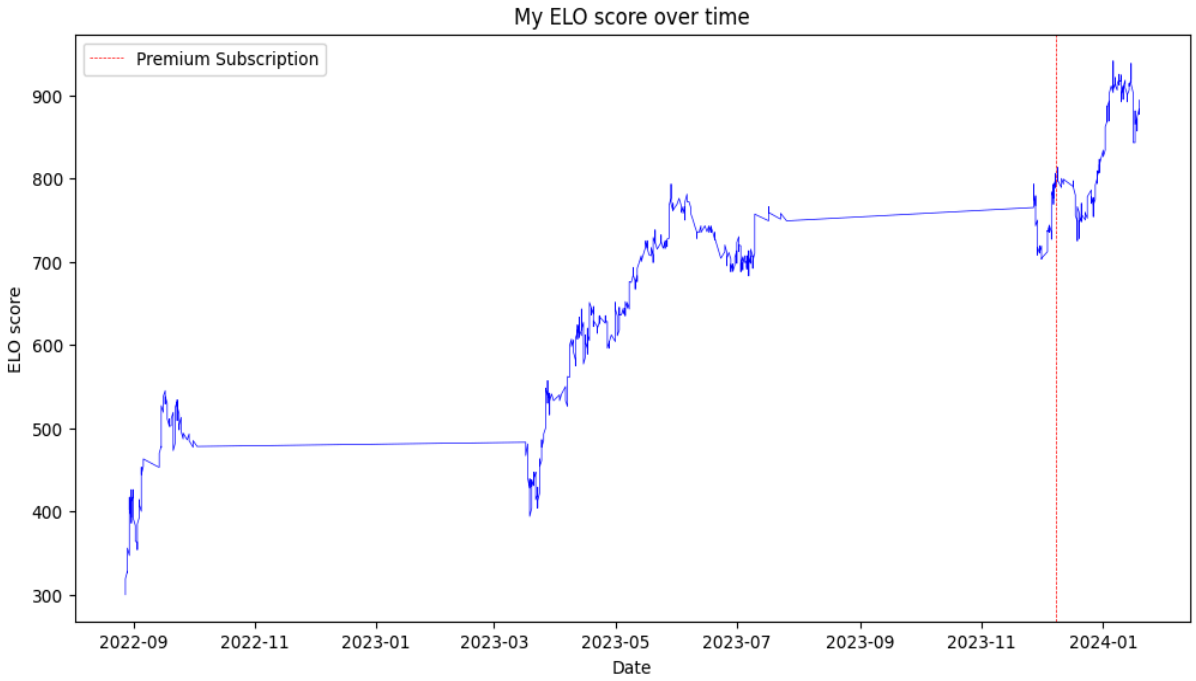


Since there are so many of them, I will focus on the ones that I faced at least 20 times.

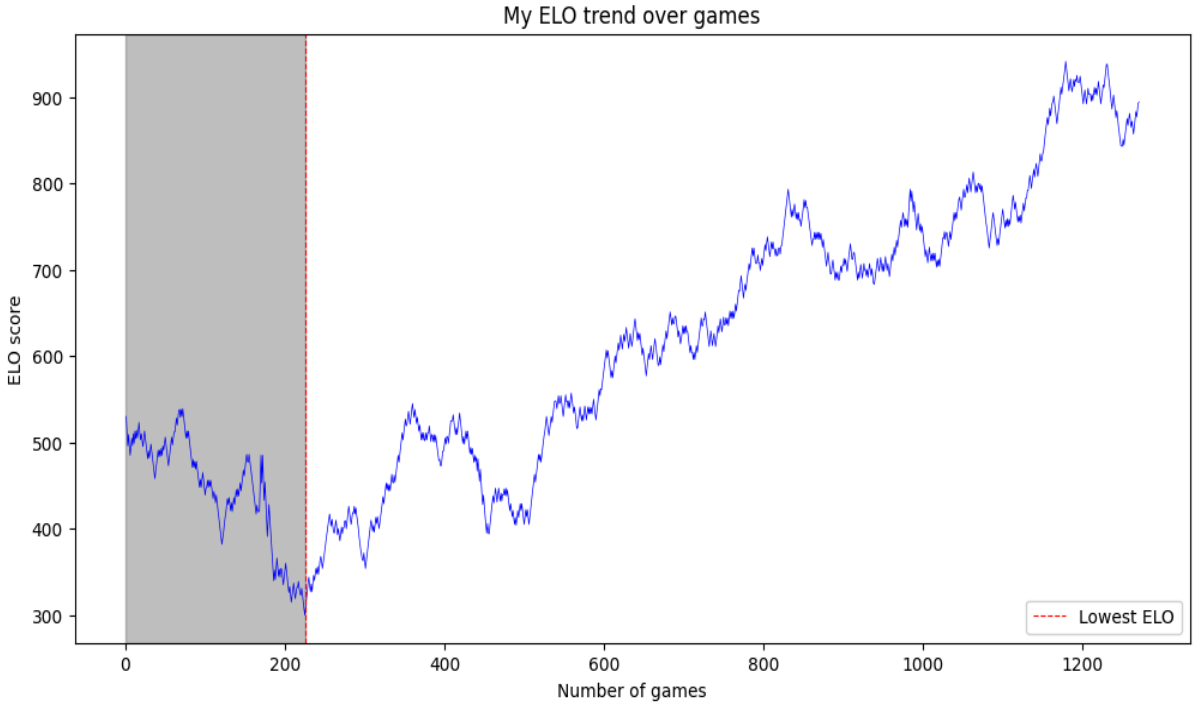




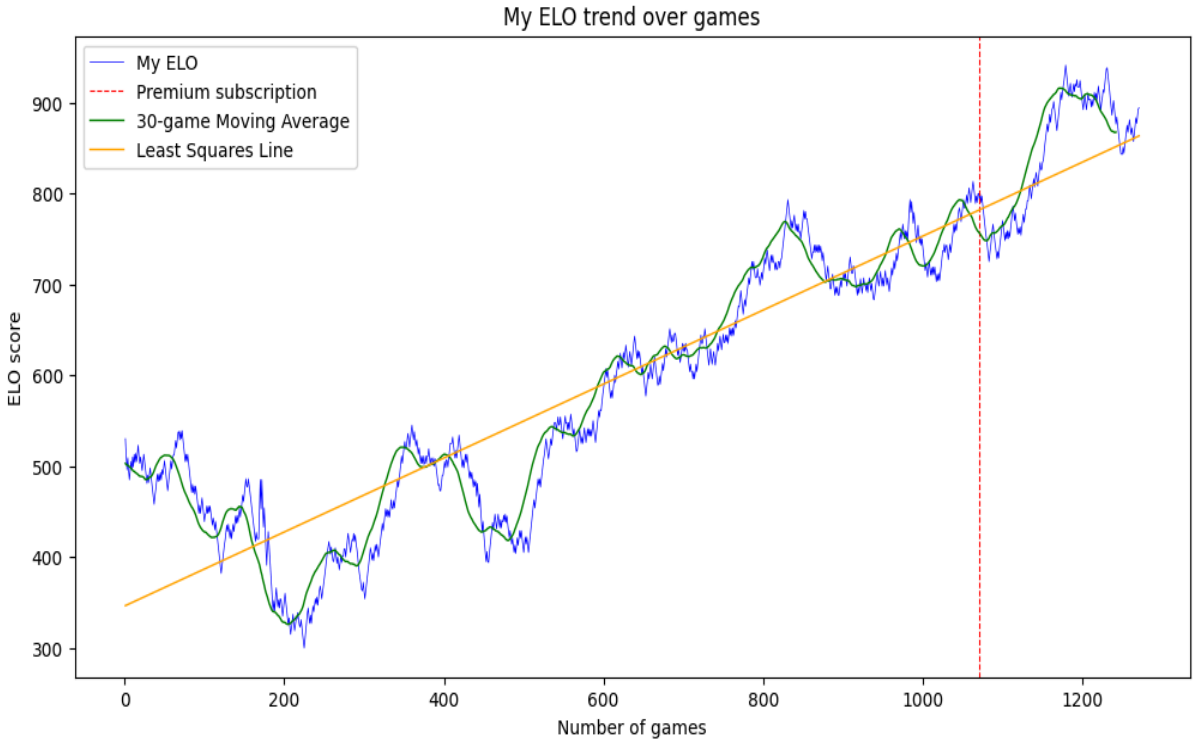
Now let’s focus on the main topic. Before going deep into research it seems reasonable to visualize my progress over time.

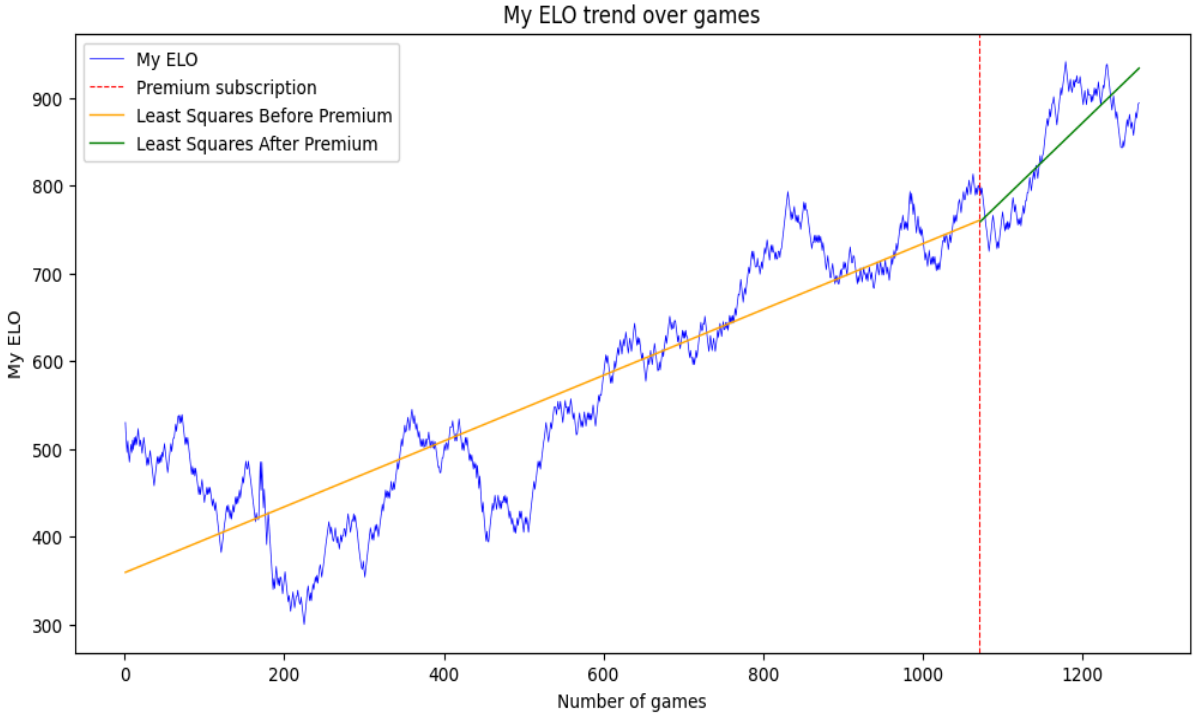


From this graph, it is not possible to understand clearly how my game evolved, as there have been long periods in which I did not play, affecting the graph understanding negatively. Therefore, I will remove those blank time spaces.



This seems better too. Additionally, for a more robust analysis, I've decided to exclude the initial games where I experienced nothing but a downtrend. Removing those will allow for a more accurate comparison without the influence of many outliers.

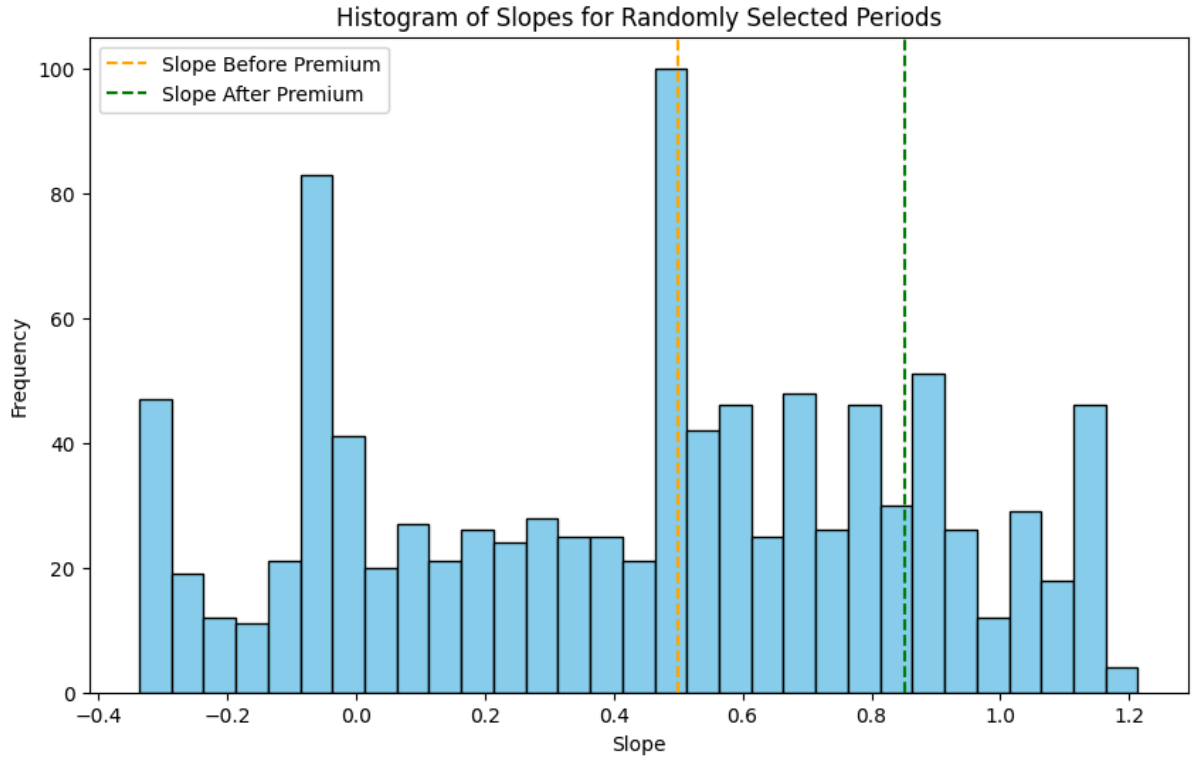




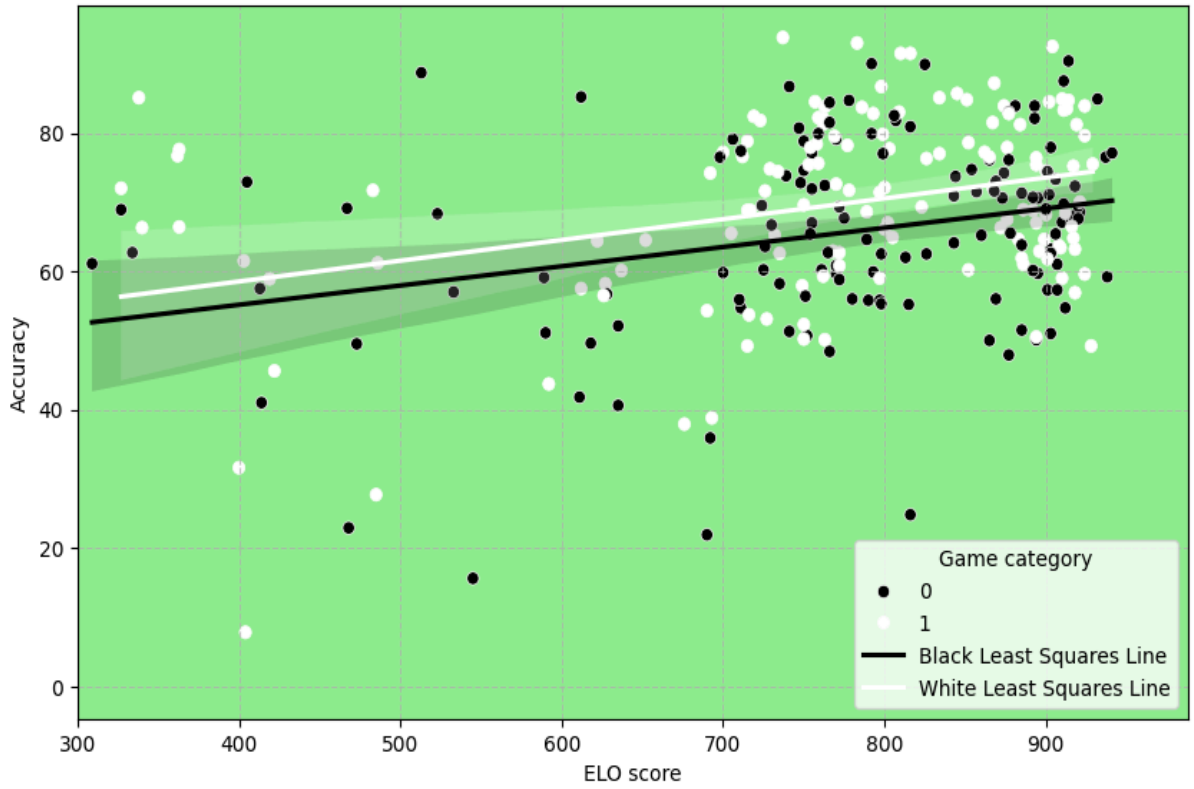
The graph above displays a comparison of the least square lines slopes between premium games and non-premium games. Specifically, the actual values of the two slopes are as follows:

- Slope Before Premium: 0.375
- Slope After Premium: 0.878

This result essentially means that before purchasing a premium account, I would gain 0.375 ELO points for every game. Now, it has increased to 0.878. To test whether this difference is significant, I will run a simulation. I will draw 1000 sample intervals from the non-premium population, each of the same size as the premium games. Then, I will compute all of them and present the results on a histogram.



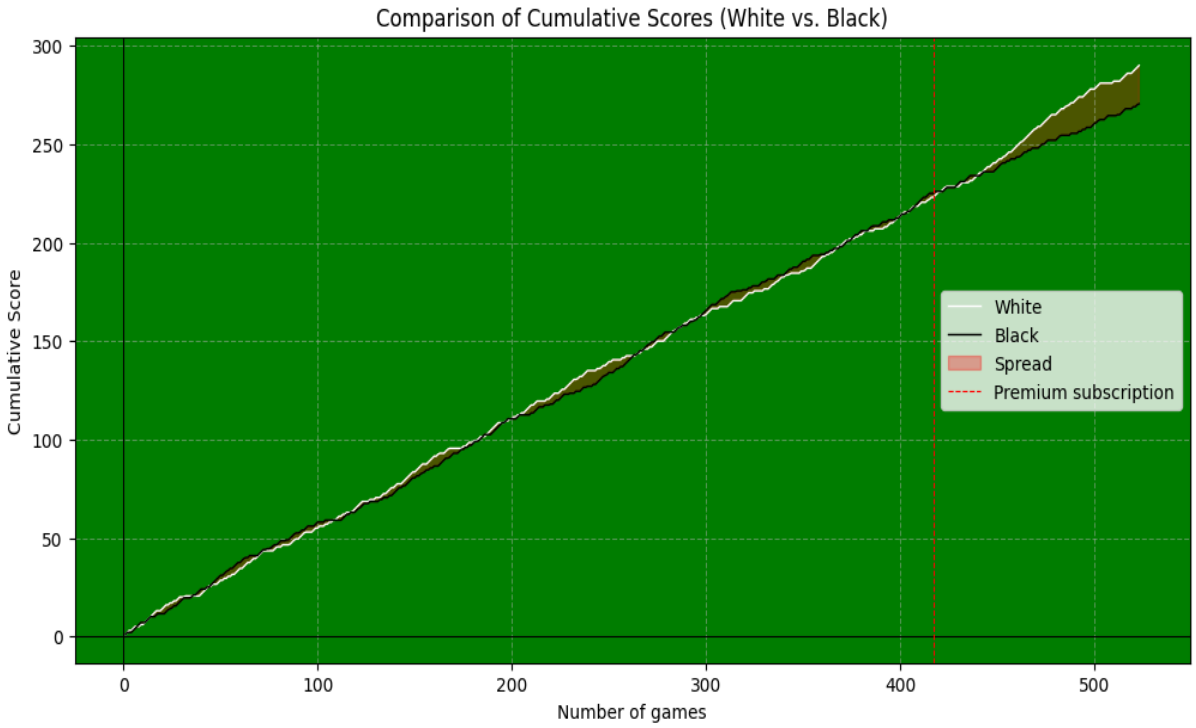
Now let's focus on the main topic of this research: the first mover advantage. First of all, I will check if on average I play more accurately as white.



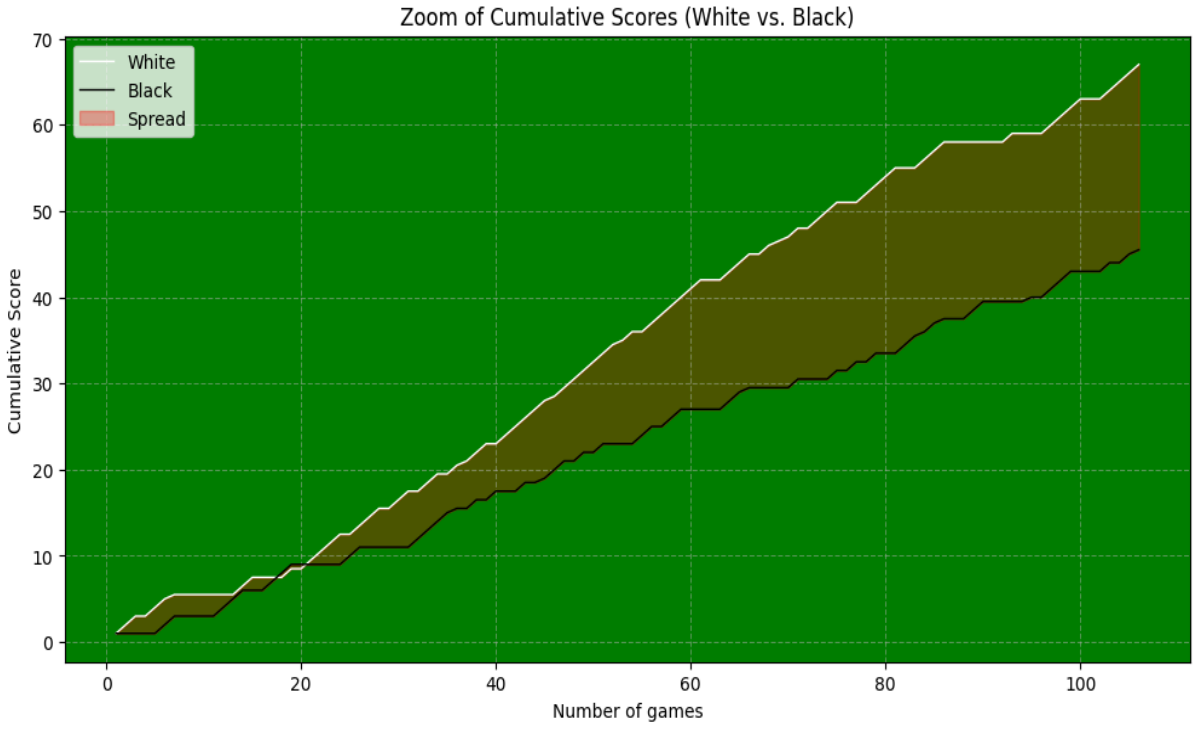
The two least squares lines, while almost parallel, show a few points of difference. On average, when I play as white, I tend to be slightly more accurate, around 2 points better:

- Black Line: Slope = 0.0278, Intercept = 44.0193
- White Line: Slope = 0.0300, Intercept = 46.5107

Let's continue the comparison between games as black and games as white. Below, a graphical comparison of the two types of games is presented. Specifically, each game result increases the line by 1 if the game is a win, 0.5 if it is a draw, and 0 if it is a defeat.

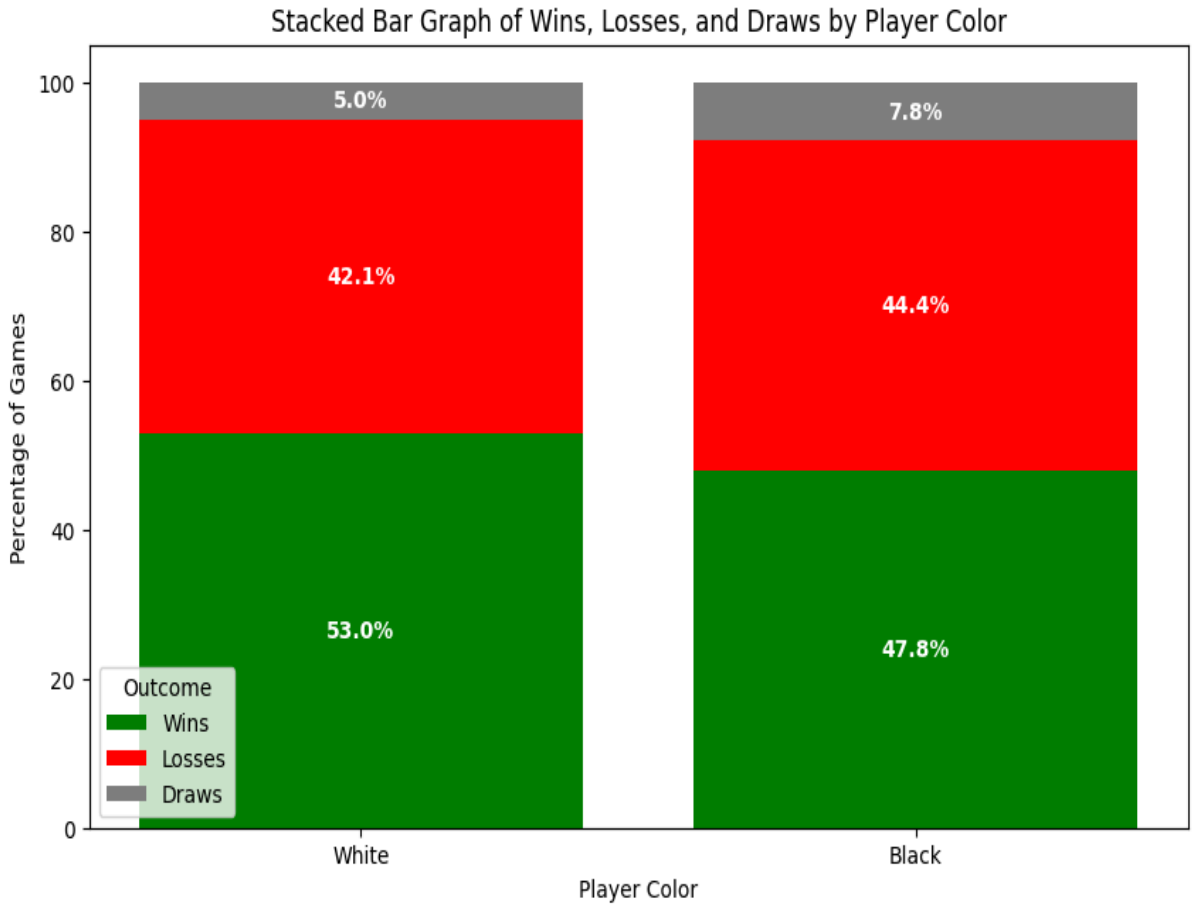


The difference between black and white results appears to widen exponentially just after the premium subscription was purchased. Let's zoom in on the period after the purchase.

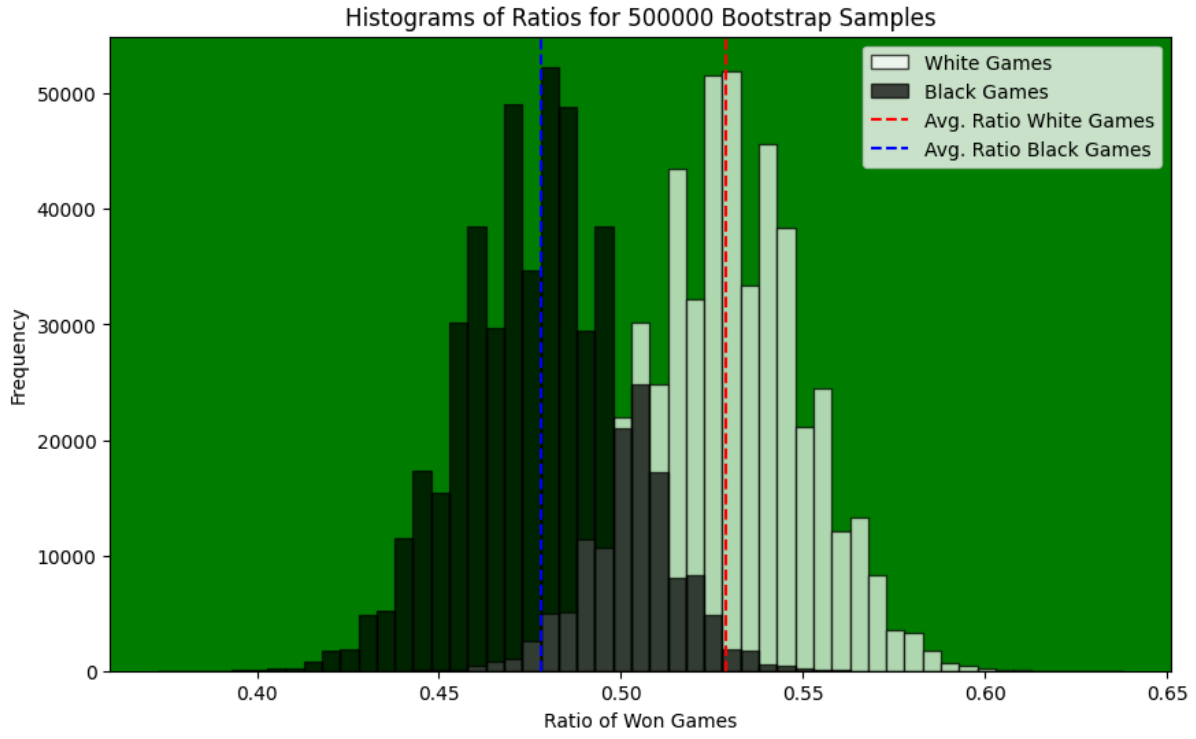


Compared to the previous period, where there is almost no difference at all over a period of more than 800 games, after the premium subscription, it appears that the gap widened progressively.

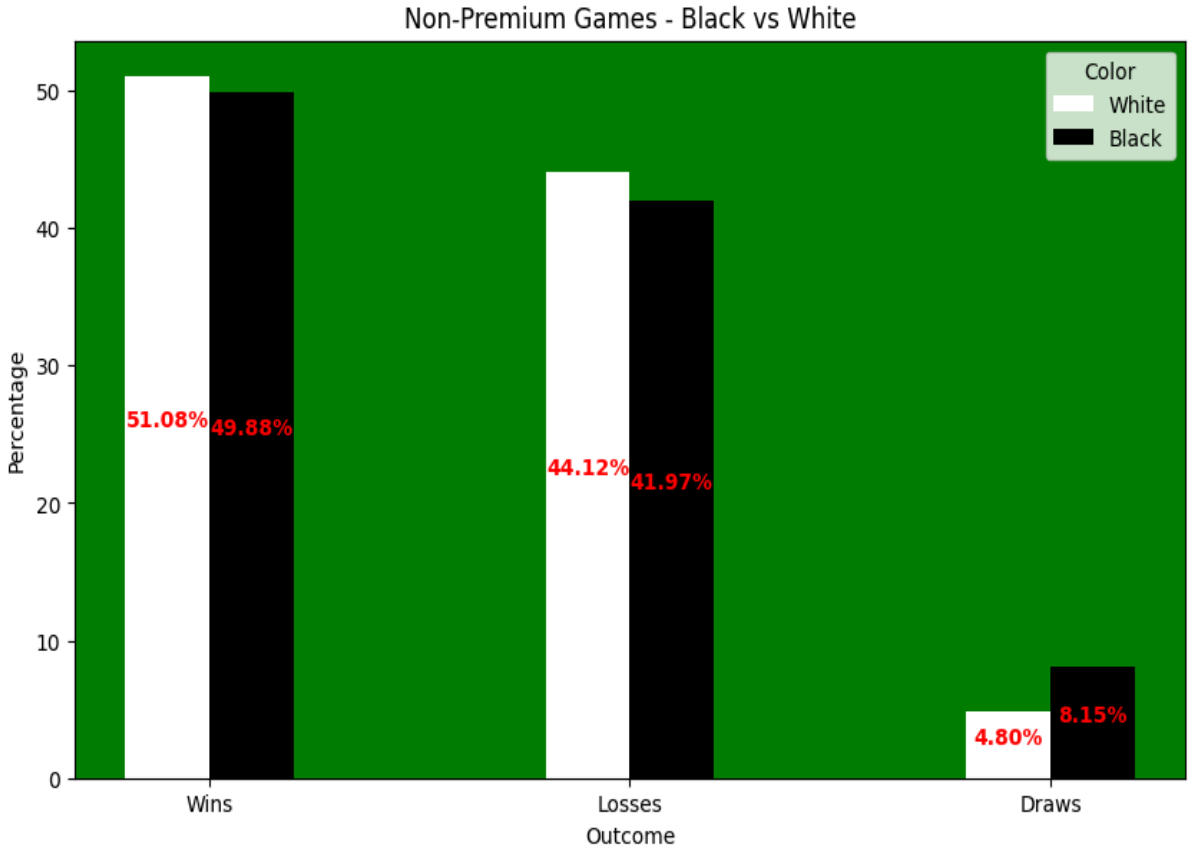
While the overall situation appears to be like this:

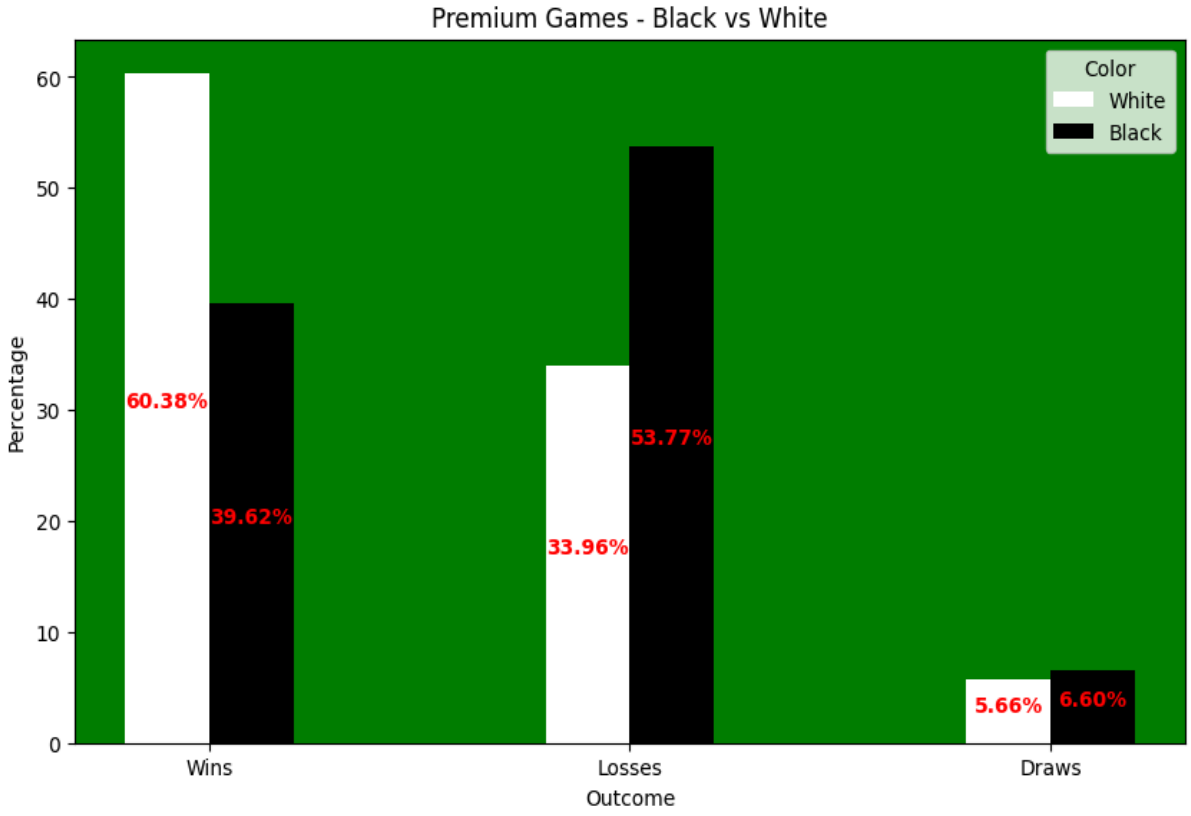


And a simulation on this type of comparison gives these significant results:



The results are even more drastic if we isolate the premium games from the non premium ones:





Looking at this new comparison it does not even seem necessary to run a statistical test: the results speak for themselves, with a stunning 20% difference in wins between black and white after premium period.
In conclusion we can say that, at least for me, playing as white gives a slight advantage. That advantage has been amplified after the premium purchase. Whether this is due to learning new openings or playing with more skilled players, it will investigated in the next project. ;)