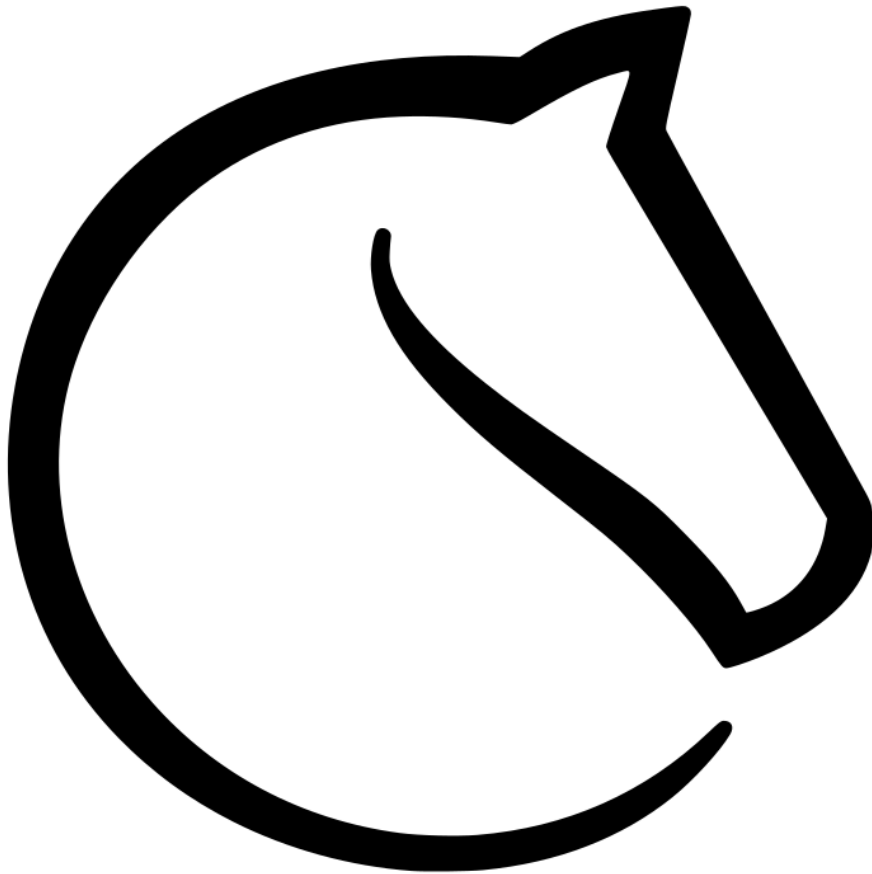


CS 210 REPORT

“MY OWN LICHESS HISTORY ANALYSIS”

Prepared By: Kağan Aydınçelebi



1. Motivation

As a FIDE Master (FM) in chess, this project offers an opportunity to dive into my personal game history and analyze my performance. By utilizing data science techniques, I aim to identify patterns, strengths, and areas for improvement in my gameplay, enabling me to refine my strategies and achieve greater success in the future.



2. Data Source

2.1 Data Extraction

The dataset for this project is a PGN file which includes detailed history of my chess games from Lichess. This file include information about each match, such as the event type, date, opponent, result, player ratings, rating changes, and the sequence of moves played. I used Lichess API to get this data. Here a sample of one Lichess game:

```
[Event "Hourly SuperBlitz Arena"]
[Site "https://lichess.org/fOiCtSQz"]
[Date "2016.05.12"]
[White "SpeedRacer"]
[Black "legend2014"]
[Result "0-1"]
[UTCDate "2016.05.12"]
[UTCTime "19:02:28"]
[WhiteElo "1722"]
[BlackElo "1879"]
[WhiteRatingDiff "-7"]
[BlackRatingDiff "+57"]
[BlackTitle "FM"]
[Variant "Standard"]
[TimeControl "180+0"]
[ECO "B90"]
[Termination "Normal"]

1. e4 c5 2. Nf3 d6 3. d4 cxd4 4. Nxd4 Nf6 5. Nc3 a6 6. Bd3 g6 7. Be3 Bg7 8. f3 O-O 9. Qd2 b5 10. g4 Bb7
11. h4 d5 12. h5 dxe4 13. fxe4 Nxe4 14. O-O-O Nxe3 15. Qxe3 Qxd4 16. Qxd4 Bxd4 17. Ne2 Bg7 18. Rdf1
Nc6 19. Nf4 Ne5 20. hxg6 hxg6 21. Be2 Rac8 22. Rh3 Bxe4 23. c3 b4 24. c4 Nxc4 0-1
```

As you can see, there are several details for each game such as “RatingDiff”, “Elo”. It was easy step to extract necessary data, thanks to API. Now, we can create a datatable that includes these informations as a column.

	Event	Site	Date	White	Black	Result	UTCDate	UTCTime	WhiteElo	BlackElo	WhiteRatingDiff	BlackRatingDiff
0	Rated bullet game	https://lichess.org/oECdV8Ew	2024.12.12	VladTheBest	legend2014	1-0	2024.12.12	18:01:22	2919	3050	+8	-12
1	Rated bullet game	https://lichess.org/f6smS9O0	2024.12.12	legend2014	sid-al_1979	0-1	2024.12.12	17:58:13	3059	2754	-9	+10
2	Rated bullet game	https://lichess.org/vi4fXJ2z	2024.12.12	SantosBarrios	legend2014	0-1	2024.12.12	15:24:25	2820	3057	-2	+2
3	Rated bullet game	https://lichess.org/csbBQelA	2024.12.12	legend2014	stdcout	1-0	2024.12.12	12:55:22	3053	2936	+4	-4
4	Rated bullet game	https://lichess.org/ZJJpFq73	2024.12.12	legend2014	WhatWouldMorphyDo	1-0	2024.12.12	12:37:25	3050	2859	+3	-3

2.2 Data Preprocessing

In the data preprocessing phase, I focused on cleaning and refining the dataset to ensure its convenience. First, I removed unnecessary columns such as "FEN", "SetUp", and "ECO", which did not provide valuable information for our analysis. These columns were irrelevant to our objectives and could clutter the dataset. Additionally, I filtered out rows where the "Moves" column was empty, as such rows do not contain any meaningful game data. This step ensured that only complete and valid game records were retained, allowing us to focus on analyzing games with sufficient information. Furthermore, I add new column “ResultStatus” shows the result of game as in “win”, “loss” or “draw”.

	UTCDate	UTCTime	WhiteElo	BlackElo	WhiteRatingDiff	BlackRatingDiff	WhiteTitle	BlackTitle	Variant	TimeControl	Termination	Moves	ResultStatus
0	2024.12.12	17:58:13	3059	2754	-9	+10	FM	FM	Standard	60+0	Normal	1. d4 Nf6 2. c4 c5 3. dxc5 e6 4. Nf3 Bxc5 5. e...	loss
2	2024.12.12	15:24:25	2820	3057	-2	+2	None	FM	Standard	60+0	Time forfeit	1. e4 c5 2. Nf3 d6 3. Bb5+ Nd7 4. O-O Nf6 5. R...	win
3	2024.12.12	12:55:22	3053	2936	+4	-4	FM	None	Standard	60+0	Normal	1. e4 c5 2. b3 Nc6 3. Bb2 Nf6 4. e5 Nd5 5. Nf3...	win
4	2024.12.12	12:37:25	3050	2859	+3	-3	FM	None	Standard	60+0	Time forfeit	1. e4 c5 2. Nf3 Nc6 3. Bb5 Nd4 4.	win

3. Data Analysis Techniques

3.1 Descriptive Analysis

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 28838 entries, 0 to 28837
```

```
Data columns (total 18 columns):
```

#	Column	Non-Null	Count	Dtype
0	Event	28838	non-null	object
1	White	28838	non-null	object
2	Black	28838	non-null	object
3	Result	28838	non-null	object
4	UTCDate	28838	non-null	object
5	UTCTime	28838	non-null	object
6	WhiteElo	28838	non-null	object
7	BlackElo	28838	non-null	object
8	WhiteRatingDiff	27437	non-null	object
9	BlackRatingDiff	27437	non-null	object
10	WhiteTitle	20590	non-null	object
11	BlackTitle	20677	non-null	object
12	Variant	28838	non-null	object
13	TimeControl	28838	non-null	object
14	ECO	28838	non-null	object
15	Termination	28838	non-null	object
16	Moves	28838	non-null	object
17	ResultStatus	28838	non-null	object

```
dtypes: object(18)
```

```
memory usage: 4.0+ MB
```

```
None
```

	Event	White	Black	Result	UTCDate	\
count	28838	28838	28838	28838	28838	
unique	113	5361	5293	3	1803	
top	Rated bullet game	legend2014	legend2014	1-0	2022.07.23	
freq	22583	14404	14434	14072	117	

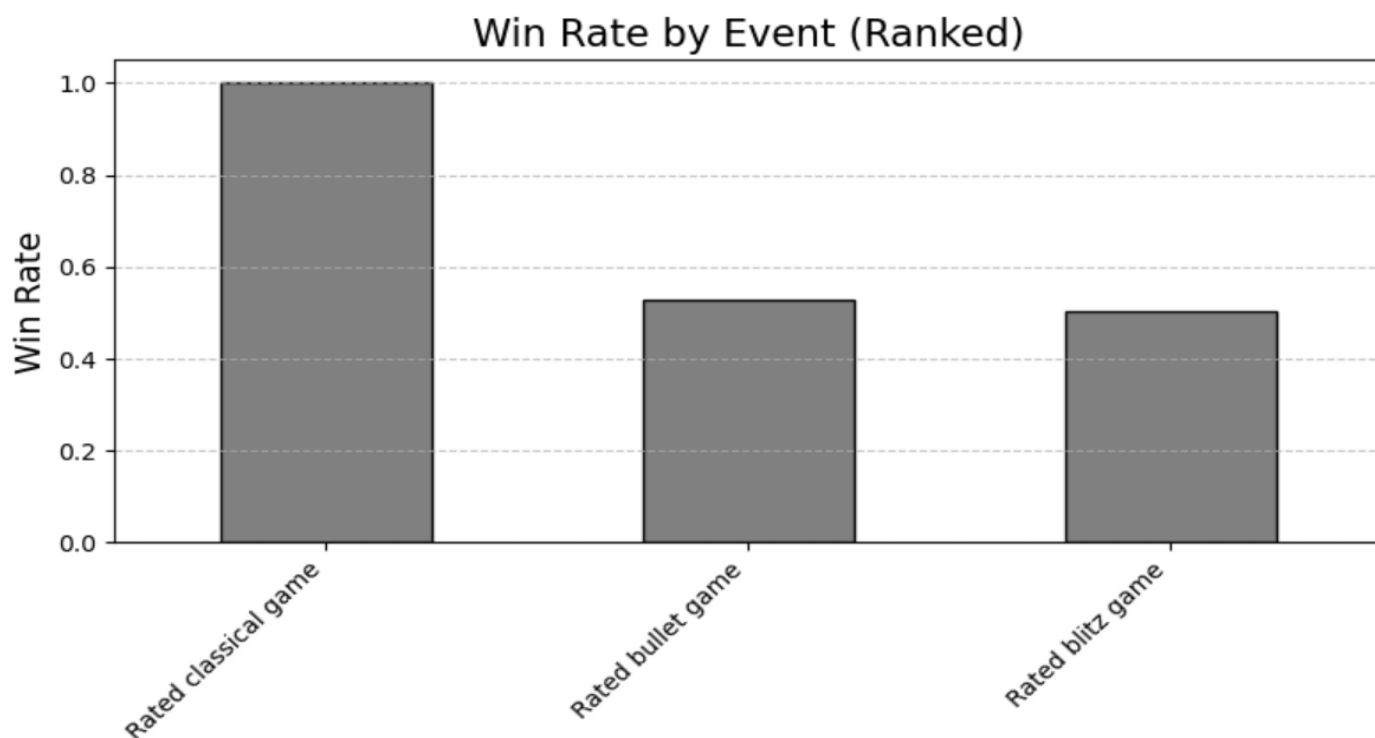
	UTCTime	WhiteElo	BlackElo	WhiteRatingDiff	BlackRatingDiff	WhiteTitle	\
count	28838	28838	28838	27437	27437	20590	
unique	22245	1480	1474	135	132	10	
top	19:00:02	1500	1500	+5	-4	FM	
freq	9	163	160	2095	2080	16018	

	BlackTitle	Variant	TimeControl	ECO	Termination	Moves	ResultStatus
count	20677	28838	28838	28838	28838	28838	28838
unique	10	10	36	345	3	28812	3
top	FM	Standard	60+0	A00	Normal	0-1	win
freq	16073	28378	24059	2132	22864	10	15442

The dataset contains 28,838 chess games with attributes like players, events, results, and moves. Most games are from "Rated bullet game" events, with "legend2014" being the most frequent player. White wins are slightly more common, with "1-0" as the most frequent result. The most popular time control is "60+0," and the most common opening falls under the ECO code "A00." The dataset highlights fast-paced games and emphasizes "legend2014's" performance.

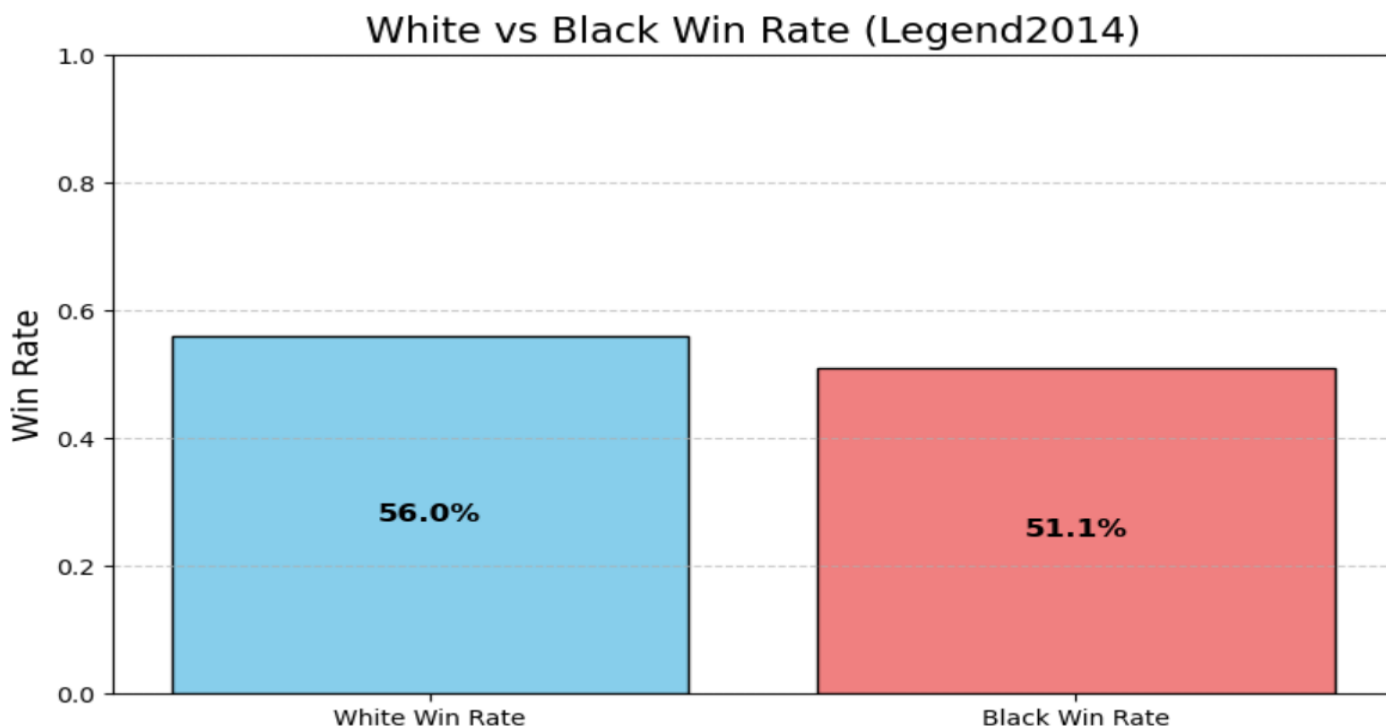
3.2 Exploratory Data Analysis (EDA)

1) Which chess mode, I perform best?



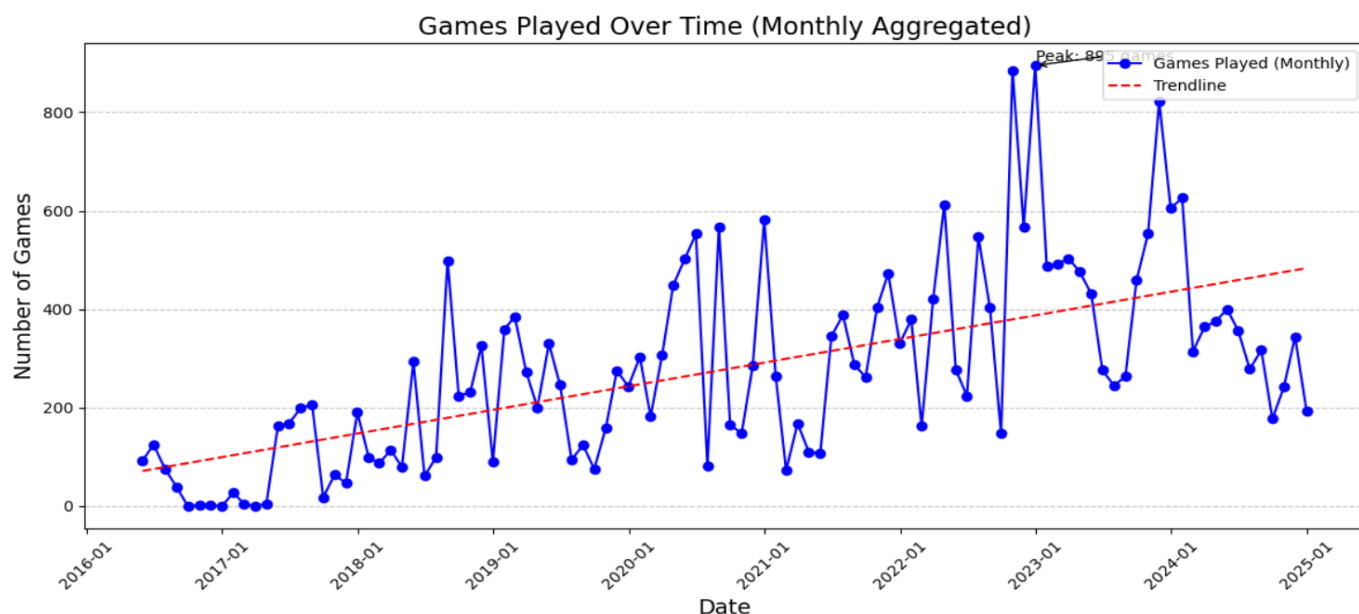
The findings show that the win rate is the highest in Rated Classical Games (nearly 100%), indicating exceptional performance in longer time formats. The reason of this result, might be due I played only few games in “Ranked Classical Game” and win all of them.

2) Black or White?



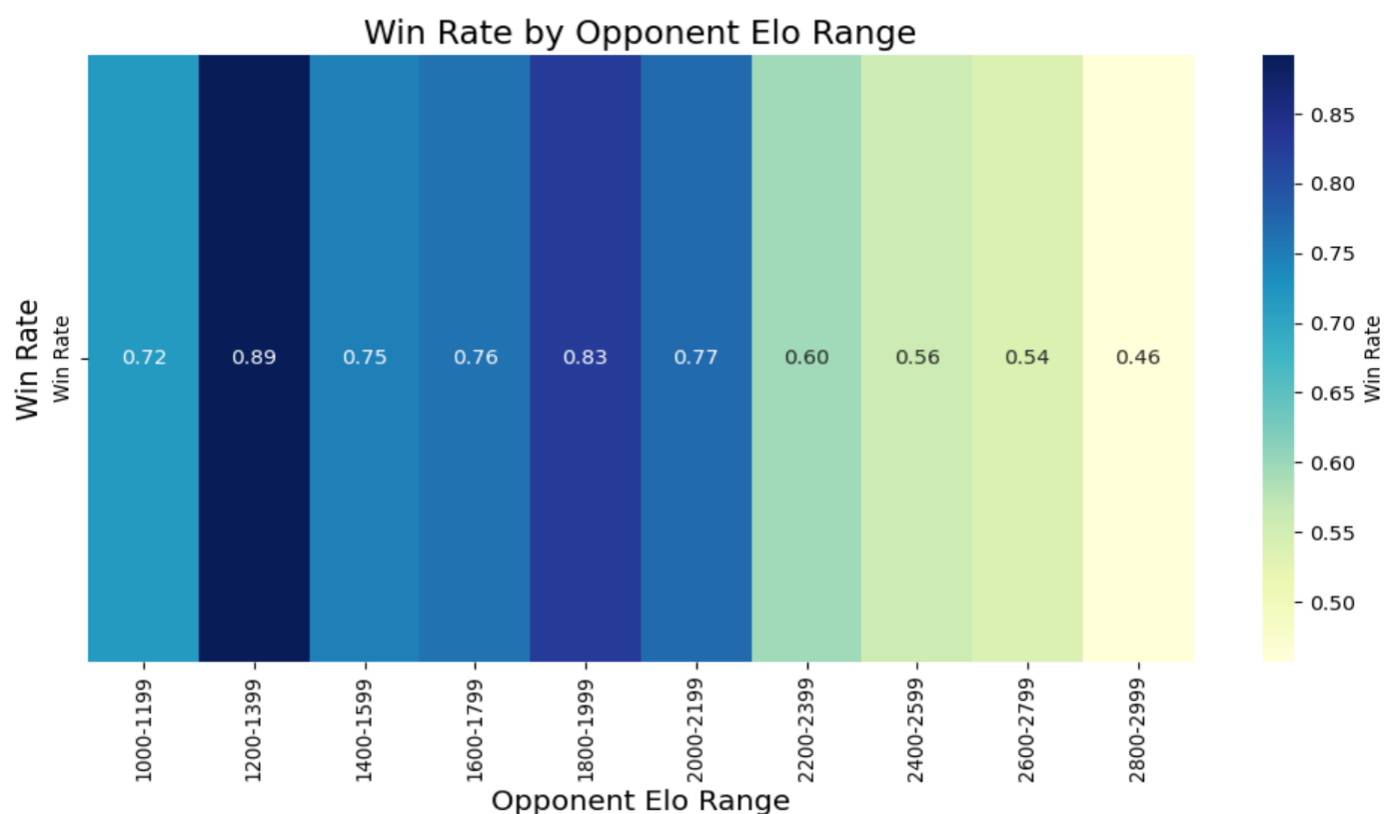
This chart shows my win rates when playing as White and Black. I win **56.0%** of my games as White and **51.1%** as Black. I do slightly better as White, probably because of the first-move advantage, but I still perform well as Black.

3) How Has My Chess Activity Changed Over the Years?



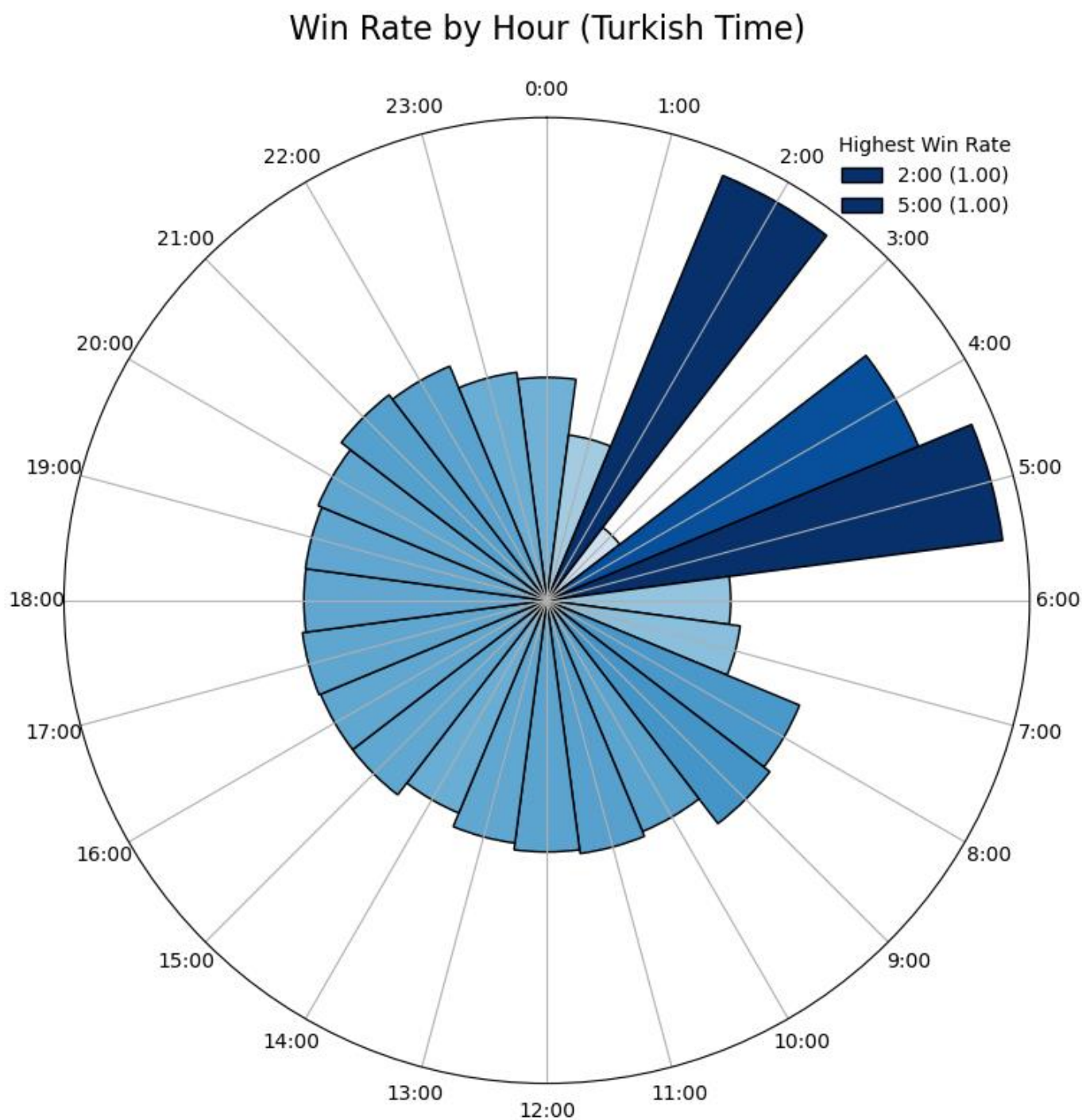
This chart shows how many games I played each month from 2016 to 2025. The blue line tracks my games, and there's a clear peak where I played 895 games in one month. The red dashed line shows that, overall, I've been playing more games over time. There are some months where I played less, but my activity **generally increased**, especially in the middle years.

4) How does my win rate vary across different opponent Elo ranges?



This heatmap shows how I perform against opponents of different Elo ranges. As the opponent Elo increases, my win rate gradually decreases, dropping to around 46% against players in the 2800-2999 range. This trend reflects the expected difficulty of playing against stronger opponents. Overall, I perform significantly better against players rated below 2000, with win rates consistently above 70%.

5) At what times of the day do I perform best in chess games?



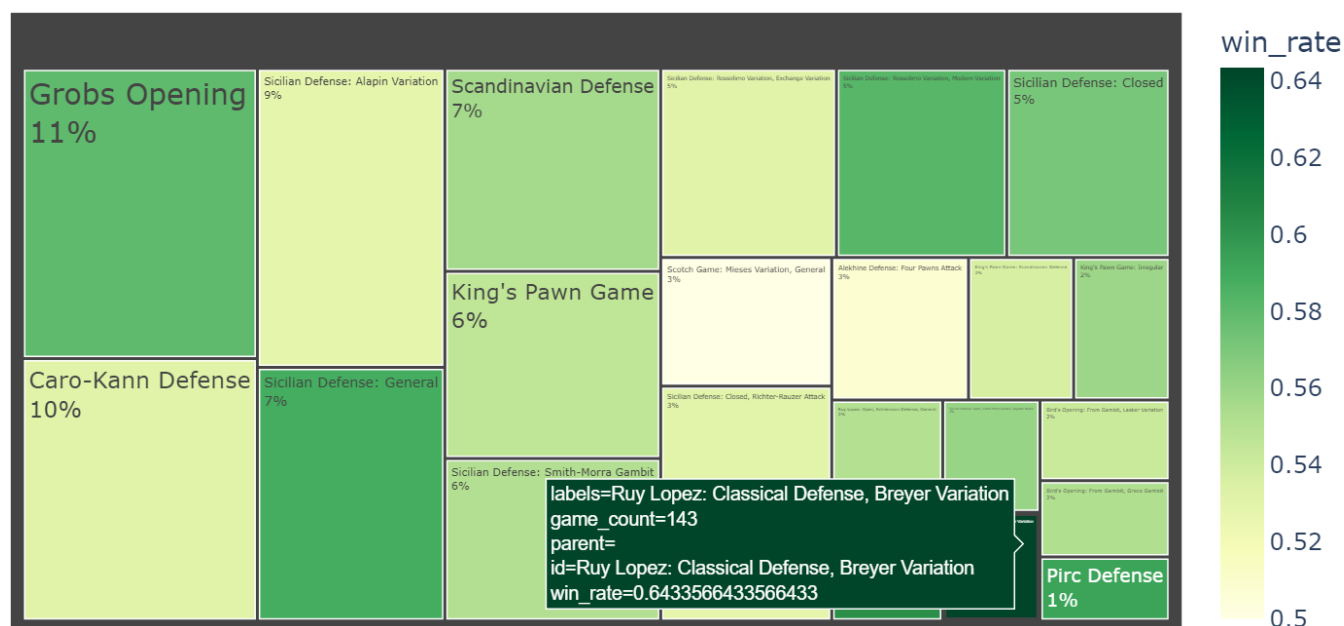
The radial visualization shows how my win rate varies across the 24 hours of the day. The larger and darker segments represent the hours when I achieve my highest win rates. For example, I perform best at 2:00 and 5:00, since the piece is dark, whereas I perform worst at 3:00.

“Ranked Bullet Game” Analysis

In LiChess, I am a Top 100 player in Bullet Game format (1+0). From now on, I will focus only my bullet games and investigate them to increase my winrate.

a. Which opening I won most, when I am playing with White?

Win Rate by Opening (White Side) - Ranked Bullet Games



This treemap visualizes my performance in ranked bullet games as White, showing the win rates for different openings. The size of each box represents how frequently I played that opening, while the color reflects the win rate (darker green indicates a higher win rate). As seen here, *Ruy Lopez: Classical Defense, Breyer Variation* stands out as my best-performing opening, with a high win rate and a significant number of games played. Although I played lots of games with Caro-Kann Defense, I have low win rate compared to other openings.

b. Which opening I won most, when I am playing with Black?

Win Rate by Opening (Black Side) - Ranked Bullet Games

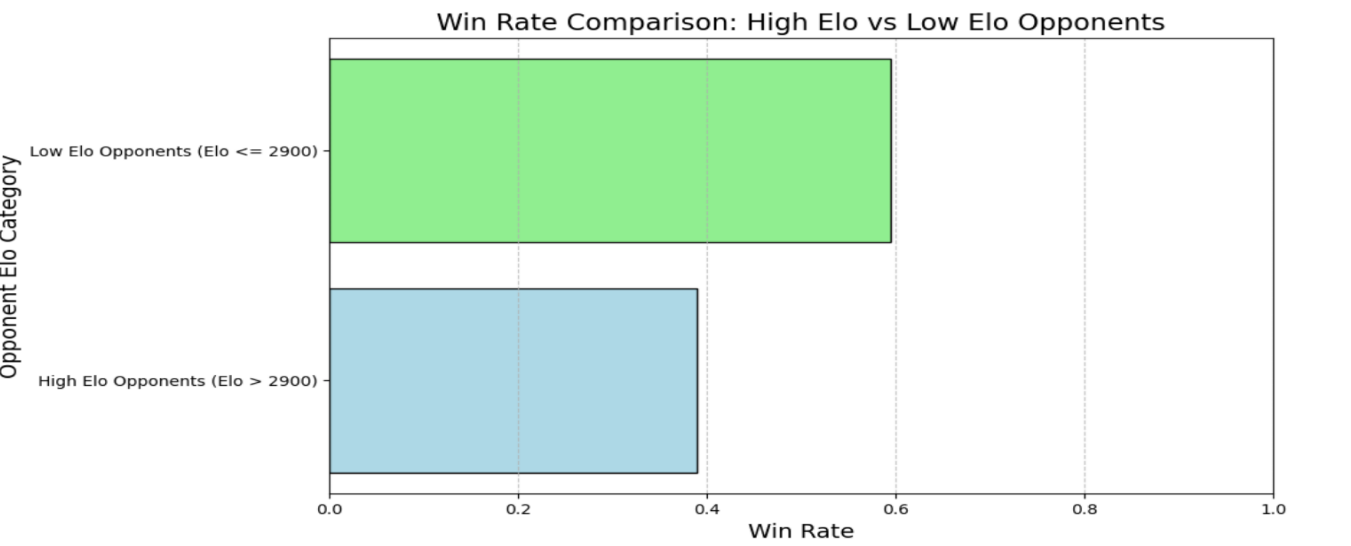


My best performance is with the Reti Opening: King's Indian Attack, where I achieved a higher win rate (%55) compared to other openings. This opening also has a significant number of games, reflecting both its effectiveness and frequent use. Although I played the “Bird’s Opening” most, I have low win rate compared to others.

c. Hypothesis Testing Part

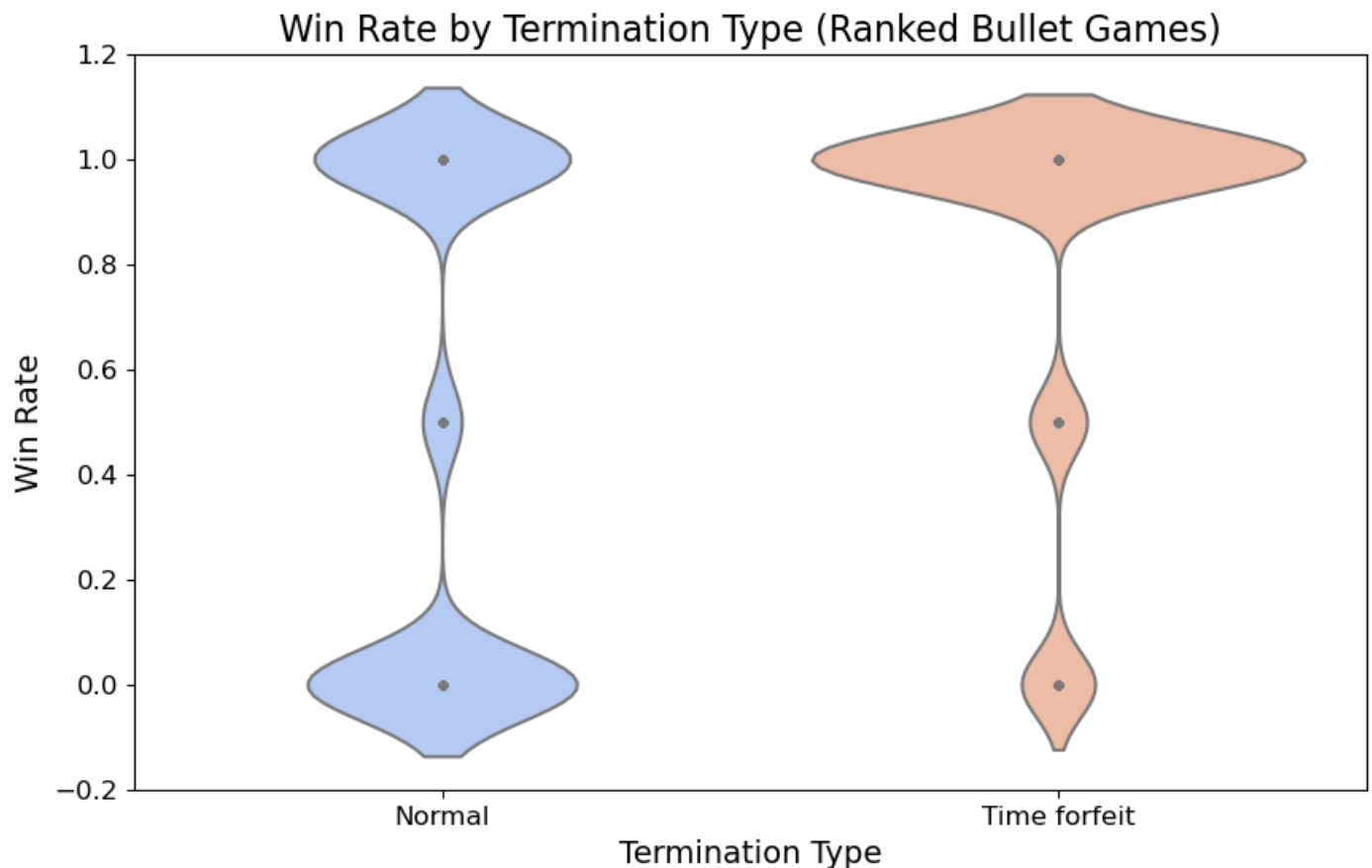
1. Null Hypothesis: I perform against the opponents whose rating is above 2900, just as good as I perform against below 2900 opponents in “Rated Bullet Game”

T-statistic: -23.388592384002916
P-value: 1.7631420696209118e-114
Reject the null hypothesis: There is a significant difference in win rates.



Result: Our hypothesis is rejected. Alternative hypothesis is accepted in this situation which is: There is a significant difference in win rates. Therefore, we can't say I don't have decrease in my performance when I am facing high elo opponents.

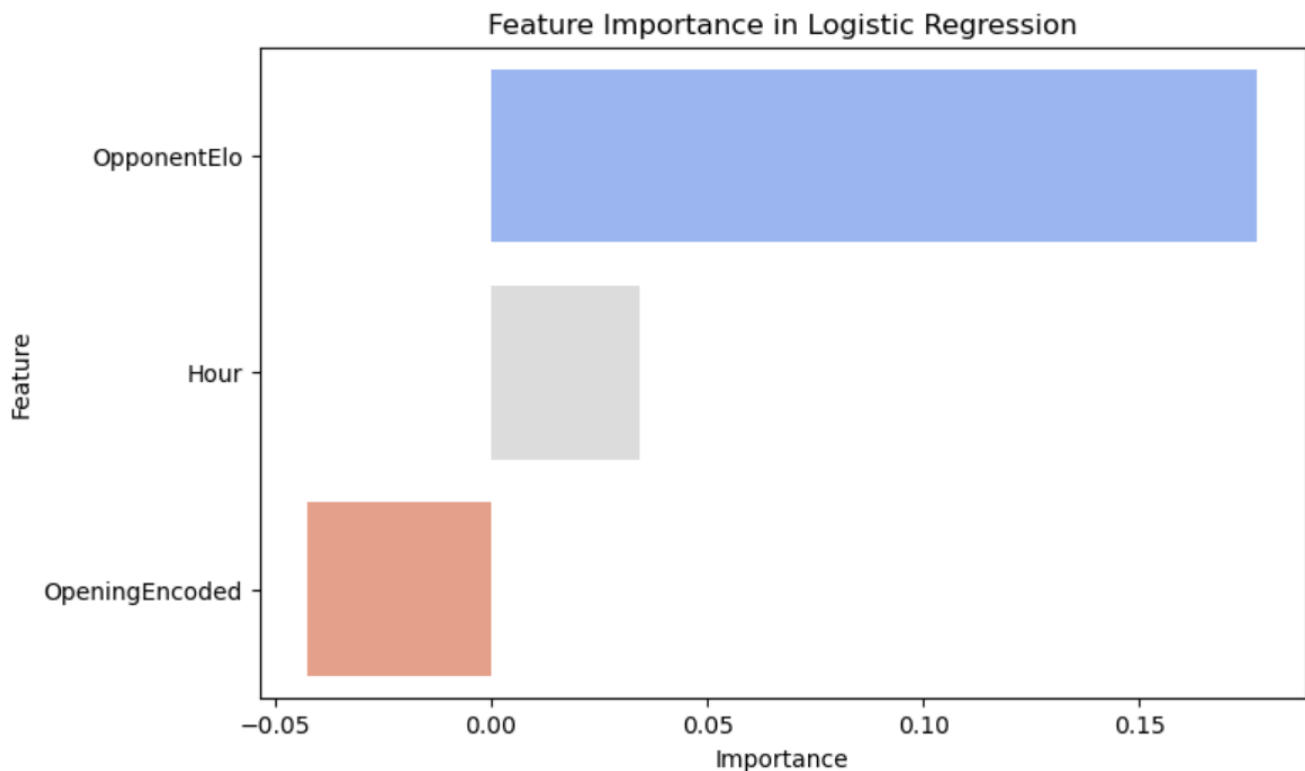
2. Null Hypothesis: The reason I have high elo in bullet game is not related to finishing my opponent's time tactic.



This plot shows how my win rates vary based on how my ranked bullet games ended—either normally or due to time forfeit. When games end normally, my win rates have a wider range, indicating that my performance can fluctuate more in these situations. On the other hand, games that end in time forfeits show a more consistent pattern, with less variation in my win rates. Overall, my win rates are high in both cases, but it seems I tend to perform slightly more consistently when games are decided by time forfeit.

Result: Hypothesis is rejected. There is a significant difference between the games I won by checkmating and the games I won by time forfeit. That mean

d. Machine Learning Model



This bar chart highlights the feature importance from the logistic regression model predicting my chess game outcomes.

- **OpponentElo:** The most influential factor, showing that opponent strength strongly impacts results.
- **Hour:** Moderately important, suggesting that the time of play affects my performance, possibly due to focus or fatigue.
- **OpeningEncoded:** Less significant, indicating that while openings matter, they have a smaller effect compared to Elo and playtime.

Overall, understanding opponent strength and timing games strategically are key to improving performance. Openings play a lesser role in my winrate.

4. Limitations and Future Work

Throughout this project, I focused on analyzing my ranked bullet chess games to uncover patterns and insights that could help improve my performance. While the project provided valuable findings, such as identifying openings where I perform best and understanding my win rates against opponents with varying Elo ratings, there are still limitations and areas for improvement. One major limitation was the reliance on available game data, which lacked deeper context, such as specific mistakes or critical moments in each game. Additionally, the project primarily focused on statistical analysis and simple machine learning models, which, while effective, do not capture the full complexity of chess strategies.

Looking forward, I plan to expand the project by incorporating advanced techniques such as game-specific data extraction, move-by-move analysis, and reinforcement learning models to simulate and improve decision-making in high-pressure scenarios. I also aim to include a time-based analysis to understand how fatigue or peak focus hours influence my performance.

Another potential direction is to benchmark my games against similar players to identify unique strengths or areas for improvement. By addressing these limitations and integrating new methodologies, I aim to turn this project into a comprehensive tool for enhancing my chess skills.

