

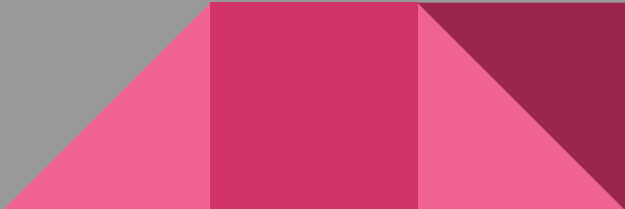


# Checkmate Your Stats

Improving your chess skills with Data

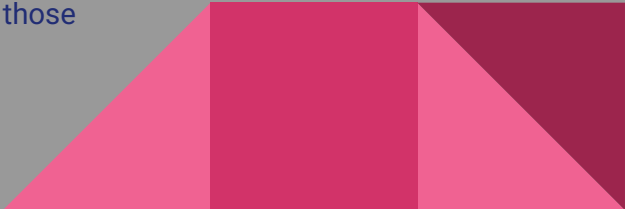
Alden Kallabat

# Contents

1. Hypothesis & Questions
  2. Approach and Analysis
  3. Challenges
  4. Exploratory Insights
  5. Q1
  6. Q2
  7. Q3 - Expected vs Reality
  8. Rating Analysis - Bias
  9. Conclusion
- 

# Hypothesis & Questions

**Hypothesis:** By analyzing trends in opening choices, time controls, and opponent rating ranges, I can discover actionable patterns that correlate with higher ratings and use them to improve my current chess rating of 1320.

1. **What are the most successful openings to play as white, and as black?**
    - a. This will help me select an opening, for both White and Black pieces, that maximizes my chances of winning.
  2. **Are there any specific time increment categories (game modes) that have an overall higher rating per player than other categories?**
    - a. This would allow me to focus on categories (game modes) that generally produce higher rated players.
  3. **Which rating difference results in the highest total rating gain over 100 games, according to the Elo system?**
    - a. Lichess.org offers filters that let me choose the rating range of my opponents, allowing me to strategically decide whether it's more beneficial to play against significantly higher-rated players or those closer to, or below, my skill level.
- 

# Approach & Analysis

## Part 1: Data Preparation in Excel

- Sourced a dataset of over 20,000 Lichess.org matches from Mysar Ahmad Bhat on Kaggle.com. Used Excel to inspect and clean the data. This included standardizing column headers by removing special characters, verifying that rows and columns were properly aligned, and ensuring consistent formatting. Using Excel's Power Query, I removed leading and trailing spaces, applied proper capitalization, and validated that each column adhered to its appropriate data type.

## Part 2: Data Exploration in SQL

- Used SQL to query the data into a usable format, and ran calculations for analysis.

## Part 3: Hypothesis Testing & Conclusion

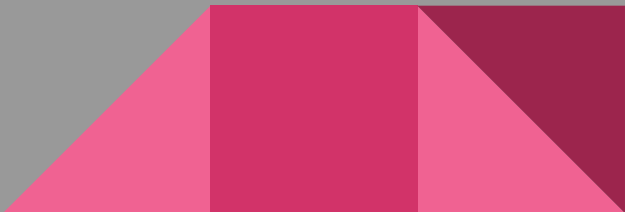
- Analyzed performance trends, win rates, and rating change to uncover patterns.

## Part 4: Data Visualization

- Built an interactive Power BI dashboard to effectively communicate data trends and patterns.
- 

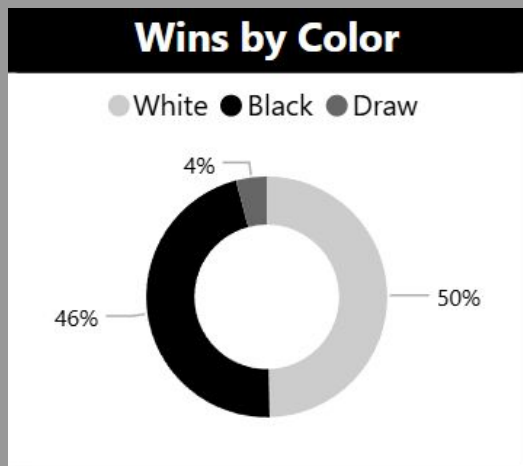
# Challenges

1. An error occurred involving the `white_rating` and `black_rating` columns. After investigation, the issue was traced to incorrect data types. Converting both columns to the proper types resolved the error and allowed the queries to execute successfully.
2. The analysis results deviated from expectations. Further investigation revealed a bias specific to online chess behavior.

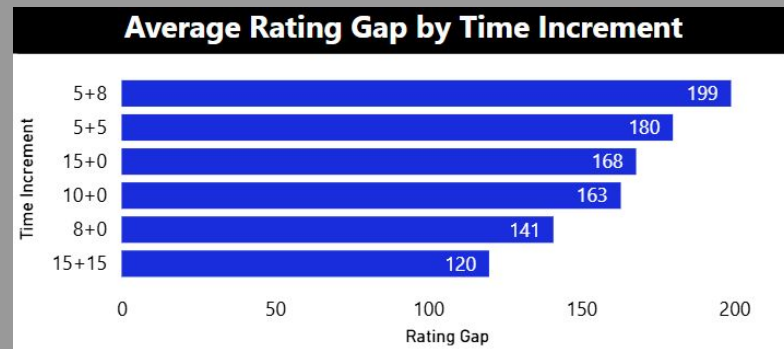


# Exploratory Insights

What color chess pieces typically win more?



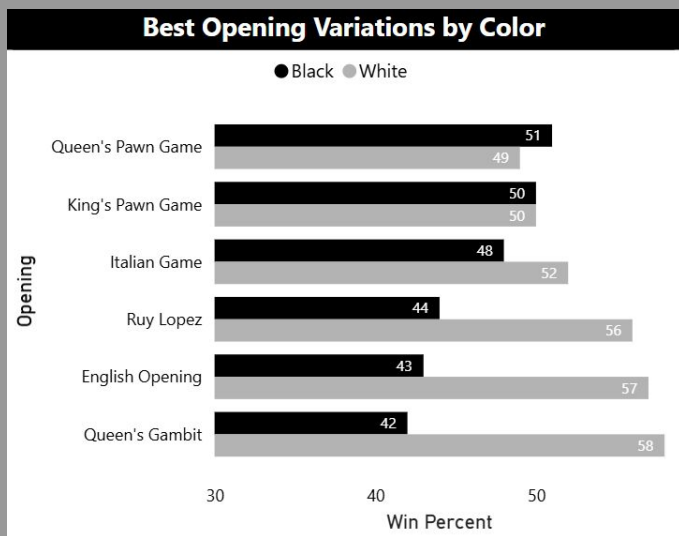
What is the average skill (rating) difference between each player?



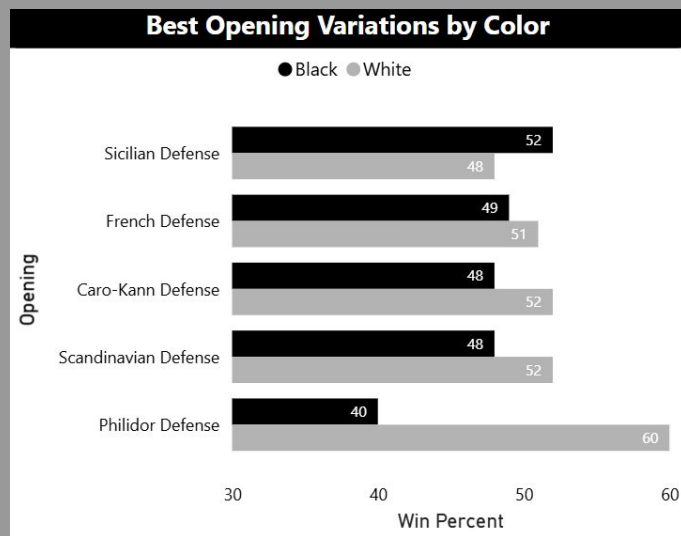
# Q1

## What are the most successful openings?

### Openings for White



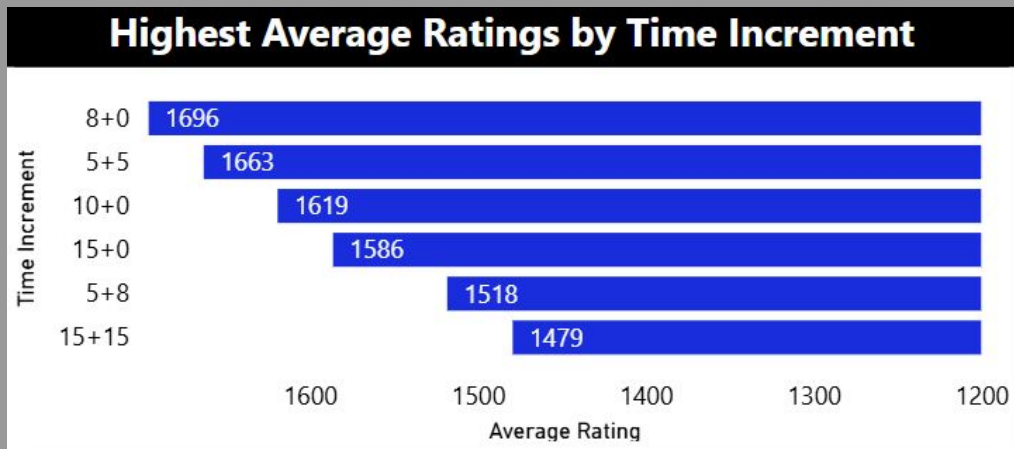
### Openings for Black



Using the Queen's Gambit, English Opening, and Ruy Lopez as White can boost my win rate by nearly 8%, while playing the Sicilian Defense, French Defense, and Caro-Kann Defense as Black can increase my chances of winning by almost 6%.

# Q2

**What time increment categories (game modes) have an overall higher rating per player?**

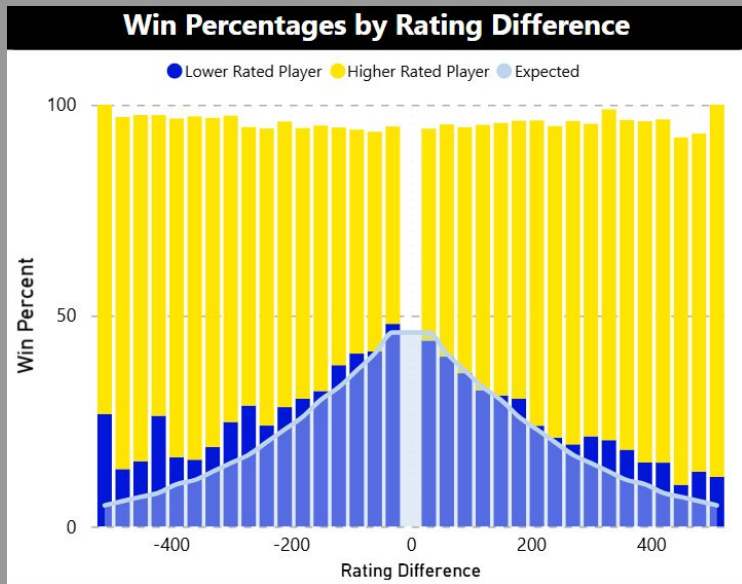


By prioritizing time controls like 8+0, 5+5, and 10+0 I can maintain a rating that is 15% higher than the lowest-performing time control, assuming consistent, average play.



# Q3

## What are the chances of winning against opponents ranging from 500 points lower to 500 points higher in 30-point intervals?



Higher-rated players (yellow) win more often as the rating gap increases. Lower-rated players (blue) win less. The curved line shows the expected win rate based on Elo, aligning closely with actual results.

### Expected vs Reality

The “Expected” line shows predicted win rates based on rating difference. Actual results often differ, as lower-rated players win more than expected.

# Rating Analysis - Bias

Rating Difference	Rating Change
-510	-433
-480	-182
-450	-194
-420	-386
-390	-169
-360	-120
-330	-147
-300	-219
-270	-277
-240	-133
-210	-146
-180	-138
-150	-98
-120	-151
-90	-132
-60	-65
-30	-98
30	24
60	23
90	33
120	25
150	70
180	120
210	55
240	71
270	78
300	171
330	160
360	176
390	151
420	174
450	135
480	210
510	134

\*The Rating Change column shows the projected rating gain or loss after playing 100 matches at each rating difference.

## Q3.1. Which rating difference yields the greatest rating gain over 100 games?

### Patterns

- Rating gains are always positive against stronger players
- Rating gains are always negative against weaker players
- As the rating gap increases, the amount of gains or losses are amplified
- The lower-rated player always wins more than expected

These patterns can be attributed to misclassification bias in player ratings. It's common for strong players to create lower-rated alternate accounts and filter to play against strong players to avoid risking their high ratings on their main accounts. Additionally, chess cheaters who use AI are often banned and create new accounts. When they return, they frequently compete against higher-rated players in an effort to rapidly gain rating.

# Conclusion

- The White pieces win 4% more than the Black pieces. Prioritizing strong openings for Black is ideal in order to mitigate rating losses considering the immediate disadvantage.
- By playing the strongest openings, the overall win rate can be increased by 14%.
- Focusing on time controls like 8+0, 5+5, and 10+0 can lead to ratings up to 15% higher than the lower performing modes, and 28% higher than my current rating, assuming average performance.
- Facing opponents around 180 rating points higher results in a gain of 120 rating points over 100 matches, which is a 9 percent increase from my current rating.

## FINAL CONCLUSION

**By analyzing and implementing trends in opening choices, time controls, and opponent rating ranges, I can improve my chess rating by up to 51%, bringing my initial rating of 1320 up to 1993 after completing 100 matches.**

