Analysis of Chess Game Dynamics

Investigating the Factors Behind Winning

Anirudh Unni

Game and Dataset overview

•**The Goal**: The objective is to "checkmate" the opponent's King, trapping it so it cannot escape.

•**The Players**: There are two sides, White and Black. **White always moves first**, which gives them a slight initial advantage that we will analyze.

•**Player Skill (ELO Rating)**: Every player has a skill score called an **ELO rating**. A higher number means a stronger player. The difference in ELO between players is a good way to predict who might win.

•**Game Phases**: A game has three phases. We will focus on the **Opening** which are the first 5-15 moves that often follows well-known strategies.

•**Dataset**: https://www.kaggle.com/datasets/datasnaek/chess

id	rated	created_at	last_move_at	turns	victory_status	winner	increment_code	white_id	te_rating	black_id	ack_rating	moves	ning_eco	opening_name
ZJHLljE	FALSE	1.50421E+12	1.50421E+12	13	outoftime	white	15+2	bourgris	1500	a-00	1191	c3 Ba5 Bf4	D10	Exchange Variation
NXvwaE	TRUE	1.50413E+12	1.50413E+12	16	resign	black	5+10	a-00	1322	kinnerua	1261	e5 c4 Bb4+	B00	Kennedy Variation
lCvQHh	TRUE	1.50413E+12	1.50413E+12	61	mate	white	5+10	ischia	1496	a-00	1500	Ka4 Bd1#	C20	Leonardis Variation
VKvrqYL	TRUE	1.50411E+12	1.50411E+12	61	mate	white	20+0	urashov	1439	nov2009	1454	Bg7 Qxg7#	D02	Zukertort Variation
Xo1AUZ	TRUE	1.50403E+12	1.50403E+12	95	mate	white	30+3	k221107	1523	nov2009	1469	Kh6 Rh7#	C41	Philidor Defense
oDV9wj	FALSE	1.50424E+12	1.50424E+12	5	draw	draw	10+0	elynn17	1250	lin14532	1002	Nf3 Qa5 a3	B27	Mongoose Variation
wU9rasv	TRUE	1.50423E+12	1.50423E+12	33	resign	white	10+0	capa_jr	1520	s_chess	1423	+ Ke7 Bd2	D00	Pietrowsky Defense
N0N3VK	FALSE	1.50368E+12	1.50368E+12	9	resign	black	15+30	s_chess	1413	soultego	2108	c3 Ng6 b4	B00	ssspringer Variation
F3DJHO	TRUE	1.50351E+12	1.50351E+12	66	resign	black	15+0	habfanri	1439	s_chess	1392	Kh2 Nxf1+	C50	illing-Kostic Gambit
oMwnLg	TRUE	1.50344E+12	1.50344E+12	119	mate	white	10+0	s_chess	1381	mirco25	1209	Kh7 Qg7#	B01	es-Kotroc Variation
KLWPsz	FALSE	1.50335E+12	1.50335E+12	39	mate	white	20+60	s_chess	1381	anaissac	1272	Bd5 Rxe8#	A00	an't Kruijs Opening
f5fKWzI	FALSE	1.50335E+12	1.50335E+12	38	resign	black	20+60	s_chess	1381	ham777	1867	Ned2 Ba6+	C02	n | Paulsen Attack
Rti5mKv	FALSE	1.50335E+12	1.50335E+12	60	resign	black	5+40	s_chess	1381	2342005	1936	Ra2 Nxe4	C00	se: Knight Variation
2fEjSei6	FALSE	1.50334E+12	1.50334E+12	31	resign	black	8+0	s_chess	1381	alkhan	1607	c3 Nc5 Be3	C00	French Defense #2
7i6dOaJ	FALSE	1.50409E+12	1.5041E+12	31	mate	white	15+15	egenius	1094	a_akshat	1141	Kh8 Qg7#	C50	ne: Italian Variation
anvMR5	FALSE	1.50409E+12	1.50409E+12	43	resign	black	15+15	a_akshat	1141	egenius	1094	Qd1+ Qf1	C57	nse | Knight Attack
kWkmm	FALSE	1.50409E+12	1.50409E+12	52	resign	black	15+15	egenius	1094	a_akshat	1141	Nxd8 Rxd8	C50	ne: Italian Variation
waK0IsE	FALSE	1.50401E+12	1.50401E+12	66	mate	black	15+16	a_akshat	1141	egenius	1094	Kh1 Qxh2#	C50	ne: Italian Variation
SDozT3	FALSE	1.50401E+12	1.50401E+12	101	resign	black	15+15	egenius	1094	m_ment	1300	c5 Qxd7 c6	C41	Philidor Defense #3
31mXlvc	FALSE	1.50376E+12	1.50376E+12	25	resign	white	11+0	g-ios	1500	egenius	1094	Nc4 Bxa8	D00	ueen's Pawn Game
klnWWp	FALSE	1.50295E+12	1.50295E+12	14	resign	black	15+15	egenius	1094	lex_v1	1676	g4 h3 Nxf2	C55	Fried Liver Defense
CZwY1f	FALSE	1.50295E+12	1.50295E+12	3	resign	white	30+60	egenius	1094	reaction	1068	d4 e6 Nc3	A40	Horwitz Defense
cgBygpI	FALSE	1.50295E+12	1.50295E+12	17	resign	white	15+5	rm28rus	1500	egenius	1094	fxe6 Bxh6	B20	se: Bowdler Attack

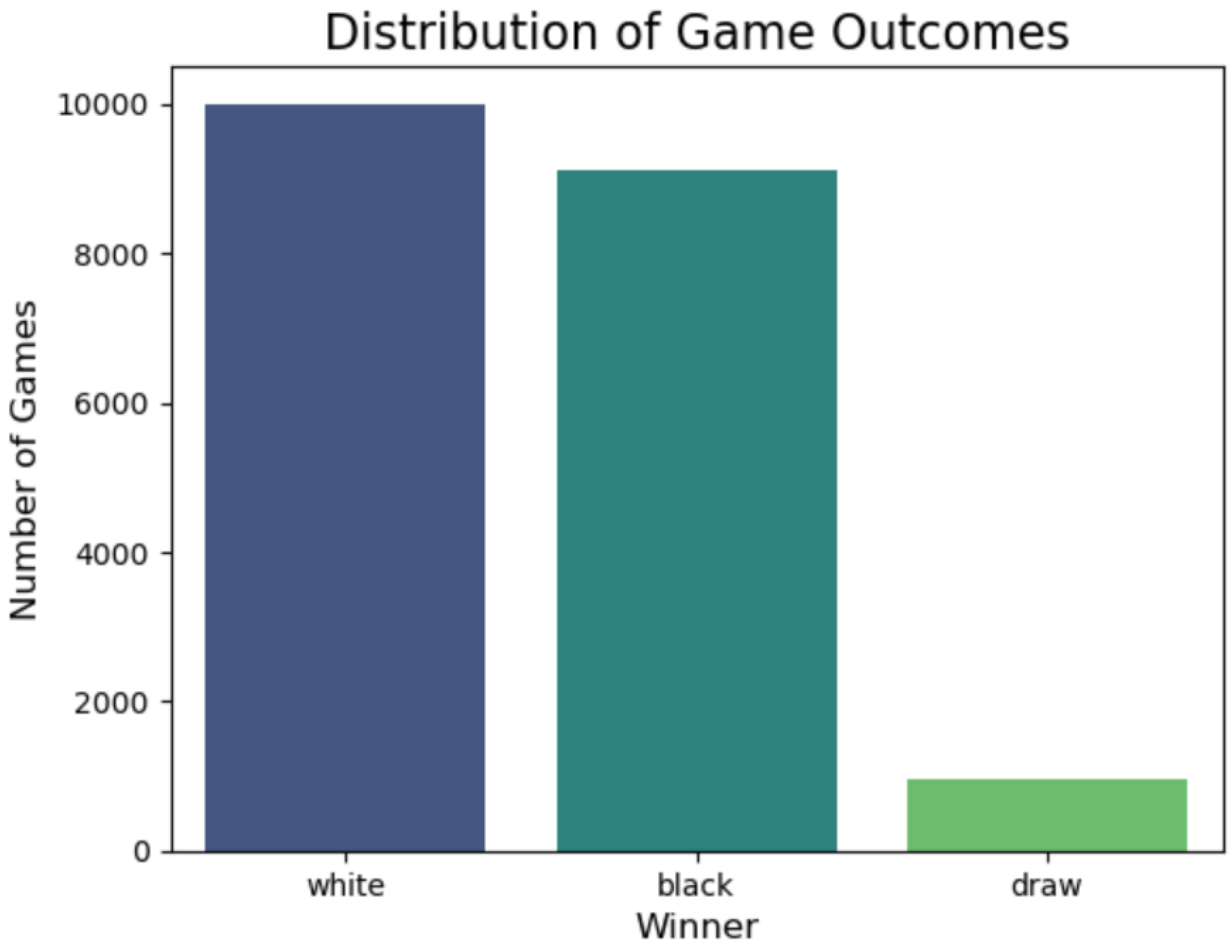
Creating new features in dataset

```
#Create a new feature that calculates difference in rating between white and black player
chessdb['ratingdiff'] = chessdb['white_rating'] - chessdb['black_rating']
```

```
def get_rating_tier(row):
    if row['white_rating'] > 2000 and row['black_rating'] > 2000:
        return 'Both High (>2000)'
    elif row['white_rating'] < 1500 and row['black_rating'] < 1500:
        return 'Both Low (<1500)'
    else:
        return 'Mixed / Mid-Rated'
```

ratingdiff	rating_tier
309	Mixed / Mid-Rated
61	Both Low (<1500)
-4	Mixed / Mid-Rated
-15	Both Low (<1500)
54	Mixed / Mid-Rated
248	Both Low (<1500)
97	Mixed / Mid-Rated
-695	Mixed / Mid-Rated
47	Both Low (<1500)
172	Both Low (<1500)
109	Both Low (<1500)
-486	Mixed / Mid-Rated
-555	Mixed / Mid-Rated
-226	Mixed / Mid-Rated
-47	Both Low (<1500)
47	Both Low (<1500)
-47	Both Low (<1500)
47	Both Low (<1500)
-206	Both Low (<1500)
406	Mixed / Mid-Rated
-582	Mixed / Mid-Rated
26	Both Low (<1500)
406	Mixed / Mid-Rated

Does White have a tangible advantage in chess?



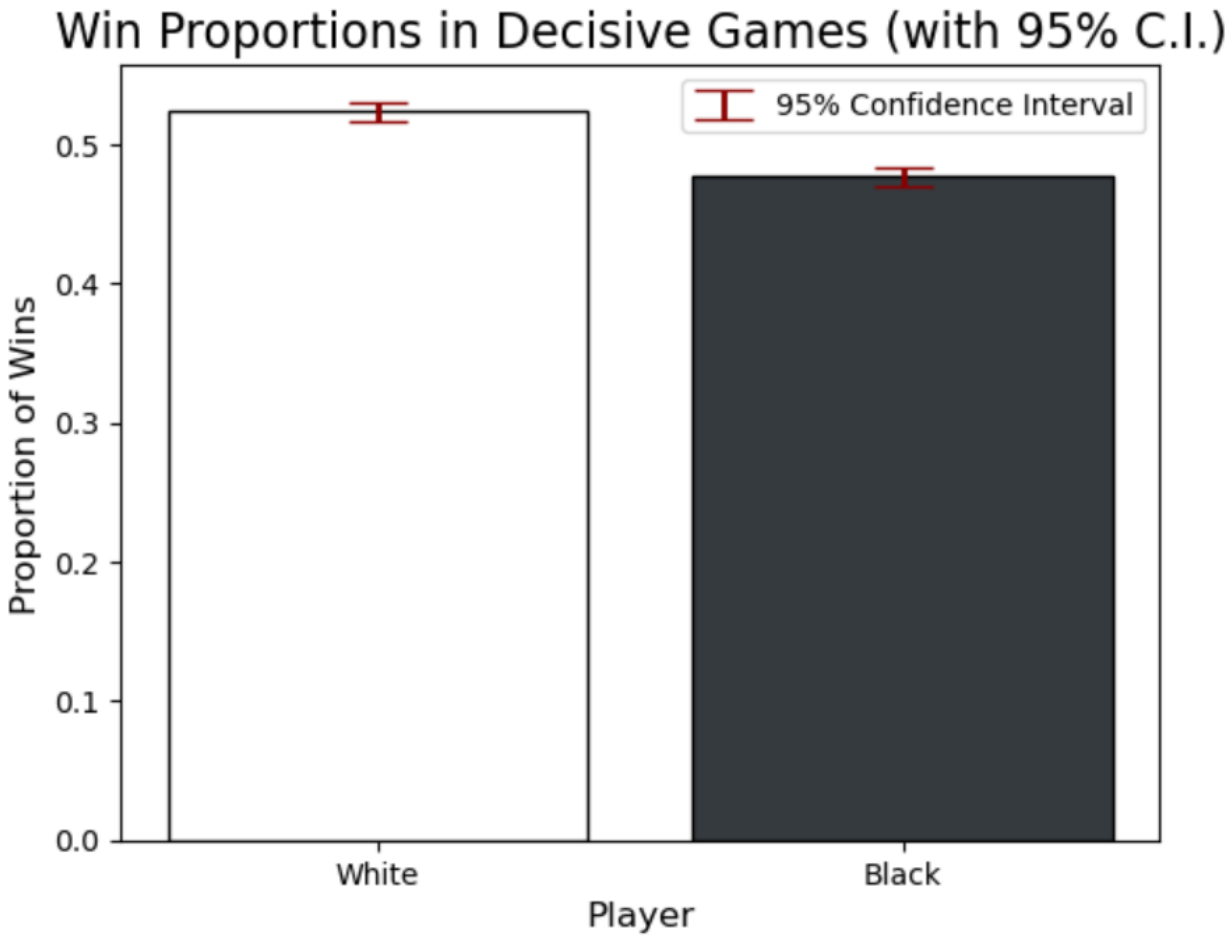

```
SELECT winner,COUNT(*) AS number_of_games FROM cleaned_games GROUP BY winner ORDER BY number_of_games DESC;
```

	winner	number_of_games
1	White	10001
2	Black	9107
3	Draw	950

- Across thousands of games, White consistently wins more often than Black.
- Is this result statistically significant?

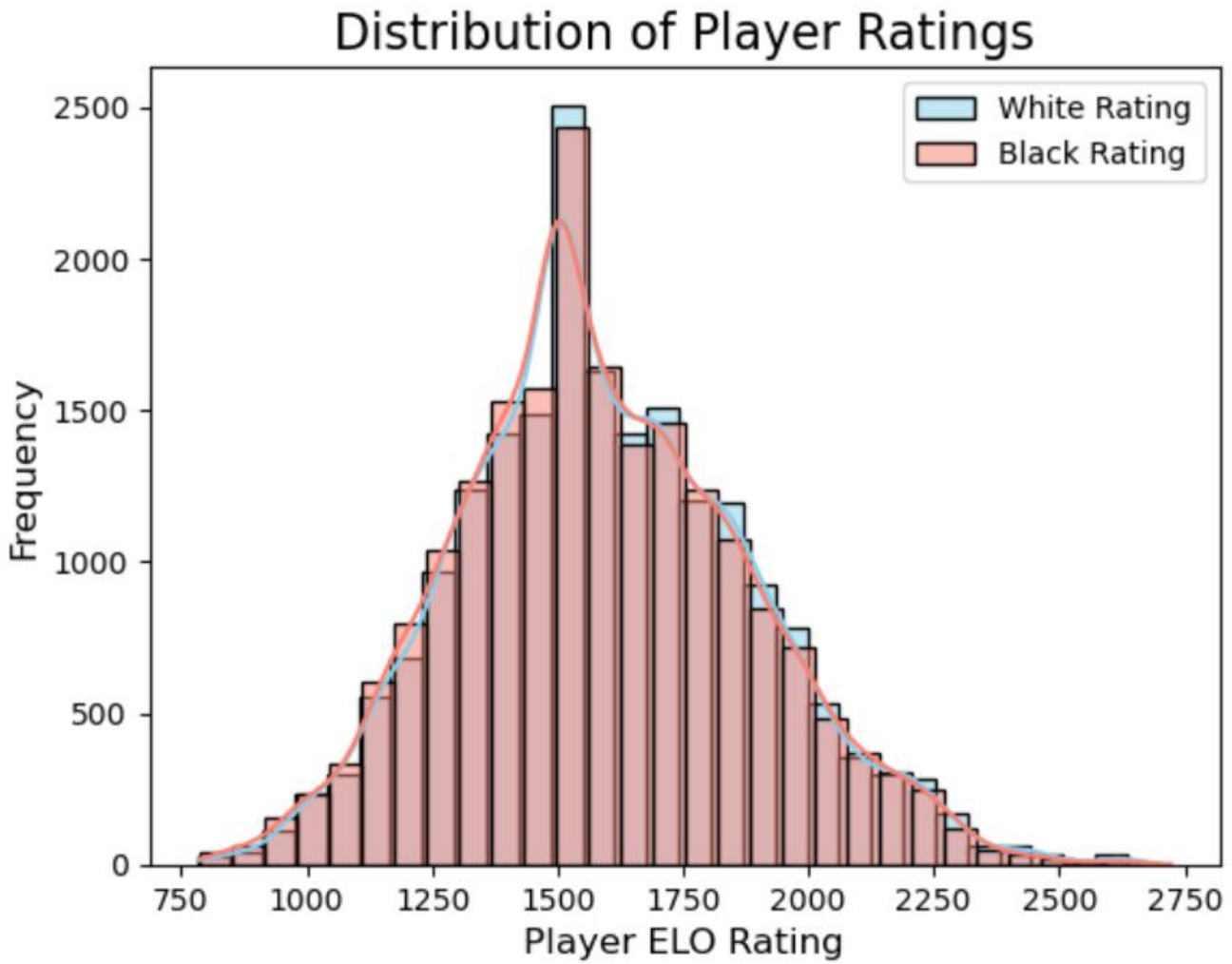
Does White win significantly more than 50% of the games that are not draws?

- By analyzing only the decisive games, we see White wins about 52% of the time.
- The 95% confidence intervals do not overlap, which tells us this is not due to random chance.

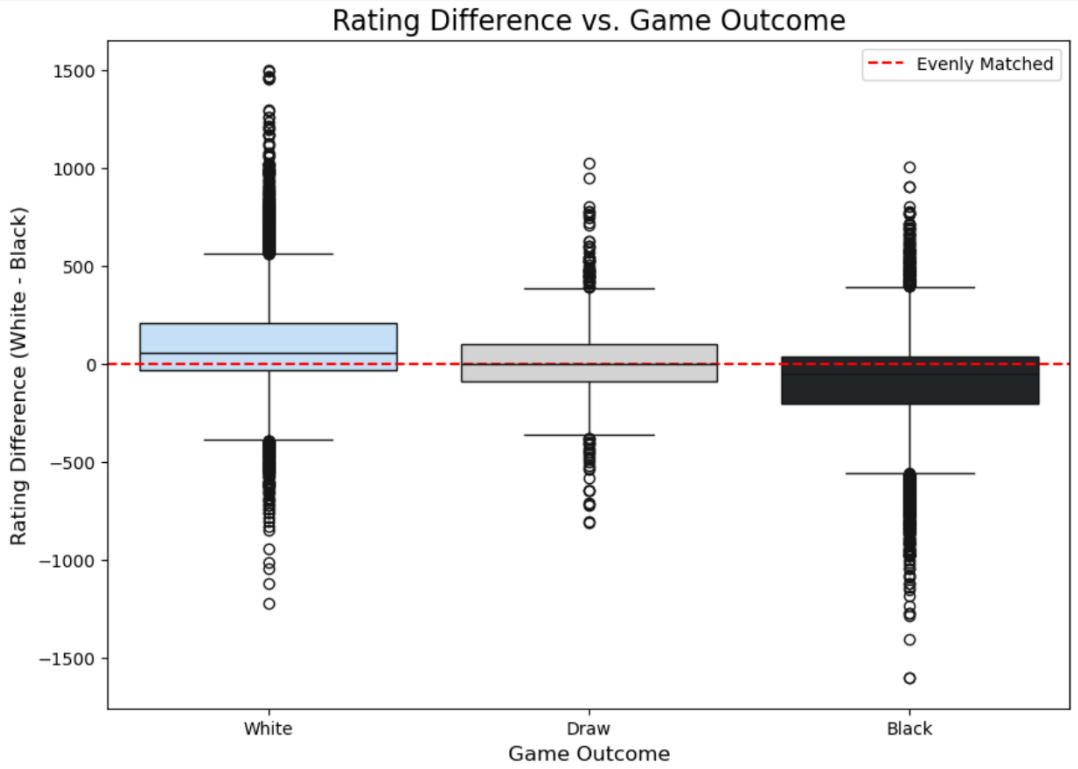


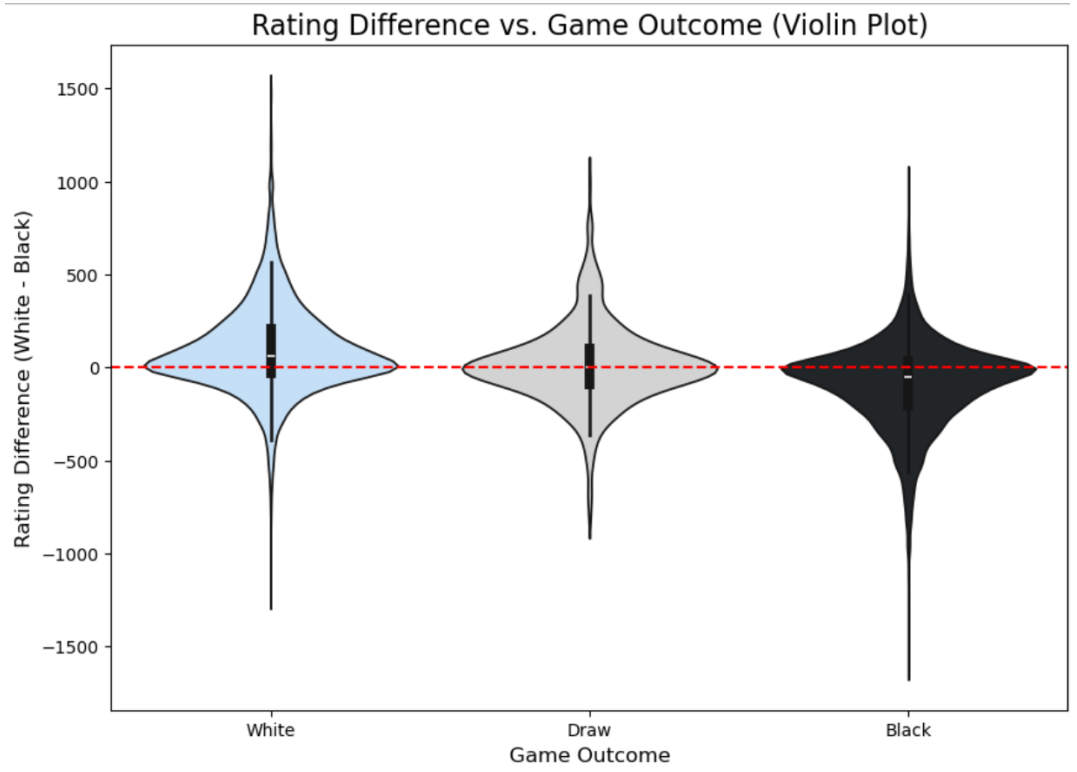
Does player rating influence the color assignment?

- The two histograms have nearly the same shape, center, and spread. The ELO ratings for players assigned White and Black are identically distributed.
- Fair and unbiased chess color assignment matching system where a player's rating does not influence the color assigned.
- Advantage is not because White players are stronger. The advantage must come from the well-known “first-move advantage”.



Do higher-rated players win more often? Is there a correlation between the difference in player strengths and the game's outcome?



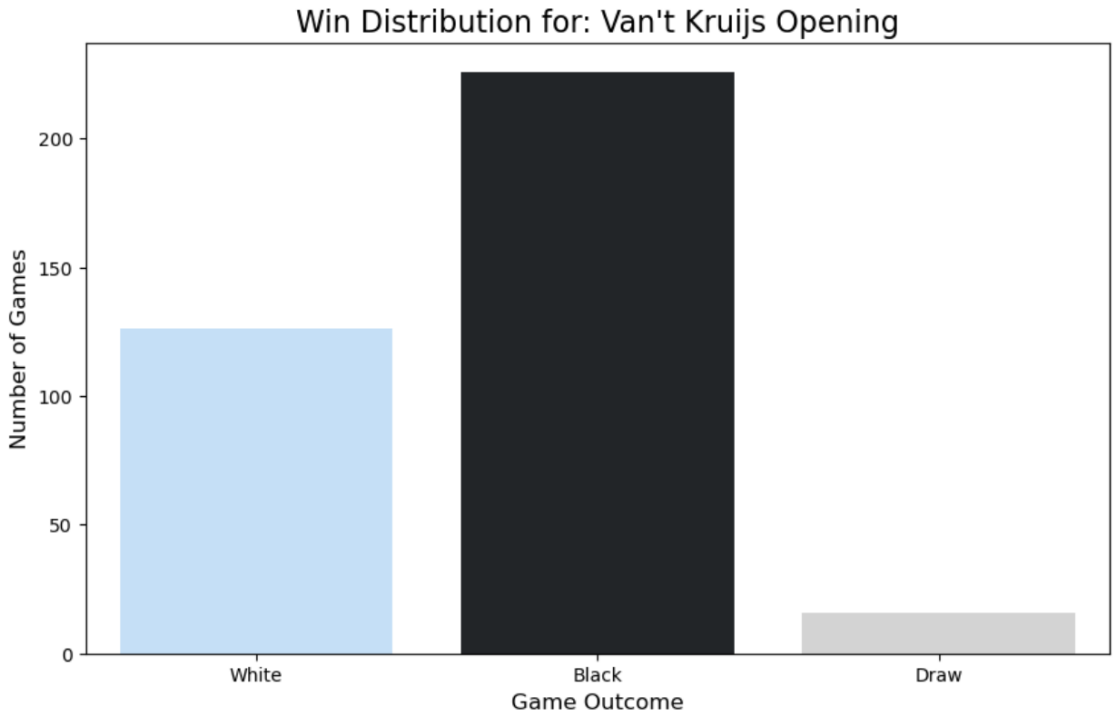


Player Skill is the Overwhelming Predictor of a Win. **Higher-rated player wins most of the time**.

'White' wins are almost all above the red line, and 'Black' wins are below it

Strategic Opening Choices Can Overcome the First-Move Advantage

- Trend reversal in win distribution with respect to color
- Van't Kruijs Opening (1.e3) is known to be a passive and unambitious opening for White. It allows Black to easily control the center of the board.
- Poor strategic opening choice can be more decisive than the first-move advantage.



	opening_name	frequency
1	Van't Kruijs Opening	368
2	Sicilian Defense	358
3	Sicilian Defense: Bowdler Attack	296
4	Scotch Game	271
5	French Defense: Knight Variation	271
6	Scandinavian Defense: Mieses-Kotroc …	259
7	Queen's Pawn Game: Mason Attack	232
8	Queen's Pawn Game: Chigorin Variation	229
9	Scandinavian Defense	223
10	Horwitz Defense	209

```
SELECT opening_name, COUNT(*) AS frequency FROM cleaned_games GROUP BY opening_name ORDER BY frequency DESC LIMIT 10;
```

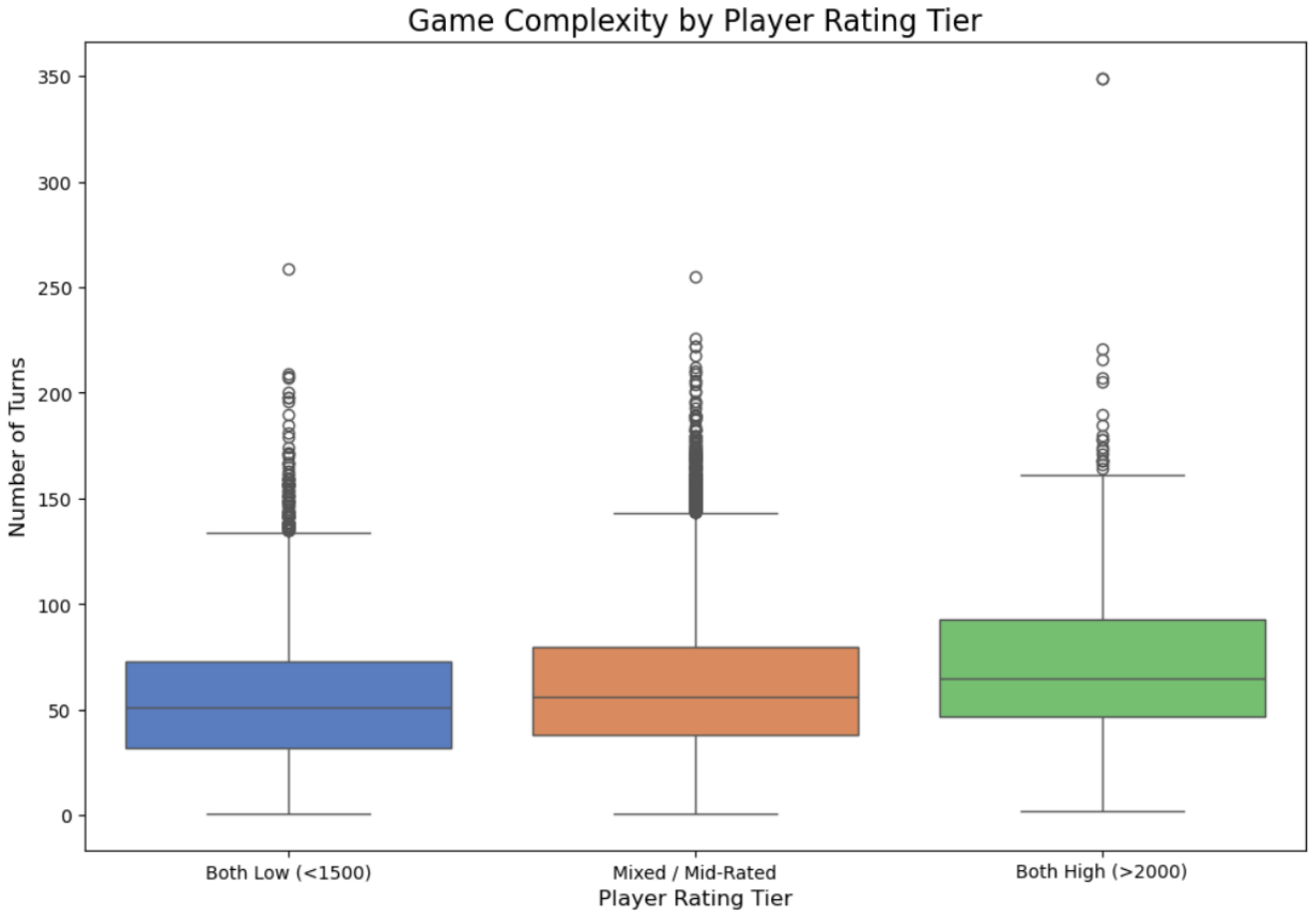
Do games between higher-rated players (e.g., both players > 2000) tend to last longer than games between lower-rated players?

Average turns by player rating tier:
Both High (>2000) 72.83
Both Low (<1500) 55.16
Mixed / Mid-Rated 61.70

A player's rating has an impact on game length and complexity.

Games between higher-rated players on average are approximately 30% longer and more complex than games between lower-rated players.

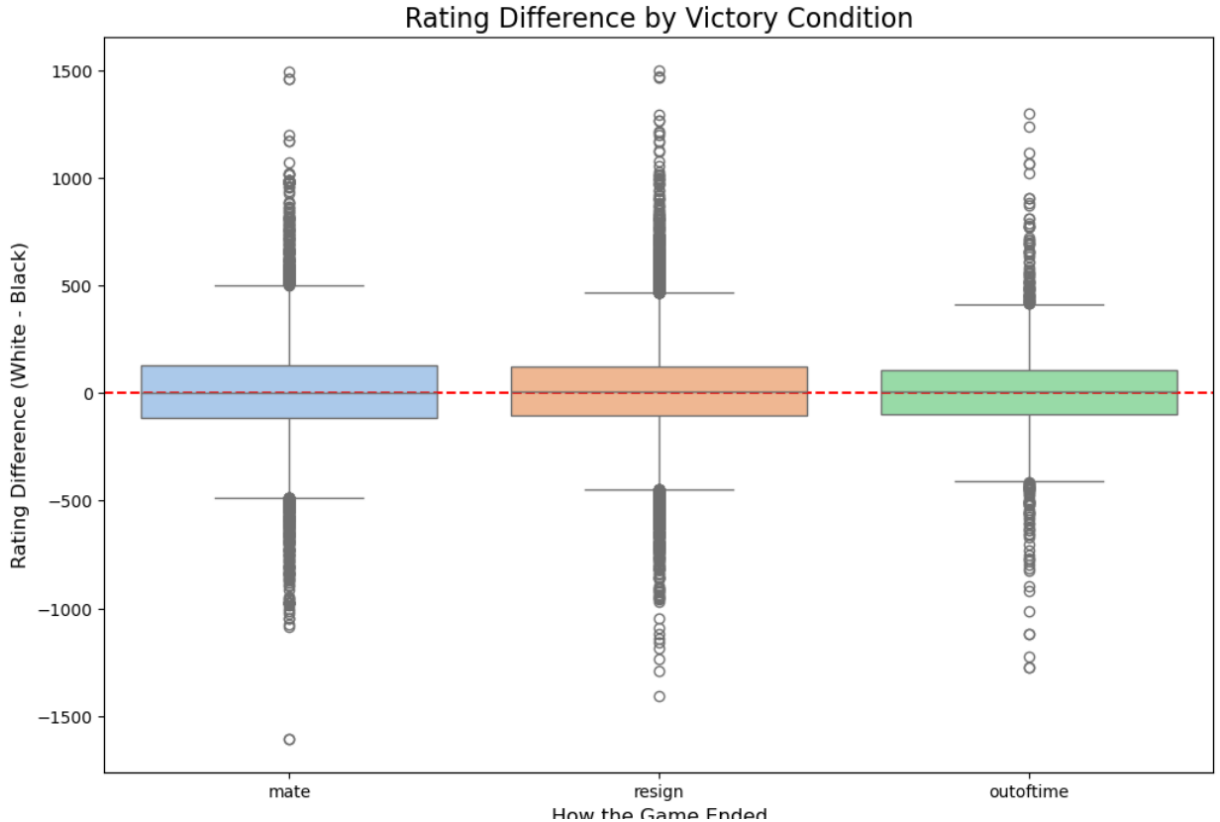
High-level chess involves fewer early blunders and more tenacious defense, leading to longer, more technical battles.



How does the method of victory (e.g., checkmate, resign, lost-on-time) correlate with the players' rating difference?

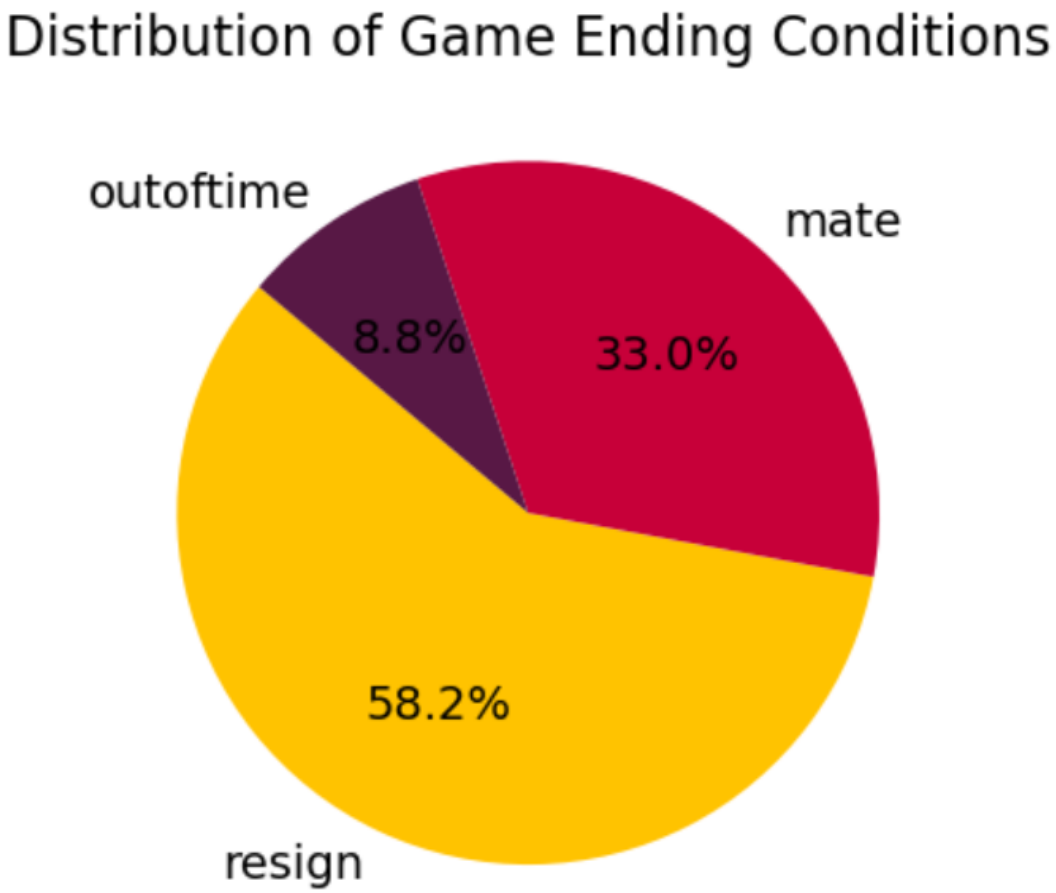
All box plots look similar.

The decision to ‘resign’ versus playing to ‘checkmate’ vs ‘outoftime’ is not strongly influenced by the initial skill gap between players.



What percentage of games end in resignation versus a checkmate or lost on time?

- The Majority of Decisive Games End by Resignation
- 1 in 3 games play out until checkmate
- Around 9% of the games end when one player loses on time



Summary

White Has a Slight Edge: Across thousands of games, the data confirms a small but statistically significant **first-move advantage** for White.

Skill is the Deciding Factor: This small advantage is minor compared to player skill. The **higher-rated player** is the overwhelming favorite to win, regardless of color.

Strategy Can Overcome the Edge: A player's **strategic choices**, like the opening, can be so critical that they can even reverse the inherent first-move advantage, as seen in the case of the "Van't Kruijs Opening.“

Complexity Increases with Skill: Games between higher-rated players are, on average, significantly **longer and more complex** than those between lower-rated players, shifting from quick tactical battles to long, strategic struggles.

THANK YOU