

Summary

Chess.com is an internet chess server and social networking website. Players can play standard chess or different chess variations. They can choose their preferred time class, design etc. By analyzing chess games we can get some important insights about players' preferences and style of game. These facts can be further used to better understand chess as a whole. Players or coaches can use information from this analysis to adjust their game preparation.

Prepare and Process phase

Used dataset

The data source used for this case study is *Chess.com Random Games*. This dataset is stored in Kaggle and contains 9,000+ Chess.com games (Randomly Sampled from Aug 2023).

Accessibility and privacy of data

The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

Information about this dataset

Author of this dataset used a Python script and cleaned data using Kaggle Notebook.

Data Organization and verification

One CSV document is available (clean_data.csv). There are 14 columns. I used Excel and Power Query to transform my data for my analysis. I deleted 5 original columns and added 5 custom columns.

New added columns:

- `average_rating`: to get the average rating of both players I summed `white_rating` and `black_rating`, then divided them by 2. Afterwards I RoundUp this new column.
- `white_first_move`: white first move was extracted from `opening_pgn` column using `Text.Trim()` and `Text.Middle()` functions.
- `white_result_simple`: transformed from `white_result` to only 3 possibilities: win, lose, draw
- `black_result_simple`: transformed from `black_result` to only 3 possibilities: win, lose, draw
- `result`: using `white_result_simple` and `if` functions to get one of 3 possible results: 1-0, 0-1, 1/2-1/2. `if [white_result_simple] = "win" then "1-0" else if [white_result_simple] = "draw" then "1/2-1/2" else "0-1"`

Data Credibility and Integrity

If we want to undergo a thorough analysis of chess games, we will need millions of games from other months and years, different chess website etc. to make sure there is no bias.

Nevertheless 9,000+ games should indicate some patterns.

Key aspects

To verify some aspects of my dataset I added my data to the Data Model and used Pivot table in Excel.

Count function in Value Field Settings showed the total number of chess games: 9784

By sorting number of games for each opening I found that there are two irregular opening names: *Uncommon Opening* and *Uncommon King's Pawn Opening*. This means that players played irregular openings which are not noted. For simpler analysis I combined these two openings into one: *Uncommon Opening*

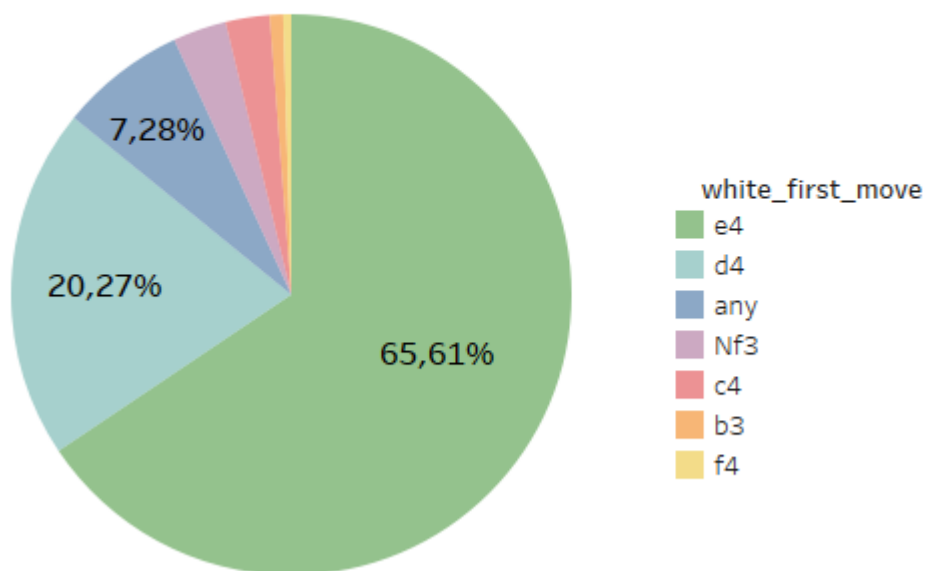
Distinct count function showed number of openings played by players: 113

One type of opening e.g. *Ruy Lopez* has many variations: *Ruy Lopez, Classical*; *Ruy Lopez, Exchange* ... To go into deeper analysis of each opening I would need bigger dataset and the whole analysis would take exponentially more time. For example the opening *Caro-Kann, Bronstein-Larsen Variation* has only one game from which I can't conclude all that much. In the analysis phase of openings, I will use only the main openings with lots of games.

Analyze and Share Phase

For analysis and visualization I used Tableau Desktop.

White First Move

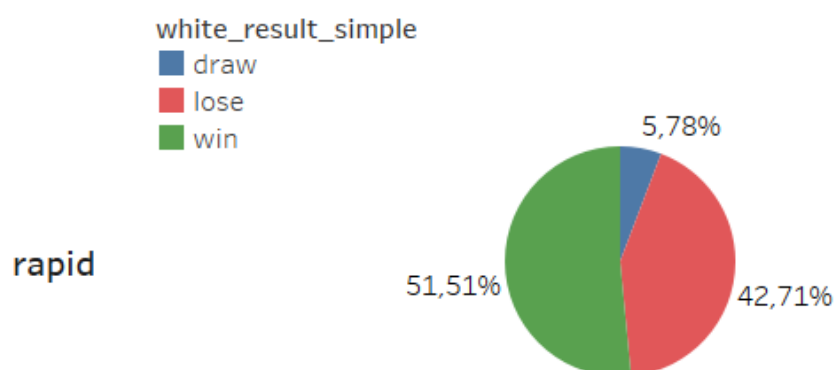
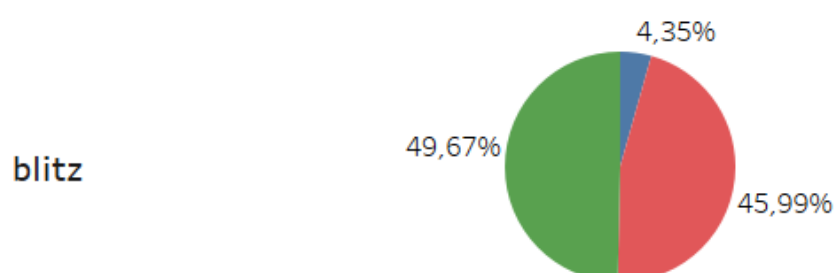
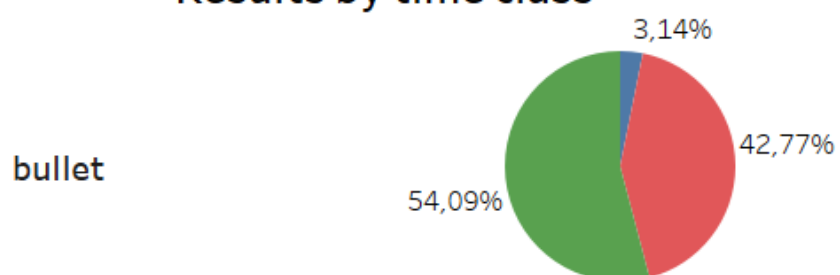


Move e4 is by far the most popular first move for white. Move d4 is second in popularity and third place goes for "any" (any=any other move not written in the list).



This dataset shows that move c4 has the highest chance of winning a game for white – 56.28%. Moves e4 and d4 have the highest percentage of games concluding in a draw. This shows that players may be used to more popular openings coming from e4 and d4, so the game is even. Move c4 on the other hand is played relatively less and the games are usually more decisive for one side, thus the draw percentage is lower. However more deep analysis and bigger dataset is needed to validate this point.

Results by time class



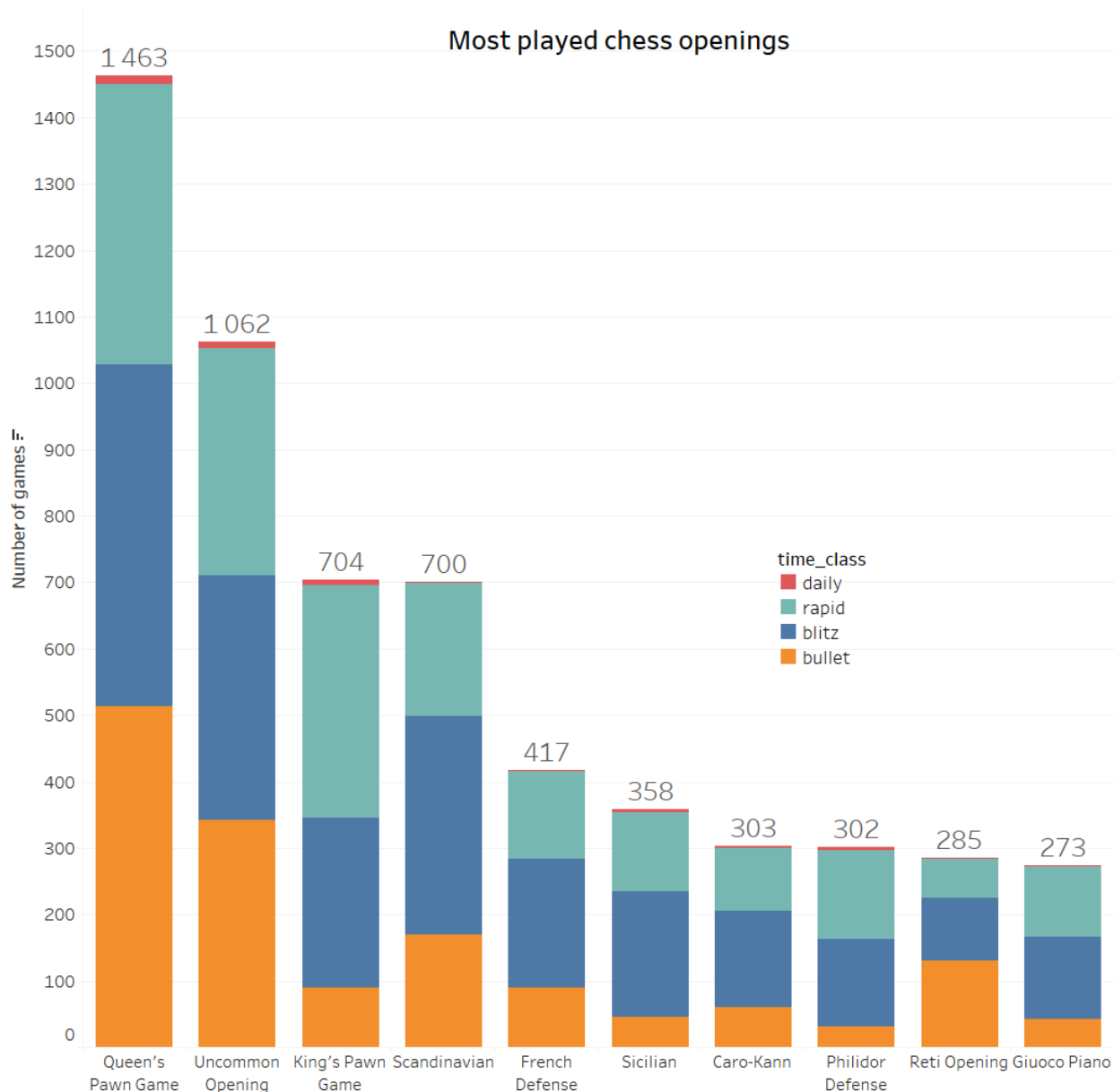
white_result_simple

draw
lose
win

This dataset contains 4 time classes:

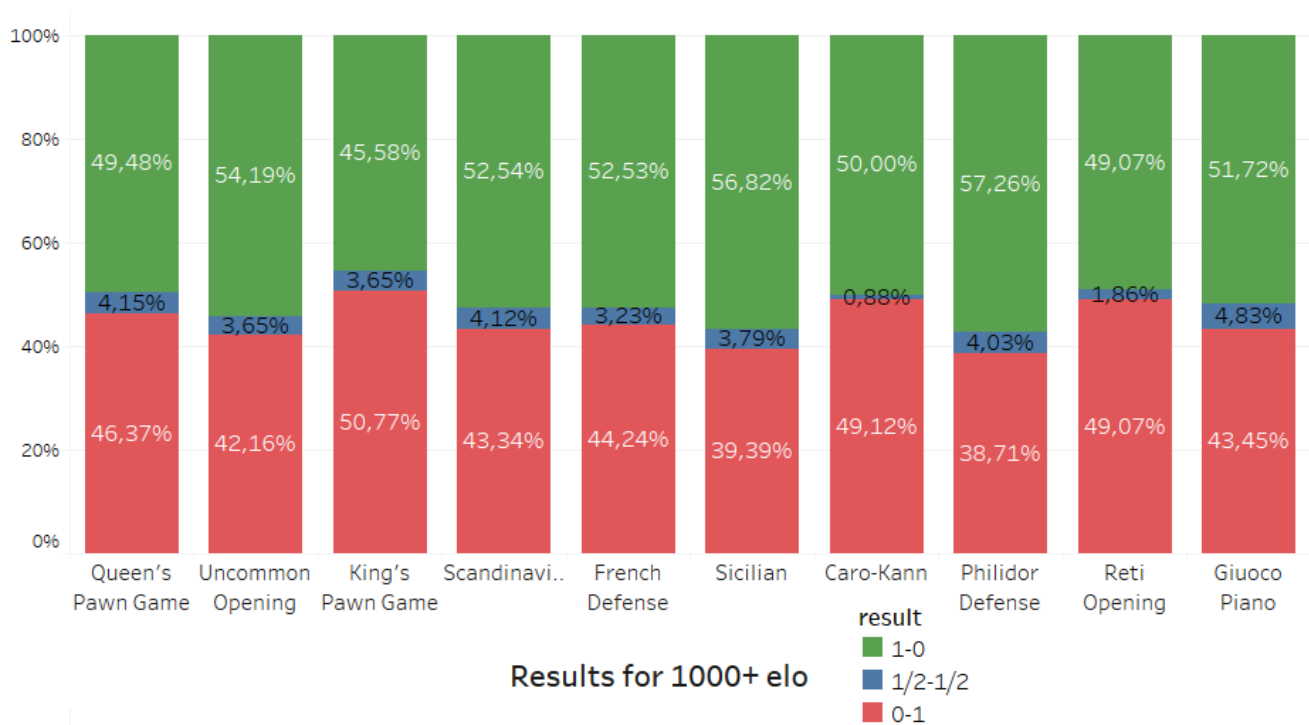
- Bullet: for games under 3 minutes
- Blitz: for games of 3 to 14 minutes
- Rapid: for games over 15 minutes
- Daily: for games lasting days

I excluded daily time class due to having only a few chess games in dataset. This visualization shows a clear pattern. With more minutes on the clock, games tend to be less decisive and end in draws. Nonetheless white always has higher win rate than black in all time classes.

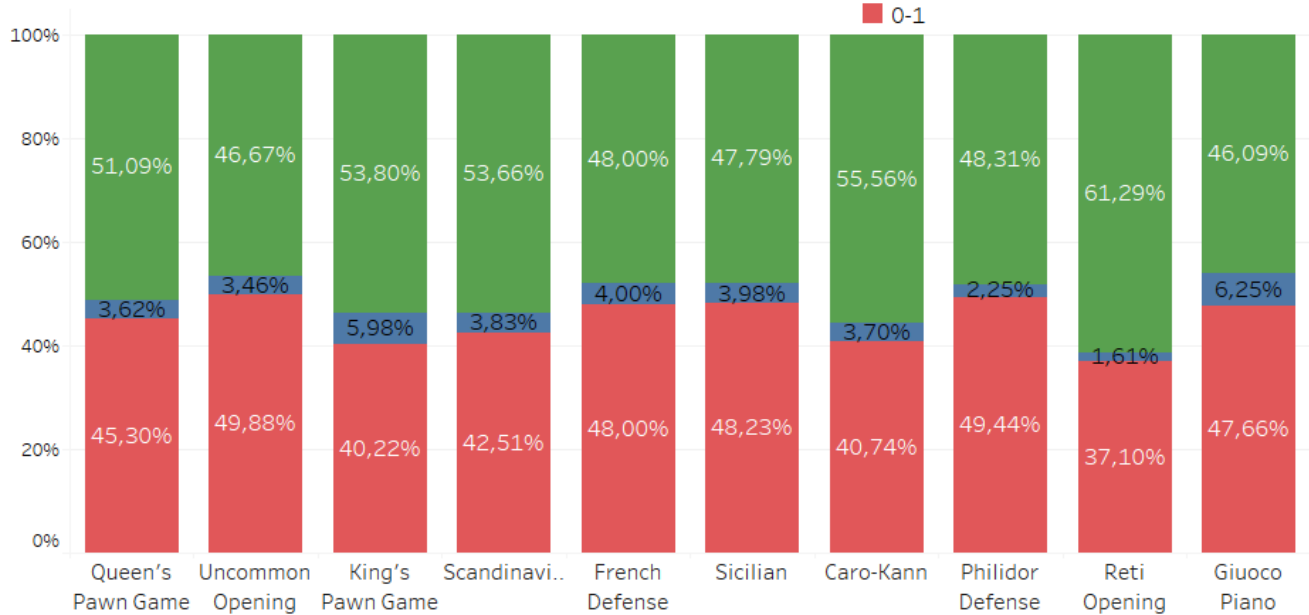


I selected top 10 openings by play rate. Queen's Pawn Game is the most played chess opening in this dataset. We can see that some openings are preferred in some time classes and neglected in others. For example, King's Pawn Game is rarely played in bullet but played very often in rapid. Scandinavian thrives in blitz, but it is almost never played in daily format.

Results for under 1000 elo



Results for 1000+ elo

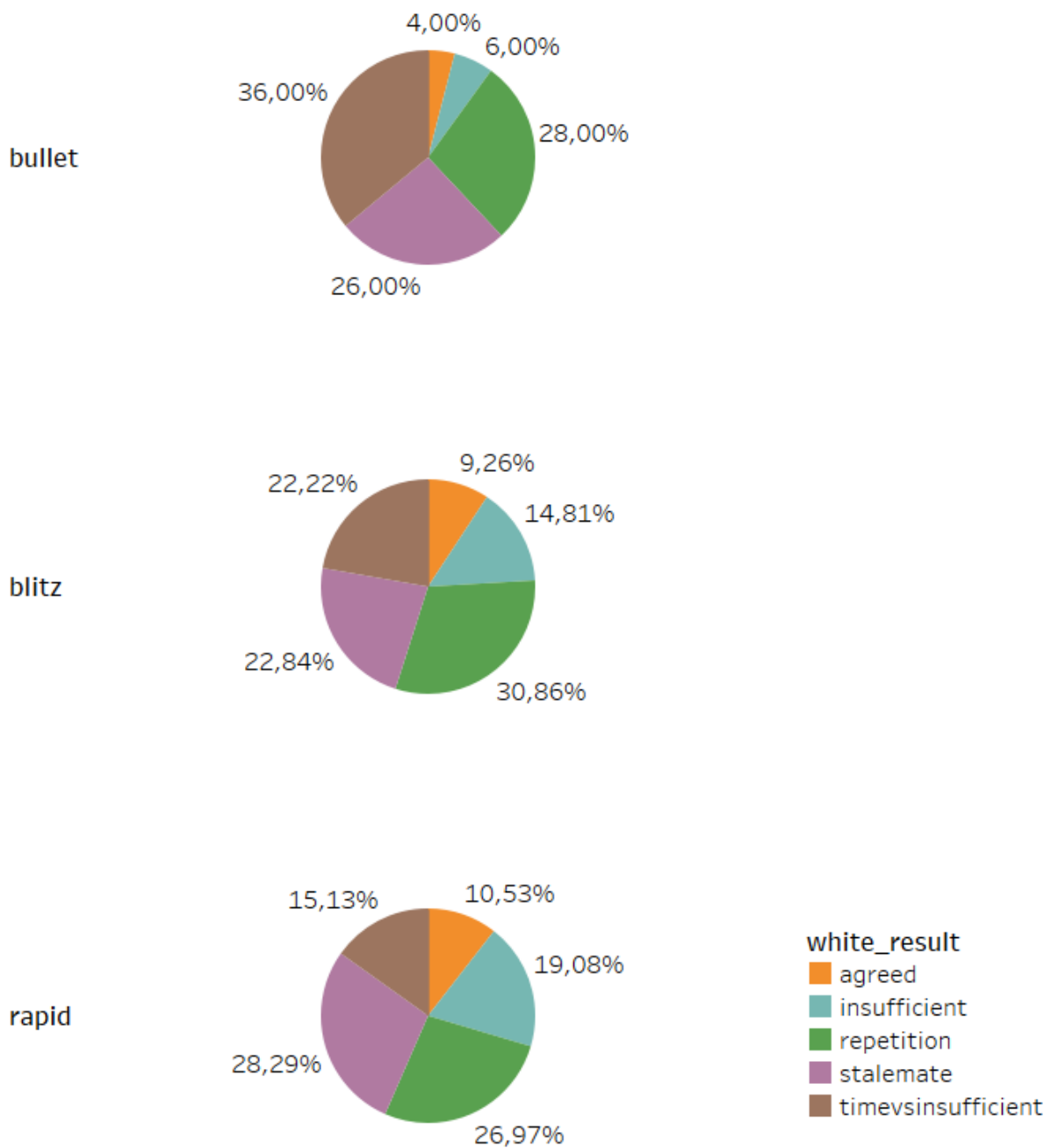


Due to skill and higher elo (elo=rating system in chess), success of chess openings varies. For this visualization, I divided this dataset into two parts. Beginners or players under 1000 elo fall into first part. More advanced players with over 1000 elo are the second part. If we look closely, there is a notable difference in win rates in almost every opening. Uncommon Opening is great for lower elo as white, but more advanced players tend to punish mistakes in irregular openings, so black is actually a favorite when it comes to playing Uncommon Opening in higher level of the game. Reti Opening is even in terms of winning and losing for beginners, but it does phenomenally in higher elo.

Analyzing more games is needed to prove these findings are correct.

time_class

Indecisive games



These three pie graphs contain data about indecisive games (draws) players had for each time class. In fastest time class – bullet, games are most likely going to end up in a draw due to one side losing all the time. But the other side doesn't have sufficient material to compete for a win. To further clarify:

- White has a winning endgame, but his clock drops down to zero. Black has only a king left. King alone cannot checkmate the other king. So the game ends in a draw

Other options for a draw are as follows:

- Agreed: if one player offers a draw and the other accepts it, the game concludes in a draw
- Draw by insufficient: both sides have no chance of checkmating the other side. Game automatically concludes in a draw
- Repetition: if a game reaches the same position three times, a draw can be claimed
- Stalemate: when one side has no legal moves to make (the king is not in check)

With more time on the starting clock, players tend to offer and accept draws more. Players also convert the winning endgame into win more often as the starting time gets longer

Conclusion

Chess is simple, yet a difficult game. Nonetheless it is well balanced and there is a lot of symmetry. White usually has an advantage due to playing first. Black has some options if he wants to get a win as well. There is a lot of openings and variations which can work in a certain elo. More time on the clock means that players can think more carefully about their moves and there are fewer mistakes – indecisive games. This and many other factors affect the result of the game.