

EXPLORING POPULARITY OF CHESS OPENINGS AND PREDICTING WINNERS

Springboard Data Science
Capstone

Elijah Seeley



Focusing on Openings

The beginning moves set the stage for the rest of the match.

These openings are studied in-depth by masters everywhere.

While tournament recorded games are available on databases, what about analytics on the casual player?



Predicting the Winner

What metrics are useful for predicting winners?

Are we able to predict the winner by opening alone?

By knowing the winning percentages of openings, will this influence the meta game?



Usefulness

By studying the billions of games played by casual players, we can collect and provide useful information to players who aren't participating at the highest level.



Opening Metrics

Knowing what the most common openings are and the success rate of each opening, players can better prepare for their casual encounters.

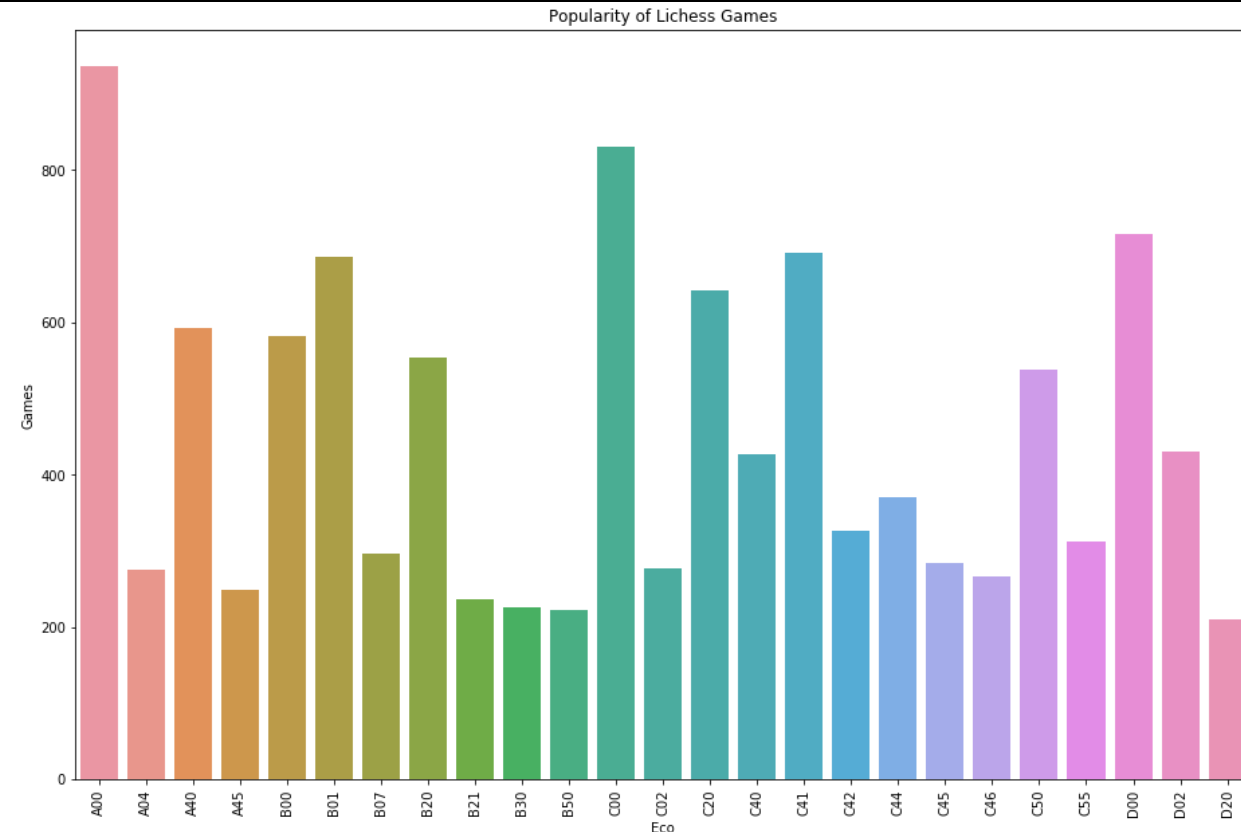
As these metrics become generalized in real-time, players could begin immerse themselves more deeply in theory as they now have focus points to study rather than aimless learning.

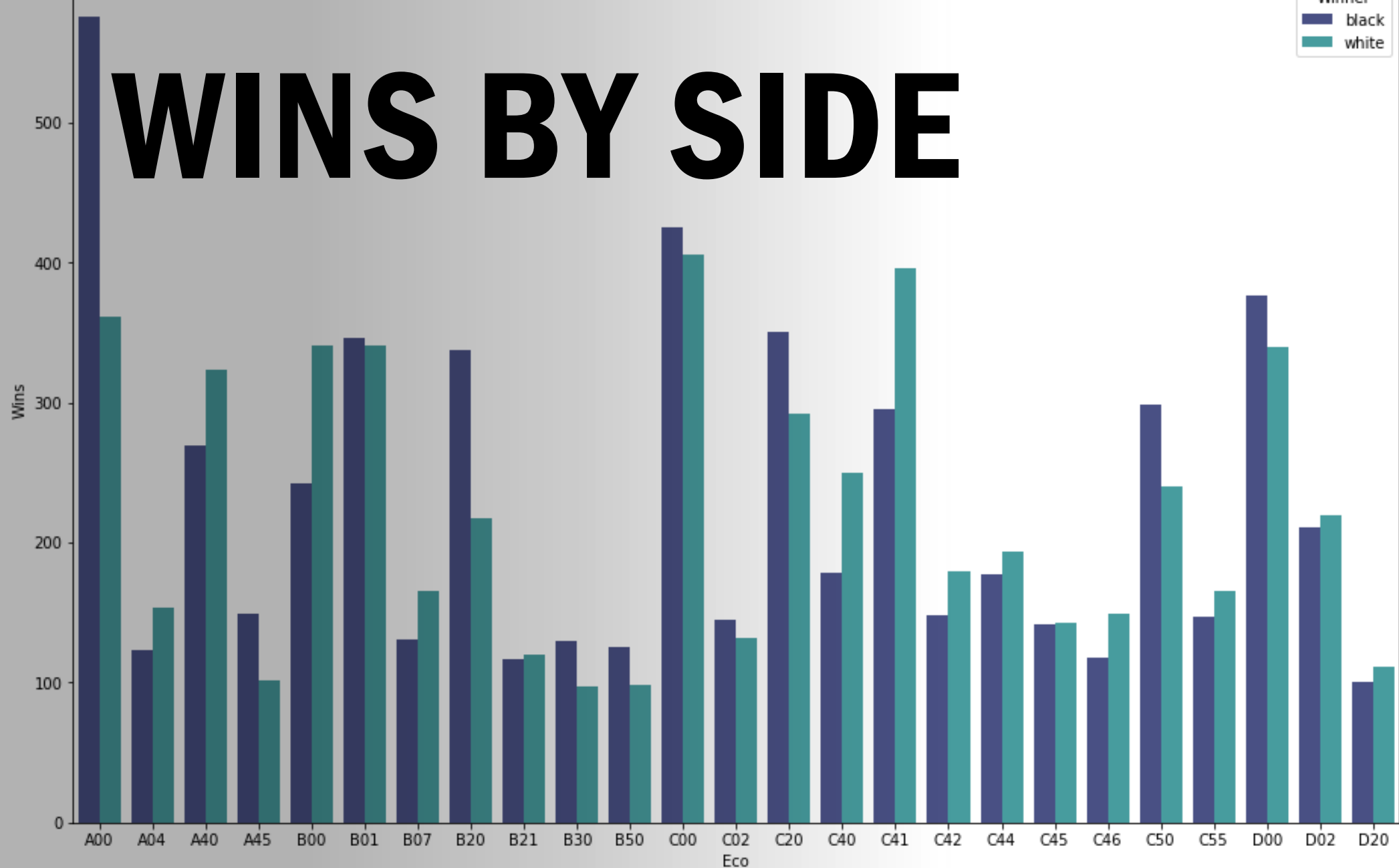


Data: Opening Popularity

As mentioned, providing players focus points such as opening popularity can give advantages in preparation. Here we see that A00, C00, and D00 are popular from our dataset.

In the future, we could even go as far as analyzing openings by demographics and provide interesting, albeit perhaps less actionable data.





Note: We have used draws as black wins under the guise of black being at a disadvantage from the beginning of the game by being behind in tempo.

```
{ '800s': 'D00', '900s': 'C20', '1000s': 'C20', '1100s': 'C20', '1200s': 'C20', '1300s': 'C20', '1400s': 'C20', '1500s': 'C41', '1600s': 'C00', '1700s': 'C00', '1800s': 'C00', '1900s': 'B01', '2000s': 'B01', '2100s': 'B01', '2200s': 'B20', '2300s': '(B23, D35) ' }
```

Most used opening by Elo ranking from our data.

Popularity by Rating

As players climb the rating ladder, it is useful to know what is to come in their new level. As ratings increase, certain openings will change in appeal.

However, this may also create openings for playing openings that are not often played, disrupting preparedness.

Chess as an E-Sport

In 2020, Hikaru Nakamura has been the first player signed on by an e-sports association.

This infers that chess is on its way to becoming popularized as a legitimate spectator sport.

Which brings us to winner predictions.



Predictions and Betting

Predicting the winner of sports games is largely enhanced by correctly analyzing player and team metrics.

Once betting on chess games becomes popularized, finding the correct metrics to predict the winners of chess games will quickly become of use.



MACHINE LEARNING - PREDICTIONS

	precision	recall	f1-score	support
Black	0.61	0.66	0.63	1579
White	0.64	0.60	0.62	1632
accuracy			0.63	3211
macro avg	0.63	0.63	0.63	3211
weighted avg	0.63	0.63	0.63	3211

Model using all parameters, such as length of game and player ratings.

	precision	recall	f1-score	support
Black	0.73	0.19	0.30	1579
White	0.54	0.93	0.69	1632
accuracy			0.57	3211
macro avg	0.64	0.56	0.49	3211
weighted avg	0.63	0.57	0.50	3211

Model using only openings

Conclusion

Chess has historically been a game about analyzation and theory, but education has been dominated by understanding high stakes games by renowned players and computer engines. By focusing on casual player metrics we will be able to provide useful information to players who do not play with high stakes. By understanding that predicting chess games will be commonplace in the coming years we may be able to get a head-start on understanding what metrics are powerful in determining which competitive players will come out on top.

