

Построение стратифицированных эталонных кривых Lichess для Elo+Chess

Elo+Chess Исследовательские заметки

31 мая 2026 г.

Abstract

Отчеты Elo+Chess строятся на основе больших стратифицированных выборок Игры Lichess.org, в которых используется образец теста производительности, содержащий 483 974 1+0 буллевых игр, 479 974 3+0 блиц-игр, 925 636 10+0 быстрых игр, и 660 540 более длительных быстрых игр. Для каждого основного типа игр конвейер расширяется. каждую игру в состояния доски и после хода, вычисляет сотни промежуточные и игровые переменные, объединяет эти переменные в показатели коучинга, а затем усредняет эти показатели по корзине Эло Lichess. выборка намеренно стратифицирована, а не пропорциональна исходному игроку пул: цель – оценить кривые устойчивого поведения на каждом уровне рейтинга, включая ведра, которые в противном случае были бы редкими, чтобы не воспроизводить естественный рейтинговое распределение. В результате получается семейство эмпирических эталонных кривых: для по каждому показателю, мы можем спросить, как часто игроки на разных уровнях Эло показывают данный привычка или принцип в реальных играх. В данной заметке описан метод, масштаб вычисления и несколько примеры дебютного принципа, включающие рокировку, развитие легкой фигуры, повторяющиеся ходы первой фигуры и раннее движение ферзя. Акцент делается на методологический: Elo+Chess не предполагает, что шахматные принципы универсальны. линейна на всех уровнях квалификации. Вместо этого эталонные кривые проверяются диапазон рейтинга, и продукт ориентирован на регион от начинающего до продвинутого уровня. где отношения наиболее сильны и наиболее интерпретируемы.

1 Цель

Цель эталонной системы — превратить общие рекомендации по шахматам в измеримые величины. Такое утверждение, как «развивайте свои второстепенные пьесы», является полезно, но становится более действенным, когда мы можем сказать, сколько мелких частей игрок обычно развивается к 10-му ходу, как это соотносится с его сверстниками того же уровня рейтинговая корзина и то, как эта привычка меняется в зависимости от рейтингового спектра.

Таким образом, Elo+Chess создает библиотеку показателей на основе реальных игр. Каждая метрика вычисляется для каждой строки «игрок-игра», суммируется по типу игры и корзине Эло, и используется в качестве эталонной кривой в отчете. Затем отчет может поместить пользователя на ту же кривую и показывать, находится ли пользователь выше, ниже или рядом с поведением наблюдается среди сопоставимых игроков.

2 Исходные данные и стратификация

Источником тестов является общедоступная база данных игры Lichess.org. Elo+Chess использует большие стратифицированные выборки по типу игры и рейтинговой группе, а не маленькая удобный образец. Текущие артефакты среды выполнения продукции охватывают четыре основных категории игр:

- 1+0 пуля,
- 3+0 блиц,
- 10+0 рапид,
- быстрые игры продолжительностью более 10 минут.

Целью стратификации является широта рейтинговых уровней. Пропорциональный случайный выборка была бы эффективна для описания общего пула игроков, но она быть неэффективным для построения эталонных кривых. Наиболее распространенные рейтинговые сегменты будет доминировать в выборке, а хвостовых ковшей с низкой опорой будет слишком мало игры для стабильных средних показателей. Elo+Chess поэтому производит выборку по средней игре Эло ведро. В текущей конструкции чистой выборки сырые игры занимают первое место. фильтруется по играм с действительным рейтингом с известными игроками, известными рейтингами, действительными результатами, и известный контроль времени. Затем игры фильтруются по игрокам, у которых достаточно игр. с такой скоростью и достаточно стабильными рейтингами, чтобы отражать последовательный уровень навыков: для каждого игрока Elo+Chess вычисляет диапазон рейтинга игрока в пределах скорости. $(\max(\text{Elo}) - \min(\text{Elo}))$ через исходный пул и сохраняет игры только тогда, когда у обоих игроков есть хотя бы настроенное минимальное количество игр и не более 300 баллов Эло за дрейф в пределах скорости. Дрейфовый фильтр не подразумевается для удаления нормального улучшения или отклонения; это уменьшает случаи, когда игрок в играх будут сочетаться существенно разные уровни навыков, например, предварительные учетные записи, вернувшиеся игроки или учетные записи, которые быстро перемещались по нескольким сегментам в течение период выборки. Наконец, подходящие игры ранжируются по каждому среднему рейтингу Эло. ведро детерминированным хешем идентификатора игры и начального числа, а также ограниченным количеством игры сохраняются из каждого сегмента.

Такая конструкция делает контрольные кривые гораздо более полезными. Каждое ведро Эло предоставляет достаточно наблюдений, чтобы оценить среднее значение каждого показателя, а редкие высоко- или низкорейтинговые ведра не заглушаются естественной формой распределения активных игроков. Тестовые кривые, используемые Elo+Chess: затем ограничивается диапазоном, наиболее соответствующим продукту, примерно для начинающих через раннюю продвинутую игру. В отчетах шкалы Lichess этот основной диапазон 600-1600. Когда пользователь просматривает отчет по шкале Chess.com, соответствующие сегменты рейтинга отображаются на один и тот же базовый Lichess. контрольные сегменты с использованием отдельного метода кросс-платформенного сопоставления рейтингов описано в сопутствующем документе: [Estimating Cross-Platform Chess Rating Mappings with Модальная регрессия](#).

3 Распределение рейтингов активных игроков

Если посмотреть на распределение рейтингов Эло игроков Lichess.org на конец По состоянию на март 2026 года по основным типам игр мы видим следующие показатели и процентиля. В этом представлении распределения каждый игрок учитывается один раз для каждого конкретного типа игры. используя последний наблюдаемый рейтинг этого игрока в соответствующем сырье ряды с контролем времени после того, как потребуется не менее пяти геймов в этой конкретной игре. тип.¹

Game type	Players	Median	75th pct.	80th pct.	% ≤1600	Exact 1500	1500 share
1+0 bullet	1,221,146	1,479	1,858	1,951	58.9%	4,756	0.4%
3+0 blitz	1,144,045	1,542	1,845	1,912	54.9%	2,315	0.2%
10+0 rapid	1,679,401	1,381	1,691	1,765	68.3%	3,472	0.2%
>10 rapid	668,255	1,400	1,686	1,753	68.2%	7,969	1.2%

Table 1: Последнее наблюдаемое распределение рейтинга активных игроков Lichess по точным тип игры после того, как потребуется как минимум пять игр именно этого типа.

Это распределение не следует путать со стратифицированной эталонной выборкой. Распределение описывает активный пул игроков Lichess, наблюдаемый в необработанных данных. сканирование. Затем эталонная выборка намеренно стратифицируется так, чтобы Эло распределял иметь достаточную поддержку для стабильных метрических кривых.

Чтобы сопоставить рейтинг Эло Lichess.org для определенного типа игры с эквивалентный рейтинг Chess.com, Elo+Chess использует модальное окно, зависящее от типа игры. уравнения регрессии, оцененные в сопутствующем документе [Elo+Chess rating-mapping paper](#):

$$\begin{aligned}
 \hat{R}_{\text{Chess.com,bullet}} &= -531.71 + 0.9871 R_{\text{Lichess}}, \\
 \hat{R}_{\text{Chess.com,blitz}} &= -551.54 + 1.0853 R_{\text{Lichess}}, \\
 \hat{R}_{\text{Chess.com,10minrapid}} &= -495.04 + 1.0741 R_{\text{Lichess}}, \\
 \hat{R}_{\text{Chess.com,>10rapid}} &= -369.02 + 0.9374 R_{\text{Lichess}}.
 \end{aligned}$$

Эти уравнения отображают значения рейтингов, а не процентильные ранги. Процентильные ранги определены внутри определенного пула игроков, а игроки Chess.com и Lichess бассейны имеют разную форму. В таблице 2 для 10-минутная быстрая игра, в первом столбце приведена небольшая выборка возможных вариантов. Chess.com Значения рейтинга Эло. Во втором столбце даны соответствующие Процентильный рейтинг Chess.com отображается на Chess.com по состоянию на 31 мая 2026 г. 71,7

ранг на Lichess.org по 10-минутному рапиду соответствует рейтингу Эло Lichess 1,644², и это значение показано в третий столбец. Если мы применим уравнение отображения Chess.com–Lichess, то 1644 на Lichess эквивалентен примерно 1271 на Chess.com. Предполагая, что предполагаемая уравнение отображения является разумным, разрыв +522 между рейтингом Chess.com в столбец 1 и сопоставленный рейтинг в столбце 4 можно объяснить большим количеством случайных игроков в пуле игроков Chess.com, раздувающих Chess.com процентильный

¹We apply a five-game minimum because lightly active accounts disproportionately remain at the 1500 starting/provisional anchor, creating an artificial spike in all-player latest-rating distributions. Figures 1–4 покажите распределения с этим фильтром и без него.

²Calculated from the cumulative distribution function of the ZXQPROT3ZXQ 10+0 rapid latest-rating sample after requiring at least five games in that exact game type; see Table 1 and Figure 3.

ранг. Другими словами, находиться в 72-м процентиле на Lichess.org — это большее достижение, чем попадание в 72-й процентиль на Chess.com, потому что Процентильные ранги не соответствуют друг другу четко из одной системы в другую. Настоящее Эло значения могут быть отображены с помощью уравнений, рассчитанных на основе результатов одного и того же игрока по две системы.

Chess.com rating	Chess.com pct.	Same-pct. Lichess	Mapped equiv.	Gap
749	71.7%	1,644	1,271	+522
1,012	86.2%	1,871	1,515	+503
1,358	95.4%	2,095	1,755	+397
2,383	99.9%	2,557	2,251	-132

Table 2: Иллюстративное сравнение 10-минутного быстрого процента Chess.com очков и такие же процентные баллы в активном рапиде Lichess 10+0. распределение после требования как минимум пяти игр со счетом 10+0. “Сопоставлен Chess.com экв.” применяет текущий Elo+Chess Быстрое преобразование рейтинговой шкалы 10+0 в рейтинг Lichess того же процента. самые верхние хвостовые строки включены, чтобы показать различие масштабов, но подобранные Рейтинговая линия наиболее надежна в основном тренерском диапазоне.

Причина пробелов в Таблице 2 заключается в том, что отображение оценивает эквивалентные рейтинги игровой силы, тогда как процентиль описывает положение среди населения одной платформы. Игрок с 71,7-м процентилем на Таким образом, Chess.com не является автоматически игроком 71,7-го процента на Lichess, и наоборот. Наблюдаемые цифры соответствуют общедоступным данным Chess.com. быстрый пул, имеющий гораздо большую массу игроков с низким рейтингом или случайных игроков, в то время как Пул активного точного контроля Lichess более сконцентрирован среди опытных повторителей игроки. Ключевым моментом является не то, что одна процентильная шкала является «правильной», а другое «неправильно»; дело в том, что процентильный ранг не переносим между платформами если только лежащие в основе распределения игроков не имеют одинаковой формы.

Note on the 1500 spike. The visible jump near 1500 присутствует в данных, а не создается в результате сглаживания.

В последнем наблюдаемое распределение рейтингов, точные 1500 оценок составляют 13,1

буллит-игроков, 5,6

17,6

стартовый или временный якорь в рейтинговой системе.

Пик в 1500 долларов во многом является результатом слабоактивных аккаунтов. Чтобы проверьте это, мы пересчитали те же распределения последних рейтингов после запроса каждый игрок должен сыграть как минимум пять игр определенного типа. Шип почти исчезает под этим фильтром: точные 1500 оценок падают с 5,6–17,6 игроков в распределении всех игроков до 0,2–1,2

Распределение указано в таблице 1.

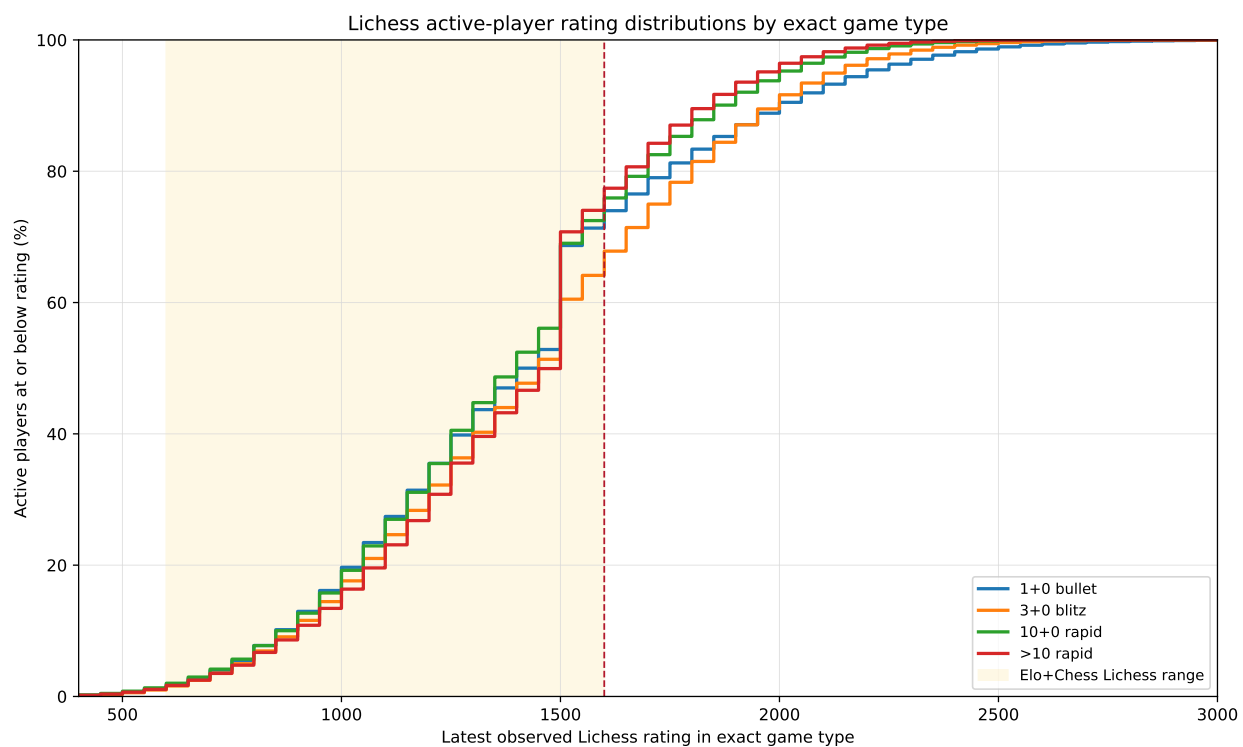


Figure 1: Распределение совокупного рейтинга активных игроков по конкретным типам игр. желтая полоса обозначает диапазон 600–1600 Lichess, используемый основным Elo+Chess. тренерские ориентиры.

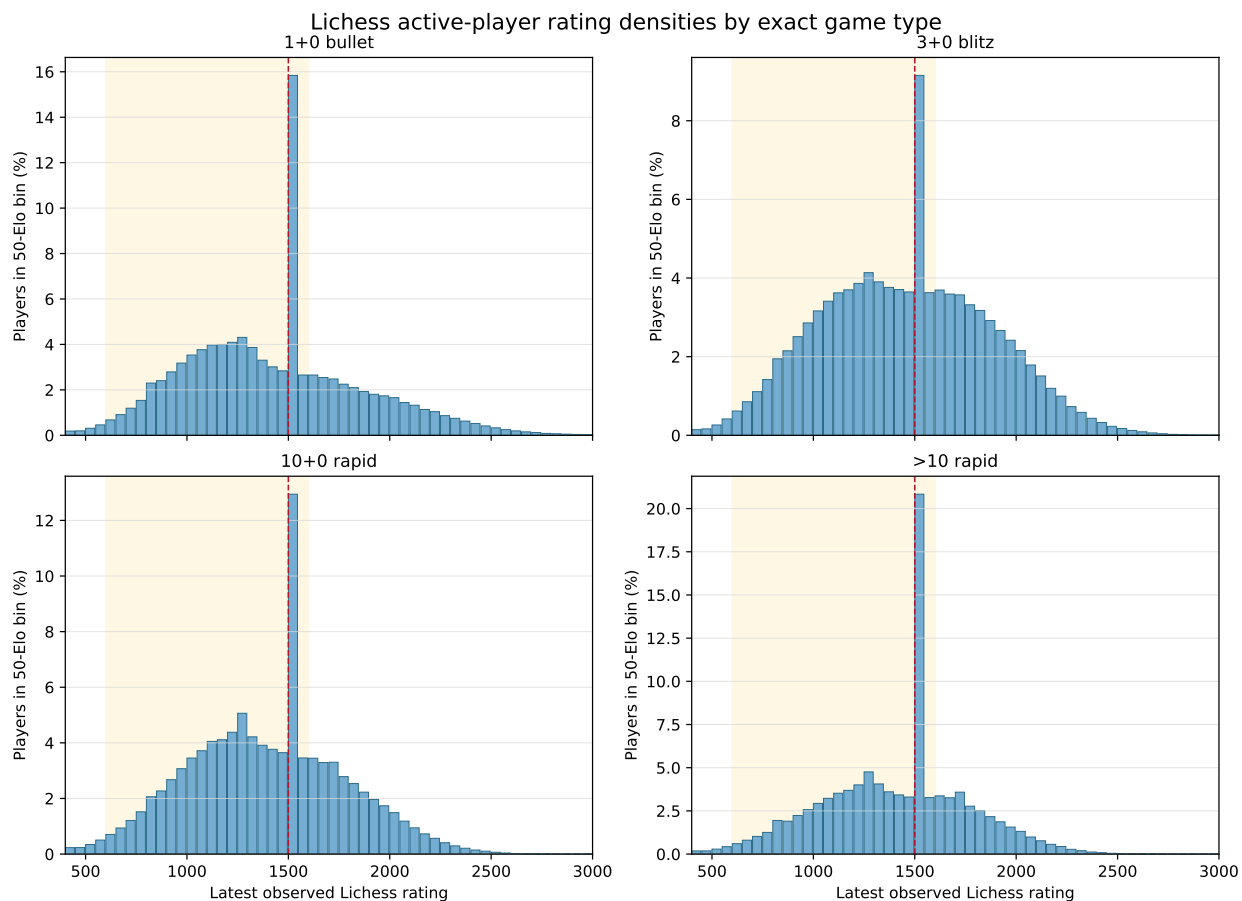


Figure 2: Представление плотности одних и тех же распределений рейтингов активных игроков. пунктирная линия отмечает 1500 Lichess, где необработанные распределения последних рейтингов большая точка массы точного номинала. Желтая полоса обозначает 600–1600 Lichess. диапазон, используемый основными отчетами о метриках Elo+Chess. Шип наиболее заметен в плотность всех игроков, поскольку многие малоактивные учетные записи остаются точно на уровне 1500; Рисунок 4 повторяет график плотности. после требования как минимум пяти игр определенного типа, что в значительной степени устраняет этот артефакт.

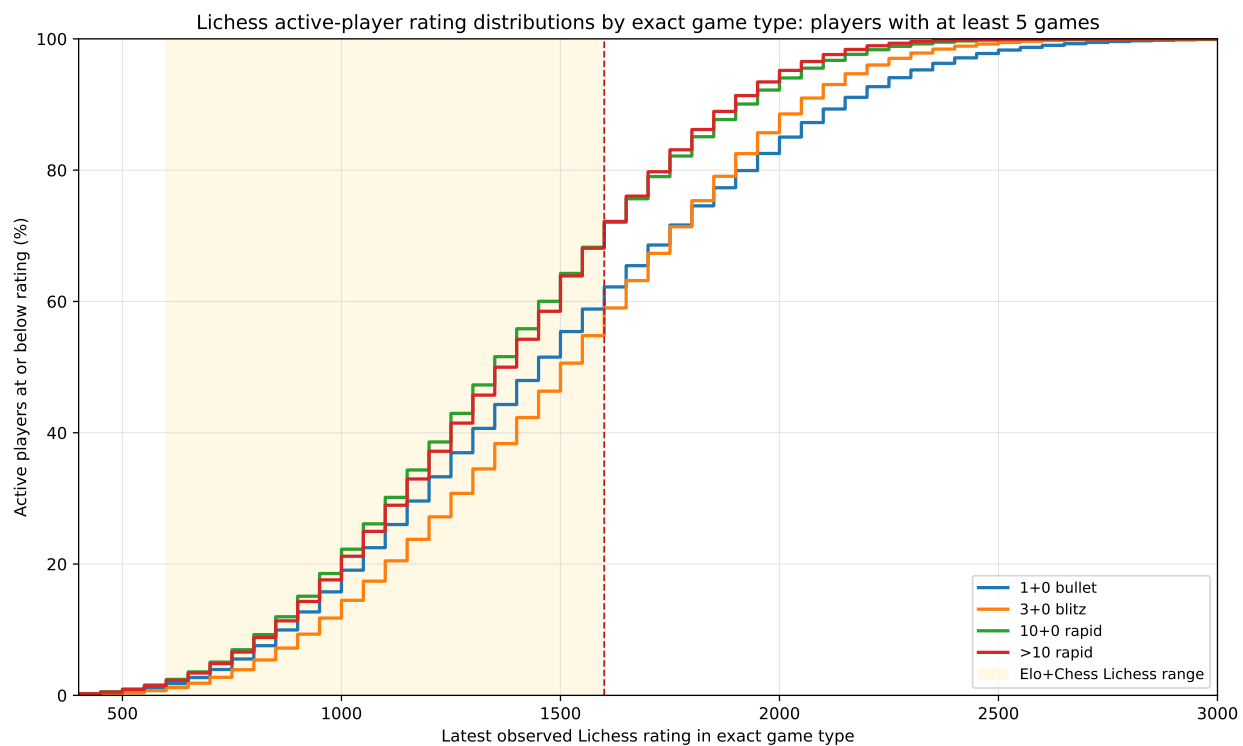


Figure 3: Распределение совокупного рейтинга активных игроков после требования не менее пяти игр определенного типа. Прыжок на 1500 значительно уменьшен.

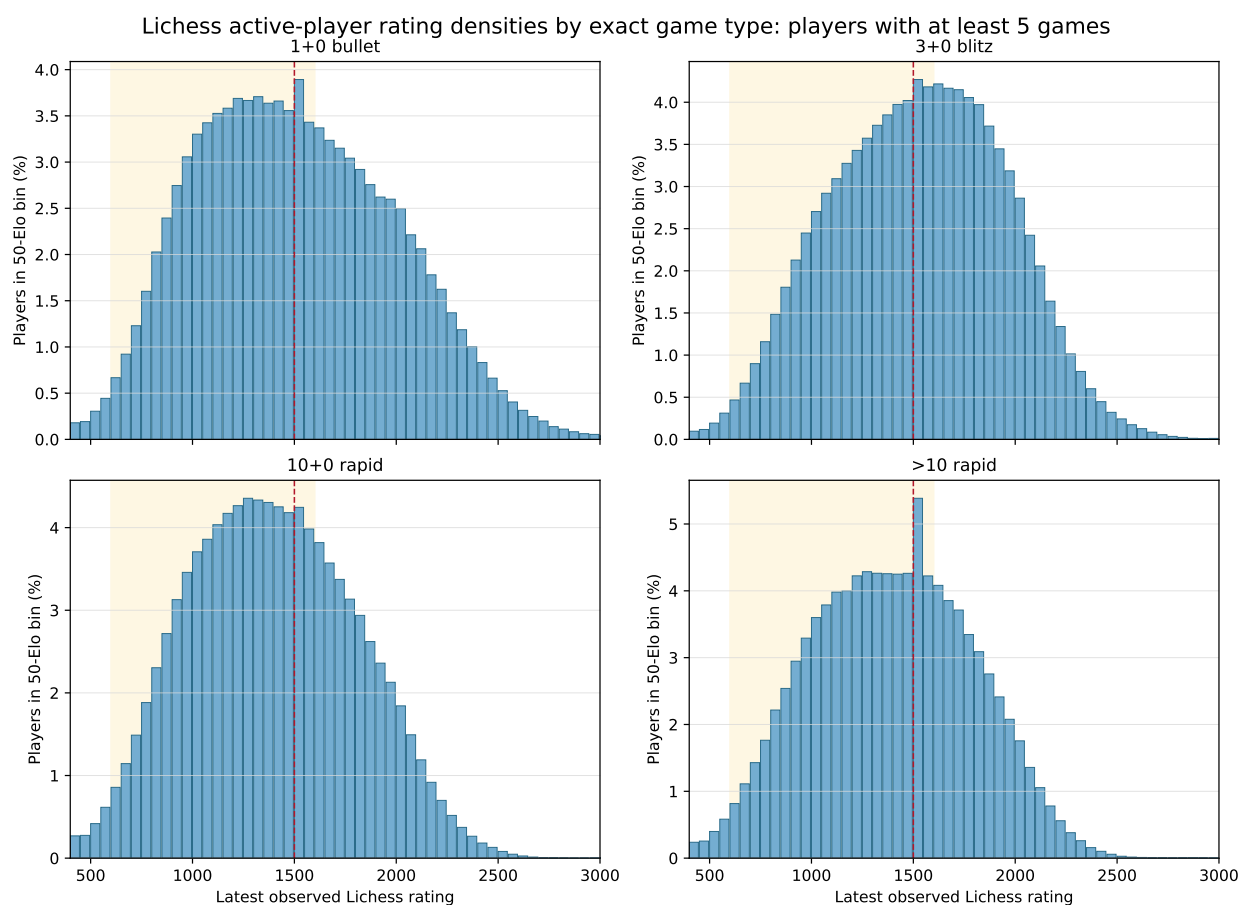


Figure 4: Просмотр плотности после запроса как минимум пяти игр в конкретной игре тип. Эта точка зрения подтверждает, что большое 1500 очков массы в общеигровом режиме Распространение в основном обеспечивается малоактивными аккаунтами. Желтая полоса отмечает диапазон 600–1600 Lichess, используемый в основных отчетах по метрикам Elo+Chess.

4 Масштаб текущих артефактов эталонного теста

Вычисление намеренно большое. Это не царапина на профиле, а небольшая набор загруженных PGN или собранная вручную коллекция поучительных примеров. Первоначальный проход Lichess охватывал полные ежемесячные файлы с января 2025 года. по март 2026 года. Нестратифицированная таблица сканирования содержит 1 382 178 087 игр. и 2 764 356 174 строки рейтинга игроков. В этих строках 6 623 438 уникальных имена пользователей игроков появляются хотя бы один раз.

Speed	Games in raw scan	Player-game rows	Unique players
Bullet	519,375,423	1,038,750,846	3,077,635
Blitz	839,823,863	1,679,647,726	5,987,634
Rapid	22,978,801	45,957,602	1,535,865
All scanned speeds	1,382,178,087	2,764,356,174	6,623,438

Table 3: Нестратифицированная шкала сканирования Lichess перед выборкой эталонов. Сканирование охватывает полные ежемесячные файлы истории игр с января 2025 года по март. 2026.

Артефакты производственного теста намного меньше, чем необработанное сканирование, потому что они выбраны для метрического построения. Они по-прежнему велики: эталон Этап представляет собой конвейер расширения и агрегирования состояний перемещения по сотням тысячи игр каждого типа.

Game type	Games	Player-game rows	Unique players	Move states	Feature columns
1+0 bullet	483,974	967,948	119,732	22,985,272	277
3+0 blitz	479,974	959,948	129,398	28,138,904	277
10+0 rapid	925,636	1,851,272	229,235	57,554,339	277
>10 rapid	660,540	1,321,080	133,661	38,580,214	277

Table 4: Масштаб стратифицированных эталонных артефактов после выравнивания по общей схеме функций среды выполнения из 277 столбцов. А «Ряд игрок-игра» означает одну игру с точки зрения одного игрока, поэтому большинство игр создайте две строки игрока-игры.

Только для блица 3+0 текущий производственный тест использует 479 974 партии. с участием 129 398 уникальных игроков и 28 138 904 различных строк состояний хода. Выровненная таблица характеристик блиц-игры содержит 277 производных переменных. Артефакты быстрого построения функций еще шире на этапе перемещения: полный Таблица состояний быстрого перемещения, используемая при построении объектов, содержит 168 столбцы состояния перемещения перед объединением в функцию игры с 277 столбцами стол. Артефакты тестов среды выполнения интерфейса для всех четырех типов игр: затем приведено в соответствие с той же схемой функций игрока-игры из 277 столбцов, включая расширенные переменные таксономии. Из этих таблиц характеристик следует В отчете выбираются и отображаются наиболее поддающиеся интерпретации показатели коучинга.

Эти подсчеты важны для интерпретации результатов. Эталонные кривые не являются оценками нескольких сотен проверенных вручную игр. Вместо этого они исходят от десятков

миллионов оцениваемых государств совета директоров и более чем четырех миллионов строк с играми в четырех текущих категориях игр.

5 Обзор трубопровода

Эталонный процесс строительства состоит из шести основных этапов:

1. Выберите игры Lichess с рейтингом в целевом типе игры и страте рейтинга.
2. Анализируйте каждую игру ход за ходом и сохраняйте ход до и после хода. состояние правления.
3. Рассчитайте промежуточные переменные состояния хода на доске, ход, точка зрения игрока и точка зрения противника.
4. Объедините переменные состояния движения в переменные функций игры игрока.
5. Преобразуйте переменные игрока в игру в метрики отчета с понятным направлением интерпретации: чем выше обычно лучше, тем ниже обычно лучше или описательно.
6. Усредните каждую метрику по корзине Эло и вручную проверьте полученный результат. отношение.

Эта структура важна, поскольку многие показатели отчета не являются напрямую виден в заголовке PGN. Они требуют реконструкции положения, определения какие фигуры двинулись с каких стартовых полей, сравнивая значения до хода и после перемещения материала и структуры, а затем агрегирование результатов из переместите уровень на уровень игрока-игры.

6 Промежуточные переменные

Промежуточные переменные предназначены для описания того, что изменилось на доске. и почему это важно с точки зрения игрока, составляющего отчет. Примеры включают в себя:

- номер хода, номер хода игрока, сторона хода, цвет хода и игрок. цвет;
- с поля, на поле, тип движущейся фигуры, тип захваченной фигуры и был ли ход взятием, шахом, рокировкой, продвижением или матом;
- материальный баланс до и после переезда;
- покинула ли второстепенная фигура свое стартовое поле на заданный номер хода;
- ходила ли одна и та же фигура при последовательных ходах игрока в открытие;
- сделал ли ферзь ход до 10-го хода;
- сторона рокировки, номер рокировки и связались ли ладьи после рокировки;
- свободные фигуры, висячие фигуры, сдвоенные пешки, изолированные пешки, проходные. пешки и отсталые пешки;

- статус линии ладьи, открытые линии, полуоткрытые линии и защита короля функции;
- тактические возможности, реализованная тактика, тактика материального давления, гарантированные последующие выгоды и реализованные выгоды.

Точный промежуточный набор зависит от стадии сборки. Базы данных отчетов времени выполнения сохраняют только столбцы, необходимые для быстрого предоставления отчетов, а полная функциональность артефакты включают более широкие таблицы состояний перемещения и патчи функций, используемые для построения последние строки функций игрока-игры.

7 Примеры метрик принципа открытия

Принципы открытия являются полезными примерами, поскольку пользователям легко их понять. понять и легко подключить к расчетам состояния движения.

7.1 Minor Pieces Developed by Move 10

Этот показатель подсчитывает, сколько из четырех коней и слонов игрока осталось. свои домашние поля к 10-му ходу игрока. Промежуточный расчет использует цвет хода, тип движущейся фигуры, стартовое поле и ход игрока. номер. Для белых домашними полями являются b1, g1, c1 и f1. Для черных они b8, g8, c8 и f8. Для каждой игры-игрока Elo+Chess учитывает отдельный дом. поля среди коней и слонов, которые передвинул игрок на 10-м ходу.

Это показатель «чем выше, тем лучше». Он не утверждает, что каждый несовершеннолетний фигура всегда должна ходить на 10-й ход, но среди развивающихся игроков эмпирический взаимосвязь очевидна: игроки, достигающие более высоких показателей Эло, как правило, развиваются более мелкие фрагменты ранее.

7.2 Repeated Piece Moves in the Opening

Этот показатель учитывает дополнительные ранние ходы, потраченные на повторное перемещение той же фигуры. разработки чего-то нового. Промежуточный расчет использует предыдущий квадрат назначения, текущий исходный квадрат и номер хода игрока. Если та же фигура была перемещена во время хода предыдущего игрока и снова перемещается в пределах После первых 10 ходов игрока количество повторных ходов увеличивается.

Это показатель «чем ниже, тем лучше». Он измеряет потерянный темп открытия в простой способ. Некоторые повторяющиеся действия тактически оправданы, но в масштабе Кривая корзины отражает общую тенденцию: более сильные игроки достигают цели уровни тратят меньше ранних ходов на многократное перемещение одной и той же фигуры.

7.3 Queen Touched Before Move 10

Этот показатель учитывает ходы ферзя до 10-го хода игрока. при расчете используется тип движущейся фигуры и номер хода игрока. Если движущаяся часть является ферзем и количество ходов игрока не превышает 10, счет увеличивается.

Это также обычно лучше для целевого диапазона Elo+Chess. Ранний ходы ферзя могут принести материальные выгоды или вызвать уступки, но многие развивающиеся игроки вывести

ферзя слишком рано, потерять темп из-за атак на ферзя и задержать развитие второстепенных фигур и безопасность короля.

7.4 Castling and King Safety

Метрики рокировки используют флаги рокировки, сторону рокировки, номер хода рокировки и Состояние доски после рокировки. Примеры: сделал ли игрок рокировку, сделал ли игрок рокировку на 10-м ходу, сделал ли игрок рокировку на королевский фланг или ферзевый фланг, соединились ли лады после рокировки и были ли ладейные вертикали были открытыми или полуоткрытыми после рокировки.

Эти показатели не все интерпретируются одинаково. Завершение рокировки и рокировка на 10-м ходу, как правило, в целевом диапазоне обычно выше, значит, лучше. Рокировка на ферзевый фланг или рокировка на открытые вертикали могут в большей степени зависеть от контекста. Вот почему Elo+Chess не просто жестко запрограммировал моральное правило; он строит ведро кривые и проверяют, какие отношения достаточно стабильны для коучинга.

8 Кривые ковша и ручной контроль

После расчета показателей игры-игрока Elo+Chess группирует их по Lichess Elo. ведро и вычисляет средние значения ведра. Среднее значение сегмента — это необработанный эмпирический результат. контрольную точку для этого показателя. Затем в отчете используется линия тренда для присваивать пользовательским значениям оценки в стиле Elo.

Кривые проверяются вручную перед использованием. Метрика более полезна когда средства сегмента демонстрируют последовательную взаимосвязь во всем диапазоне рейтингов обслуживается продуктом. Метрика менее полезна, если она шумная, плоская, сильно немонотонны или обусловлены особыми случаями, которые трудно объяснить пользователю.

Этот этап ручной проверки особенно важен над верхней частью Elo+Chess. отсечка. Более сильные игроки знают, когда следует нарушить дебютные принципы, и делают это. эффективно. Выше уровня «новичок» и «ранний продвинутый уровень» соотношение между простыми привычками и рейтингом часто выравнивается или меняет форму. Продукт поэтому подчеркивает диапазон, в котором эмпирические связи сильны и педагогически полезно.

9 Пример полнодиапазонных кривых открытия

На рисунке 5 показаны четыре кривые принципа дебюта 3+0 для блица. Серые точки показывают более широкий диапазон рейтингов Lichess. Синие точки и синий тренд линия показывает диапазон 600–1600 Lichess, используемый основным тренером Elo+Chess. ориентиры. Пунктирная красная линия отмечает верхнюю границу.

Та же картина наблюдается в быстром и более быстром темпе 10+0 на полной дистанции. артефакты. Основное решение моделирования остается неизменным: использовать эмпирически стабильный диапазон требований коучинга и не притворяйтесь, будто это простая кривая привычки. остается одинаково информативным навсегда.

Тот же шаблон численно появляется в таблице 5. Slopes are сообщается как изменение значения показателя на 100 баллов Эло. Castling, minor-piece развитие, повторяющееся движение фигур и раннее движение ферзя — все это ясно демонстрирует движение по горизонтали 600–1600. Выше 1600 уклоны здесь гораздо меньше. примеров, что согласуется

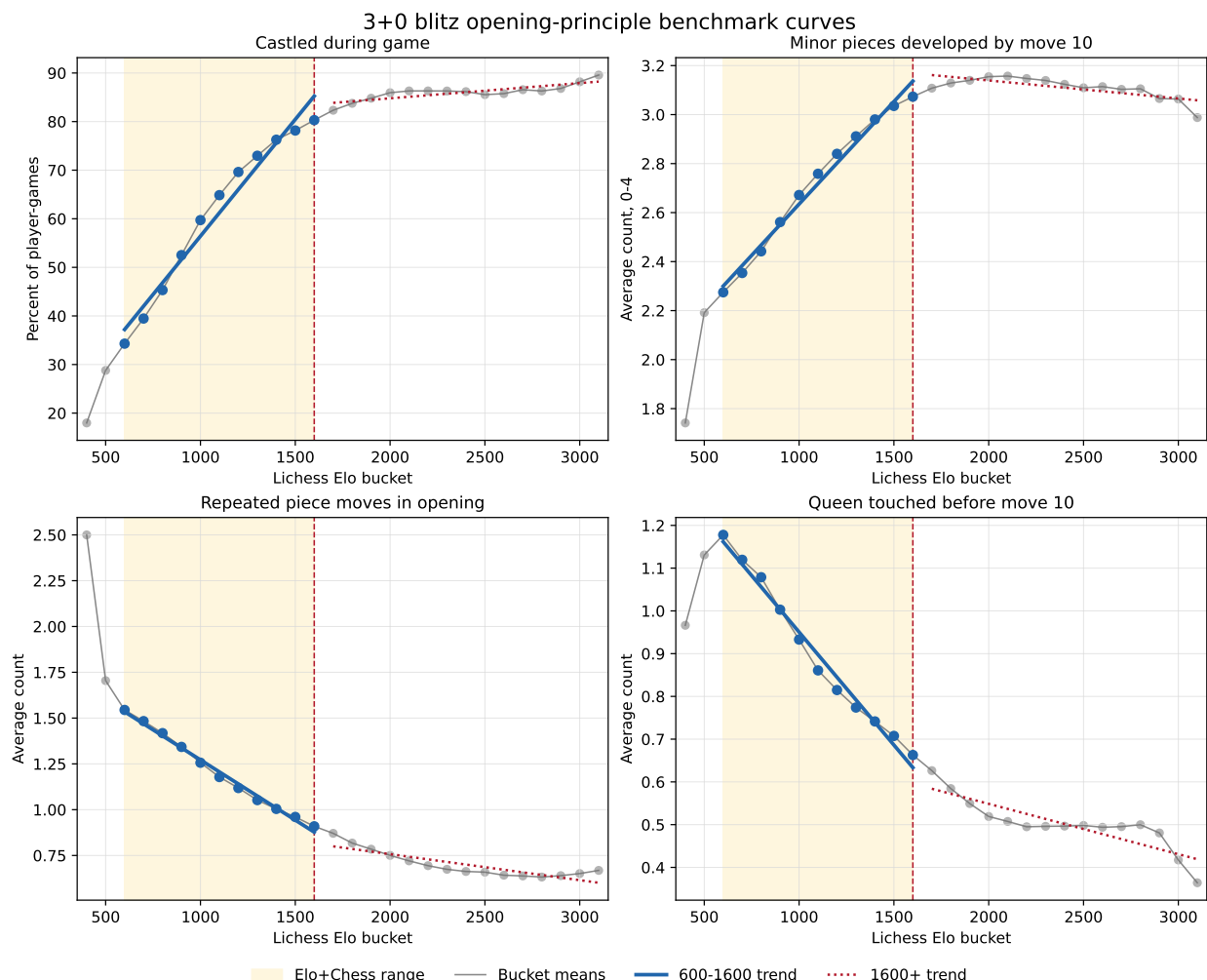


Figure 5: Избранные контрольные кривые принципа дебюта для блица 3+0. отношения самые крепкие и полезные в коучинговом диапазоне Elo+Chess. Выше верхнего порога некоторые отношения существенно выравниваются.

с идеей о том, что игроки с более высоким рейтингом нарушают основные принципы более избирательно и более успешно. A second mechanism is saturation. Для некоторых привычек показатель достигает практического потолка или минимального уровня. спот: сильные игроки уже усвоили эту привычку, поэтому нет причин для кривая продолжает резко двигаться вверх. Больше также не всегда лучше после that point. Таким образом, данные выравниваются, поскольку более сильные игроки могут сломаться. правила в конкретных положениях и потому, что многие основные привычки уже достигли их полезный рабочий диапазон.

10 Почему верхний предел имеет значение

Целевой пользователь Elo+Chess — игрок от начинающего до продвинутого уровня. Для этих игроков, основные привычки тесно связаны с рейтингом: развивайте фигуры, избегайте повторяющихся ранних ходов фигуры, избегайте ненужных ранних ходов ферзя, замок вовремя

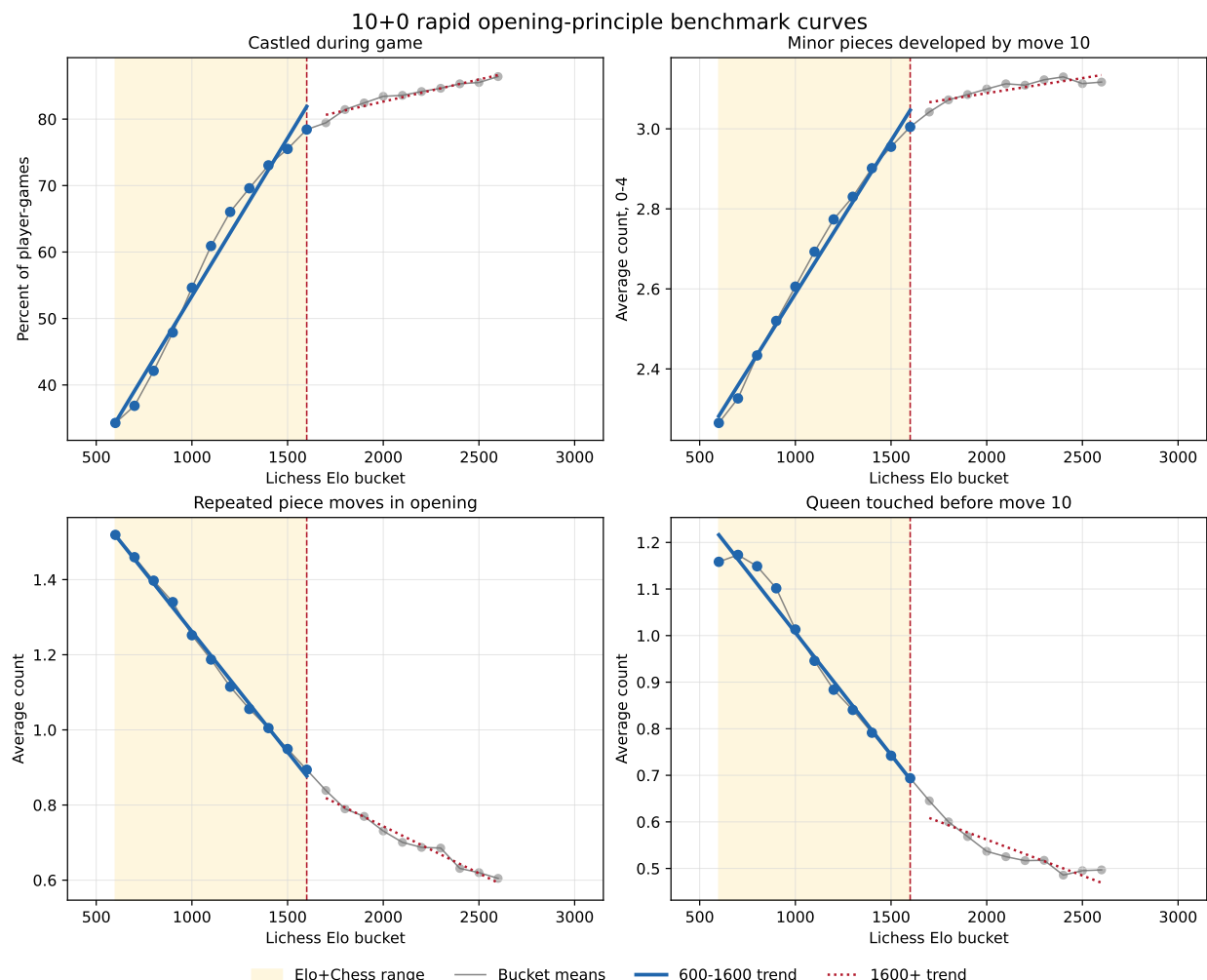


Figure 6: Избранные контрольные кривые принципа дебюта для быстрых игр 10+0. отношения самые крепкие и полезные в коучинговом диапазоне Elo+Chess. Выше верхнего порога некоторые отношения существенно выравниваются.

соединяйте ладьи, управляйте материалом и не оставляйте фигуры незакрепленными или висит.

При более высоких рейтингах те же самые поверхностные привычки становятся менее диагностическими. Сильный игрок может отложить рокировку, потому что центр закрыт, двигайте ферзя раньше потому что позиция содержит конкретную тактику, или дважды переместить одну и ту же фигуру потому что он выигрывает в темпе или вынуждает пойти на структурные уступки. Сырая привычка все еще имеет смысл, но связь между привычкой и рейтингом становится более условной. Вот почему Elo+Chess использует верхний порог для основного обучения. претензии и почему пользователи, находящиеся выше этого диапазона, предупреждаются о том, что тесная связь между принципиальными привычками и рейтингом рушится.

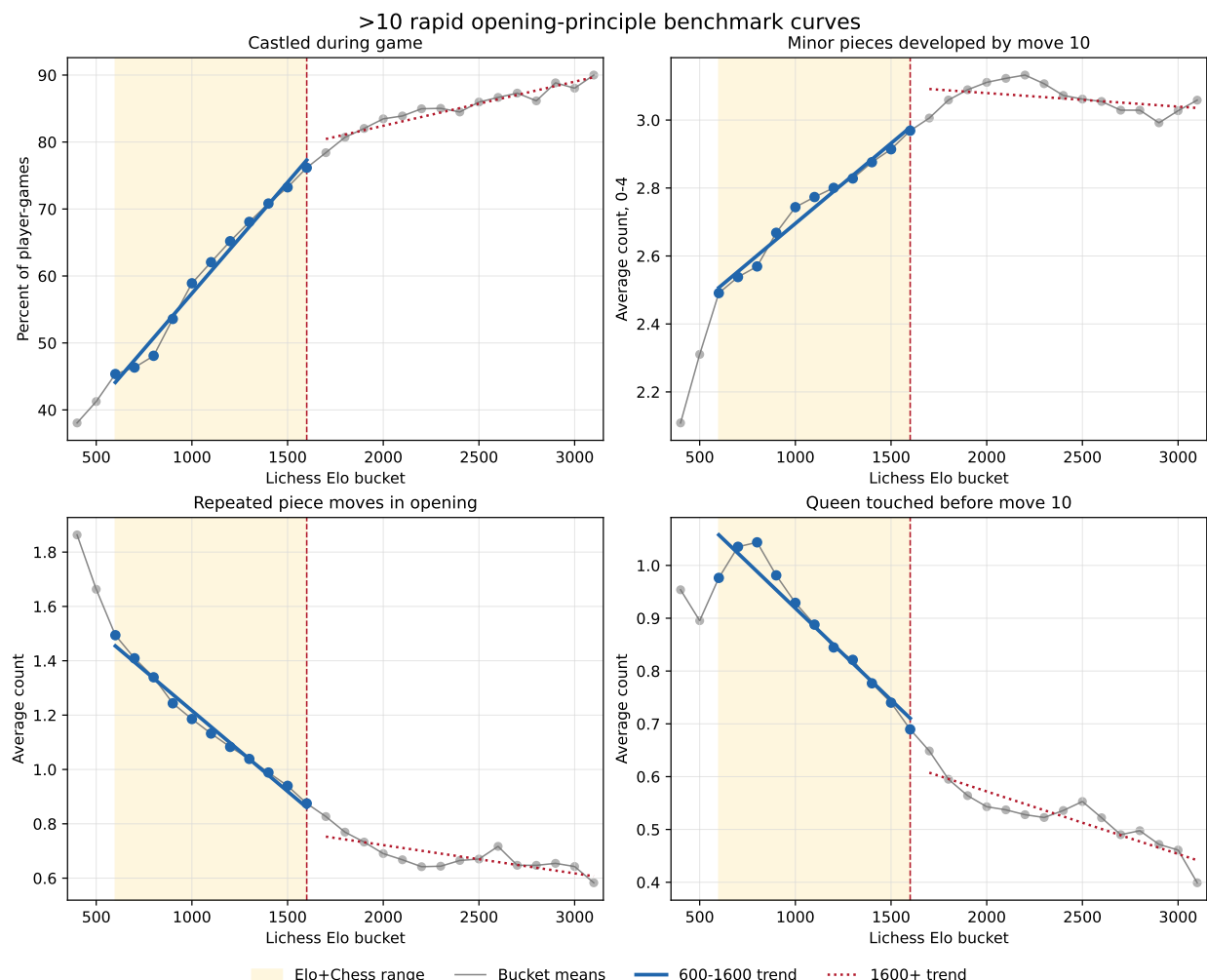


Figure 7: Избранные контрольные кривые принципа открытия для длительных быстрых игр.

11 Двусторонние тактические показатели и процентильные цели

Некоторые метрики имеют форму, противоположную примерам начального принципа. Вилки являются полезным примером, поскольку событие по своей сути двустороннее. Игрок, который находит и конвертирует больше полезных форков, чем одноранговые узлы, делает что-то хорошее, но количество конвертированных форков падает с рейтингом, потому что сильнее оппоненты допускают меньше возможностей для форков и защищаются от тактических угроз более активно.

Рисунок 8 демонстрирует этот эффект. Обе конвертированные вилки игрок-отчетник и реализованные оппонентом вилки становятся реже по мере того, как рейтинги растут. Это не значит, что выполнение форков — это плохо. Это означает сырое событие Скорость частично контролируется частотой ошибок противника и текстурой игры. Для метрик таким образом, Elo+Chess использует процентиля одноранговой истории внутри собственную корзину Эло игрока, а не полагаться только на глобальную тенденцию среднего значения корзины.

To build these percentile targets, Elo+Chess samples players by game type and Elo bucket, then uses recent qualifying games for each sampled player. The current target set uses 11 buckets

Game type	Metric	600–1600 slope per 100 Elo	1600+ slope per 100 Elo
1+0 bullet	Castled during game	4.137	0.550
	Minor pieces developed by move 10	0.063	0.009
	Repeated piece moves in opening	-0.051	-0.026
	Queen touched before move 10	-0.040	-0.011
3+0 blitz	Castled during game	4.804	0.313
	Minor pieces developed by move 10	0.084	-0.007
	Repeated piece moves in opening	-0.066	-0.014
	Queen touched before move 10	-0.053	-0.012
10+0 rapid	Castled during game	4.753	0.665
	Minor pieces developed by move 10	0.076	0.008
	Repeated piece moves in opening	-0.064	-0.025
	Queen touched before move 10	-0.052	-0.015
>10 minute rapid	Castled during game	3.320	0.659
	Minor pieces developed by move 10	0.047	-0.004
	Repeated piece moves in opening	-0.059	-0.010
	Queen touched before move 10	-0.035	-0.012

Table 5: Приблизительные наклоны ковшовой кривой для четырех примеров принципа открытия метрики по типам производственных игр.

from 600 through 1600, 200 players per bucket, and up to 50 recent qualifying games per player for each production game type. This gives 2,200 player samples per game type and approximately 110,000 player-game rows per game type for within-bucket percentile distributions.

Game type	Elo buckets	Players per bucket	Games per player	Player samples
1+0 bullet	11	200	49–50	2200
3+0 blitz	11	200	27–50	2200
10+0 rapid	11	200	37–50	2200
>10 minute rapid	11	200	47–50	2200

Table 6: Текущие целевые наборы процентилей по истории коллег. Это выборки истории игроков, используемые для оценки распределения внутри корзины для метрики в стиле процентилей.

Рисунок 9 иллюстрирует полученную цель. Распределение коэффициента конвертации вилки выбранных игроков в блиц-формате 3+0. Каждая коробка внутригрупповое распределение по выбранным игрокам, а не среднее значение по корзине индивидуальные игры. В пользовательском интерфейсе отчета значение пользователя сравнивается с соответствующим поле для сопоставленного сегмента Эло пользователя. На этом основаны такие утверждения, как является ли коэффициент конвертации форков игрока высоким или низким по сравнению с его коллегами на данный момент том же уровне, даже если общая частота событий среди населения снижается с ростом рейтинга.

12 Стоимость репликации

Метод в принципе воспроизводим, но не является легковесным. восстановление эталонные кривые требуют:

- доступ к большим ежемесячным файлам истории игр Lichess;

Two-sided converted-fork rates by Elo bucket

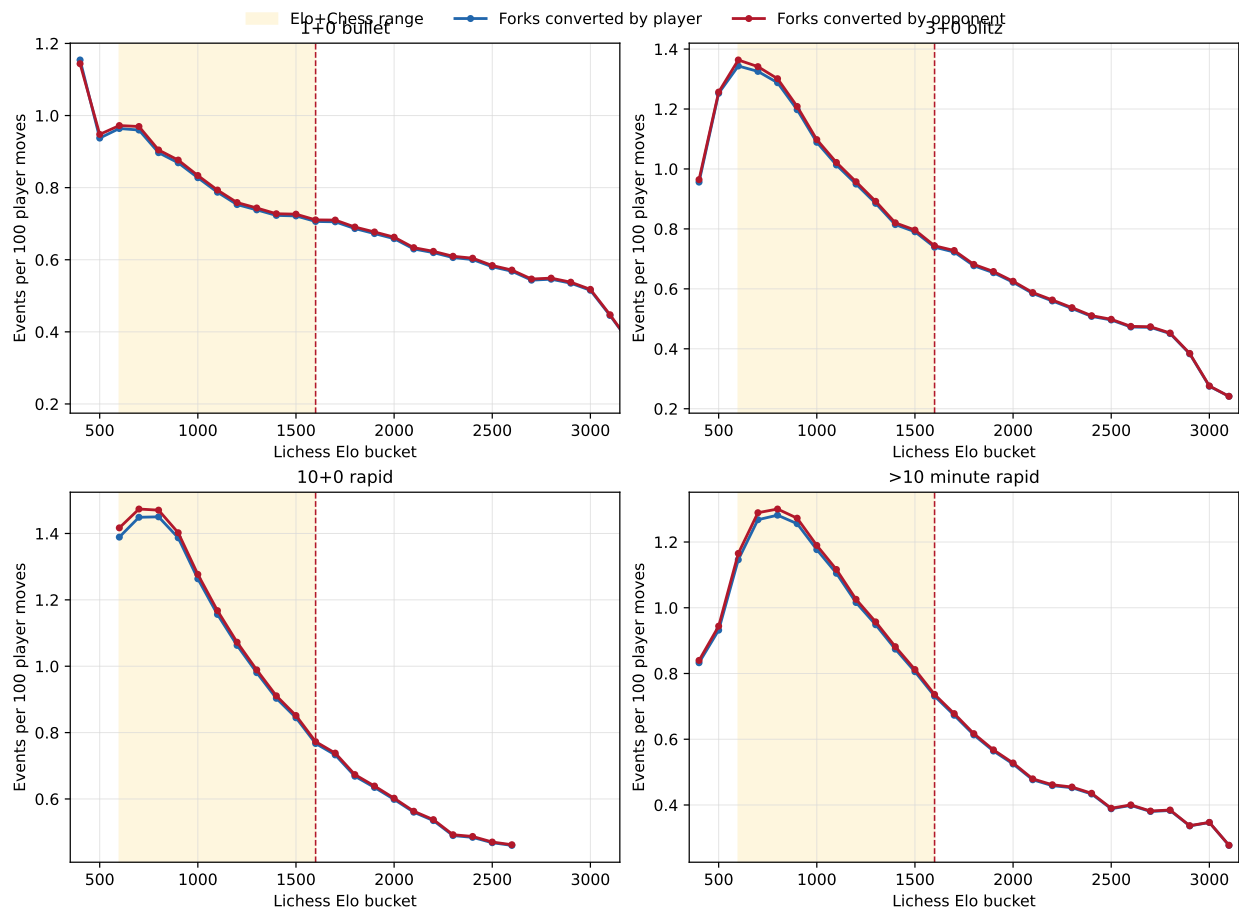


Figure 8: Двусторонние конвертированные вилки с ведром Elo. Преобразованные события форка падение при более высоких рейтингах, потому что обе стороны защищаются лучше и меньше чистых вилок конверсии доступны.

- стратифицированная фильтрация по типу игры, рейтинговой группе и критериям участия правила;
- легальный анализатор ходов и оценщик состояния доски, способный расширять десятки миллионов перемещающихся состояний;
- логика дебюта, материал, пешечная структура, безопасность короля, ладья активность, тактика, тайм-менеджмент и результаты игры;
- агрегирование состояния движения к функциям игры игрока;
- сводные данные на уровне сегмента для каждого показателя отчета;
- ручная проверка каждой потенциальной эталонной кривой перед ее использованием в тренерский отчет.

Для блица 3+0 это означает почти полмиллиона партий, почти миллион строк игрока-игры и более 28 миллионов строк состояния хода. Через четыре тестовые типы игр, текущие

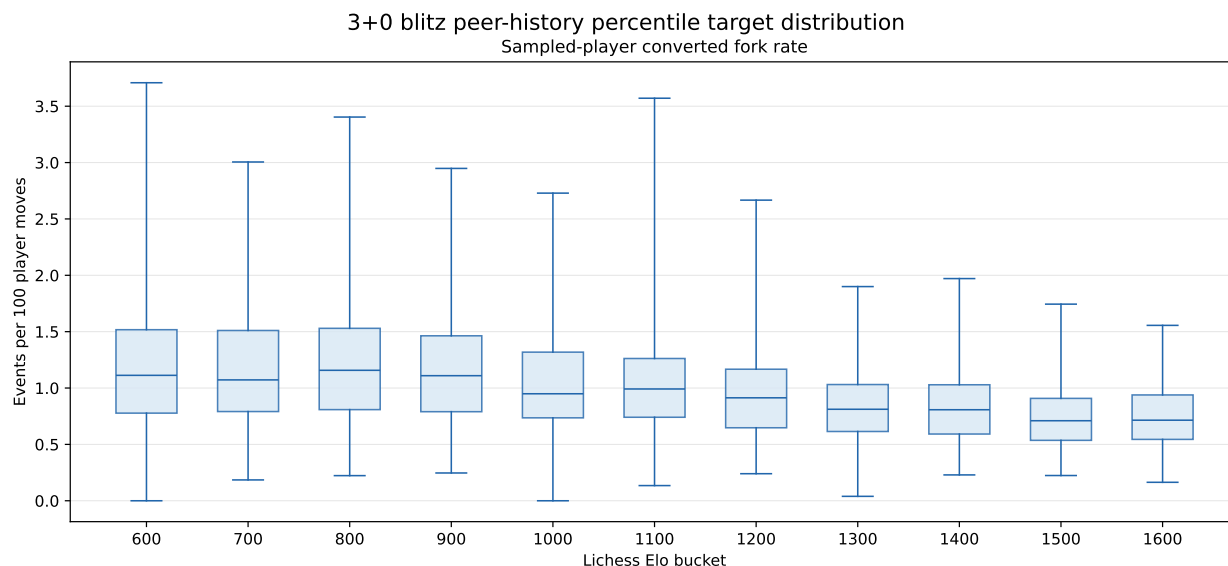


Figure 9: Пример целевого распределения по истории сверстников 3+0 для выбранного игрока конвертированная скорость вилки. Процентили рассчитываются в сегменте Эло, поэтому отчет сравнивает игрока со сверстниками, имеющими схожие тактические возможности ставки.

тестовые артефакты содержат более 2,55 миллионов игр, более 5,10 миллионов строк между игроками и более 147 миллионов строк состояния перемещения. Масштаб имеет решающее значение для ценности эталонный тест: он позволяет отчету показывать пользователям, что на самом деле происходит в играх, в которые они играют. игроками близкого к их уровню, а не полагаться только на общие шахматные лозунги.

13 Ограничения

Эталонные кривые представляют собой эмпирические описания, а не причинно-следственные доказательства. Игрок не получает рейтинг просто за счет соответствия среднему значению сегмента. Метрики также несовершенное описание сложных позиций: ранний ход ферзя может быть отличным в одном положении и безрассуден в другом. Цель кривых состоит в том, чтобы выявить повторяющиеся привычки и сравнить их с поведением сверстников, а не заменить расчет или тренерское суждение.

Кривые также являются кривыми в масштабе Lichess. Когда пользователи приносят игру Chess.com истории, Elo+Chess сопоставляет соответствующий сегмент рейтинга с тестом Lichess. масштабировать с помощью отдельной кросс-платформенной модели сопоставления. Этот слой отображения описано в сопутствующей исследовательской заметке Elo+Chess по Lichess-to-Chess.com. конвертация рейтинга: [Estimating Cross-Platform Chess Rating Mappings with Модальная регрессия](#). Это сопоставление не следует рассматривать как сохранение процентилей. ранг. Он сохраняет расчетную рейтинговую эквивалентность между двумя рейтингами. системы; процентильный ранг остается специфичным для платформы, поскольку активный игрок дистрибутивы разные.

14 Графика производственного отчета

Производственный отчет превращает артефакты эталонного тестирования в панели на уровне показателей. это можно прочитать, не зная полного конвейера сборки. Цифры 10 and 11 покажите три примера из живого дизайна отчета Elo+Chess.

Панели с принципом открытия на рисунке 10. используйте обычное представление эталонной кривой. Каждая карточка начинается с семейство метрики и имя метрики, затем приводится краткое правило интерпретации. большой сорт в правом верхнем углу намеренно отделен от сырья значение показателя: он отвечает на вопрос пользователя: «Как оценивается эта привычка?» относительно эталонного значения?», в то время как в строках ниже сохраняются измеренные значения. В четырех строках сводки показаны «За все время», «Последние 10», «Выигрыши» и «Поражения». Это делает панель диагностическая, а не просто декоративная. Если время жизни и последние 10 совпадают, привычка, вероятно, устойчива. Если выигрыши и проигрыши расходятся, метрика может быть зависит от результата или привязан к состоянию игры, а не к простой привычке.

В области графика используется сдержанный темный фон, золотая линия тренда и цветные маркеры для четырех срезов. В аннотации тренда явно указано, эталонные отношения растут или падают в зависимости от Эло. Это важно, потому что некоторые показатели имеют принцип «чем выше, тем лучше», например, разработка мелких деталей, в то время как другие - чем меньше, тем лучше, например, раннее движение ферзя. Карта поэтому не просит пользователя определить направление только по уклону. Финал Строки «Что это значит» и «Совет тренера» переводят результаты теста. в конкретную шахматную инструкцию, сохраняя при этом эмпирический контекст.

На рисунке 11 показан процентиль. представление, используемое для показателей, исходный показатель которых плохо представлен единая глобальная тенденция. Выигрышные вилки — тактический пример: уровень населения Количество конвертированных вилок меняется в зависимости как от навыков игрока, так и от частоты ошибок противника. Таким образом, карта сравнивает ценность пользователя с историями одноранговых игроков в сопоставлено ведро Эло. На диаграмме показано распределение одноранговых узлов, включая межквартильный размах, медиана и усы. Маркер пользователя наносится на той же шкалы и помечены как исходным значением, так и процентильным рангом. Соседний сегменты сравнения остаются видимыми в приглушенном виде, чтобы пользователь мог видеть, что одноранговый сегмент имеет локальный, а не универсальный масштаб.

Такая конструкция делает процентильные показатели проверяемыми. Пользователь видит необработанное значение, процентиль, одноранговая группа, используемая для сравнения, и переведенный Сопоставление сегментов Chess.com с Lichess. То же самое время жизни, последние 10, победы и Строки потерь сохраняются, поэтому можно проверить тактические проценты. для эффектов новизны и зависимости от результата.

15 Заключение

Отчеты Elo+Chess построены на основе большого расчета состояния перемещения стратифицированные образцы Lichess. Каждая метрика начинается с набора конкретных состояний доски. и переменные состояния движения, становятся особенностью игры игрока, а затем усредняются. по корзине Эло для формирования эмпирической эталонной кривой. Принцип открытия примеры показывают, почему этот подход полезен: простые шахматные привычки имеют сильные, видимые связи с рейтингом в целевом диапазоне, а те После верхнего предела

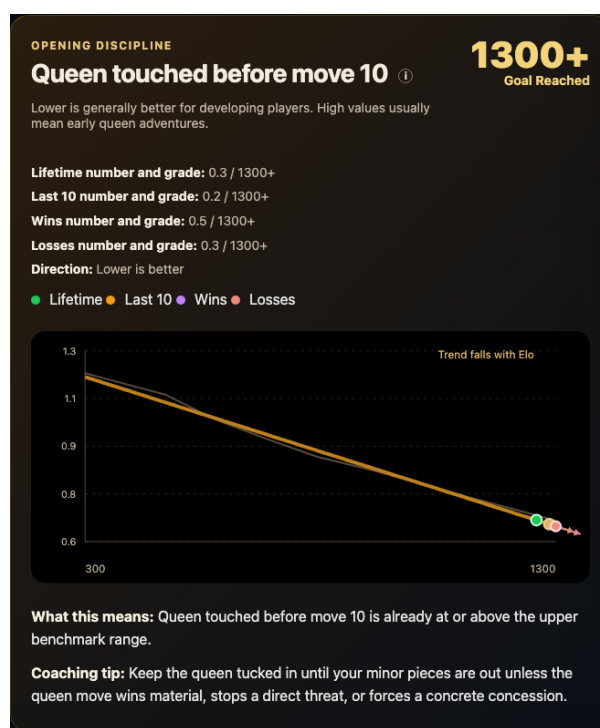
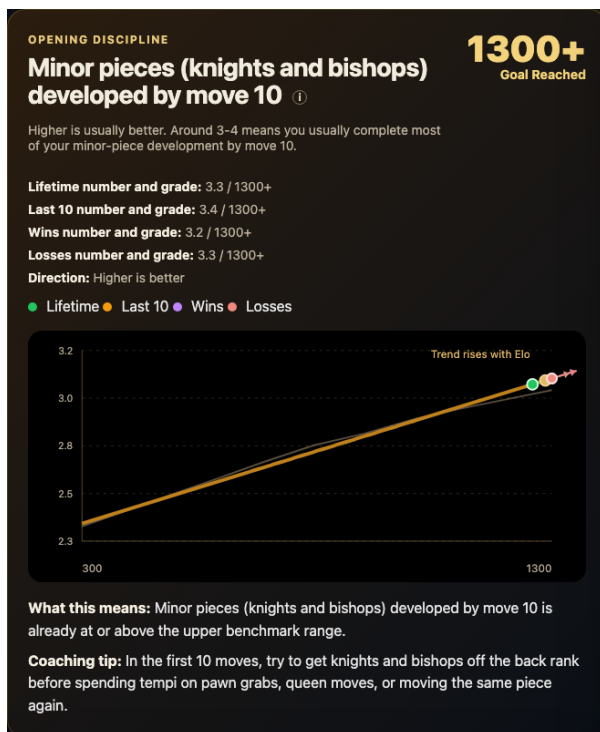


Figure 10: Производство панелей открытого типа. Та же визуальная грамматика обрабатывает метрика «чем выше, тем лучше», разработка второстепенных частей и метрика «чем ниже, тем лучше», раннее движение ферзя. Обе панели отображают время жизни, последние 10, победы и поражения. чтобы в отчете можно было отделить устойчивые привычки от недавних или зависящих от результата. узоры.

отношения часто выравниваются или становятся более условными. Это основная причина, по которой Elo+Chess фокусирует свои тренерские претензии на рейтинге. диапазон, где данные подтверждают их наиболее явно.

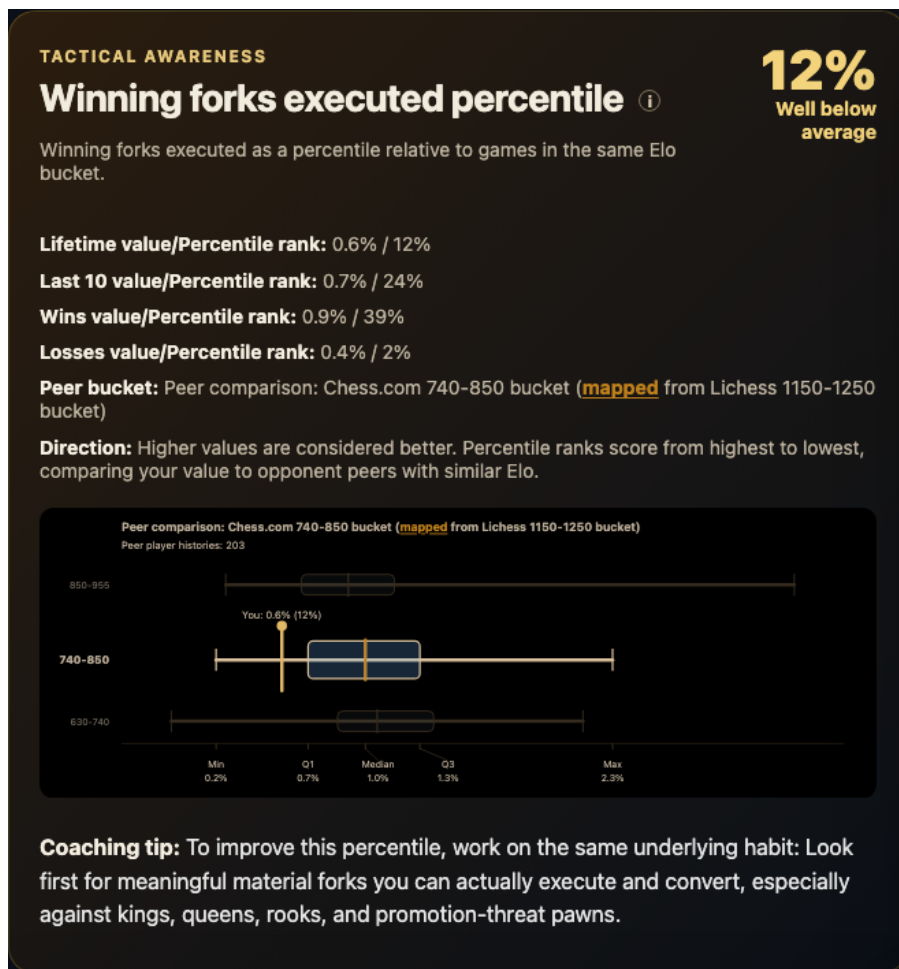


Figure 11: Производственная процентильная панель для выигрышных вилок выполнена. Сюжет показывает распределение аналогов по одному и тому же сегменту, использованное в отчете, а золотой Маркер показывает исходное значение пользователя и процентильный рейтинг.