

Tworzenie warstwowych krzywych wzorcowych Lichess dla Elo+Chess

Notatki badawcze Elo+Chess

31 maja 2026 r

Abstract

Raporty porównawcze Elo+Chess są tworzone na podstawie dużych warstwowych próbek Gry Lichess.org, z próbką porównawczą wykorzystaną w tej nocy, zawierającą 483 974 gry typu bullet 1+0, 479 974 gry błyskawiczne 3+0, 925 636 gry szybkie 10+0, i 660 540 dłuższych szybkich gier. Dla każdego głównego typu gier potok się rozszerza w każdej grze stan plan-szy przed i po ruchu, oblicza setki zmienne pośrednie i zmienne gry gracza, agreguje te zmienne metryki coachingowe, a następnie uśrednia te metryki za pomocą wiadra Lichess Elo. The próbka jest celowo stratyfikowana, a nie proporcjonalna do surowego odtwarzacza pula: celem jest oszacowanie stabilnych krzywych zachowania na każdym poziomie ratingu, łącznie z wiadrami, które w przeciwnym razie byłyby rzadkie, a nie w celu odtworzenia naturalnego rozkład ocen. Rezultatem jest rodzina empirycznych krzywych wzorcowych: dla dla każdej metryki możemy zapytać, jak często gracze na różnych poziomach Elo wykazują daną metrykę nawyk lub zasada w prawdziwych grach. W notatce opisano metodę, skalę obliczenia i kilka przykłady zasady początkowej obejmujące roszady, rozwój drobnych elementów, powtarzane ruchy otwierające i wczesny ruch hetmana. Nacisk jest metodologiczne: Elo+Chess nie zakłada, że zasady szachowe są uniwersalne liniowy na wszystkich poziomach umiejętności. Zamiast tego sprawdzane są krzywe wzorcowe zakresie ocen, a produkt koncentruje się na regionie od początkujących do wczesno zaawansowanych gdzie relacje są najsilniejsze i najłatwiejsze do interpretacji.

1 Zamiar

Celem systemu wzorcowego jest przekształcenie ogólnych porad szachowych w mierzalne ilości. Stwierdzenie takie jak „rozwijaj swoje mniejsze elementy” jest przydatne, ale staje się bardziej praktyczne, gdy możemy powiedzieć, ile mniejszych elementów a Gracz zwykle rozwija się w ruchu 10, jak to wypada w porównaniu z rówieśnikami w tym samym segment ratingowy i jak ten nawyk zmienia się w całym spektrum ratingowym.

Dlatego Elo+Chess tworzy bibliotekę metryczną na podstawie rzeczywistych gier. Każda metryka jest obliczana dla każdego wiersza gry gracza, podsumowane według typu gry i segmentu ELO, oraz wykorzystana w raporcie jako krzywa odniesienia. Raport może następnie umieścić użytkownika tę samą krzywą i pokazują, czy użytkownik znajduje się powyżej, poniżej czy blisko zachowania obserwowane wśród porównywalnych graczy.

2 Dane źródłowe i stratyfikacja

Źródłem testów porównawczych jest publiczna baza danych gier Lichess.org. Elo+Chess wykorzystuje raczej duże stratyfikowane próbki według rodzaju gry i kategorii ocen, a nie małe próbki

wygody. Obecne artefakty środowiska wykonawczego obejmują cztery główne kategorie gier:

- Pocisk 1+0,
- Błyskawica 3+0,
- 10+0 szybki,
- szybkie gry dłuższe niż 10 minut.

Celem stratyfikacji jest szeroki zakres poziomów oceny. Losowość proporcjonalna próbka byłaby skuteczna do opisania ogólnej puli graczy, ale tak by było być nieefektywne przy tworzeniu krzywych wzorcowych. Najpopularniejsze segmenty ocen dominowałyby w próbie, podczas gdy łyżki tylne o niskim wsparciu byłyby ich zbyt mało gry dla stabilnych średnich metrycznych. Dlatego Elo+Chess pobiera próbki z przeciętnej gry Wiadro Elo. W obecnej konstrukcji czystej próbki surowe gry są na pierwszym miejscu filtrowane do ważnych ocenianych gier ze znanymi graczami, znanymi ocenami, ważnymi wynikami, i znane kontrole czasu. Gry są następnie filtrowane do graczy, którzy mają wystarczającą liczbę gier przy tej prędkości i wystarczająco stabilnych ocenach, aby reprezentować spójny poziom umiejętności: dla każdego gracza Elo+Chess oblicza zakres prędkości gracza ($\max(\text{Elo}) - \min(\text{Elo})$) w puli źródłowej i utrzymuje gry tylko wtedy, gdy obaj gracze mają co najmniej skonfigurowaną minimalną liczbę gier i nie więcej niż 300 punktów Elo driftu w ramach prędkości. Filtr dryfu nie jest tu przeznaczony w celu usunięcia normalnej poprawy lub odchylenia; zmniejsza liczbę przypadków, w których gracz gry łączyłyby zasadniczo różne poziomy umiejętności, takie jak konta tymczasowe, powracający gracze lub konta, które w trakcie szybko przechodziły przez kilka segmentów próbkowany okres. Wreszcie, kwalifikujące się gry są klasyfikowane w ramach każdego średniego Elo wiadro przez deterministyczny skrót identyfikatora gry i materiału siewnego oraz ograniczoną liczbę gry są zachowywane z każdego segmentu.

Dzięki takiemu rozwiązaniu krzywe wzorcowe są znacznie bardziej przydatne. Każde wiadro Elo wnosi wystarczającą liczbę obserwacji, aby oszacować średnią wartość każdego miernika, a rzadkie wiadra o wysokiej lub niskiej wartości nie są zagłuszane przez naturalny kształt dystrybucji dla aktywnych graczy. Krzywe wzorcowe używane przez Elo+Chess to następnie ograniczone do zakresu najbardziej odpowiedniego dla produktu, mniej więcej dla początkujących poprzez wczesną zaawansowaną grę. W raportach w skali Lichess jest to zakres podstawowy wynosi 600-1600. Gdy użytkownik przegląda raport w skali Chess.com, plik odpowiednie segmenty ratingów są mapowane na ten sam bazowy Lichess segmenty testów porównawczych przy użyciu oddzielnej metody mapowania ocen między platformami opisano w artykule towarzyszącym: [Estimating Cross-Platform Chess Rating Mappings with Regresja modalna](#).

3 Rozkład ocen aktywnych graczy

Jeśli spojrzymy na rozkład ocen Elo graczy Lichess.org na koniec z marca 2026 r., według głównego typu gry, widzimy następujące liczby i percentyle. W tym widoku rozkładu każdy gracz jest liczony raz dla danego typu gry, używając ostatniej zaobserwowanej oceny tego gracza w odpowiednim surowym formacie rzędy kontroli czasu, po wymaganiu co najmniej pięciu gier w tej samej grze typ.¹

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 pokaż rozkłady z i bez tego filtra.

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: Ostatni zaobserwowany dokładny rozkład ocen aktywnego gracza Lichess typ gry po wymaganiu co najmniej pięciu gier w tym konkretnym typie gry.

Rozkładu tego nie należy mylić ze warstwową próbą porównawczą. Dystrybucja opisuje aktywną pulę graczy Lichess obserwowaną w stanie surowym skanuj. Próbką wzorcowa jest następnie celowo stratyfikowana, tak aby Elo czerpało z niej korzyści mają wystarczające wsparcie dla stabilnych krzywych metrycznych.

Aby zmapować ocenę Elo Lichess.org dla określonego typu gry na odpowiednik oceny Chess.com, Elo+Chess używa modalu specyficznego dla typu gry równania regresji oszacowane w dodatku [Elo+Chess rating-mapping paper](#):

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

Te równania odwzorowują wartości ocen, a nie rangi percentylowe. Rangi percentylowe są zdefiniowane w określonej puli graczy, a odtwarzacze Chess.com i Lichess baseny mają różne kształty. W tabeli 2 dla 10-minutowy typ szybkiej gry, pierwsza kolumna zawiera małą próbkę możliwych Chess.com Wartości znamionowe Elo. Druga kolumna podaje odpowiedni Ranga percentylowa Chess.com wyświetlana w Chess.com na dzień 31 maja 2026 r. A 71,7 pozycja w Lichess.org w 10-minutowym trybie szybkim odpowiada ocenie Elo Lichess wynoszącej 1,644² i ta wartość jest pokazana w trzecia kolumna. Jeśli zastosujemy równanie odwzorowania Chess.com–Lichess, 1644 na Lichess odpowiada około 1271 w Chess.com. Zakładając szacunkowo równanie mapowania jest rozsądne, różnica +522 pomiędzy wartością znamionową Chess.com w kolumnę 1 i przyporządkowaną ocenę w kolumnie 4 można wytłumaczyć dużą liczbą przypadkowych graczy w puli graczy Chess.com zawyżających Chess.com ranga percentylowa. Innymi słowy, bycie w 72. percentylu na Lichess.org to to większe osiągnięcie niż znalezienie się na 72. percentylu w Chess.com, ponieważ rangi percentylowe nie są w sposób przejrzysty mapowane z jednego systemu na drugi. Prawdziwy Elo wartości można mapować za pomocą równań oszacowanych na podstawie wyników tego samego gracza oba systemy.

Przyczyną luk w Tabeli 2 jest to, że mapowanie szacuje równoważne oceny siły gry, podczas gdy percentyl opisuje pozycję w populacji jednej platformy. Gracz z 71,7 percentylem na boisku Dlatego Chess.com nie jest automatycznie graczem na poziomie 71,7 percentyla na Lichess, i odwrotnie. Zaobserwowane liczby są zgodne z opinią publiczną Chess.com Rapid Pool mający znacznie większą masę graczy o niższych ocenach lub zwykłych graczy, podczas gdy Aktywna, dokładna kontrola puli Lichess jest bardziej skoncentrowana wśród doświadczonych powtórzeń gracze. Kluczową kwestią nie jest to, że jedna skala percentylowa jest „właściwa”, a inny jest „zły”;

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.

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: Przykładowe porównanie 10-minutowego szybkiego percentyla Chess.com punktów i te same punkty percentyla w aktywnym szybkim Lichess 10+0 dystrybucja po wymaganiu co najmniej pięciu gier 10+0. "Mapowane Chess.com odpowiednik" dotyczy bieżącego Elo+Chess Szybka konwersja skali ocen 10+0 na ocenę Lichess o tym samym percentylu. Uwzględniono rzędy z najwyższym ogonem, aby pokazać rozróżnienie skali, ale dopasowane linie ocen jest najbardziej niezawodna w głównym zakresie coachingowym.

chodzi o to, że rangi percentylowej nie można przenosić między platformami chyba że podstawowe rozkłady graczy mają ten sam kształt.

Note on the 1500 spike. The visible jump near Wartość 1500 jest obecna w danych, a nie utworzona w wyniku wygładzania. W najnowszym zaobserwowany rozkład ocen, dokładnie 1500 ocen stanowi 13,1 gracze typu bullet, 5,6 17,6

początkowa lub tymczasowa kotwica w systemie ratingowym.

Skok wynoszący dokładnie 1500 jest w dużej mierze artefaktem słabo aktywnych kont. Do sprawdź to, po zażądaniu ponownie obliczyliśmy te same rozkłady najnowszych ocen aby każdy gracz rozegrał co najmniej pięć partii dokładnie tego samego rodzaju. Skok prawie znika pod tym filtrem: dokładne 1500 ocen spada z 5,6–17,6 graczy w dystrybucji all-player do 0,2–1,2 dystrybucja podana w tabeli [1](#).

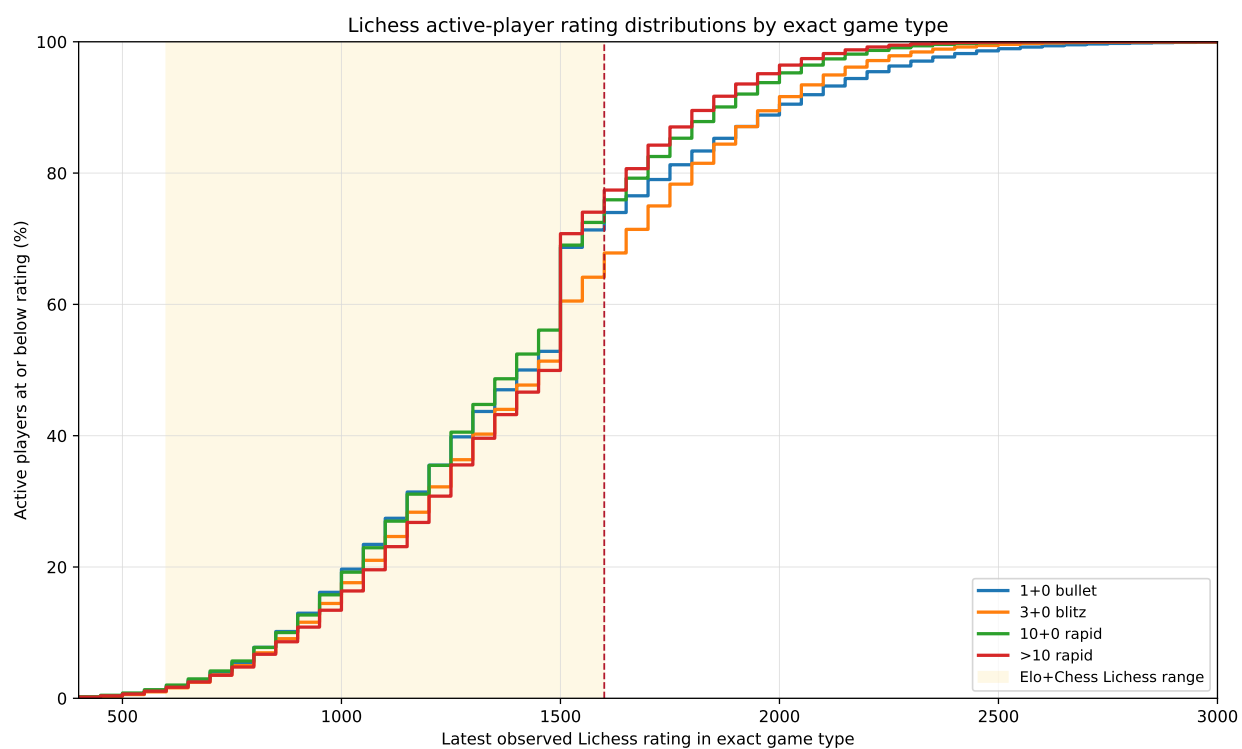


Figure 1: Skumulowany rozkład ocen aktywnych graczy według dokładnego typu gry. The żółty pasek oznacza zakres 600-1600 Lichess używany przez główny Elo+Chess benchmarki coachin-gowe.

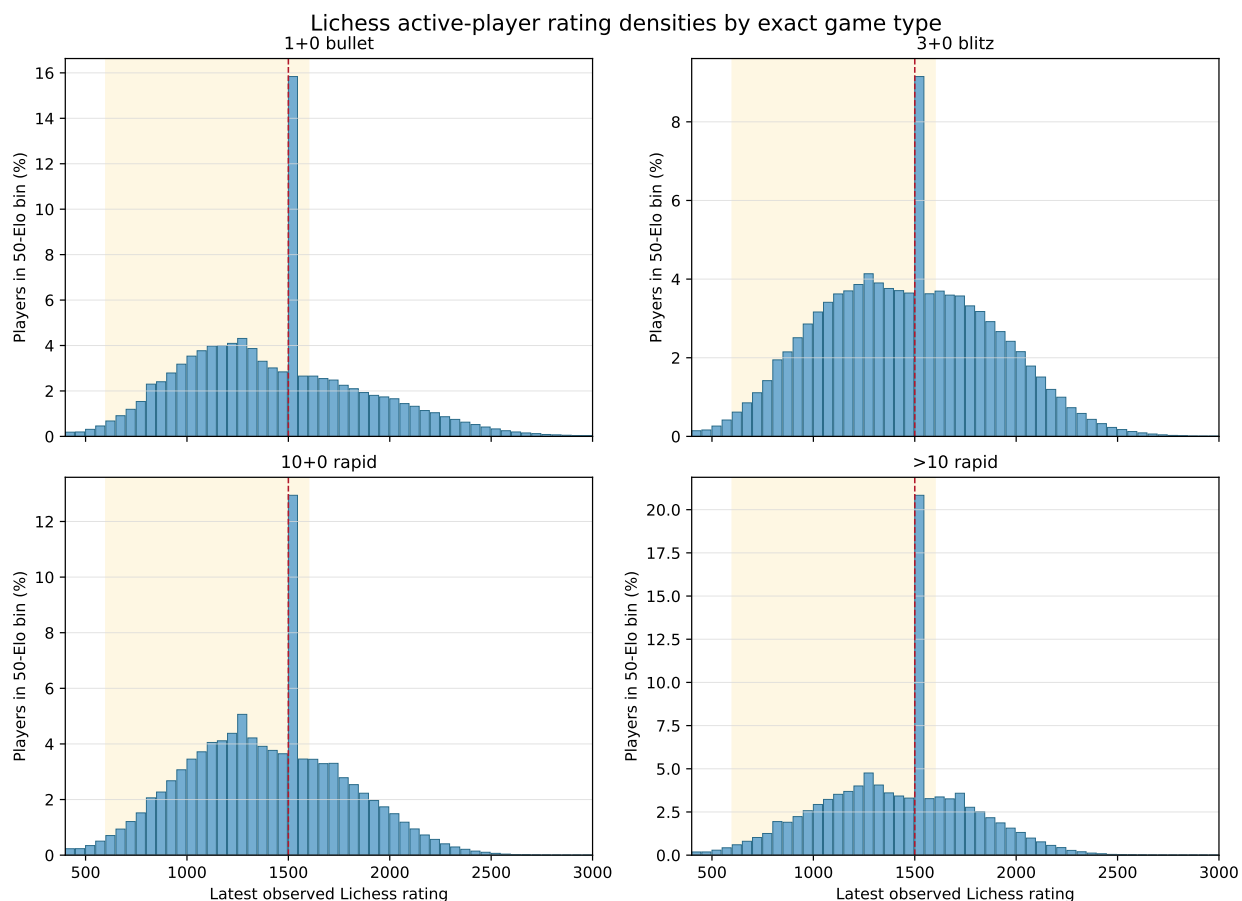


Figure 2: Widok gęstości tych samych rozkładów ocen aktywnych graczy. The linia przerywana oznacza 1500 Lichess, gdzie znajdują się surowe rozkłady najnowszych ocen duży, dokładny punkt masy. Żółty pasek oznacza 600-1600 Lichess zakres używany przez główne raporty metryczne Elo+Chess. Kolec jest najbardziej widoczny w zagęszczeniu wszystkich graczy, ponieważ wiele mniej aktywnych kont pozostaje dokładnie na tym samym poziomie 1500; Rysunek 4 powtarza wykres gęstości po wymaganiu co najmniej pięciu gier dokładnie tego typu, co w dużej mierze eliminuje ten artefakt.

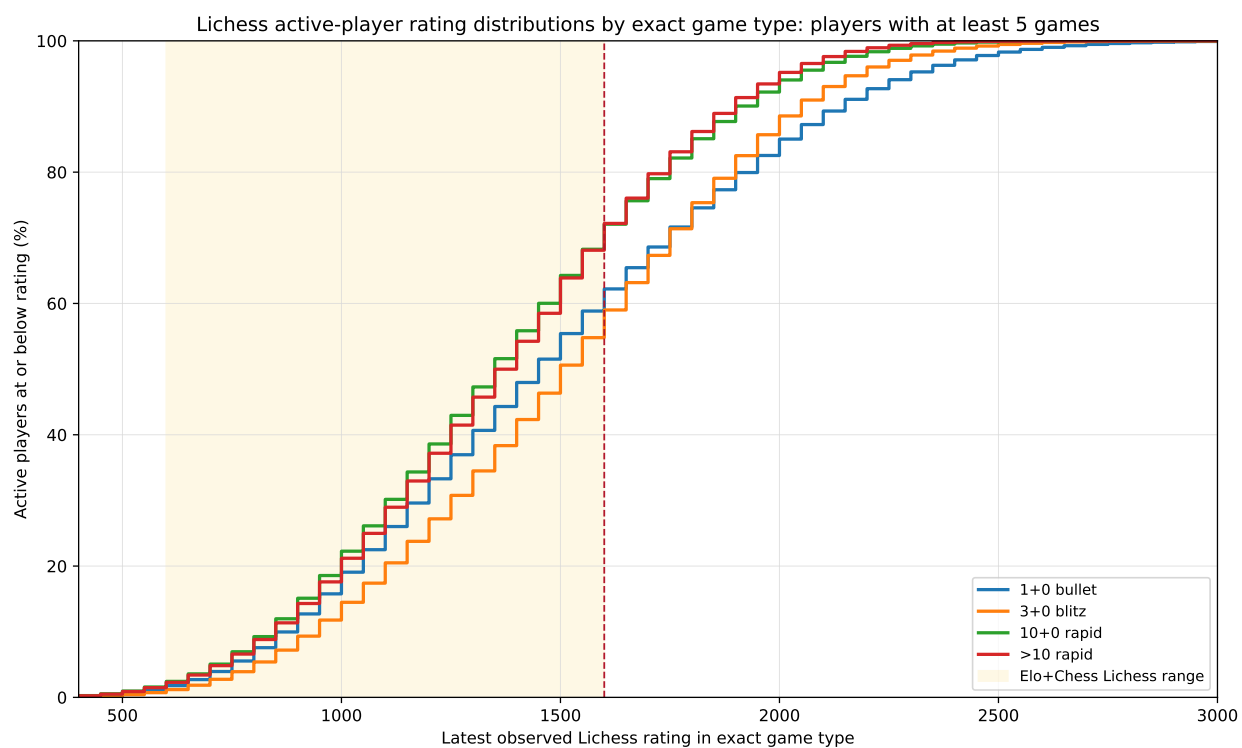


Figure 3: Skumulowany rozkład ocen aktywnych graczy po wymogu at co najmniej pięć gier dokładnie tego typu. Skok na 1500 metrów jest znacznie zmniejszony.

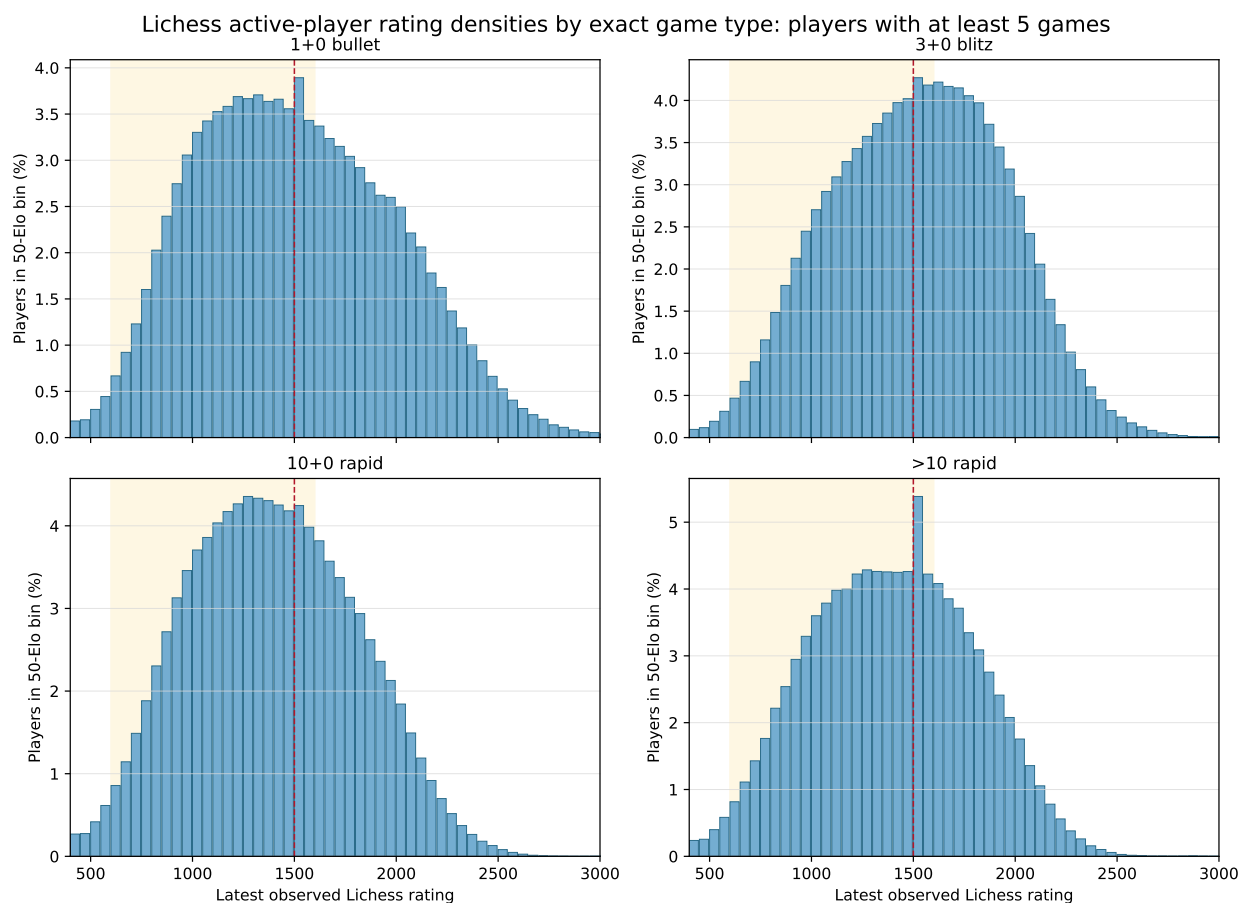


Figure 4: Widok gęstości po wymaganiu co najmniej pięciu gier w danej grze wpis. Pogląd ten potwierdza, że duży punkt masy 1500 w trybie all-player dystrybucja opiera się głównie na mało aktywnych kontach. Znaki żółtego paska zakres 600-1600 Lichess używany w głównych raportach metrycznych Elo+Chess.

4 Skala bieżących artefaktów wzorcowych

Obliczenia są celowo duże. To nie jest zadrapanie profilu, małe zestaw pobranych PGN lub ręcznie wybrany zbiór pouczających przykładów. Początkowa przepustka Lichess obejmowała pełne pliki miesięczne od stycznia 2025 r. do marca 2026 r. Niestratyfikowana tabela skanowania zawiera 1 382 178 087 gier i 2 764 356 174 wierszy ocen graczy. W tych rzędach 6 623 438 unikalnych nazwy użytkowników graczy pojawiają się przynajmniej raz.

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: Niestratyfikowana skala skanowania Lichess przed próbkowaniem porównawczym. Skan obejmuje pełne miesięczne pliki historii gier od stycznia 2025 r. do marca 2026.

Artefakty testów porównawczych produkcji są znacznie mniejsze niż surowy skan, ponieważ są wybierane do konstrukcji metrycznej. Nadal są duże: punkt odniesienia Etap to potok ekspansji i agregacji stanu przeniesienia obejmujący setki tysiące gier każdego typu.

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: Skala warstwowych artefaktów wzorcowych po dostosowaniu do wspólny schemat funkcji środowiska wykonawczego składający się z 277 kolumn. A „rząd-gra” oznacza jedną grę z perspektywy jednego gracza, a więc większość gier utwór dwa wiersze gier gracza.

Tylko w przypadku trybu 3+0 blitz bieżący test porównawczy wykorzystuje 479 974 gier z udziałem 129 398 unikalnych graczy i 28 138 904 różnych rzędów stanów ruchu. The wyrównana tabela funkcji gry błyskawicznej zawiera 277 zmiennych pochodnych. The Artefakty szybkiego tworzenia funkcji są jeszcze szersze na etapie stanu ruchu: pełne Tabela stanów szybkiego ruchu używana podczas konstruowania funkcji zawiera 168 kolumny stanu ruchu przed agregacją w 277-kolumnową funkcję gry gracza stół. Artefakty testów porównawczych środowiska wykonawczego frontendu dla wszystkich czterech typów gier to: następnie dostosowane do tego samego 277-kolumnowego schematu funkcji gry gracza, w tym rozszerzone zmienne podziału taktyki i taksonomii. Z tych tabel funkcji wynika, że raport wybiera i wyświetla metryki coachingowe, które są najbardziej zinterpretowane.

Liczby te są ważne dla interpretacji wyników. Krzywe wzorcowe nie są szacunkami na podstawie kilkuset ręcznie sprawdzonych gier. Zamiast tego oni pochodzą z dziesiątek milionów ocenianych państw zarządu i z ponad czterech milionów wierszy gier graczy w czterech bieżących kategoriach gier produkcyjnych.

5 Przegląd rurociagu

Wzorcowy rurociąg budowy składa się z sześciu głównych etapów:

1. Wybierz oceniane gry Lichess w docelowym typie gry i warstwach rankingowych.
2. Analizuj każdy ruch w grze, ruch po ruchu i zapisz ruch przed i po ruchu stan zarządu.
3. Oblicz pośrednie zmienne stanu ruchu z planszy, ruchu, perspektywa gracza i perspektywa przeciwnika.
4. Agreguj zmienne stanu ruchu w zmienne funkcji gry gracza.
5. Konwertuj zmienne gry gracza na metryki raportu za pomocą narzędzia Clear kierunek interpretacji: wyższy jest zwykle lepszy, niższy jest zwykle lepiej lub opisowo.
6. Uśrednij każdą metrykę za pomocą wiadra Elo i ręcznie sprawdź wynik relacja.

Ta struktura jest ważna, ponieważ wiele metryk raportów nie jest bezpośrednich widoczne w nagłówku PGN. Wymagają zrekonstruowania stanowiska, ustalenia które figury przesunęły się z jakich pól początkowych, porównując przed ruchem i po przesunięciu materiału i struktury, a następnie agregowanie wyników z przenies poziom na poziom gry gracza.

6 Zmienne pośrednie

Zmienne pośrednie mają na celu opisanie tego, co zmieniło się na planszy i dlaczego jest to istotne z punktu widzenia osoby zgłaszającej. Przykłady obejmują:

- numer ruchu, numer ruchu gracza, strona do ruchu, kolor poruszającego się i gracz kolor;
- z kwadratu, do kwadratu, typ ruchomej figury, typ zbitej figury i czy posunięciem było zabicie, szach, zamek, awans czy mat;
- bilans materiałowy przed i po przesunięciu;
- czy mniejsza figura opuściła swoje pole startowe o podany numer ruchu;
- czy ta sama figura poruszyła się w kolejnych ruchach gracza otwór;
- czy hetman poruszył się przed ruchem 10;
- strona roszady, numer ruchu roszady i to, czy wieże zostały połączone po roszowie;
- figury luźne, figury wiszące, pionki zdublowane, pionki izolowane, spasowane pionki i pionki do tyłu;
- status pliku rook, pliki otwarte, pliki półotwarte i bezpieczeństwo króla cechy;
- możliwości taktyczne, taktyka wykonana, taktyka nacisku materiałowego, gwarantowane dalsze zyski i zrealizowane zyski.

Dokładny zestaw pośredni różni się w zależności od etapu budowy. Przechowywane są bazy danych raportów wykonawczych tylko kolumny potrzebne do szybkiego udostępniania raportów, podczas gdy pełna kompilacja funkcji artefakty obejmują szersze tabele stanów przenoszenia i poprawki funkcji używane do konstruowania ostatnie wiersze funkcji gry gracza.

7 Przykłady metryk zasady otwarcia

Zasady otwierania są użytecznymi przykładami, ponieważ są łatwe dla użytkowników zrozumiałe i łatwe do podłączenia do obliczeń stanu ruchu.

7.1 Minor Pieces Developed by Move 10

Ta metryka określa, ilu z czterech rycerzy i gońców pozostało graczowi swoje pola domowe do 10. ruchu gracza. Kalkulacja pośrednia używa koloru poruszającego się, rodzaju ruchomego elementu, pola startowego i ruchu gracza numer. Dla białych kwadratami głównymi są b1, g1, c1 i f1. Dla Blacków tak b8, g8, c8 i f8. Dla każdej gry gracza Elo+Chess zlicza odrębny dom pola pomiędzy rycerzami i gońcami przesunięte przez ruch gracza 10.

Jest to wskaźnik „wyższy znaczy zwykle lepszy”. Nie twierdzi, że każdy nieletni Kawałek musi zawsze poruszać się w ruchu 10, ale wśród rozwijających się graczy jest to empiryczne zależność jest jasna: gracze, którzy osiągają wyższe poziomy Elo, mają tendencję do rozwoju więcej mniejszych kawałków wcześniej.

7.2 Repeated Piece Moves in the Opening

Ta metryka uwzględnia zamiast tego dodatkowe wczesne ruchy wydane na ponowne przeniesienie tej samej figury opracowania czegoś nowego. Obliczenie pośrednie wykorzystuje poprzednie pole docelowe, obecne pole początkowe i numer ruchu gracza. Jeśli ten sam element został przesunięty podczas ruchu poprzedniego gracza i ponownie przesunięty w obrębie pierwszych 10 ruchów gracza, liczba powtórzonych ruchów wzrasta.

Jest to wskaźnik „niższy” oznacza „zwykle lepszy”. Mierzy zmarnowane tempo otwarcia w a prosty sposób. Niektóre powtarzające się ruchy są uzasadnione taktycznie, ale w skali Krzywa wiadra oddaje ogólną tendencję: silniejsi gracze u celu Poziomy wydają mniej wczesnych ruchów na wielokrotne przesuwanie tej samej figury.

7.3 Queen Touched Before Move 10

Ta metryka liczy ruchy hetmana przed ruchem gracza 10. Średnio zaawansowany obliczenia wykorzystują typ ruchomego elementu i numer ruchu gracza. Jeśli ruchomy element jest królową, a liczba ruchów gracza wynosi maksymalnie 10, liczba wzrasta.

Jest to również wartość niższa-zwykle-lepsza dla zakresu docelowego Elo+Chess. Wcześniej ruchy hetmana mogą zyskać ustępstwa materialne lub siłowe, ale wielu rozwijających się graczy wyprowadzić królową zbyt wcześnie, stracić tempo w wyniku ataków na królową i opóźnić rozwój drobnych elementów i bezpieczeństwo króla.

7.4 Castling and King Safety

Metryki roszady wykorzystują flagi roszady, stronę roszady, numer ruchu roszady i stan planszy po roszowaniu. Przykładami mogą być: czy gracz wykonał roszadę, czy gracz roszował w ruchu 10, czy gracz roszował na skrzydle królewskim, czy też Queenside, czy wieże zostały połączone po roszadzie i czy pliki wież po roszczeniu były otwarte lub półotwarte.

Nie wszystkie te wskaźniki są interpretowane w ten sam sposób. Zakończenie roszady i roszady w ruchu 10 są na ogół wyższe-zwykle-lepsze w docelowym zakresie. Roszada Queenside lub roszada do otwartych plików może być bardziej zależna od kontekstu. Właśnie dlatego Elo+Chess

nie ogranicza się po prostu do zakodowania na stałe zasady moralnej; buduje wiadro krzywe i sprawdza, które relacje są wystarczająco stabilne dla coachingu.

8 Krzywe łyżek i kontrola ręczna

Po obliczeniu wskaźników gry gracza Elo+Chess grupuje je według Lichess Elo wiadro i oblicza średnie wiadra. Średnia koszyka jest surową wartością empiryczną punkt odniesienia dla tego wskaźnika. Następnie w raporcie używana jest linia trendu przypisz oceny w stylu Elo do wartości użytkownika.

Krzywe są sprawdzane ręcznie przed użyciem. Metryka jest bardziej przydatna gdy środki segmentu wykazują spójną zależność w całym zakresie ocen podawane przez produkt. Metryka jest mniej przydatna, gdy jest zaszumiona, płaska i bardzo wysoka niemonotoniczne lub kierowane specjalnymi przypadkami, które są trudne do wyjaśnienia użytkownikowi.

Ten etap ręcznej kontroli jest szczególnie ważny w przypadku górnej części Elo+Chess odcięcie. Silniejsi gracze wiedzą, kiedy złamać zasady otwarcia i to robią skutecznie. Powyżej zakresu od początkującego do wczesnego zaawansowanego, relacja między prostymi nawykami a oceną często się spłaszcza lub zmienia kształt. Produkt dlatego podkreśla zakres, w którym zależności empiryczne są silne i pedagogicznie przydatne.

9 Przykładowe krzywe otwarcia w pełnym zakresie

Rysunek 5 przedstawia cztery krzywe otwarcia błyskawicznego 3+0. Szare punkty pokazują szerszy zakres wartości znamionowych Lichess. Niebieskie punkty i niebieski trend linia pokazuje gamę 600-1600 Lichess używaną przez główny trener Elo+Chess punkty odniesienia. Przerywana czerwona linia oznacza górną granicę.

Ten sam schemat jest widoczny w pełnym zakresie 10+0 Rapid i dłużej-szybkim artefakty. Centralna decyzja dotycząca modelowania pozostaje niezmienną: użyj empirycznie stabilny zakres roszczeń coachingowych i unikaj udawania, że jest to prosta krzywa nawyku pozostaje na zawsze równie pouczający.

Ten sam wzór pojawia się numerycznie w tabeli 5. Stoki są raportowana jako zmiana wartości metryki na 100 punktów Elo. Roszada, drobny element rozwój, powtarzalne ruchy figur i wczesny ruch hetmanów są wyraźnie widoczne ruch w zakresie 600-1600. Powyżej 1600 zbiega się w nich znacznie mniejsze przykładów, co jest zgodne z koncepcją, że gracze z wyższymi ocenami psują się podstawowych zasad w sposób bardziej selektywny i skuteczniejszy. Drugi mechanizm to nasycenie. W przypadku niektórych nawyków metryka sięga praktycznego pułapu lub słodczy spot: silni gracze mają już przyzwyczajenie, więc nie ma powodu krzywą, aby nadal gwałtownie przesuwając się w górę. Więcej też nie zawsze znaczy lepiej później ten punkt. Dane spłaszczają zatem oba rynki, ponieważ silniejsi gracze mogą się przełamać zasady na konkretnych stanowiskach i ponieważ wykształciło się już wiele podstawowych nawyków ich użyteczny zakres działania.

10 Dlaczego górna wartość odcięcia ma znaczenie

Docelowy użytkownik Elo+Chess to gracz od początkującego do średnio zaawansowanego. Dla tych graczy, podstawowe nawyki są silnie powiązane z oceną: rozwijaj figury, unikaj powtarzających się wczesnych ruchów figurami, unikaj niepotrzebnych wczesnych ruchów hetmana, zamku z czasem połącz wieże, zarządzaj materiałem i unikaj pozostawiania luźnych elementów lub wiszące.

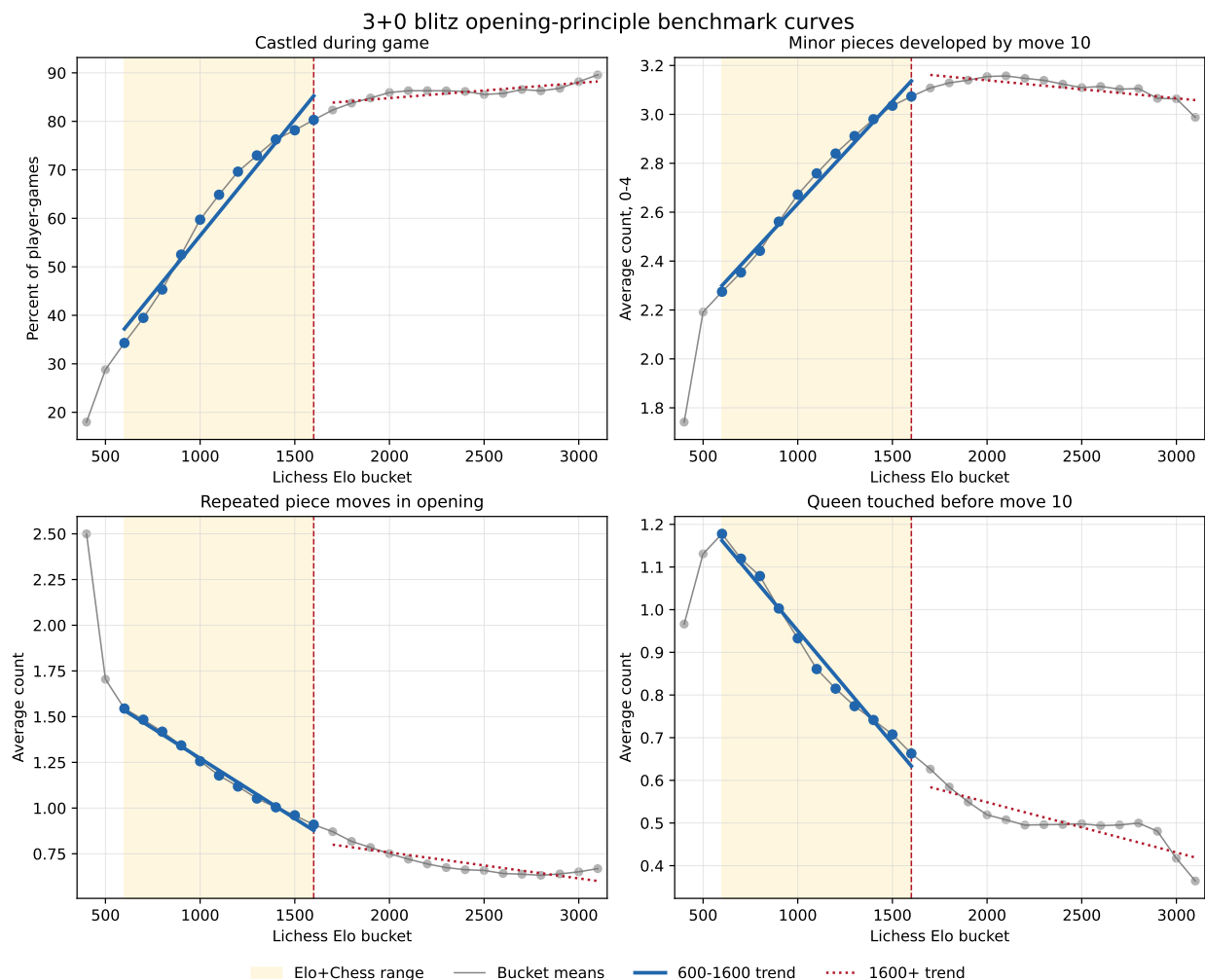


Figure 5: Wybrane krzywe wzorcowe zasady otwarcia dla ataku 3+0. The relacje są najsilniejsze i najbardziej przydatne w ofercie coachingowej Elo+Chess. Powyżej górnej granicy kilka zależności ulega znacznemu spłaszczeniu.

Przy wyższych ocenach te same powierzchowne nawyki stają się mniej diagnostyczne. Silny gracz może opóźnić roszowanie, ponieważ centrum jest zamknięte, przesunąć hetmana wcześniej ponieważ pozycja zawiera konkretną taktykę, lub przesunąć tę samą figurę dwa razy ponieważ wygrywa tempo lub wymusza strukturalne ustępstwo. Nadal surowy nawyk ma znaczenie, ale związek między nawykiem a oceną staje się coraz większy warunkowy. Właśnie dlatego Elo+Chess wykorzystuje górną wartość odcięcia dla głównego coachingu twierdzi i dlatego użytkownicy powyżej tego zakresu są ostrzegani, że istnieje ścisły związek między zasadniczymi nawykami a oceną załamuje się.

11 Dwustronne wskaźniki taktyczne i cele procentowe

Niektóre metryki mają kształt odwrotny do przykładów z zasadą otwarcia. Widelce są użytecznym przykładem, ponieważ wydarzenie jest z natury dwustronne. Gracz, który znajduje i konwertuje

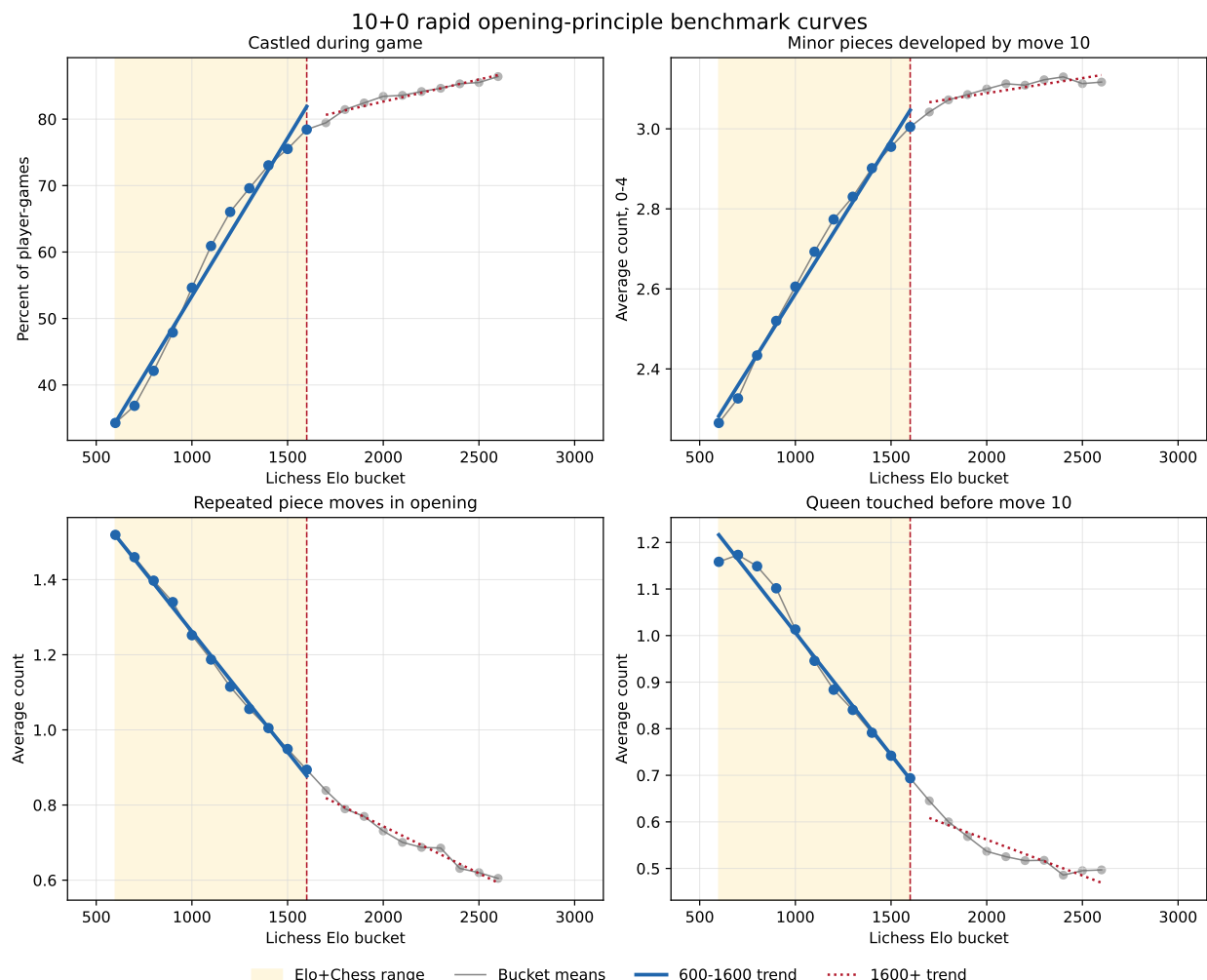


Figure 6: Wybrane krzywe porównawcze zasady otwarcia dla szybkich gier 10+0. The relacje są najsilniejsze i najbardziej przydatne w ofercie coachingowej Elo+Chess. Powyżej górnej granicy kilka zależności ulega znacznemu spłaszczeniu.

bardziej przydatne forki niż koledzy, robi coś dobrego, ale wskaźnik populacji przekonwertowanych widelców spada wraz z oceną, ponieważ jest silniejszy przeciwnicy dają mniej okazji do rozwidlenia i bronią się przed zagrożeniami taktycznymi bardziej aktywnie.

Rysunek 8 pokazuje ten efekt. Oba przekonwertowały widelce według raport gracza i skonwertowane widelce przez przeciwnika stają się rzadsze, ponieważ oceny rosną. Nie oznacza to, że wykonywanie forków jest złe. Oznacza to surowe wydarzenie szybkość jest częściowo kontrolowana przez współczynnik błędów przeciwnika i teksturę gry. Dla metryk w ten sposób Elo+Chess używa percentyli historii elementów równorzędnych wewnątrz własnego koszyka Elo gracza, zamiast polegać wyłącznie na globalnym trendzie średnich koszyków.

Aby zbudować te cele percentylowe, Elo+Chess pobiera próbki graczy według typu gry i Wiadro Elo, a następnie wykorzystuje ostatnie gry kwalifikacyjne dla każdego gracza objętego próbą. The bieżący zestaw docelowy wykorzystuje 11 segmentów od 600 do 1600, po 200 graczy na wiadro i do 50 ostatnich gier kwalifikacyjnych na gracza dla każdej produkcji typ gry. Daje to około 2200 próbek graczy na każdy rodzaj gry 110 000 wierszy gier graczy na typ gry dla percentyla

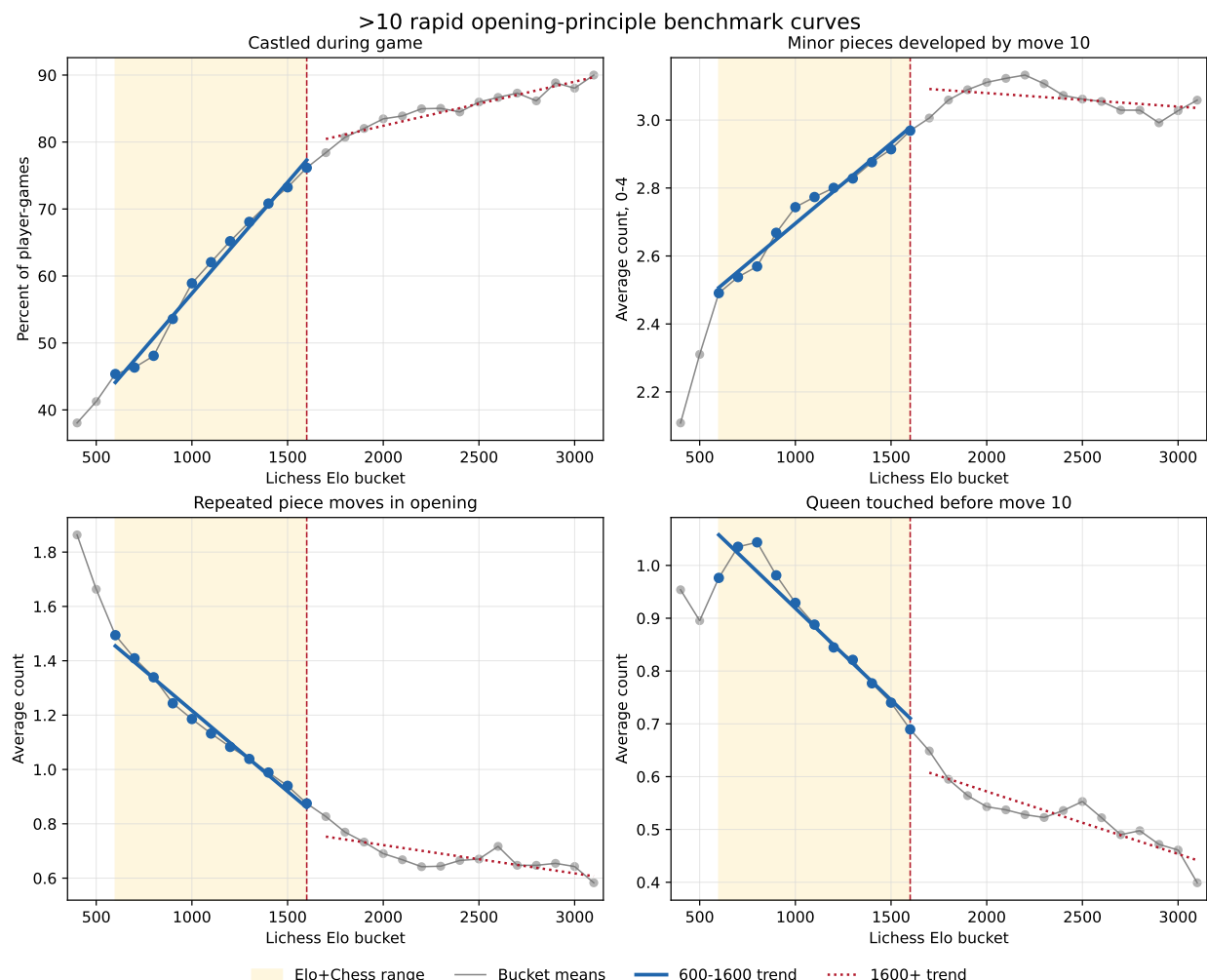


Figure 7: Wybrane krzywe porównawcze zasady otwarcia dla dłuższych, szybkich gier.

mieszczącego się w koszyku dystrybucje.

Rysunek 9 ilustruje wynikowy cel dystrybucja dla współczynnika rozwidlenia przeliczonego przez próbkowanego gracza w błyskawicznym tempie 3+0. Każde pudełko jest rozkład w obrębie koszyka wśród wybranych graczy, a nie średnia w całym koszyku gry indywidualne. W interfejsie raportu wartość użytkownika jest porównywana z odpowiednią pole na zmapowany wiadro Elo użytkownika. Jest to podstawa do stwierdzeń takich jak czy przeliczony współczynnik fork gracza jest wysoki czy niski w porównaniu z innymi graczami na tym samym poziomie, nawet jeśli ogólny wskaźnik zdarzeń w populacji spada wraz z oceną.

12 Koszt replikacji

Metoda jest w zasadzie powtarzalna, ale nie jest lekka. Odbudowa krzywe wzorcowe wymagają:

- dostęp do dużych miesięcznych plików historii gier Lichess;
- warstwowe filtrowanie według typu gry, przedziału ocen i kwalifikowalności zasady;

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: Przybliżone nachylenie krzywej kubelkowej dla czterech przykładowych zasad otwierania wskaźniki dla różnych typów gier produkcyjnych.

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: Bieżące zestawy docelowe percentyli historii równorzędnej. To są próbki historii gracza użyte do oszacowania rozkładów wewnątrz koszyka dla metryki w stylu percentyla.

- parser legalnych ruchów i ewaluator stanu tablicy, który można rozszerzać dziesiątki milionów stanów ruchu;
- logika funkcji otwarcia, materiału, struktury pionków, bezpieczeństwa króla i wieży aktywność, taktyka, zarządzanie czasem i wyniki gier;
- agregacja ze stanu ruchu do funkcji gry gracza;
- podsumowania na poziomie segmentu dla każdej metryki raportu;
- ręczny przegląd każdej potencjalnej krzywej odniesienia przed jej użyciem raport coachin-gowy.

W przypadku blitzu 3+0 oznacza to prawie pół miliona gier, prawie milion wiersze dotyczące gier gracza i ponad 28 milionów wierszy stanu ruchu. Przez całą czwórkę typów gier porównawczych, obecne artefakty testu porównawczego zawierają więcej niż 2,55 milionów gier, ponad 5,10 miliona wierszy gier graczy i ponad 147 milionów wierszy stanu przeniesienia. Skala ma kluczowe znaczenie dla wartości benchmark: pozwala raportowi pokazać użytkownikom, co faktycznie dzieje się w granych grach przez graczy zbliżonych do ich poziomu, zamiast opierać się wyłącznie na ogólnych szachowych hasłach.

Two-sided converted-fork rates by Elo bucket

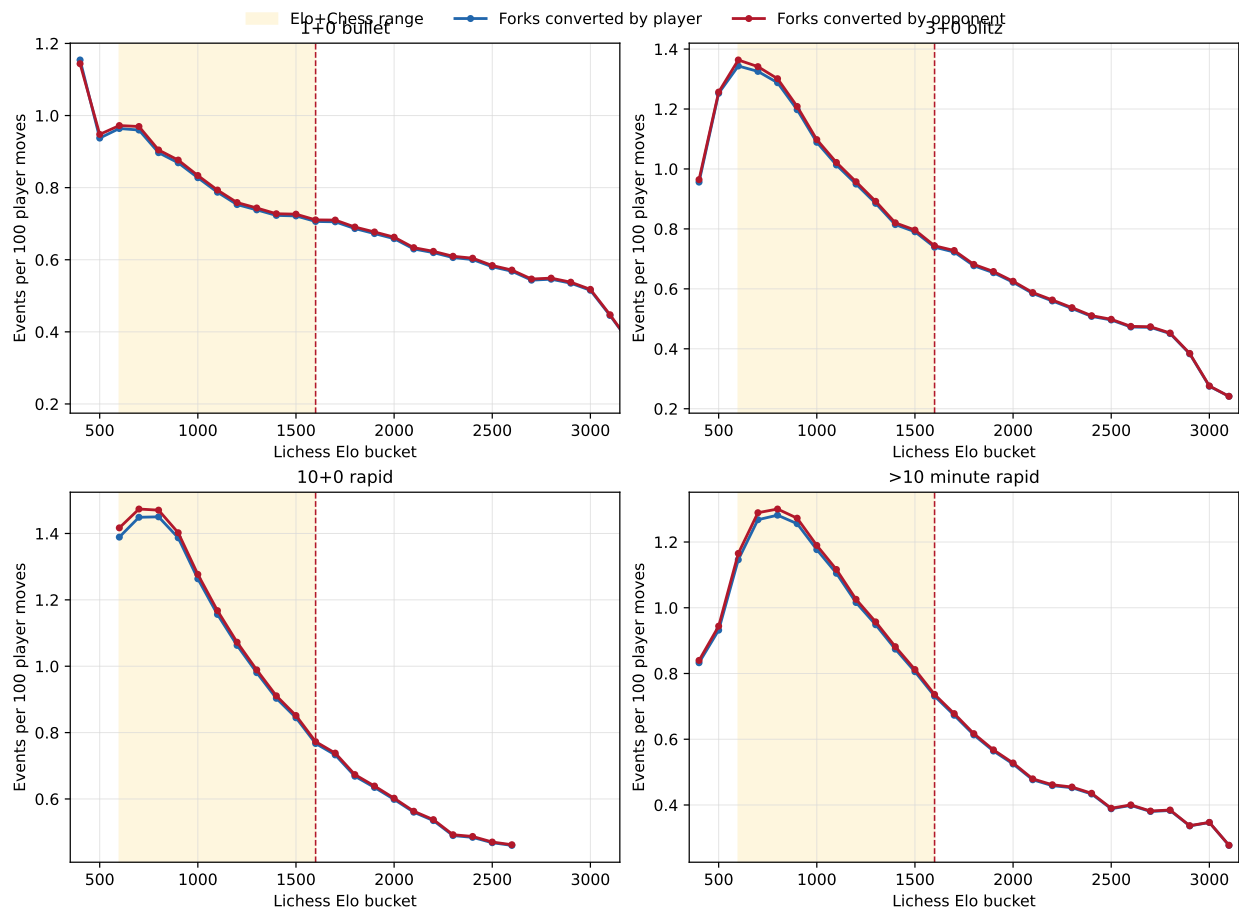


Figure 8: Dwustronne przeliczane stawki widel według łyżki Elo. Konwertowane zdarzenia rozwidlania spadają przy wyższych ocenach, ponieważ obie strony bronią lepiej i mniej czystego widelca dostępne są konwersje.

13 Ograniczenia

Krzywe wzorcowe są opisami empirycznymi, a nie dowodem przyczynowym. Gracz nie uzyskuje oceny jedynie poprzez dopasowanie średniej koszyka. Metryki też są niedoskonałe podsumowanie złożonych pozycji: wczesny ruch hetmana może być doskonały na jednym stanowisku i lekkomyślność na innym. Celem krzywych jest identyfikuj powtarzające się nawyki i porównuj je z zachowaniami rówieśników, a nie z nimi zastąpić kalkulację lub osąd coachingowy.

Krzywe są również krzywymi w skali Lichess. Gdy użytkownicy przynoszą grę Chess.com historii, Elo+Chess mapuje odpowiednią grupę ocen na benchmark Lichess skalować przy użyciu oddzielnego, wieloplatformowego modelu mapowania. Ta warstwa mapowania to opisane w towarzyszącej notatce badawczej Elo+Chess na temat Lichess-to-Chess.com konwersja ocen: [Estimating Cross-Platform Chess Rating Mappings with Regresja modalna](#). To mapowanie nie powinno być odczytywane jako zachowujący percentyl ranga. Zachowuje szacunkową równowagę ocen pomiędzy dwoma ocenami systemu; Ranga percentylowa pozostaje specyficzna dla platformy, ponieważ aktywny gracz dystrybucje są różne.

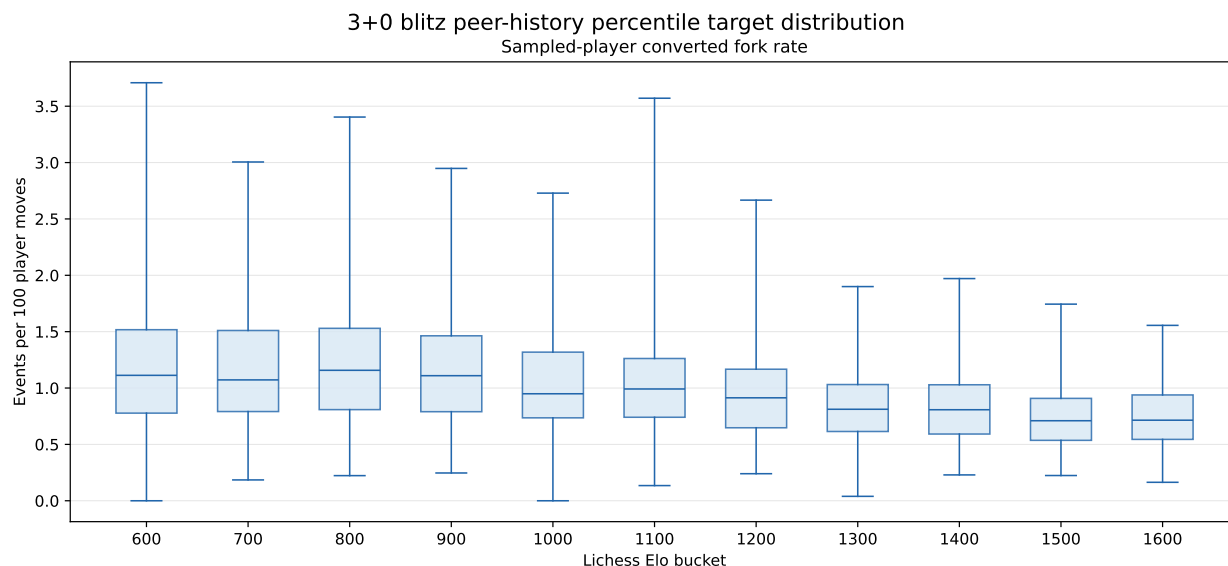


Figure 9: Przykład docelowego rozkładu historii równorzędnych 3+0 blitz dla gracza z próbką przeliczona stawka widel. Percentyle są obliczane w ramach segmentu Elo, czyli raportu porównuje gracza z rówieśnikami, którzy mają zasadniczo podobne możliwości taktyczne stawki.

14 Grafika raportu produkcyjnego

Raport produkcyjny przekształca artefakty wzorcowe w panele na poziomie metrycznym które można odczytać bez znajomości pełnego potoku kompilacji. Liczby 10 and 11 pokaz trzy przykłady z projektu raportu na żywo Elo+Chess.

Panele z zasadą otwierania na rysunku 10 użyj zwykłej prezentacji krzywej wzorcowej. Każda karta zaczyna się od rodzina metryki i nazwa metryki, a następnie podaje krótką regułę interpretacji. The duża odmiana w prawym górnym rogu jest celowo oddzielona od surowej Wartość metryki: odpowiada na pytanie użytkownika: „Jak ocenia ten nawyk”. względem wzorca?”, podczas gdy w wierszach poniżej znajdują się zmierzone wartości. Cztery wiersze podsumowania pokazują czas życia, ostatnie 10, wygrane i przegrane. To sprawia, że panel diagnostyczny, a nie tylko dekoracyjny. Jeśli Lifetime i Last 10 zgodzą się, nawyk jest prawdopodobnie stały. Jeśli wygrane i przegrane są rozbieżne, metryka może być zależne od wyniku lub powiązane ze stanem gry, a nie zwykłym nawykiem.

Obszar wykresu wykorzystuje powściągliwe ciemne tło, złotą linię trendu i kolor znaczniki dla czterech plasterków. Adnotacja trendu wyraźnie stwierdza, czy relacja wzorcowa rośnie lub spada w przypadku Elo. Jest to ważne, ponieważ niektóre wskaźniki są wyższe, tym lepsze, na przykład rozwój mniejszych fragmentów, podczas gdy inne są niższe-znacznie lepsze, jak na przykład wczesny ruch królowej. Karta zatem nie prosi użytkownika o wywnioskowanie kierunku na podstawie samego nachylenia. Finał Wiersze „Co to oznacza” i „Wskazówka trenerska” tłumaczą wynik testu porównawczego w konkretną instrukcję szachową, zachowując jednocześnie kontekst empiryczny.

Rysunek 11 przedstawia styl percentyla prezentacja używana dla metryk, których surowy wskaźnik nie jest dobrze reprezentowany przez a jednolity trend globalny. Zwycięskie widelce są przykładem taktycznym: wskaźnik populacji skonwertowanych widelców zmienia się zarówno w zależności od umiejętności gracza, jak i współczynnika błędów przeciwnika. The Dlatego karta porównuje

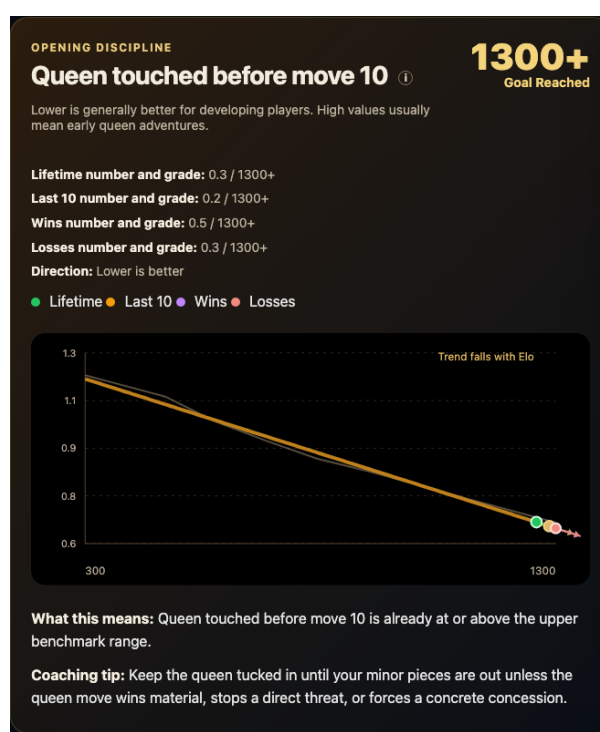
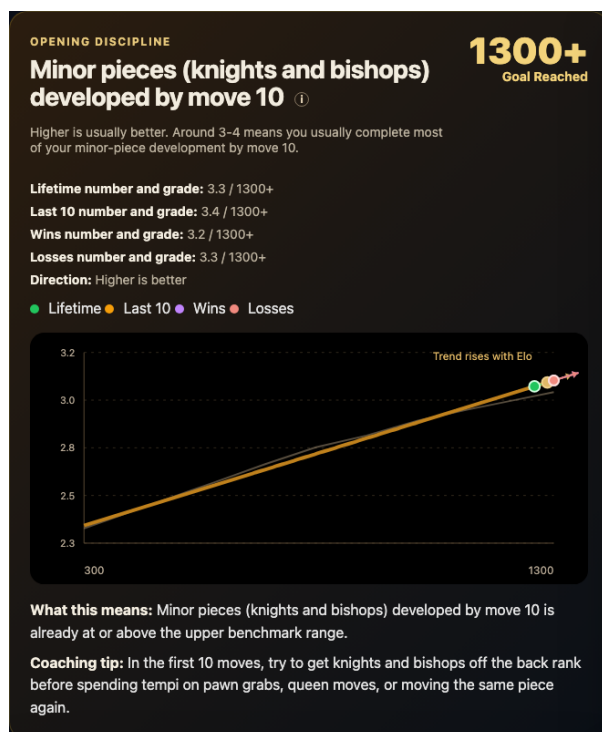


Figure 10: Panele z zasadą otwarcia produkcyjnego. Ta sama gramatyka wizualna obsługuje a metryka wyższa oznacza lepszą, rozwój drobnych elementów i metryka niższa oznacza lepszą, wczesny ruch królowej. Oba panele przedstawiają czas życia, ostatnie 10, wygrane i przegrane dzięki czemu w raporcie można oddzielić nawyki stabilne od niedawnych lub zależnych od wyników wzory.

wartość użytkownika z historiami innych graczy w zmapowany wiadro Elo. Wykres pudełkowy przedstawia dystrybucję równorzędną, w tym rozstęp międzykwartyłowy, mediana i wąsy. Znacznik użytkownika jest nanoszony na tej samej skali i oznaczone zarówno wartością surową, jak i rangą percentylową. Sąsiadujące segmenty porównawcze pozostają widoczne w wyciszonej formie, dzięki czemu użytkownik może zobaczyć, że segment peer ma charakter lokalny, a nie uniwersalny.

Dzięki takiemu rozwiązaniu możliwe jest kontrolowanie metryk percentylowych. Użytkownik widzi wartość surową, percentyl, segment równorzędny użyty do porównania i przetłumaczony Mapowanie zasobników Chess.com do Lichess. To samo życie, ostatnie 10, wygrane i Wiersze dotyczące strat są zachowywane, dzięki czemu nadal można sprawdzić taktyczne wyniki percentylowe dla efektów aktualności i zależności od wyników.

15 Wniosek

Raporty porównawcze Elo+Chess są tworzone na podstawie dużych obliczeń stanu ruchu stratyfikowane próbki Lichess. Każda metryka zaczyna się jako zbiór konkretnych stanów planszy i zmienne stanu ruchu, stają się funkcją gry gracza i są następnie uśredniane przez wiadro Elo w celu utworzenia empirycznej krzywej odniesienia. Zasada otwarcia przykłady pokazują, dlaczego to podejście jest przydatne: proste nawyki szachowe mają silne, widoczne powiązania z oceną w przedziale docelowym, natomiast te relacje często spłaszczają się lub stają się bardziej warunkowe powyżej górnej granicy. Jest to główny powód, dla którego Elo+Chess koncentruje swoje

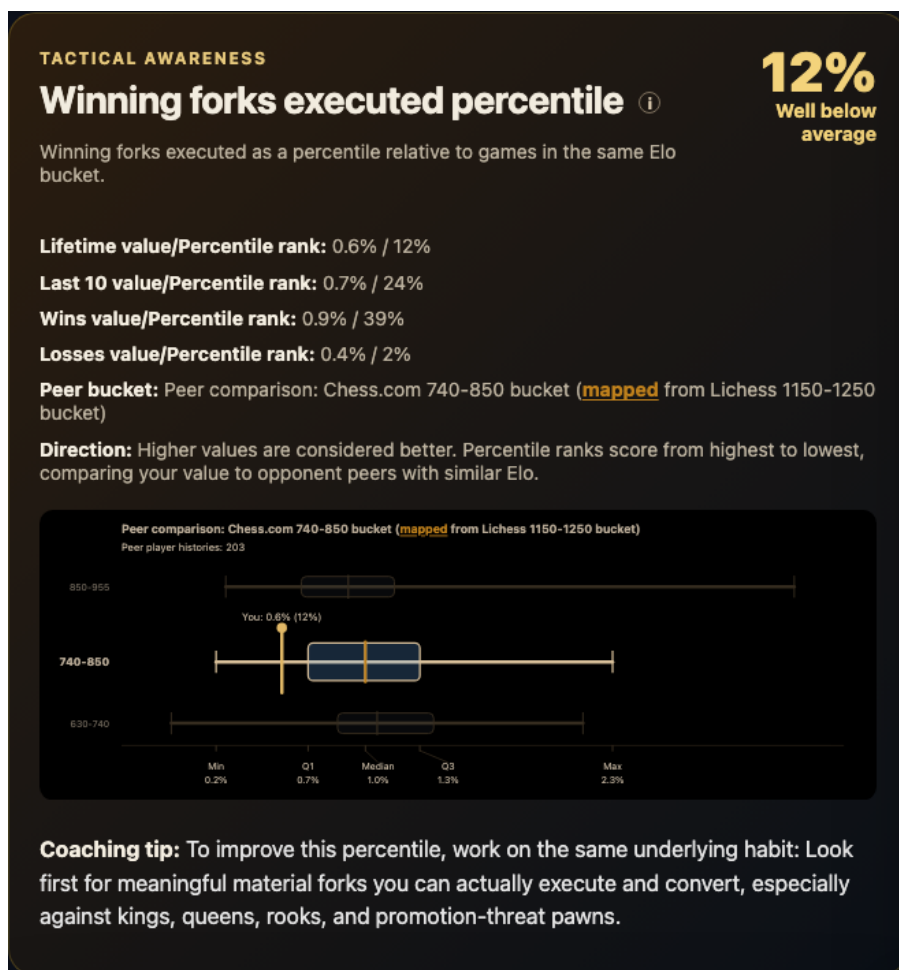


Figure 11: Wykonany panel percentyli produkcji dla zwycięskich wideł. Fabuła pudełkowa pokazuje rozkład równorzędny tego samego segmentu używany w raporcie, podczas gdy złoto znacznik pokazuje surową wartość użytkownika i pozycję percentylową.

twierdzenia coachingowe na ocenie zakresie, w którym dane potwierdzają je najwyraźniej.