

# Raport 1

Martyna Bielec, Maciej Karczewski

2022-11-23

## Wprowadzenie

Strona internetowa Lichess jest drugą co do wielkości platformą na świecie, która umożliwia grę w szachy online. Użytkownicy mogą grać z komputerem lub między sobą. Każdy gracz ma przypisany ranking, który jest ustalany na podstawie jego umiejętności. Jeśli użytkownik rozgrywa partię “rankingową”, w jej wyniku jego ranking rośnie lub maleje.

Dane, które zostaną poddane analizie, zawierają informacje o ponad 20 000 rozgrywkach szachowych ze strony Lichess Games. Dostępne są na witrynie kaggle. Składają się z następujących kolumn:

- id (zawiera elementy typu “character”) - numer ID rozgrywki
- rated (“character”) - wartość “True” lub “False” w zależności od tego, czy partia jest “rankingowa”
- created\_at (“numeric”) - czas rozpoczęcia gry
- last\_move\_at (“numeric”) - czas zakończenia gry
- turns (“numeric”) - liczba rozegranych tur
- victory\_status (“character”) - status zakończenia gry: “mate” dla zakończenia matem, “resign” dla zakończenia przez poddanie się jednej ze stron, “draw” dla remisu lub “outoftime” dla rozgrywki zakończonej z powodu upływu czasu
- winner (“character”) - zwycięzca gry: “white”, jeśli wygrał gracz biały, “black”, jeśli czarny lub “draw” w przypadku remisu
- increment\_code (“character”) - składa się z dwóch liczb rozdzielonych znakiem “+”; pierwsza oznacza liczbę minut przypadających na gracza na wszystkie jego ruchy, druga liczbę sekund, o którą zwiększa się czas gracza po wykonaniu ruchu
- white\_id (“character”) - ID gracza białego
- white\_rating (“numeric”) - ranking gracza białego
- black\_id (“character”) - ID gracza czarnego
- black\_rating (“numeric”) - ranking gracza czarnego
- moves (“character”) - wszystkie ruchy zapisane w standardowej notacji szachowej
- opening\_eco (“character”) - standaryzowany kod dla rozgranego otwarcia
- opening\_name (“character”) - nazwa rozgranego otwarcia
- opening\_ply (“character”) - liczba ruchów w fazie otwarcia

Celem analizy będzie odpowiedzenie na pytanie - jakie jest prawdopodobieństwo różnych zakończeń gry: wygranej białego gracza, czarnego lub remisu? Będziemy rozważać ten problem pod kątem rozgranego otwarcia, rankingu graczy, różnicy w rankingu między graczem białym i czarnym oraz liczby tur w partii.

## Obróbka danych

### Otwarcia (?)

W raporcie przeanalizujemy prawdopodobieństwo różnych zakończeń rozgrywki także ze względu na wybrane otwarcie. Aby uzyskać rzetelne wyniki, nie możemy uwzględnić otwarć, które rozegrała zbyt mała liczba

graczy. Okazuje się jednak, że kolumna `opening_name` zawiera 1477 unikalnych wartości. Dzieje się tak, ponieważ dane rozróżniają wiele wariantów poszczególnych otwarć. Przykładowo “obrona sycylijska” została rozegrana w 181 różnych wariantach, takich jak:

```
## [1] "Sicilian Defense: Mongoose Variation"
## [2] "Sicilian Defense: Bowdler Attack"
## [3] "Sicilian Defense: Smith-Morra Gambit #2"
## [4] "Sicilian Defense: Canal Attack | Main Line"
## [5] "Sicilian Defense: Dragon Variation | Yugoslav Attack | Main Line"
```

W celu umożliwienia analizy otwarć stworzyliśmy nową kolumnę, zawierającą jedynie nazwę otwarcia (bez uwzględnienia wariantu). Nazwaliśmy ją “`openings_general`”. Zawiera ona 149 wartości unikalnych. Przeanalizujemy sześć najczęściej rozgrywanych otwarć, czyli:

```
## # A tibble: 6 x 2
##   openings_general total_count
##   <chr>           <int>
## 1 Sicilian Defense      2581
## 2 French Defense       1377
## 3 Queen's Pawn Game    1203
## 4 Italian Game          952
## 5 King's Pawn Game      897
## 6 Ruy Lopez             832
```

gdzie kolumna “`total_count`” oznacza liczbę rozegranych gier z wykorzystaniem danego otwarcia.

## Analiza danych

### Prawdopodobieństwo różnych zakończeń rozgrywki bez warunkowania.

W celu rozważenia od czego zależy prawdopodobieństwo różnych zakończeń rozgrywki, najpierw oszacowaliśmy to prawdopodobieństwo w ogólności. Liczba rozegranych partii wynosi:

```
games_count = sum(df$winner != "a")
print(games_count)
```

```
## [1] 19618
```

Z czego gracz biały wygrał następującą liczbę razy:

```
white_winner_count = sum(df$winner == "white")
print(white_winner_count)
```

```
## [1] 9782
```

Natomiast czarny:

```
black_winner_count = sum(df$winner == "black")
print(black_winner_count)
```

## [1] 8916

Wynika z tego, że szacowane prawdopodobieństwo wygrania gracza białego wynosi w przybliżeniu 49,85%, gracza czarnego 45,40%, a remisu 4,75%.

W dalszej części analizy sprawdzimy, jak to prawdopodobieństwo się zmienia, jeśli odpowiednio uwarunkujemy rozgrywkę.