

## Pierwszy etap: Zrozumienie problemu + Zrozumienie danych

# League of Legends Early Game Statistics

AUTOR: MICHAŁ TARNOŃSKI

## Ogólny opis zbioru danych

Zbiór danych zawiera statystyki z pierwszych 10 minut gier League of Legends (LoL) rozgrywanych w randze Diamond I do Master. League of Legends to popularna gra typu MOBA (Multiplayer Online Battle Arena), w której dwie drużyny, niebieska (blue) i czerwona (red), rywalizują o zniszczenie bazy przeciwnika. Pierwsze 10 minut gry, zwane "early game", jest kluczowym okresem, który często determinuje wynik całej rozgrywki.

Dane pochodzą z Riot API. Zbiór składa się z **9879** unikalnych gier, gdzie każdy wiersz reprezentuje jedną grę z unikalnym identyfikatorem gameId. Zbiór danych zawiera **40 kolumn (atrybutów)**.

## Określenie celu eksploracji i kryteriów sukcesu

Celem eksploracji jest identyfikacja kluczowych czynników z pierwszych 10 minut gry, które najlepiej przewidują zwycięstwo drużyny niebieskiej. Docelowo rozwiązywanym problemem będzie klasyfikacja binarna (zwycięstwo lub przegrana drużyny niebieskiej).

### Kryteria sukcesu:

- Zidentyfikowanie statystycznie istotnych korelacji między zmiennymi a wynikiem gry (blueWins).
- Określenie najważniejszych czynników wpływających na zwycięstwo drużyny niebieskiej z pierwszych 10 minutach.
- Zbadanie różnic w statystykach między wygrywającymi i przegrywającymi drużynami.
- Ocena jakości danych i ich przydatności do budowy przyszłego modelu predykcyjnego.

## Opis atrybutów

Zbiór danych zawiera 40 atrybutów. Kolumna blueWins jest zmienną docelową, przyjmującą wartość 1 dla zwycięstwa drużyny niebieskiej i 0 dla jej przegranej. Pozostałe 39 atrybutów to statystyki z pierwszych 10 minut gry, podzielone na statystyki drużyny niebieskiej (blue prefix) i czerwonej (red prefix).

### Przykładowe atrybuty dla każdej drużyny (blue/red):

- WardsPlaced: Liczba postawionych totemów.
- WardsDestroyed: Liczba zniszczonych totemów przeciwnika.
- FirstBlood: Czy drużyna zdobyła pierwsze zabójstwo (1=tak, 0=nie).
- Kills: Liczba zabójstw.
- Deaths: Liczba śmierci.

- `Assists`: Liczba asyst przy zabójstwach.
- `EliteMonsters`: Liczba zabitych elitarnych potworów (smoki i heroldzi).
- `Dragons`: Liczba zabitych smoków.
- `Heralds`: Liczba zabitych heroldów.
- `TowersDestroyed`: Liczba zniszczonych wież przeciwnika.
- `TotalGold`: Całkowita ilość złota zdobytego przez drużynę.
- `AvgLevel`: Średni poziom bohaterów w drużynie.
- `TotalExperience`: Całkowite doświadczenie zdobyte przez drużynę.
- `TotalMinionsKilled`: Całkowita liczba zabitych stworów (minions).
- `TotalJungleMinionsKilled`: Całkowita liczba zabitych potworów z dżungli.
- `CSPerMin`: Średnia liczba zabitych stworów na minutę.
- `GoldPerMin`: Średnia ilość zdobytego złota na minutę.

Dodatkowo, dla obu drużyn istnieją atrybuty różnicowe: `GoldDiff` (różnica w złocie) i `ExperienceDiff` (różnica w doświadczeniu), które są kluczowe w ocenie przewagi.

**Jednostki miary:** Złoto, doświadczenie i poziom to wewnętrzne waluty i punkty doświadczenia w grze. Zabójstwa, śmierci, asysty, wieże, smoki, heroldzi, stwory i potwory z dżungli są liczone w sztukach. Czas jest mierzony w minutach.

## Eksploracyjna analiza danych (EDA)

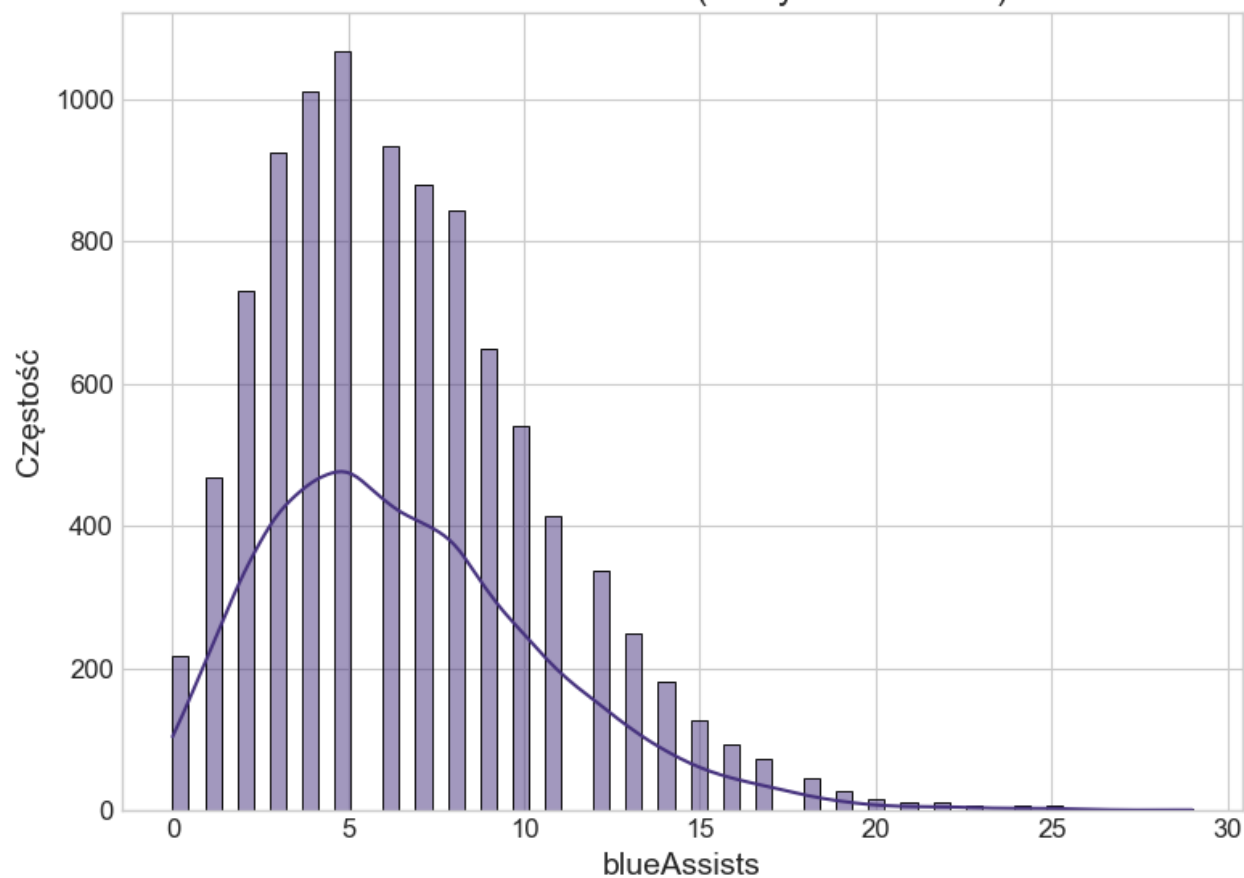
### Sprawdzenie i obsługa brakujących wartości

W zbiorze danych **nie stwierdzono żadnych brakujących wartości** (wszystkie kolumny mają 9879 `non-null` wpisów), co pozytywnie wpływa na jego jakość i ułatwia dalszą analizę.

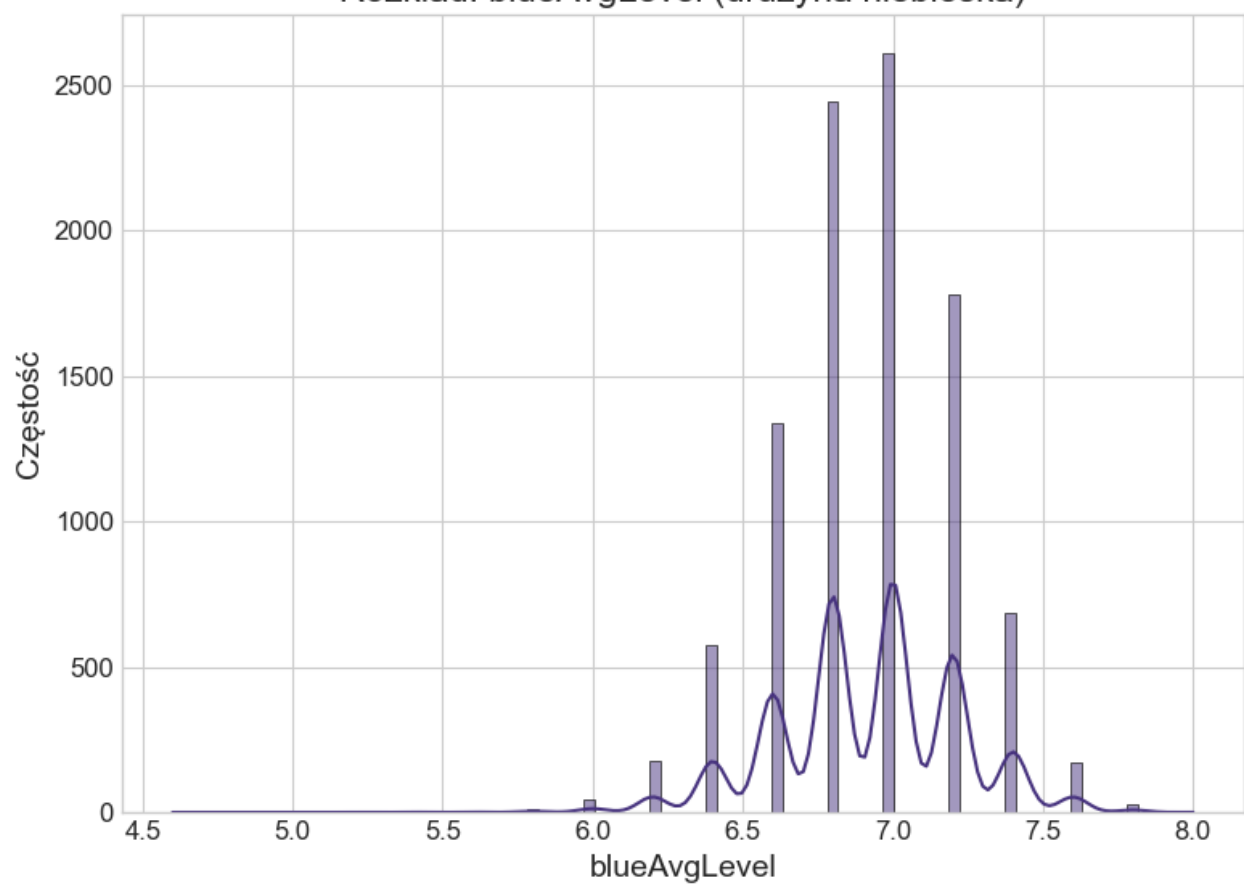
### Rozkłady wartości atrybutów

Analiza rozkładów poszczególnych atrybutów (przedstawiona za pomocą histogramów) wykazała zróżnicowane wzorce. Wiele atrybutów, takich jak `Kills`, `TotalGold`, `AvgLevel`, `EliteMonsters`, `Dragons`, `Heralds`, `TowersDestroyed` ma rozkłady skośne, co jest typowe dla danych z gier.

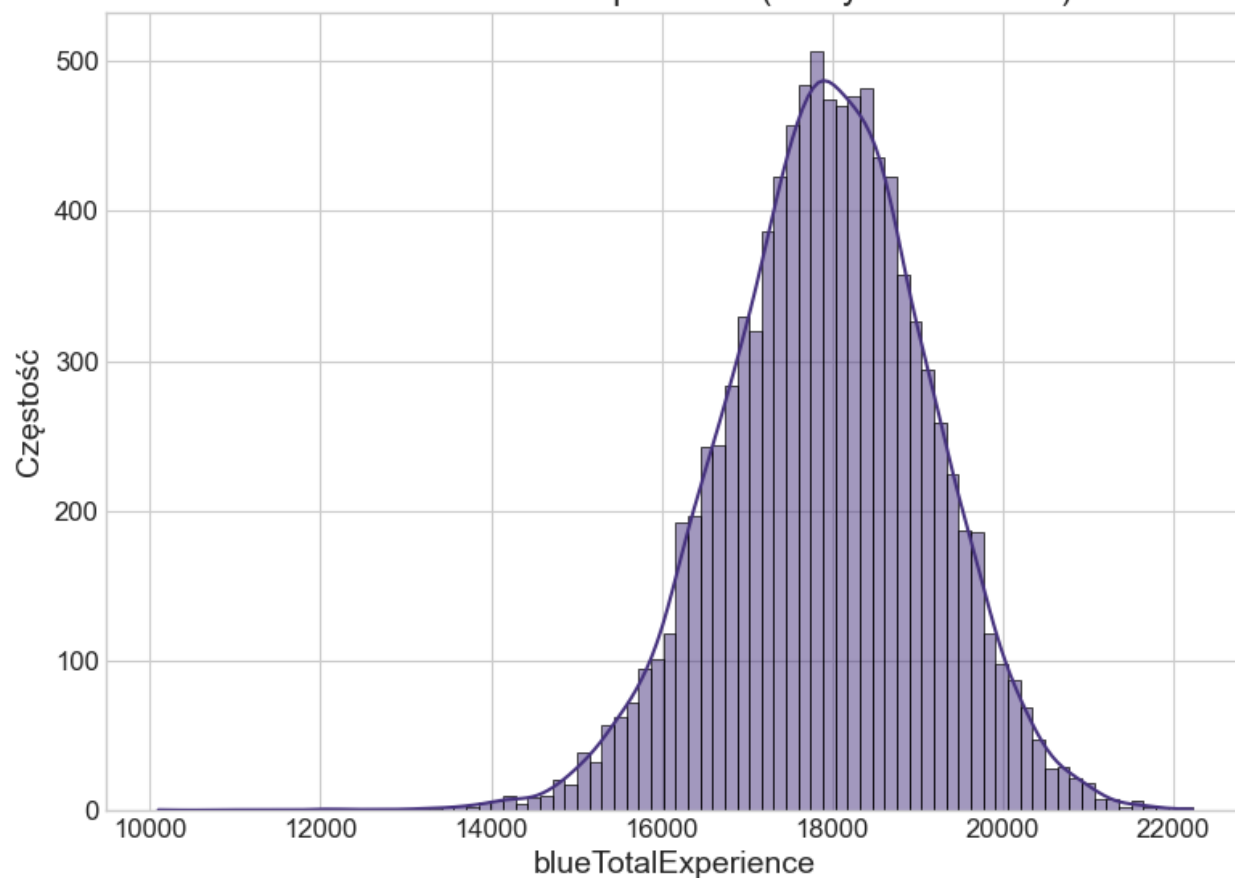
Rozkład: blueAssists (drużyna niebieska)



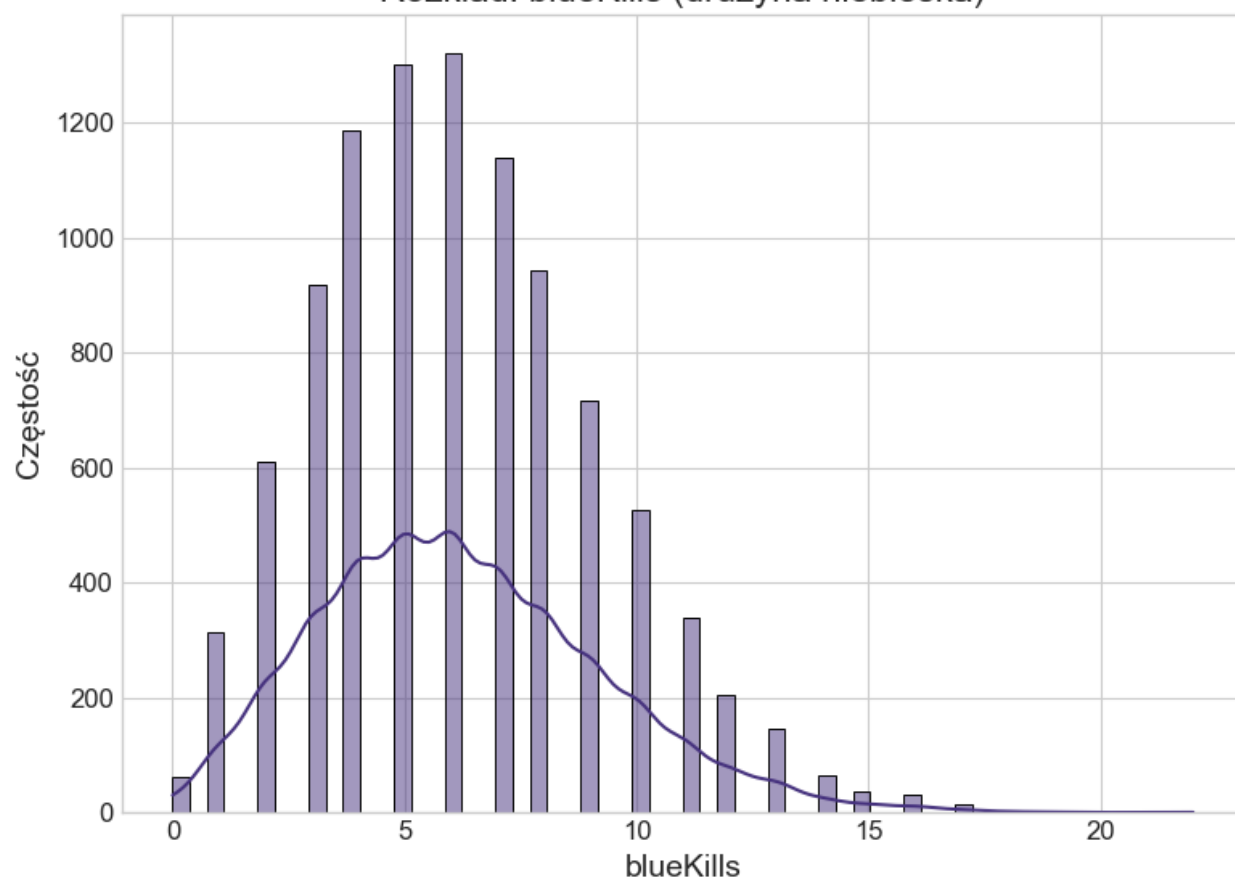
Rozkład: blueAvgLevel (drużyna niebieska)



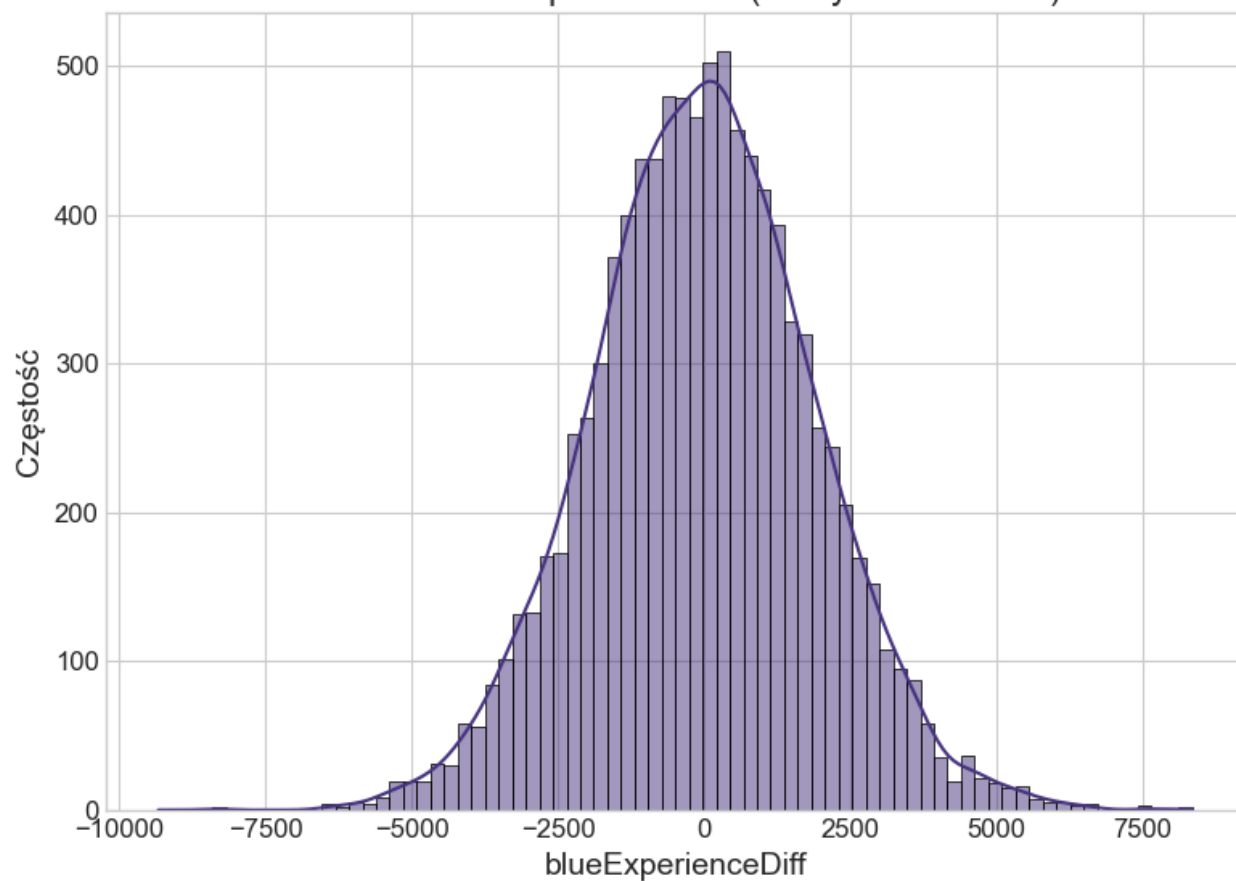
Rozkład: blueTotalExperience (drużyna niebieska)



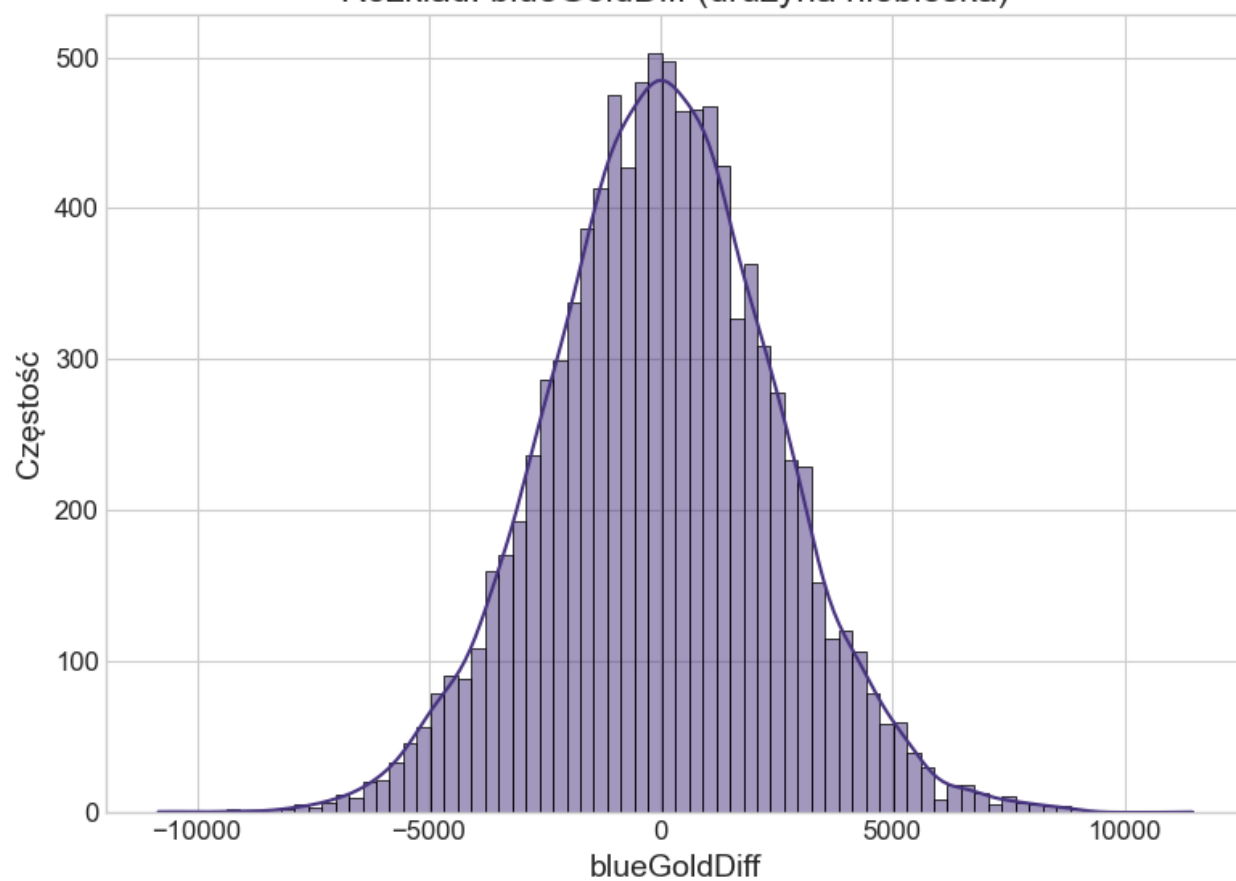
Rozkład: blueKills (drużyna niebieska)

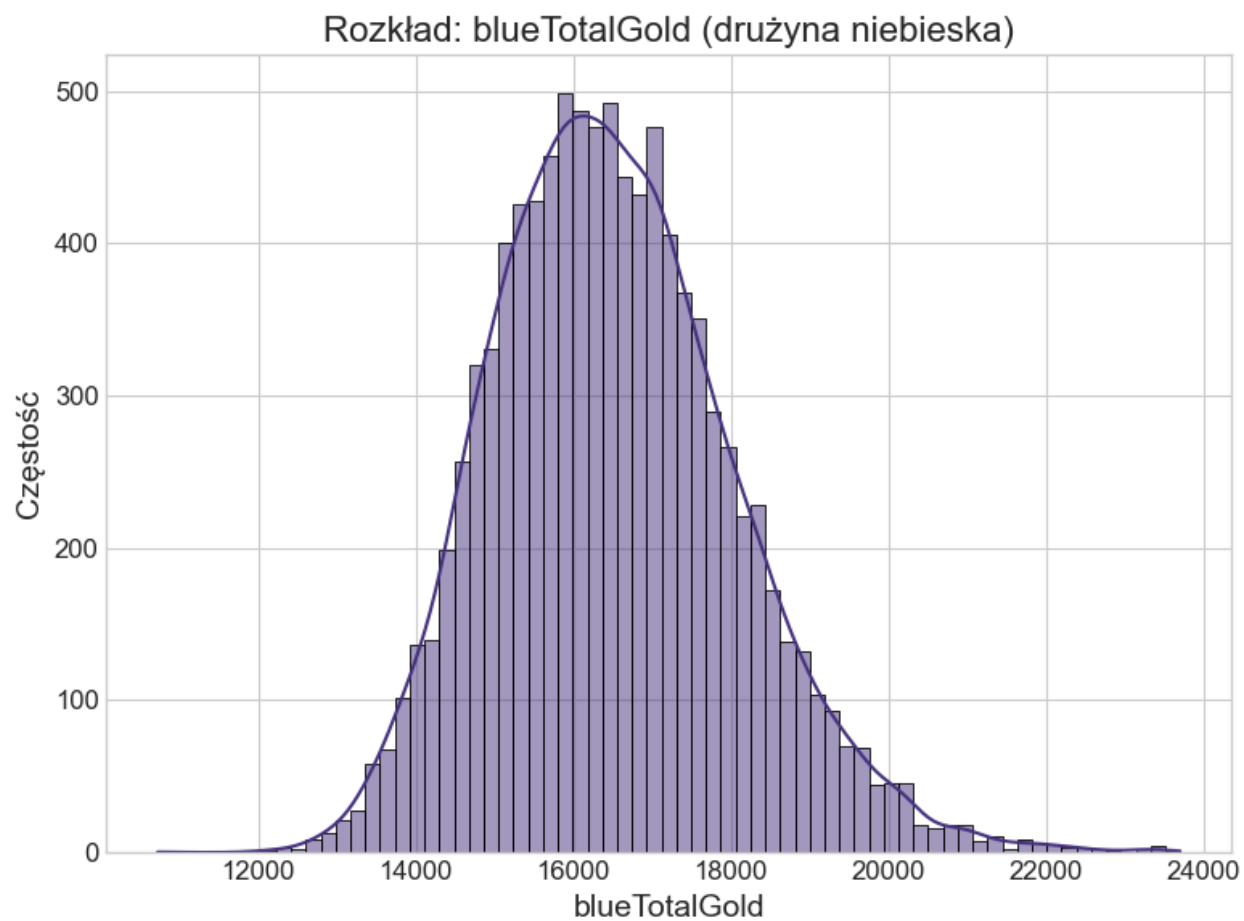
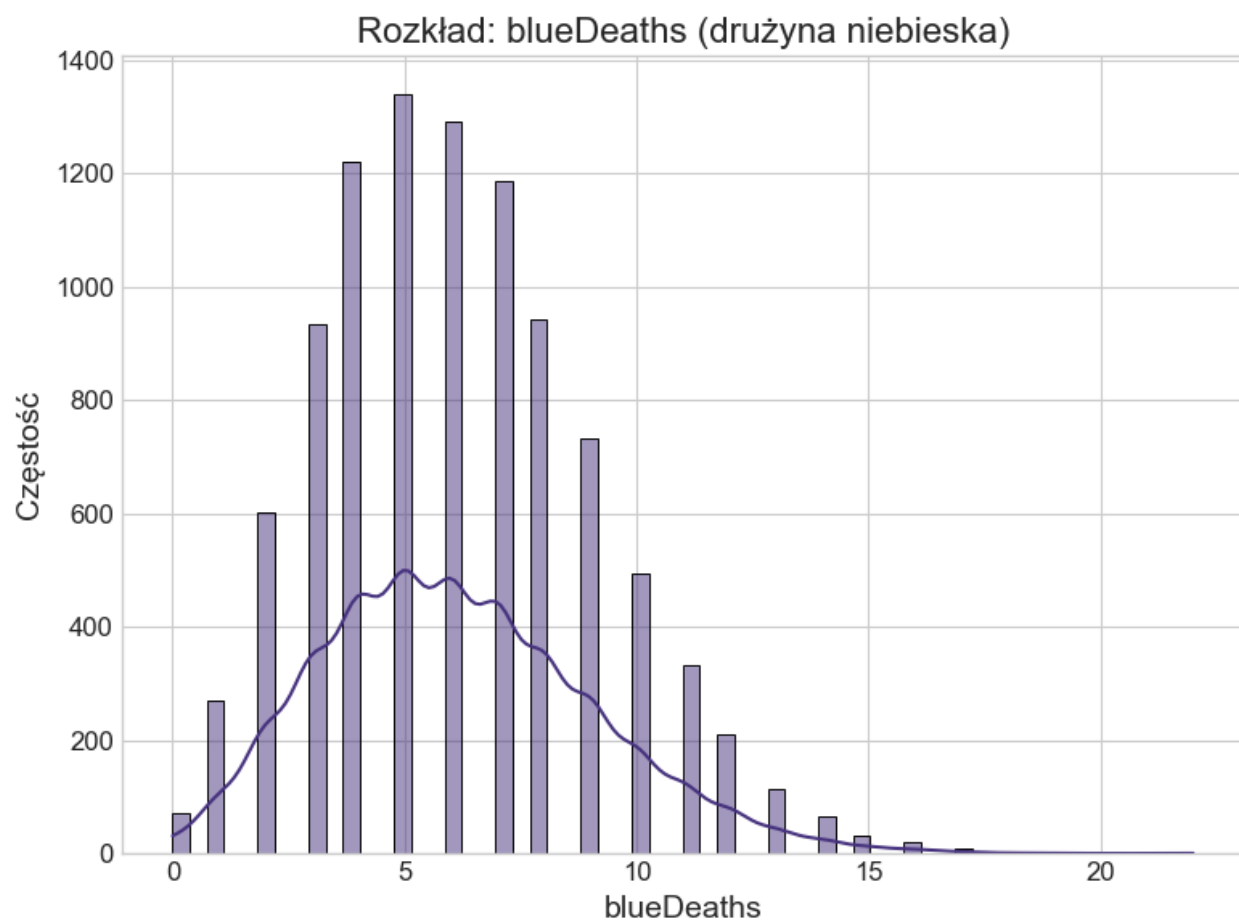


Rozkład: blueExperienceDiff (drużyna niebieska)

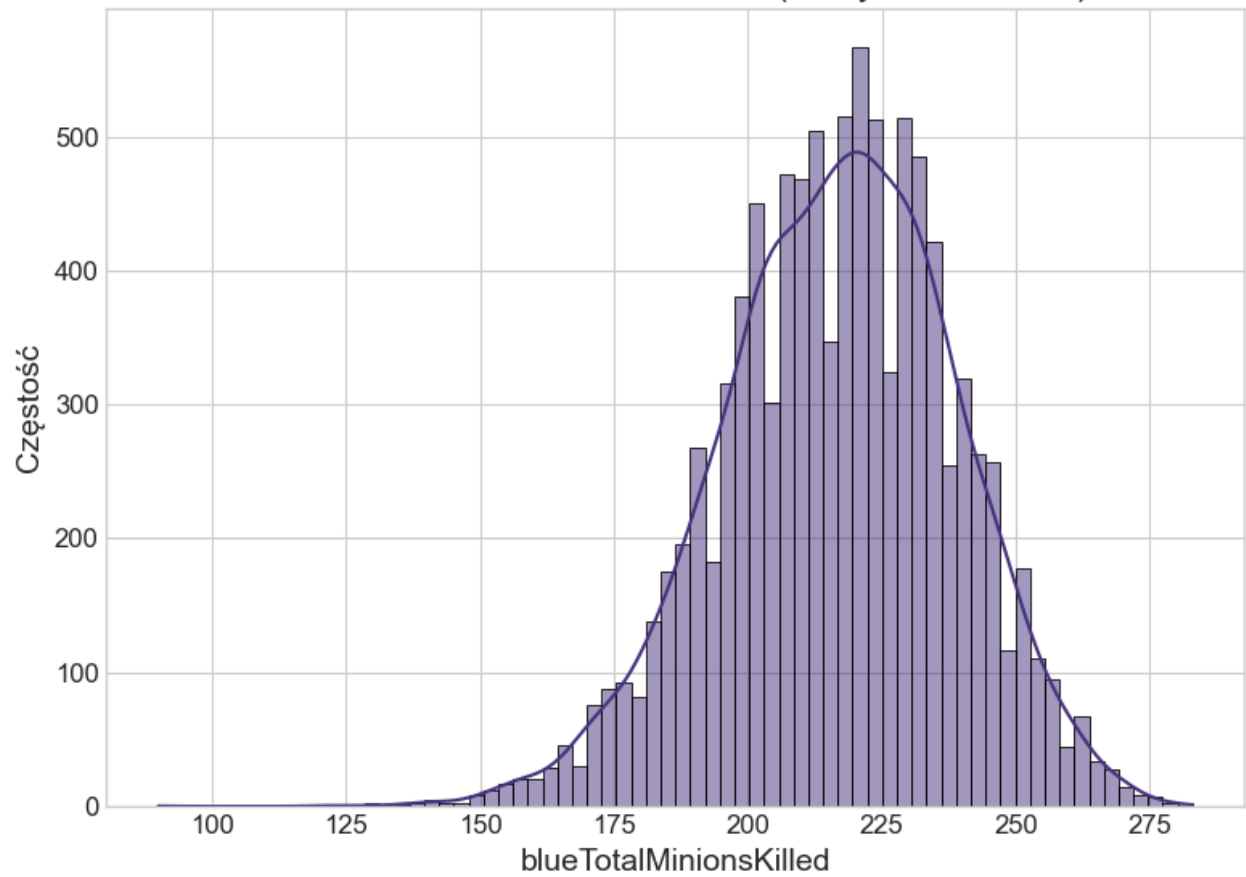


Rozkład: blueGoldDiff (drużyna niebieska)





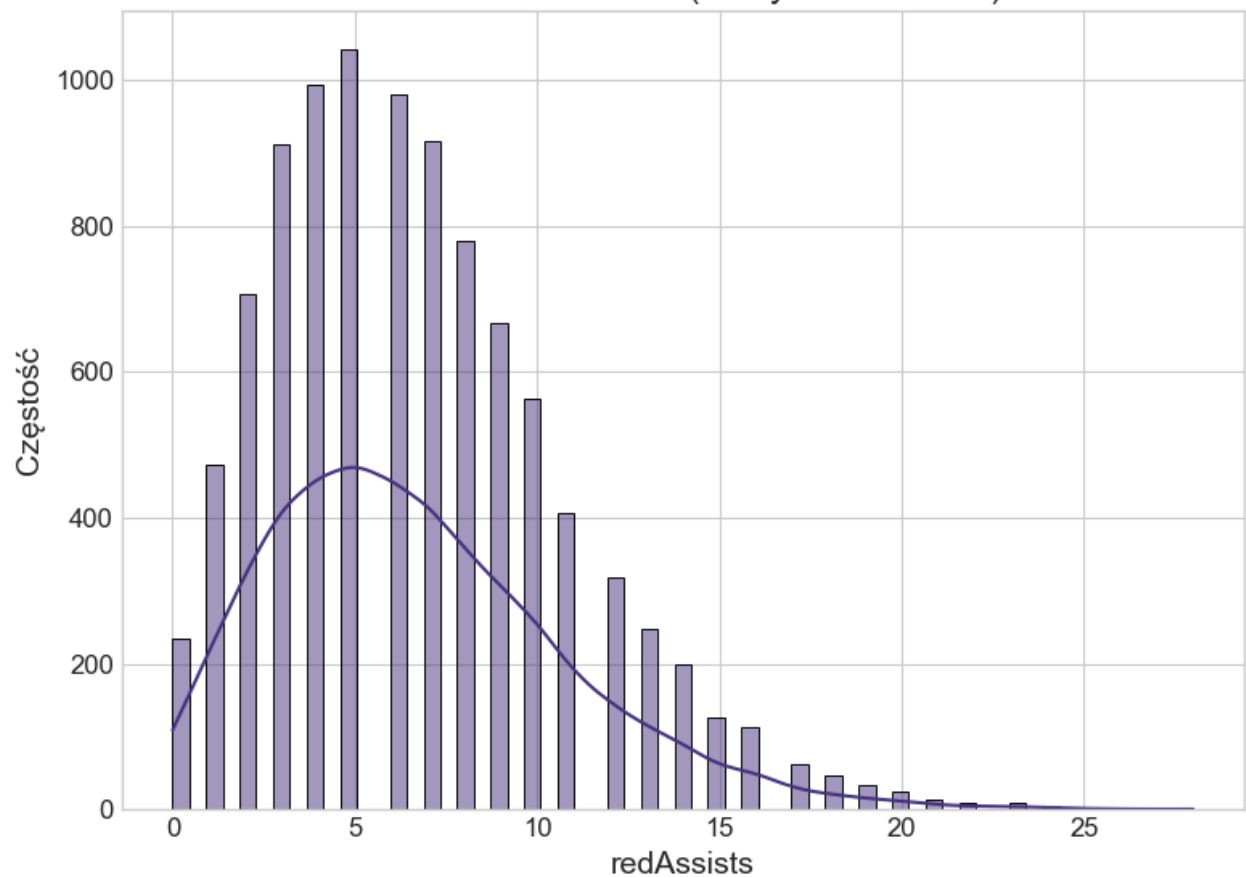
Rozkład: blueTotalMinionsKilled (drużyna niebieska)



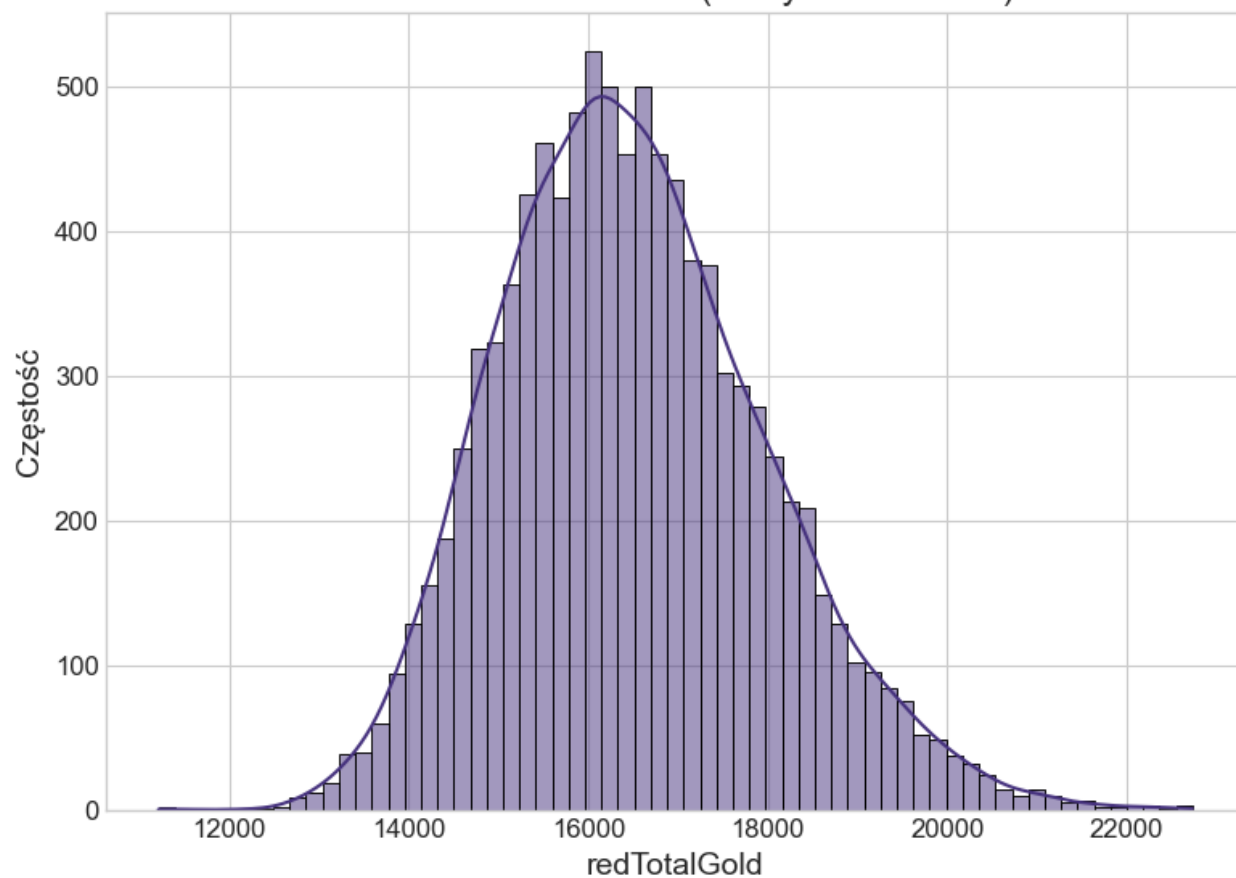
---

Dane drużyny czerwonej

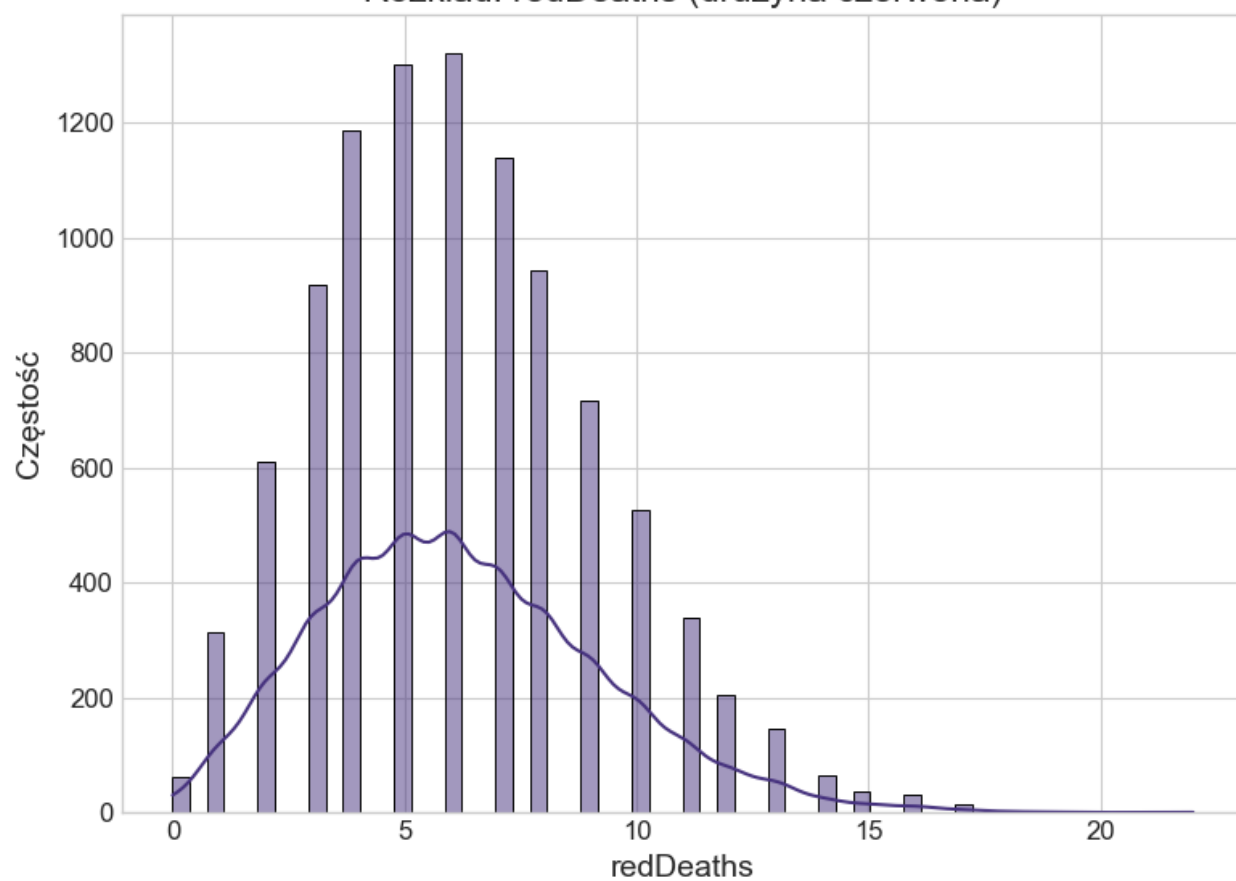
Rozkład: redAssists (drużyna czerwona)



Rozkład: redTotalGold (drużyna czerwona)

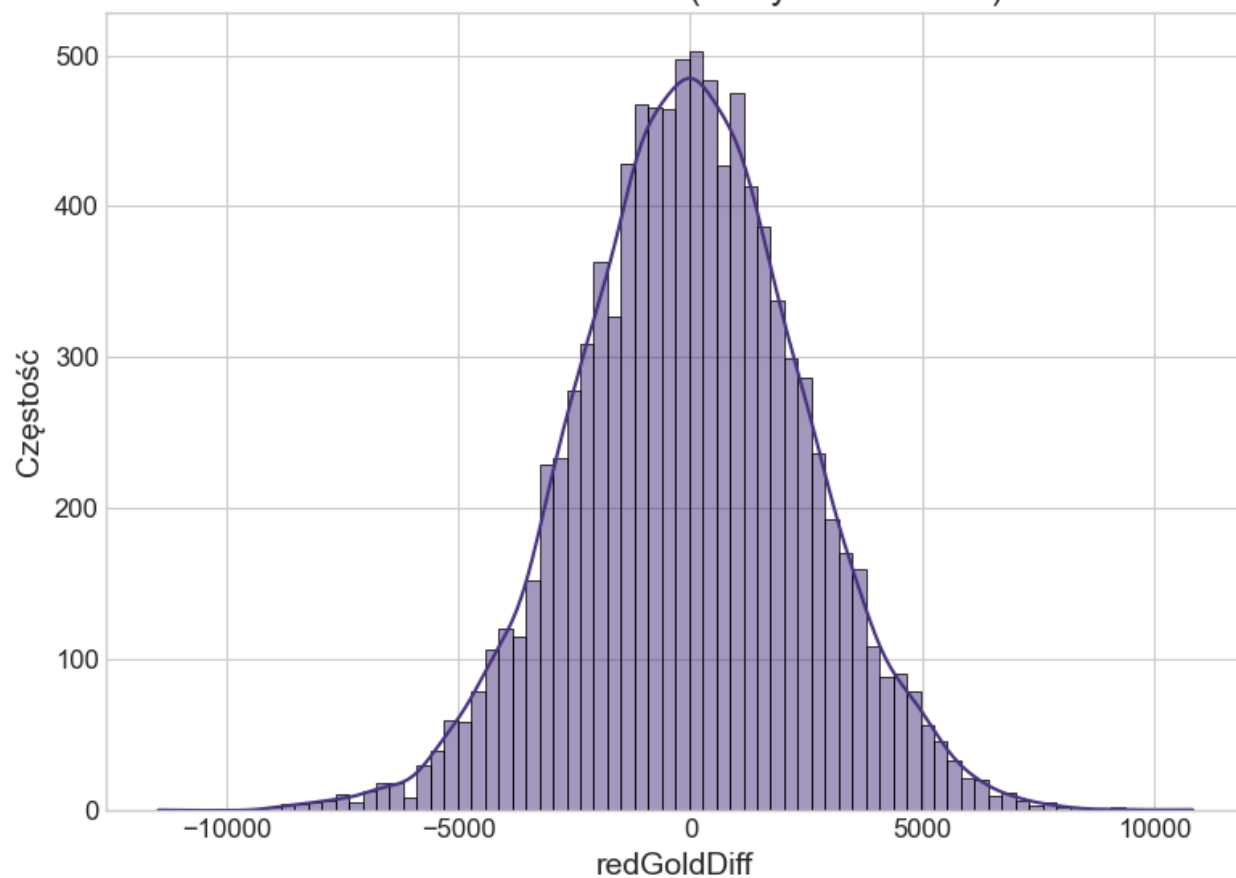


Rozkład: redDeaths (drużyna czerwona)

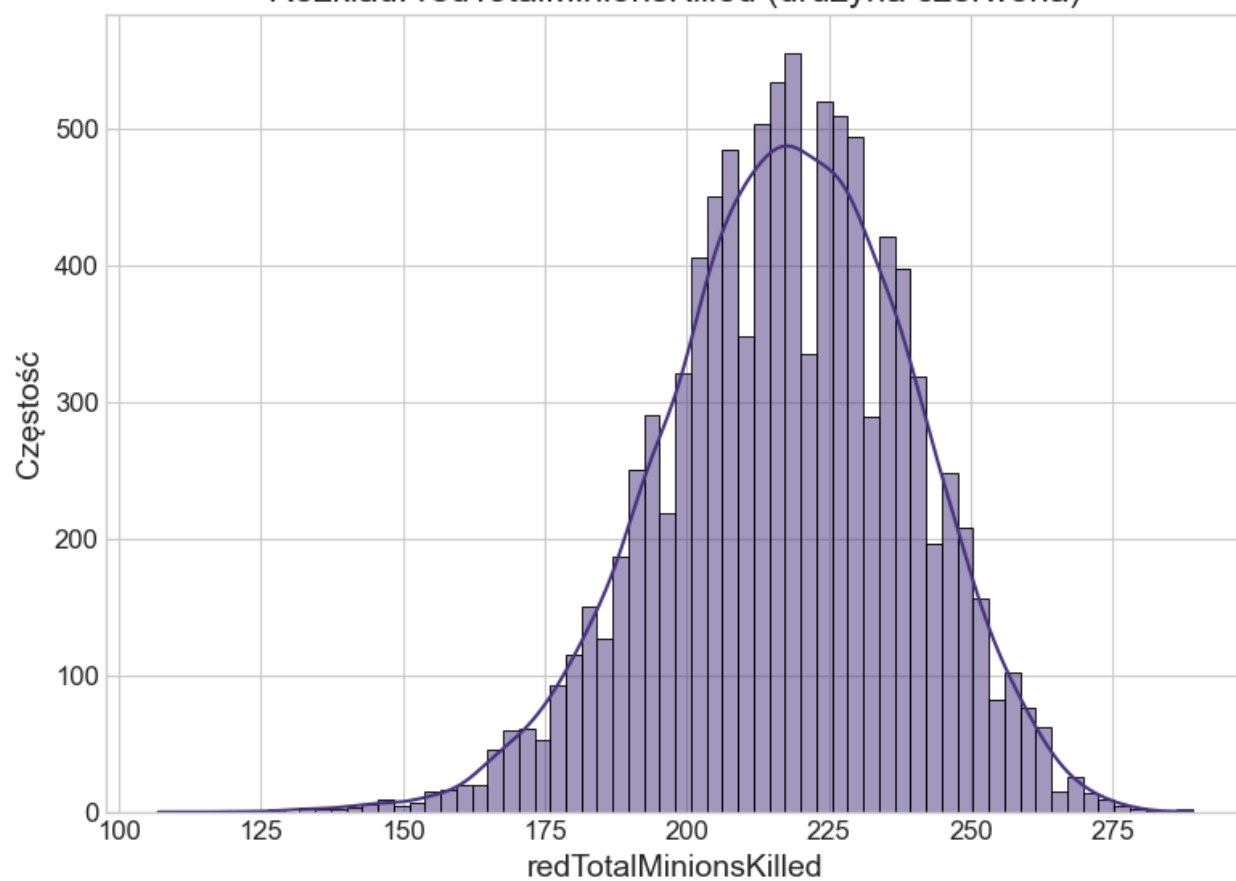




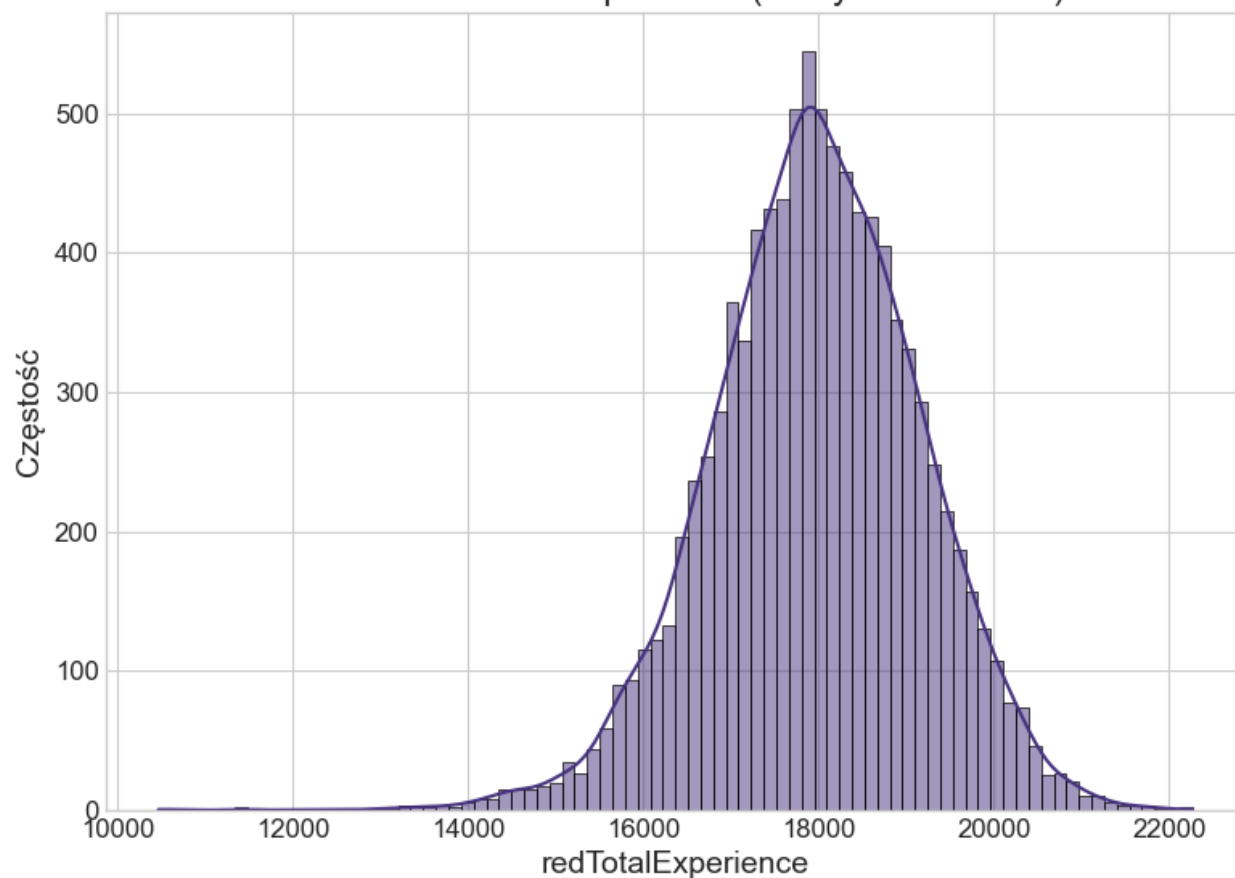
Rozkład: redGoldDiff (drużyna czerwona)



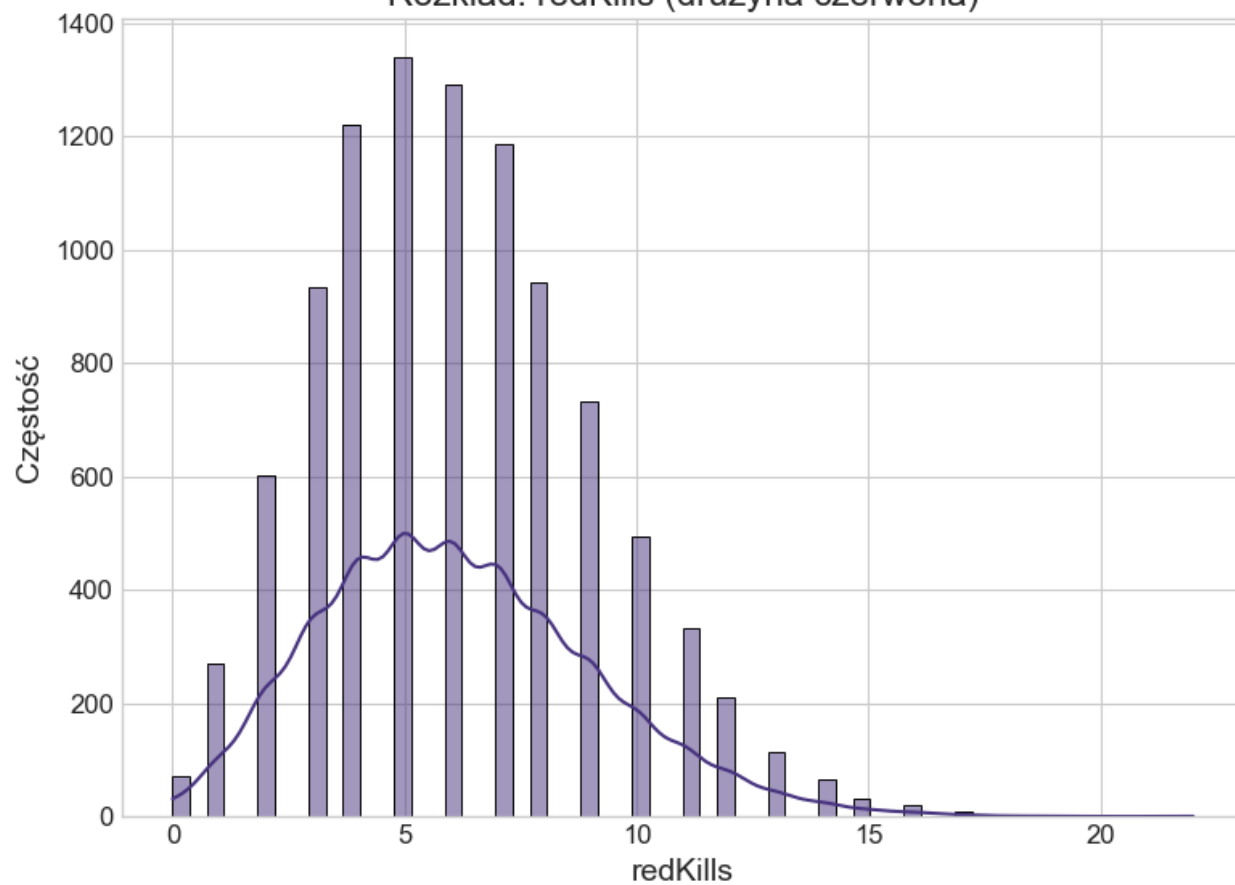
Rozkład: redTotalMinionsKilled (drużyna czerwona)

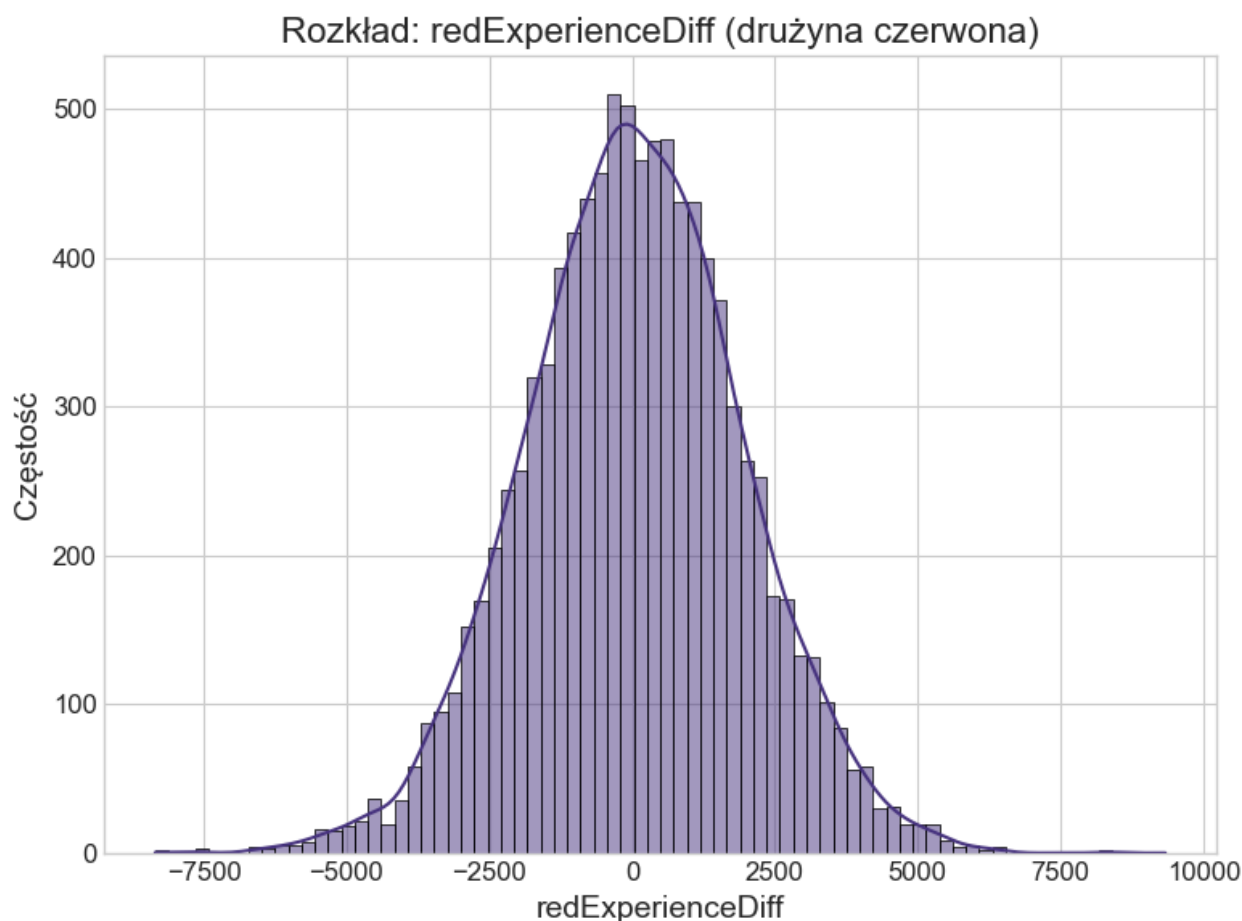


Rozkład: redTotalExperience (drużyna czerwona)



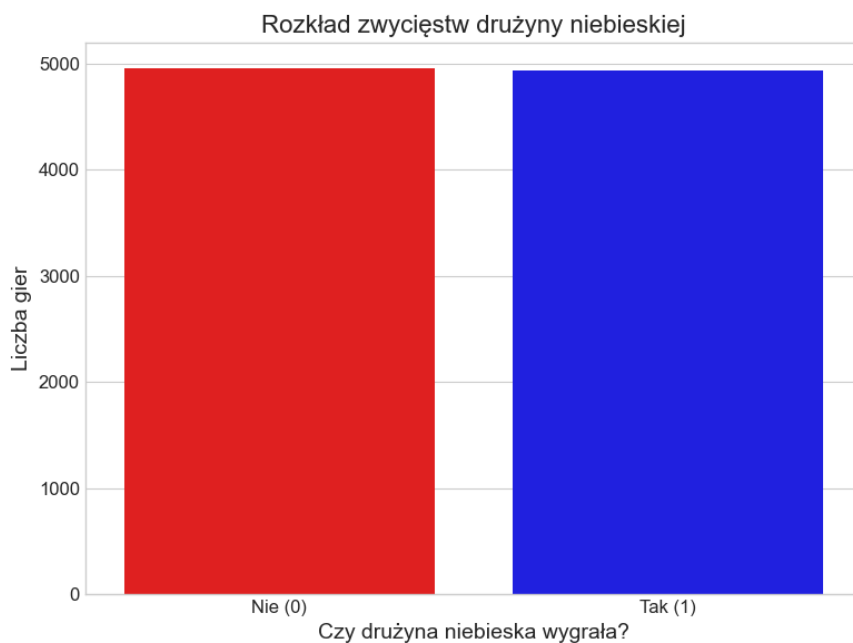
Rozkład: redKills (drużyna czerwona)



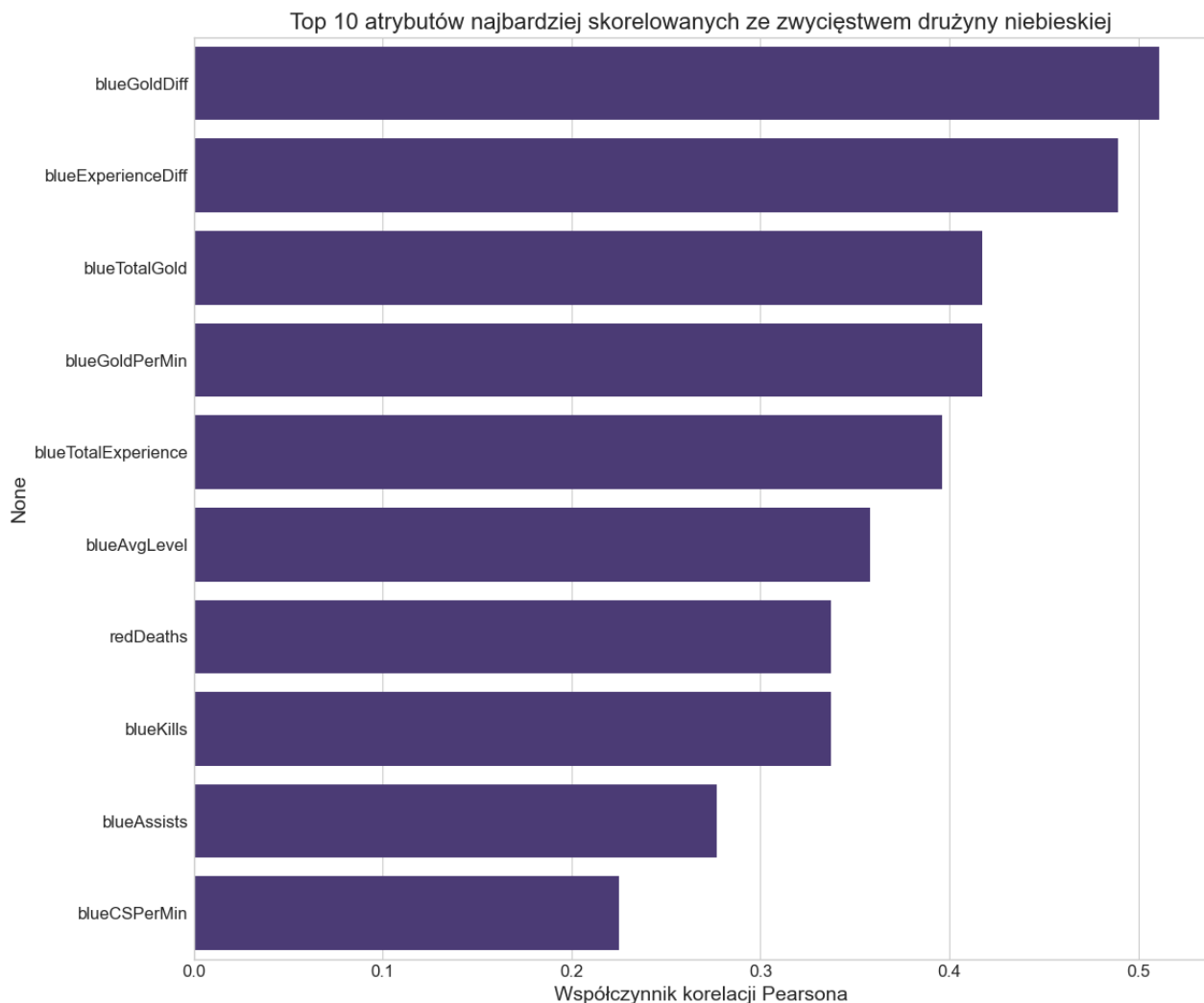


## Rozkład zmiennej docelowej

Rozkład zmiennej blueWins jest praktycznie zbalansowany. Drużyna niebieska wygrywa w **4930 grach (49.90%)**, a przegrywa w **4949 grach (50.10%)**. Ten balans jest korzystny dla budowy modeli klasyfikacyjnych, ponieważ zapobiega problemom związanym z niebalansowanymi zbiorami danych.



## Korelacje pomiędzy atrybutami



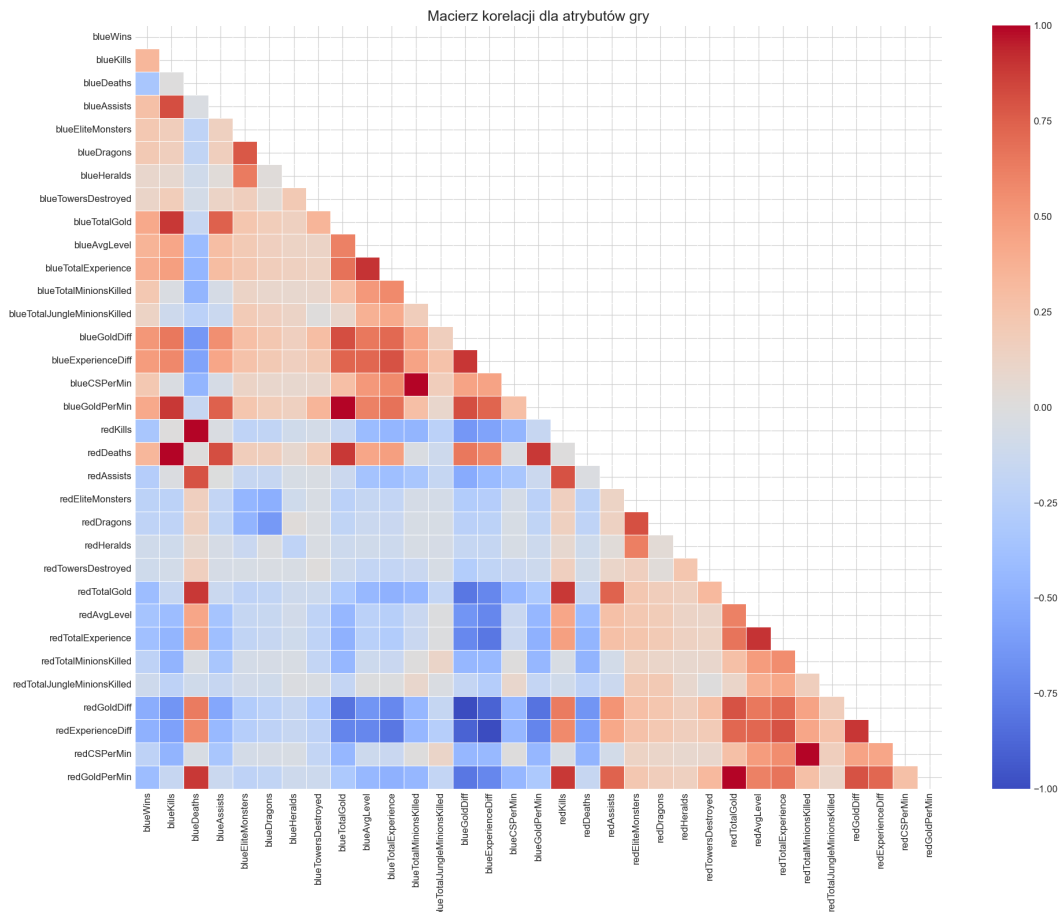
### Najsilniejsze korelacje ze zmienną docelową (**blueWins**) wykazują:

- **blueGoldDiff: 0.511119** (silna pozytywna korelacja)
- **blueExperienceDiff: 0.489558** (silna pozytywna korelacja)
- **blueTotalGold: 0.417213**
- **blueGoldPerMin: 0.417213**
- **blueTotalExperience: 0.396141**
- **blueAvgLevel: 0.357820**
- **blueKills: 0.337358** (umiarkowana pozytywna korelacja)
- **blueEliteMonsters: 0.221944** (słaba pozytywna korelacja)
- **blueDragons: 0.213768** (słaba pozytywna korelacja)
- **blueTowersDestroyed: 0.115566** (słaba pozytywna korelacja)
- **blueHeralds: 0.092385** (bardzo słaba pozytywna korelacja)

Istnieją również silne negatywne korelacje z analogicznymi statystykami drużyny czerwonej, na przykład **redGoldDiff** ma korelację **-0.511119**, a **redExperienceDiff** **-0.489558**. To

potwierdza, że przewaga jednej drużyny w złocie i doświadczeniu jest bezpośrednio związana z deficytem drugiej.

Macierz korelacji dla wszystkich istotnych atrybutów została zwizualizowana za pomocą heatmapy.



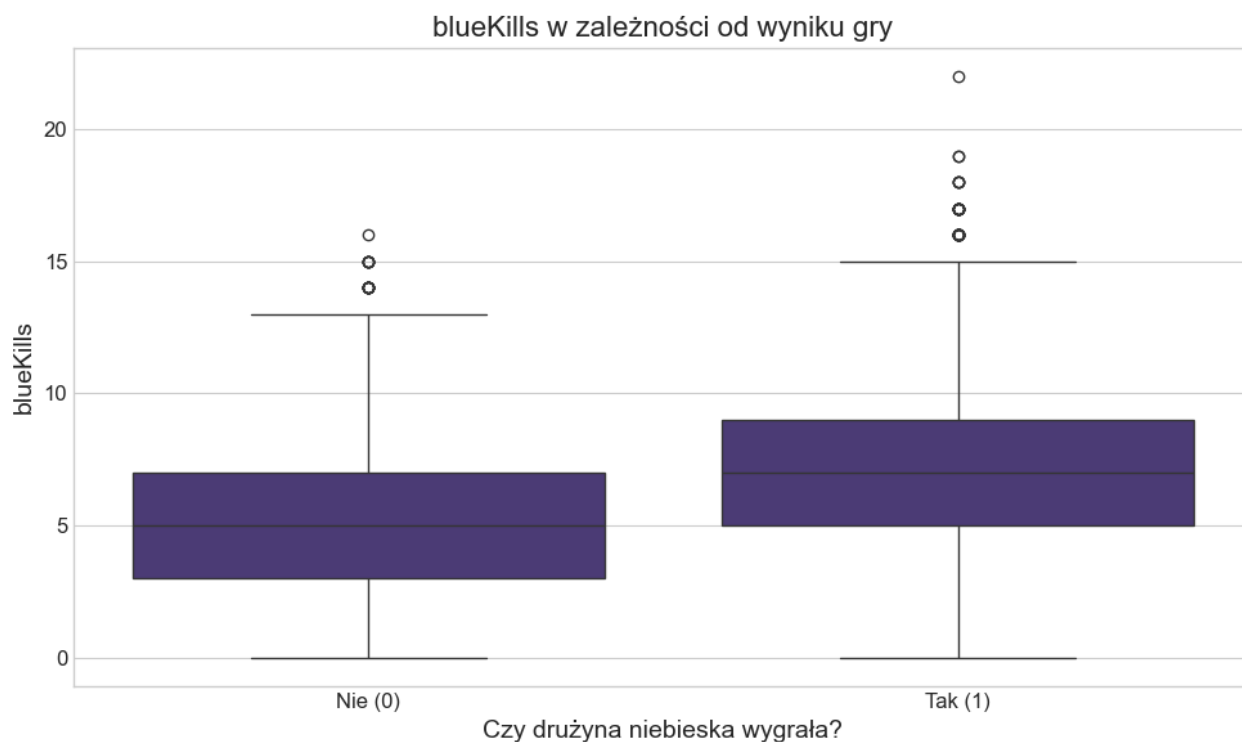
## Porównanie statystyk dla wygranych i przegranych gier

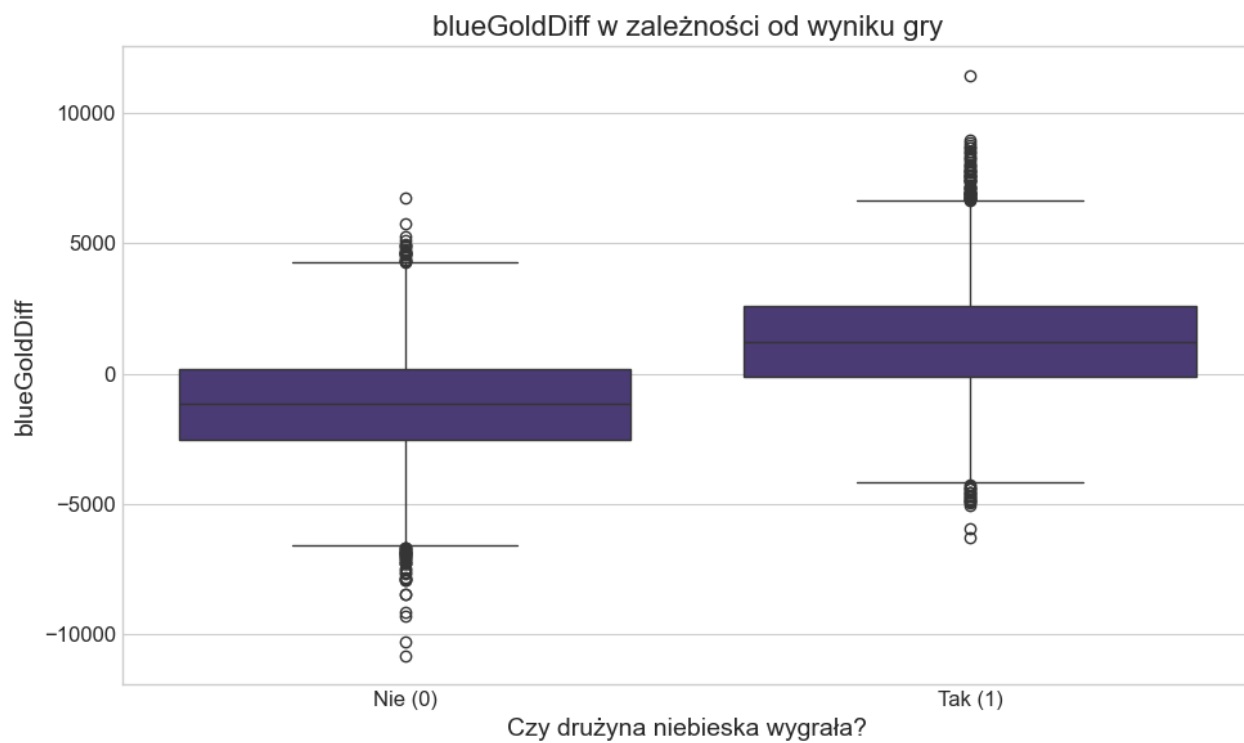
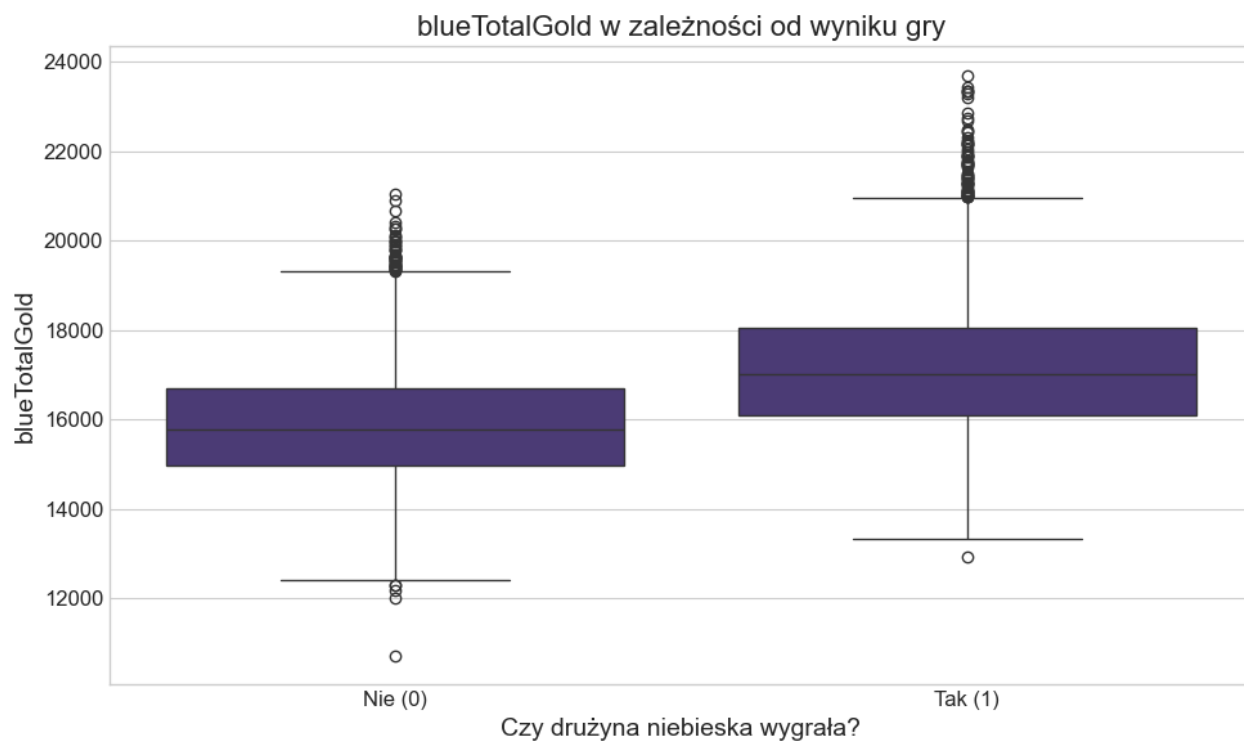
Analiza boxplotów oraz porównanie średnich wartości atrybutów dla wygranych i przegranych gier drużyny niebieskiej wykazało wyraźne różnice:

Cecha	Wygrane (średnia)	Przegrane (średnia)	Różnica (%)
blueKills	7.20	5.17	39.29
blueDeaths	5.14	7.13	-27.92
blueAssists	7.77	5.52	40.72
blueEliteMonsters	0.69	0.41	67.49
blueDragons	0.46	0.26	79.19
blueHeralds	0.22	0.15	47.51
blueTowersDestroyed	0.08	0.02	243.06
blueTotalGold	17145.26	15864.11	8.08
blueAvgLevel	7.03	6.81	3.21

blueTotalExperience	18404.58	17453.47	5.45
blueGoldDiff	1270.72	-1237.07	-202.72
blueExperienceDiff	908.27	-971.90	-193.45

- **Złoto i Doświadczenie:** Drużyna niebieska, która wygrała, miała znacznie wyższe średnie wartości blueGoldDiff (**1270.72** vs **-1237.07**) i blueExperienceDiff (**908.27** vs **-971.90**), co wskazuje na kluczową rolę przewagi ekonomicznej w początkowej fazie gry. Średnie wartości blueTotalGold (**17145.26** vs **15864.11**) i blueTotalExperience (**18404.58** vs **17453.47**) były również znacznie wyższe dla wygranych gier.
- **Zabójstwa:** Wygrywające drużyny niebieskie miały średnio więcej blueKills (**7.20** vs **5.17**) i mniej blueDeaths (**5.14** vs **7.13**) niż przegrywające.
- **Kontrola Obiektów:** Zwycięstwa niebieskiej drużyny często wiązały się z większą liczbą blueDragons (**0.46** vs **0.26**), blueHeralds (**0.22** vs **0.15**) i blueTowersDestroyed (**0.08** vs **0.02**), co podkreśla znaczenie kontroli nad celami na mapie. Różnica w liczbie zniszczonych wież jest szczególnie znacząca (**243.06%**).
- **Poziom:** Średni poziom drużyny (blueAvgLevel) był również wyższy dla wygrywających drużyn (**7.03** vs **6.81**), co jest pochodną zdobytego doświadczenia.



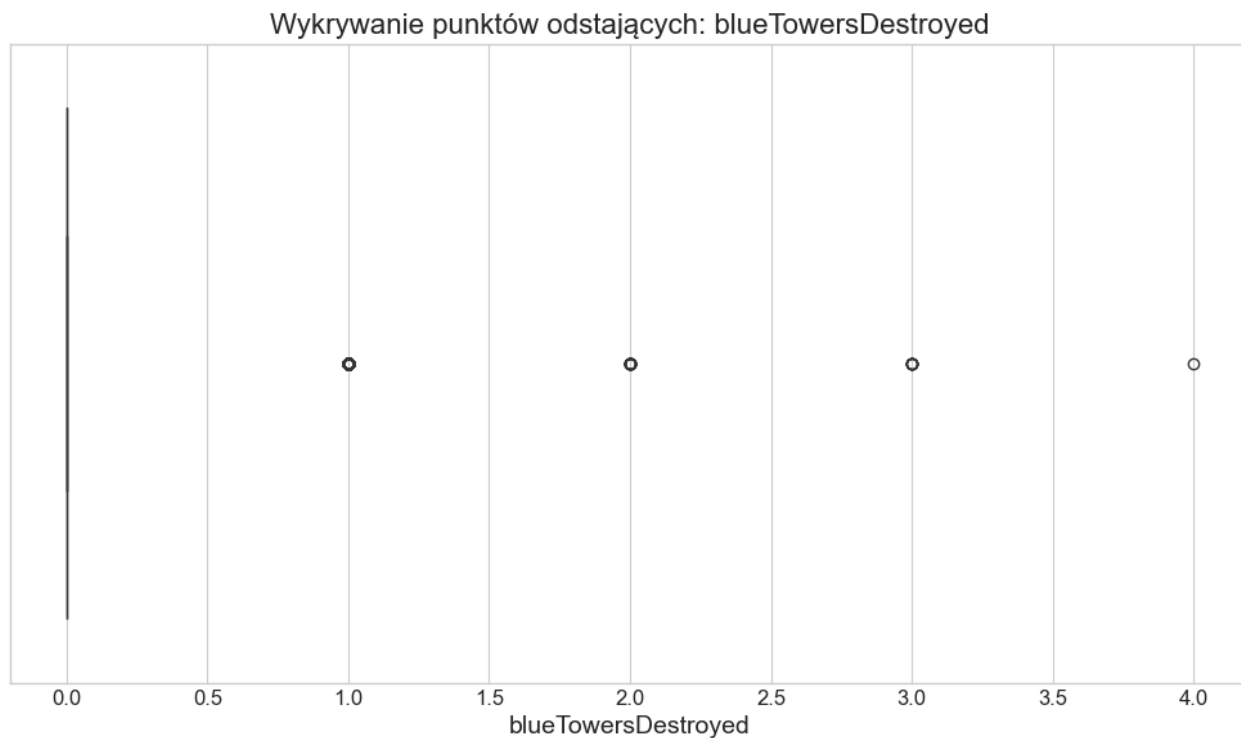


## Analiza punktów odstających

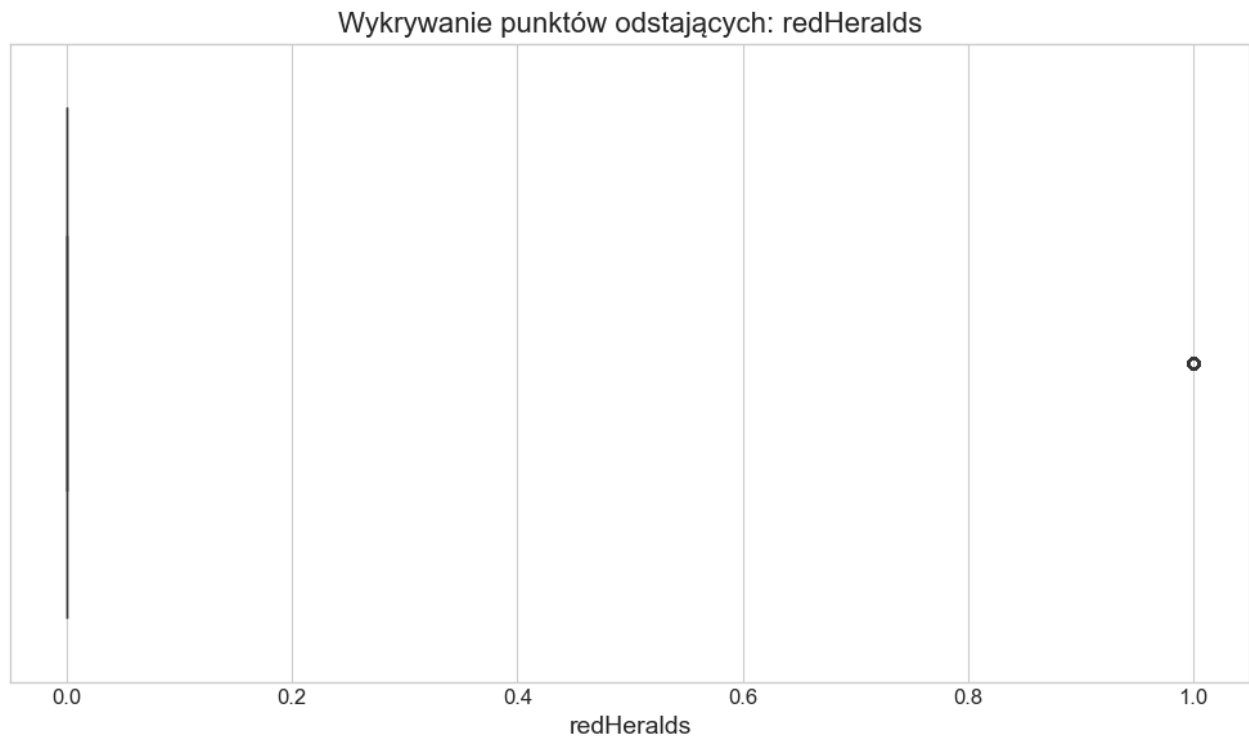
Stwierdzono obecność punktów odstających (outlierów) w kilku atrybutach, co jest naturalne w kontekście danych z gier. Ich szczegółowe zestawienie przedstawia się następująco:

- blueHeralds: **1857** outlierów (**18.80%**)
- blueTowersDestroyed: **464** outliery (**4.70%**)
- redHeralds: **1581** outlierów (**16.00%**)
- redTowersDestroyed: **396** outlierów (**4.01%**)
- blueKills: **88** outlierów (**0.89%**)
- redKills: **65** outlierów (**0.66%**)
- blueDragons i redDragons nie mają outlierów (0.00%).

Wartości odstające dla obiektów (heroldzi, wieże) są naturalne, ponieważ ich zdobycie lub zniszczenie nie zawsze następuje w każdej grze.







## Spójność danych

Weryfikacja spójności danych potwierdziła wysoką jakość zbioru:

- Suma zabójstw jednej drużyny odpowiada liczbie śmierci drugiej drużyny: **Prawda**.
- Średnia bezwzględna różnica pomiędzy `blueGoldDiff` i `redGoldDiff` wynosi: **0.0**.
- Średnia bezwzględna różnica pomiędzy `blueExperienceDiff` i `redExperienceDiff` wynosi: **0.0**.
- Dokładnie jedna drużyna ma `FirstBlood` w każdej grze: **Prawda**.

## Weryfikacja atrybutów czasowych

Przeprowadzono również weryfikację atrybutów czasowych, potwierdzając, że:

- `blueCSPerMin * 10` jest równe `blueTotalMinionsKilled` dla 100% gier.
- `blueGoldPerMin * 10` jest równe `blueTotalGold` dla 100% gier. Te zależności są zgodne z oczekiwaniami, co świadczy o prawidłowym obliczeniu tych cech.

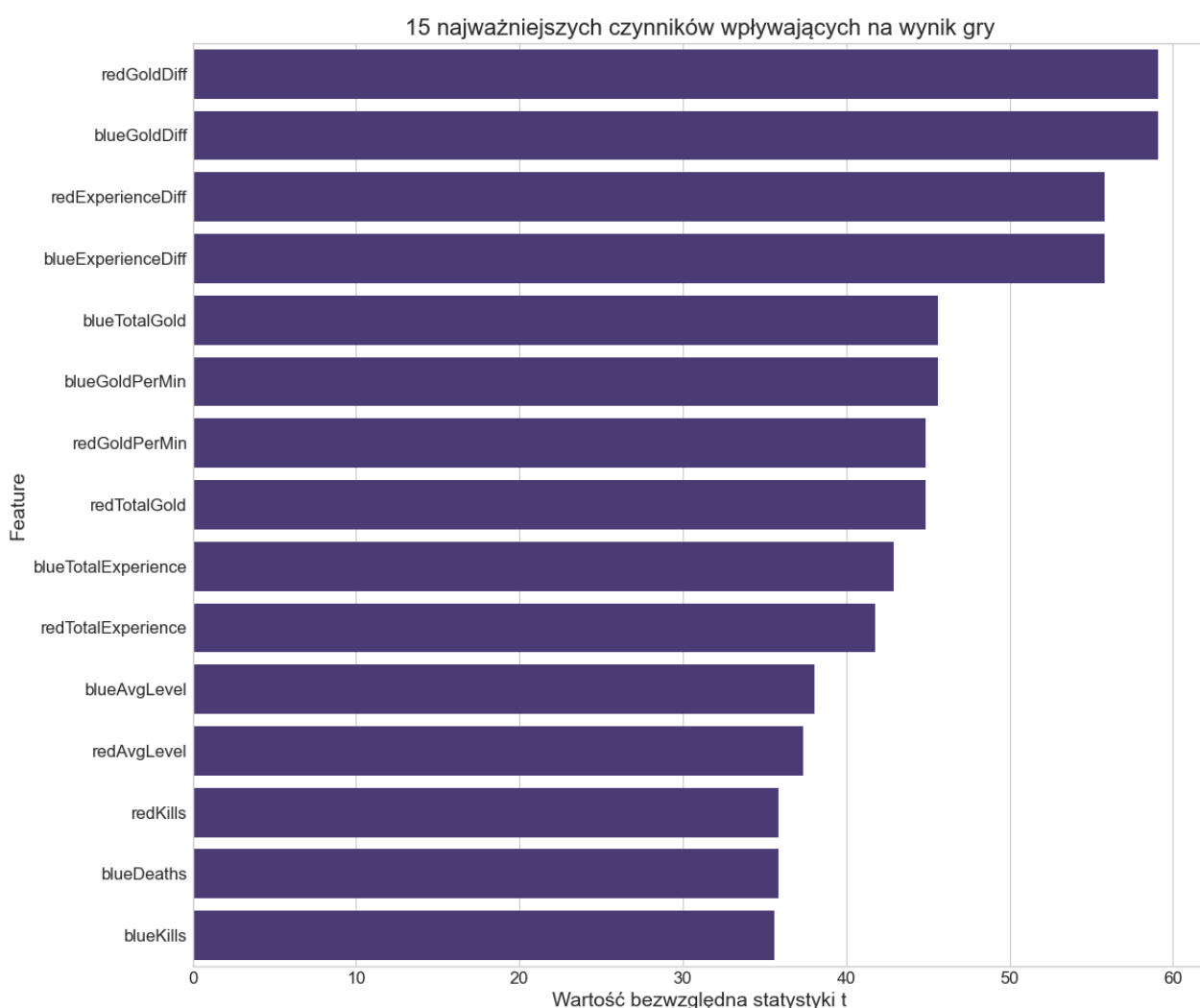
## Wnioski i ocena jakości danych

1. **Główne czynniki wpływające na zwycięstwo:** Przewaga w złocie (`blueGoldDiff`), przewaga w doświadczeniu (`blueExperienceDiff`), liczba zabójstw (`blueKills`) oraz kontrola nad obiektami (`blueDragons`, `blueHeralds`, `blueTowersDestroyed`) w pierwszych 10 minutach gry są kluczowymi czynnikami predykcyjnymi dla zwycięstwa drużyny niebieskiej. Testy statystyczne potwierdziły ich wysoką istotność.

## 2. Jakość danych: Dane są wysokiej jakości:

1. Brak wartości brakujących.
2. Wartości odstające są obecne, ale wydają się być naturalne dla dynamiki gry (np. rzadkie sytuacje z bardzo dużą liczbą zabitych potworów lub zniszczonych wież) i nie wymagają specjalnej obróbki.
3. Spójność danych została zweryfikowana (np. Kills jednej drużyny odpowiadają Deaths drugiej, sumy różnic są bliskie zeru, FirstBlood jest zawsze tylko jeden).
4. Atrybuty czasowe (CSPerMin, GoldPerMin) są spójne, potwierdzając prawidłowe obliczenia tych cech.

## 3. Przydatność danych do celów predykcyjnych: Dane z pierwszych 10 minut gry zawierają wystarczająco dużo informacji, aby budować skuteczne modele predykcyjne wyniku całej gry. Silne korelacje z zmienną docelową oraz zbalansowany rozkład blueWins czynią ten zbiór danych bardzo wartościowym.



## Podsumowanie i dalsze kroki

Przeprowadzona eksploracyjna analiza danych ze statystyk z pierwszych 10 minut gier League of Legends rangi Diamond I-Master miała na celu zidentyfikowanie kluczowych czynników wpływających na zwycięstwo drużyny niebieskiej. Zbiór danych, obejmujący **9879 gier i 40 atrybutów**, okazał się być **wysokiej jakości**, charakteryzując się brakiem wartości brakujących

oraz wewnętrzną spójnością (np. zgodność zabójstw/śmierci, różnic w złocie i doświadczeniu). Rozkład zmiennej docelowej (blueWins) był niemal idealnie zbalansowany (**49.90% zwycięstw niebieskich**), co stanowi solidną podstawę do dalszych analiz predykcyjnych.

## Kluczowe wnioski z analizy:

1. **Dominująca rola przewagi ekonomicznej:** Najsilniejsze korelacje ze zwycięstwem drużyny niebieskiej (blueWins) wykazują blueGoldDiff (0.51) oraz blueExperienceDiff (0.49). To jasno wskazuje, że akumulacja złota i doświadczenia w początkowej fazie gry jest **najważniejszym predyktorem** sukcesu. Drużyny wygrywające miały średnio ponad **1200 złota i 900 doświadczenia przewagi**.
2. **Agresja i kontrola obiektów:** Liczba zabójstw (blueKills, korelacja 0.34) oraz asyst (blueAssists, korelacja 0.28) jest również istotna. Co więcej, kontrola nad elitarnymi potworami (blueEliteMonsters, korelacja 0.22), a w szczególności smokami (blueDragons, korelacja 0.21) i zniszczonymi wieżami (blueTowersDestroyed, korelacja 0.12), znacząco przyczynia się do zwycięstwa. Wygrywające drużyny niebieskie niszczyły średnio **243% więcej wież** niż przegrywające.
3. **Spójność metryk:** Dane są logicznie powiązane; na przykład, zwiększona liczba zabójstw drużyny niebieskiej wiąże się ze wzrostem liczby śmierci drużyny czerwonej (korelacja blueKills z redDeaths jest taka sama jak redKills z blueDeaths co jest naturalne).
4. **Punkty odstające:** Odnotowano obecność punktów odstających, szczególnie w atrybutach związanych z liczbą Heroldów (około 18% dla blue team) i zniszczonych wież (około 4-5%). Z uwagi na kontekst gry (nie zawsze Herold czy wieża są zdobywane w pierwszych 10 minutach), są one uznane za naturalne i nie wymagają specjalnej obróbki.

## Implikacje dla strategii w grze:

Analiza silnie sugeruje, że we wczesnej fazie gry (pierwsze 10 minut) kluczowe jest agresywne zdobywanie przewagi ekonomicznej i doświadczenia, a także konsekwentna kontrola nad celami mapy takimi jak wieże i smoki. Skuteczna koordynacja w celu zdobywania zabójstw i asyst bezpośrednio przekłada się na lepszą pozycję do zwycięstwa.