

League of Legends Diamond Ranked Games

Eksploracja Danych Projekt 2025

Autorzy projektu: *Jan Krupiniewicz, Mateusz Fydrych, Marcin Araśniewicz*

Data opracowania: maj 2025

Źródło danych: *Kaggle – High Diamond Ranked Games (10 min)*

Link do zbioru: <https://www.kaggle.com/datasets/bobbyscience/league-of-legends-diamond-ranked-games-10-min>

Spis treści

League of Legends Diamond Ranked Games	1
Spis treści	2
Ogólny opis zbioru	3
Określenie celu eksploracji i kryteriów sukcesu	3
Charakterystyka zbioru danych	3
Wyniki eksploracyjnej analizy danych (EDA):	5
Rozkład zmiennej docelowej	5
Rozkłady atrybutów dla drużyny niebiskiej	5
Rozkłady atrybutów dla drużyny czerwonej	10
Analiza Korelacji	15
Macierz korelacji	15
Korelacje ze zmienną docelową	18
Opis wyników EDA w odniesieniu do celów eksploracji	18
Porównanie statystyk dla meczy wygranych i przegranych	18
Praktyczne wskazówki dla graczy	21
Uwagi nt. jakości danych:	24
Dane brakujące	24
Dane niespójne	24
Dane niezrozumiałe	24
Punkty oddalone	25
Wybrane wykresy punktów oddalonych	26
Ewentualna rewizja celów	31
Kluczowe wnioski z analizy	31
Dalsze kroki	31

Ogólny opis zbioru

Niniejszy projekt opiera się na analizie zbioru danych z gry **League of Legends (LoL)** — jednej z najpopularniejszych gier z gatunku **MOBA** (Multiplayer Online Battle Arena). W grze rywalizują ze sobą dwie pięcioosobowe drużyny: **niebieska (blue)** i **czerwona (red)**. Celem gry jest zniszczenie głównej struktury przeciwnika – **Nexusa**. Rozgrywka toczy się na mapie podzielonej na trzy linie, strefę dżungli oraz przypisane role dla każdej postaci.

Określenie celu eksploracji i kryteriów sukcesu

Celem analizy jest przewidzenie wyniku meczu w grze **League of Legends** na podstawie danych zbieranych po **pierwszych 10 minutach rozgrywki**. Interesuje nas binarny wynik: czy drużyna **blue** wygra mecz ($\text{blueWins} = 1$) czy nie ($\text{blueWins} = 0$). Zbiór danych zawiera **9 879 gier rankingowych** na wysokim poziomie umiejętności (od DIAMOND I do MASTER), gdzie poziom graczy w obu drużynach jest względnie wyrównany.

Kryteria sukcesu:

- Przygotowanie danych pod implementację modelu predykcyjnego`.
- Analiza znaczenia cech – **zidentyfikowanie statystycznie istotnych korelacji** między zmiennymi a wynikiem gry.
- Opracowanie wniosków - jasne wytyczne które można przełożyć na **praktyczne wskazówki dla graczy**.
- Wskazanie **różnic w statystykach między wygrywającymi a przegrywającymi** drużynami.

Charakterystyka zbioru danych

Dane pochodzą z publicznie dostępnego zbioru opublikowanego na Kaggle, w ramach licencji **CC0 (Public Domain)**. Dane zostały zebrane przy pomocy **API Riot Games**, które umożliwia pobieranie informacji o meczach rozgrywanych w grze League of Legends.

Każdy wiersz w pliku `high_diamond_ranked_10min.csv` reprezentuje jedną rozgrywkę. Klucz główny to `gameId`, który jednoznacznie identyfikuje mecz i umożliwia pobranie dodatkowych danych z Riot API.

Każda obserwacja reprezentuje jeden mecz i zawiera:

1. **38 cech (features)** – po **19 dla każdej drużyny**, m.in. liczba zabójstw, śmierci, zdobytego złota, doświadczenia, poziomu postaci, liczby zabitych stworów (minions), potworów z dżungli, smoków, heroldów oraz zniszczonych wież.
2. **Zmienna docelowa (blueWins)** – wartość binarna wskazująca, czy drużyna niebieska wygrała mecz (1) lub nie (0).

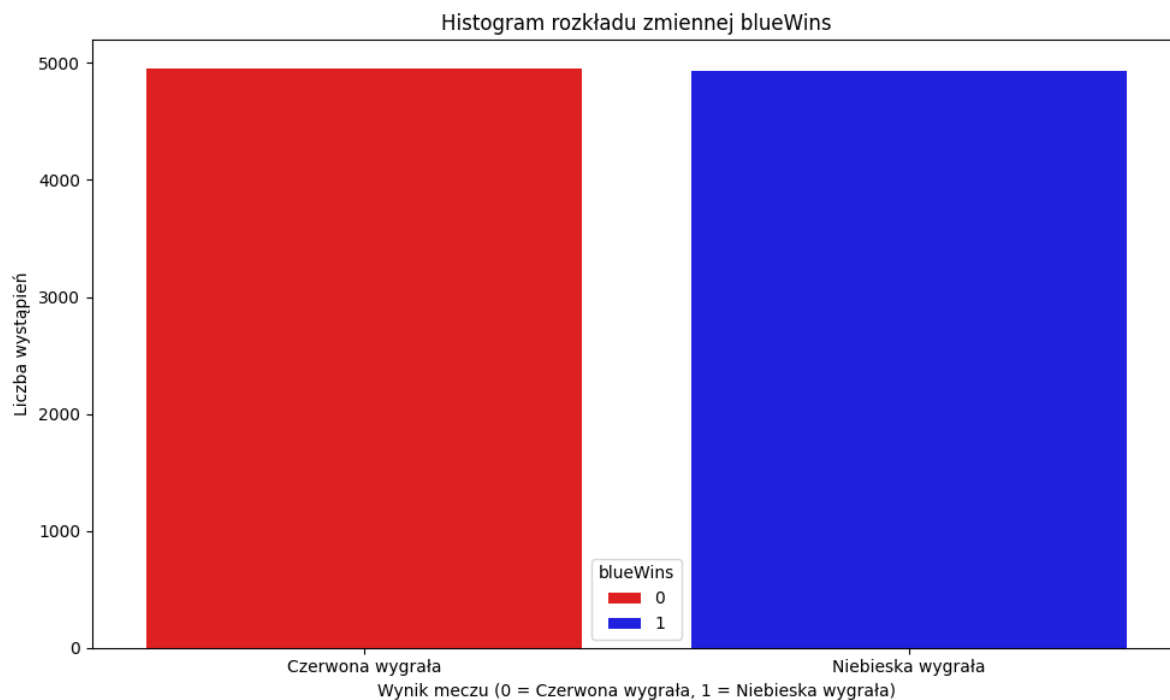
Poniżej opisze kilka przykładowych atrybutów danych które będą przytaczane najczęściej w późniejszej analizie. Jako wartości najciekawsze:

Nazwa kolumny	Opis
<code>gameId</code>	Unikalny identyfikator meczu, możliwy do powiązania z Riot API.
<code>blueWins</code>	Zmienna docelowa: 1 – wygrana drużyny niebieskiej, 0 – przegrana.
<code>blueAssists</code>	Liczba asyst przy zabójstwach w drużynie niebieskiej.
<code>blueAvgLevel</code>	Średni poziom doświadczenia graczy drużyny niebieskiej.
<code>blueDeaths</code>	Liczba śmierci graczy drużyny niebieskiej.
<code>blueTotalExperience</code>	Łączna ilość zdobytego doświadczenia przez drużynę niebieską.
<code>blueTotalGold</code>	Łączna ilość zdobytego złota przez drużynę niebieską.
<code>blueGoldPerMin</code>	Ilość złota zdobywana przez drużynę niebieską na minutę.
<code>blueWardsPlaced</code>	Liczba totemów wizji (wardów) ustawionych przez drużynę niebieską.
<code>blueCsPerMin</code>	Ilość zabitych stworów i jednostek neutralnych na minutę przez drużynę niebieską (Creep Score).
<code>blueTotalMinions Killed</code>	Łączna liczba zabitych stworów i jednostek neutralnych przez drużynę niebieską.

Wyniki eksploracyjnej analizy danych (EDA):

Rozkład zmiennej docelowej

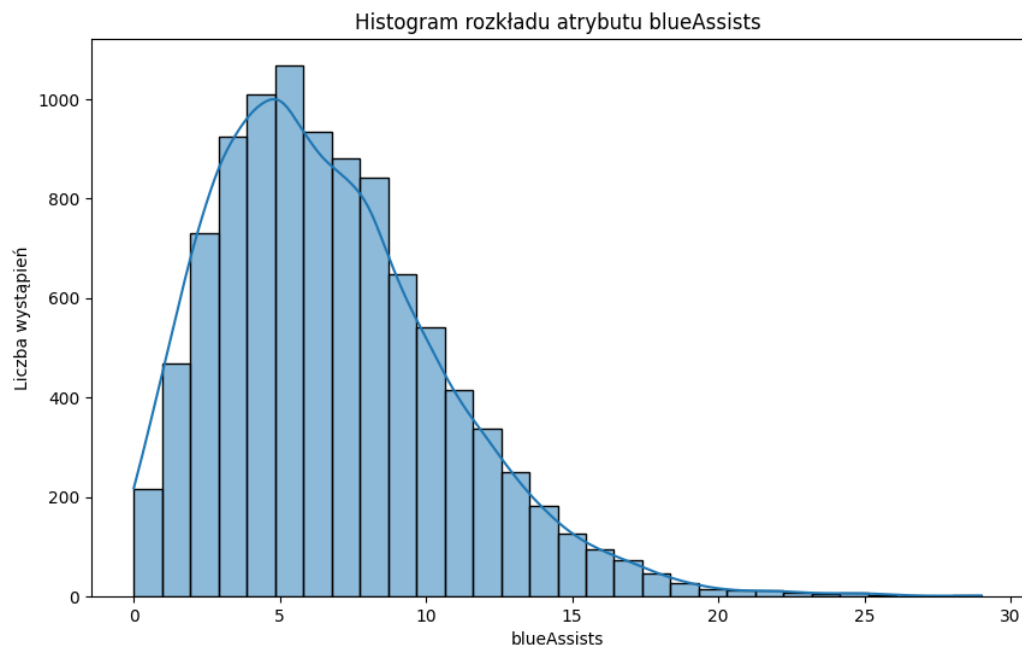
Rozkład zmiennej docelowej przedstawiono poniżej. Według danych z podanego zbioru, **drużyna czerwona wygrywała minimalnie częściej** - w sumie **w 4949 meczach**. Z drugiej strony drużyna **niebieska dominowała 4930 razy**.



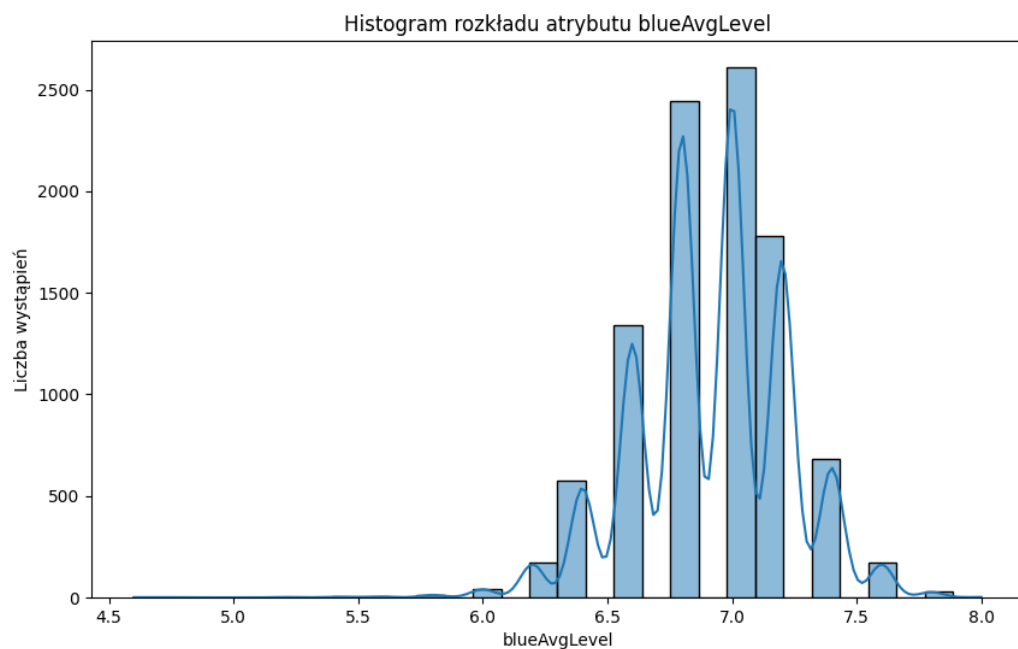
Rozkłady atrybutów dla drużyny niebieskiej

Rozkład najczęstszej liczby asyst po pierwszych 10 minutach rozgrywki.

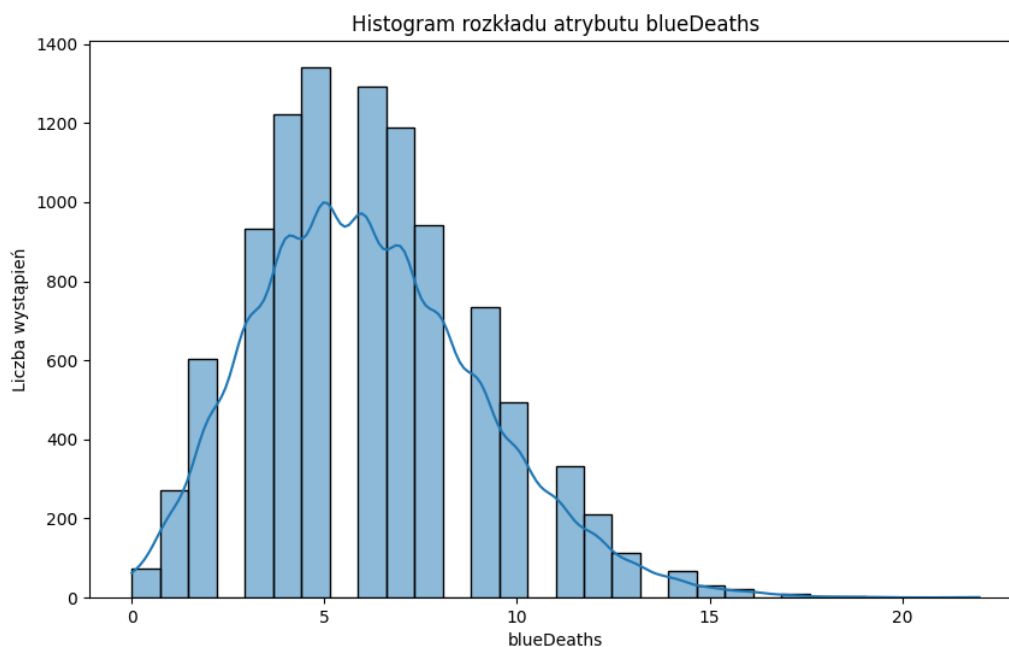
Według danych z podanego zbioru, liczba asyst po pierwszych 10 minutach rozgrywki najczęściej oscylowała w okolicy 6.



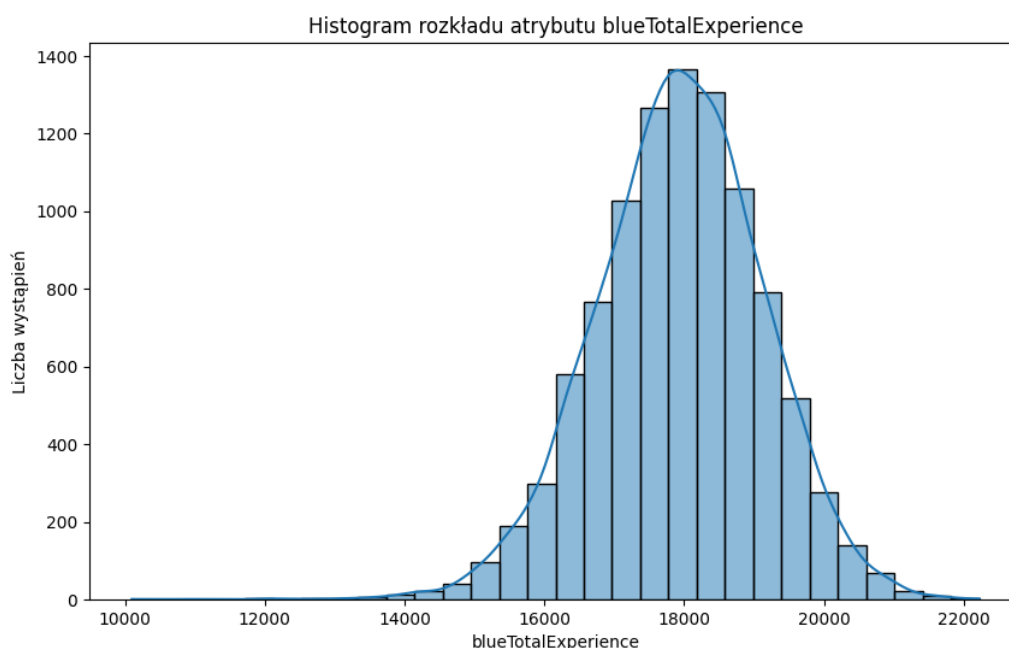
Rozkład średniego poziomu graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, średni poziom najczęściej oscylował w okolicy 7.



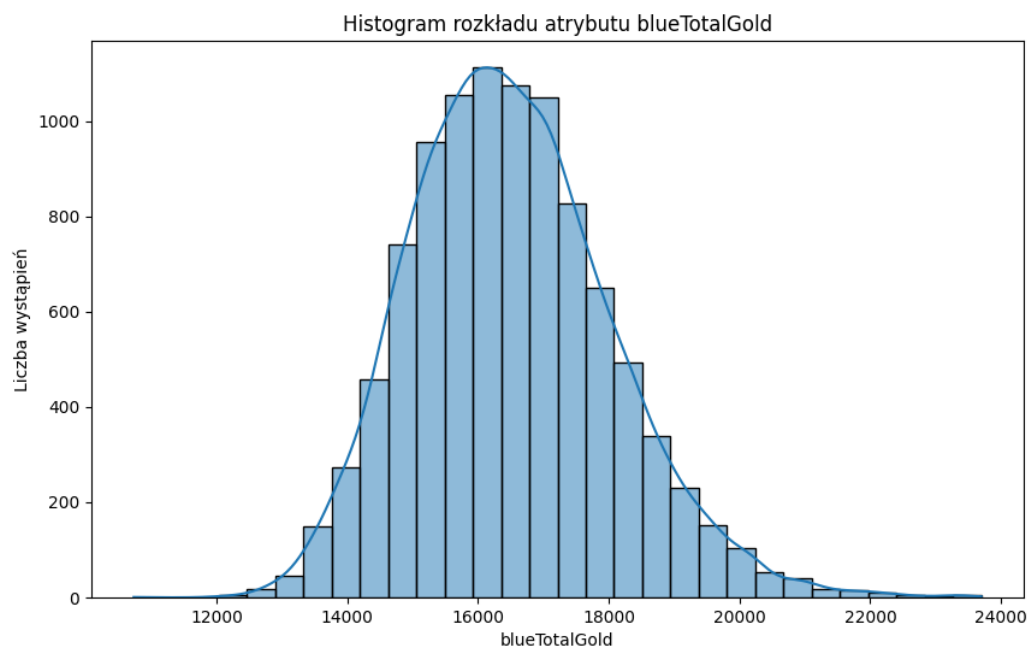
Rozkład ilości śmierci graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość śmierci najczęściej oscylowała w okolicy od 4 do 7.



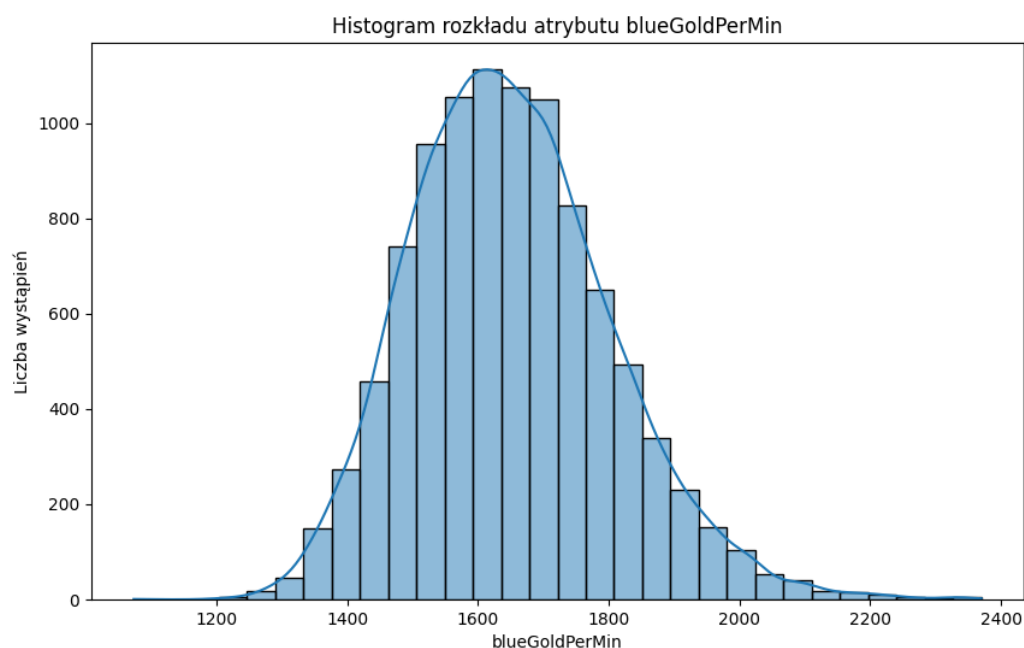
Rozkład całkowitej ilości doświadczenia graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, całkowita ilość doświadczenia najczęściej oscylowała w okolicy 18 tysięcy.



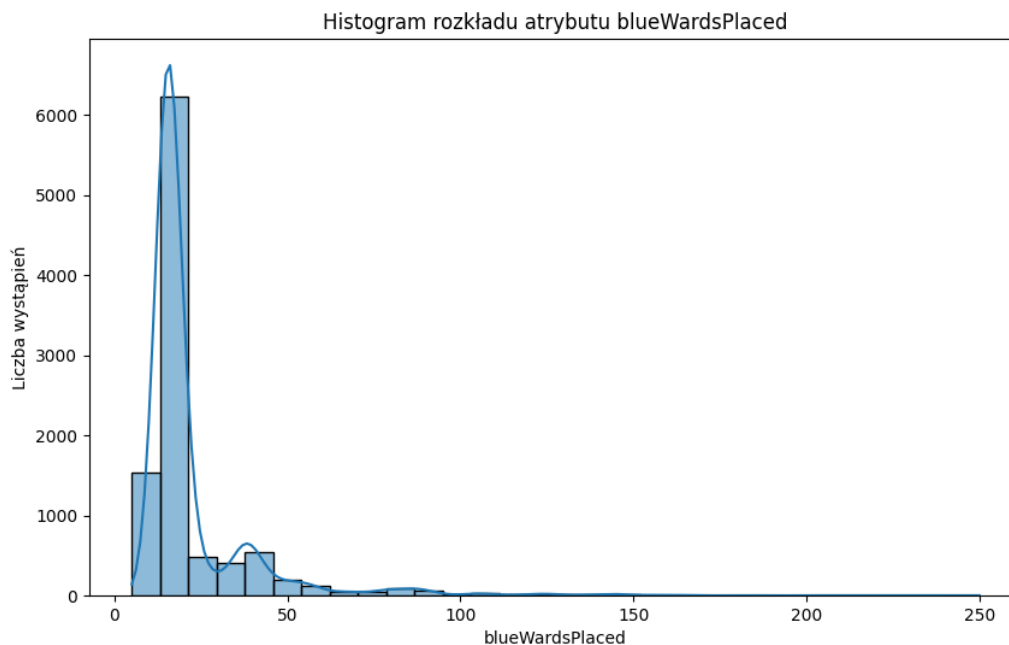
Rozkład całkowitej ilości złota graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, całkowita ilość złota najczęściej oscylowała w okolicy 16 tysięcy.



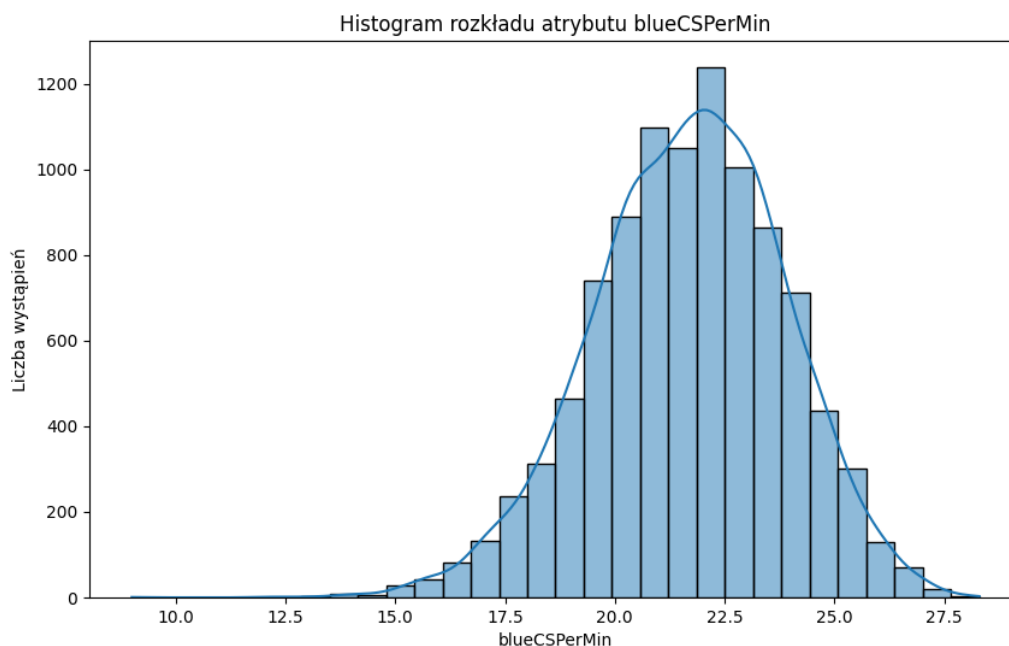
Rozkład złota graczy po stronie niebieskiej na minutę po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość złota na minutę najczęściej oscylowała w okolicy 1.6 tysiąca.



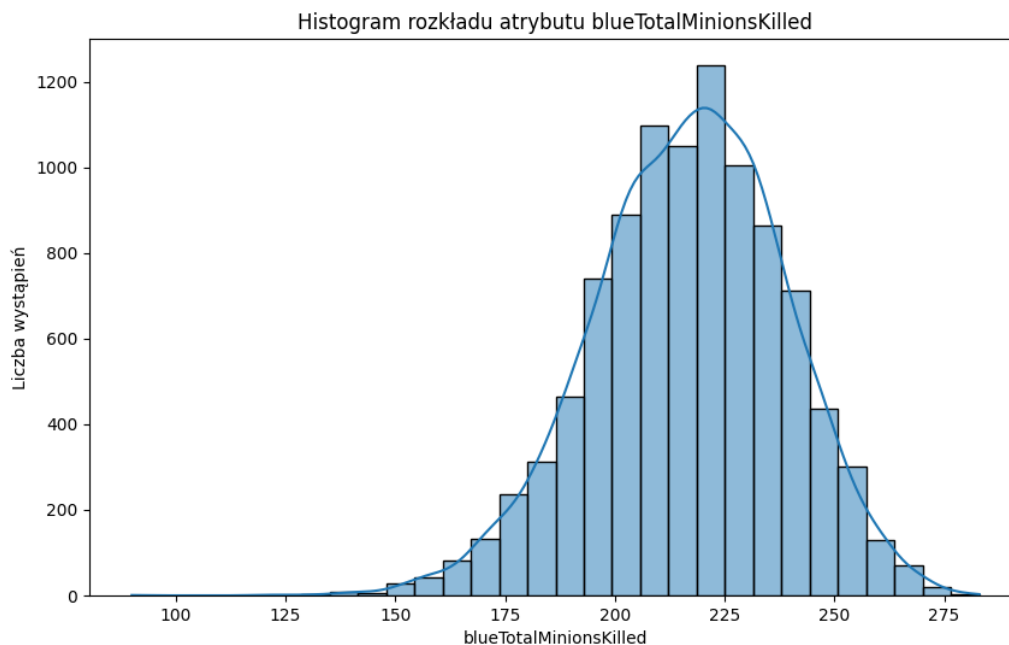
Rozkład ilości położonych totemów przez graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość położonych totemów najczęściej oscylowała w okolicy 15.



Rozkład ilości zabitych stworów na minutę przez graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość zabitych stworów na minutę najczęściej oscylowała w okolicy 22.

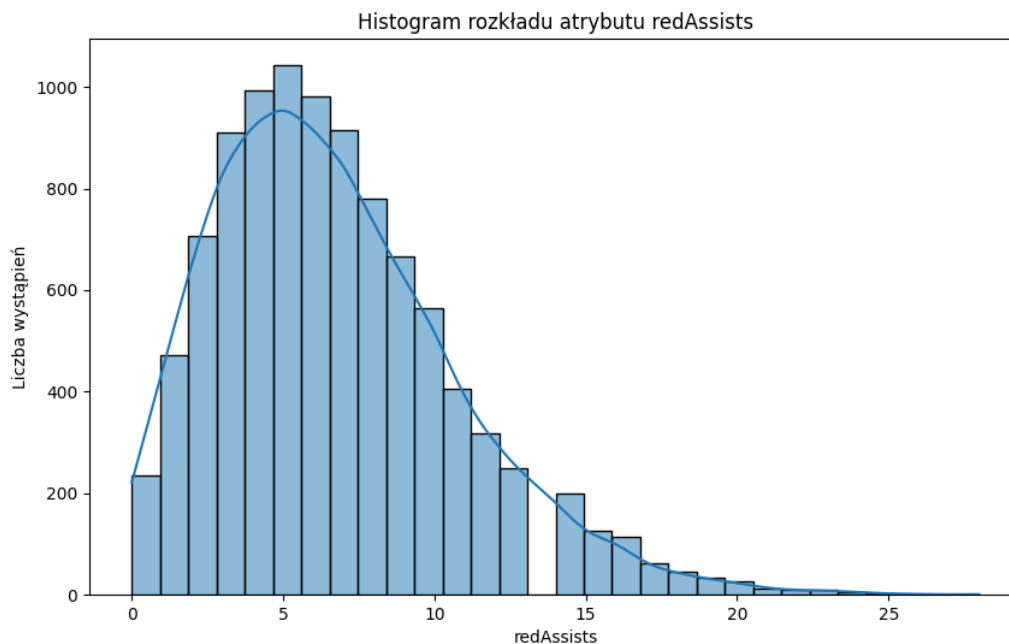


Rozkład całkowitej ilości zabitych stworów przez graczy po stronie niebieskiej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, całkowita ilość zabitych stworów najczęściej oscylowała w okolicy 224.

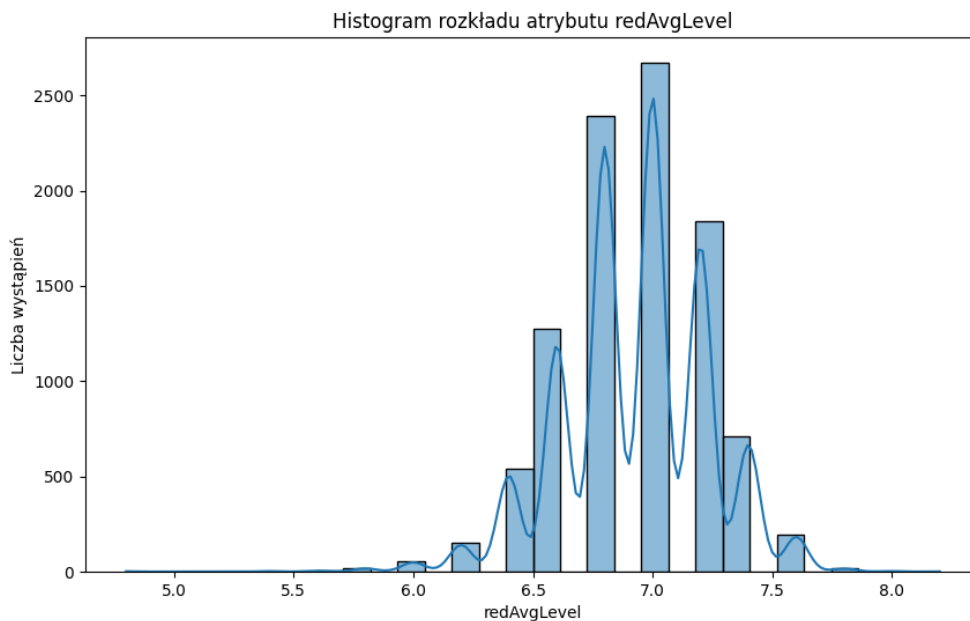


Rozkłady atrybutów dla drużyny czerwonej

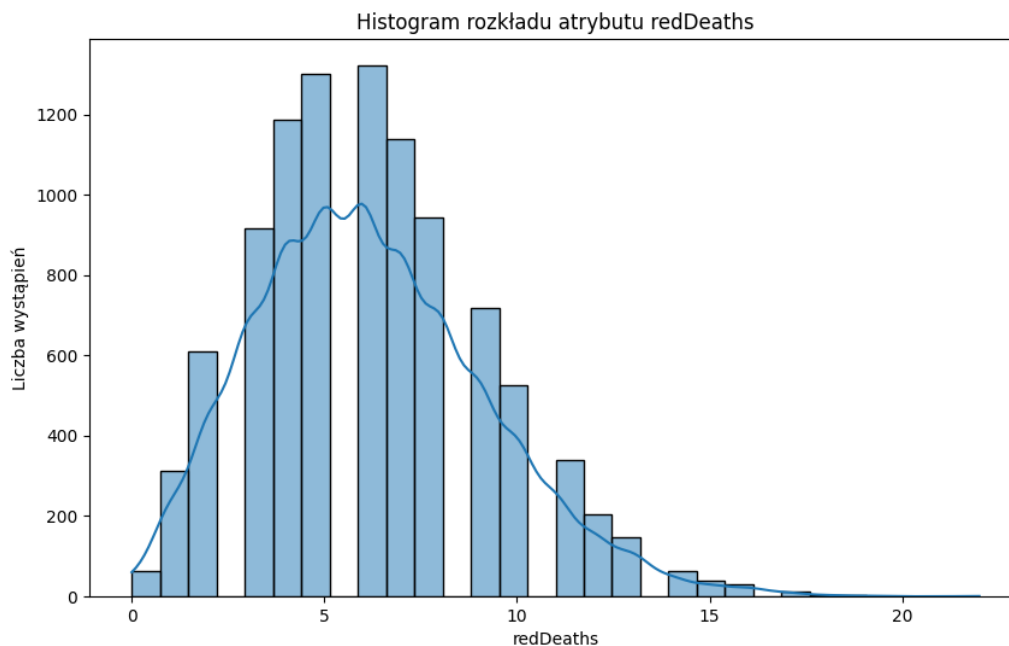
Rozkład najczęstszej liczby asyst graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, liczba asyst po pierwszych 10 minutach rozgrywki najczęściej oscylowała w okolicy 6.



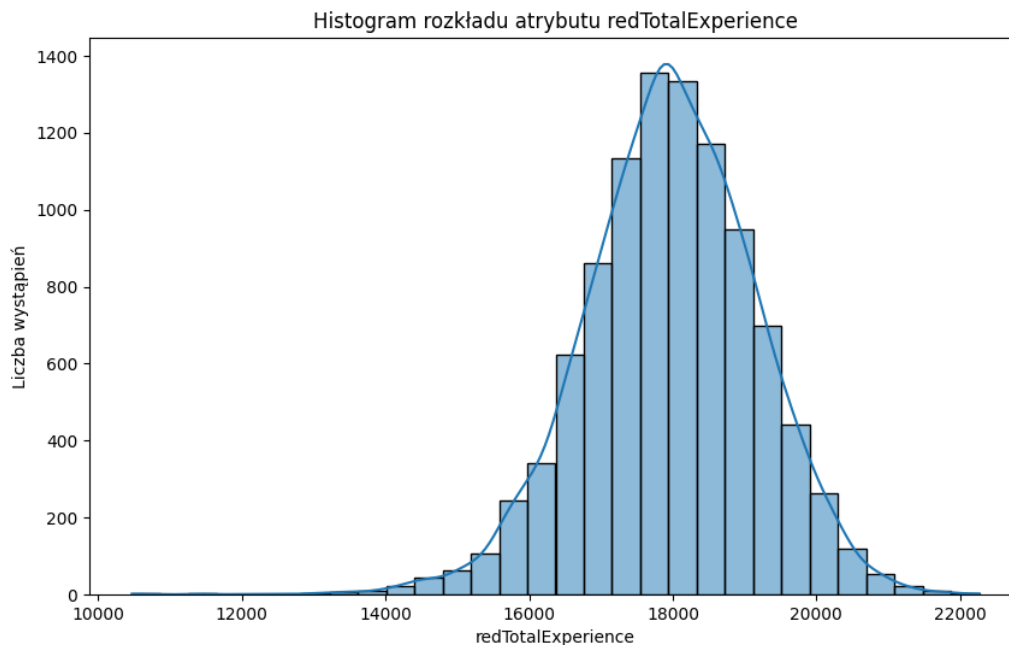
Rozkład średniego poziomu graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, średni poziom najczęściej oscylował w okolicy 7.



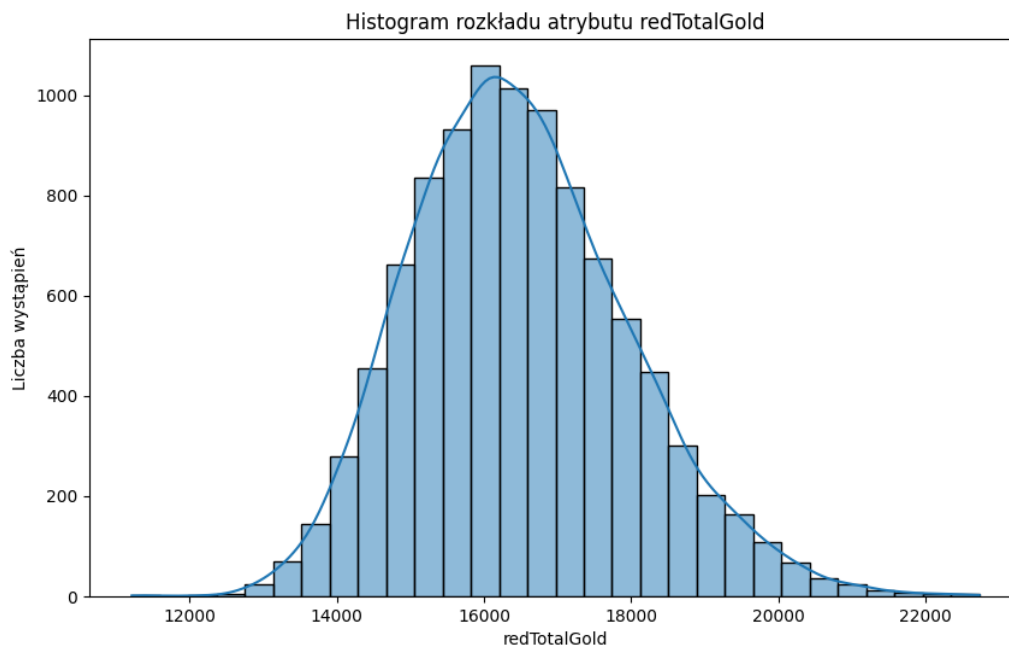
Rozkład ilości śmierci graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość śmierci najczęściej oscylowała w okolicy od 4 do 7.



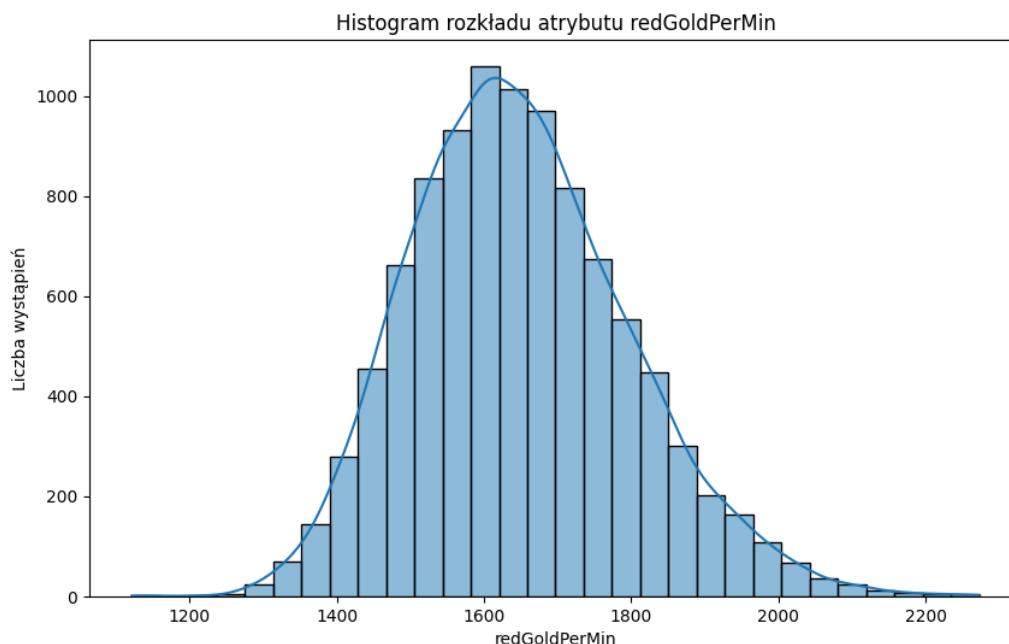
Rozkład całkowitej ilości doświadczenia graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, całkowita ilość doświadczenia najczęściej oscylowała w okolicy 18 tysięcy.



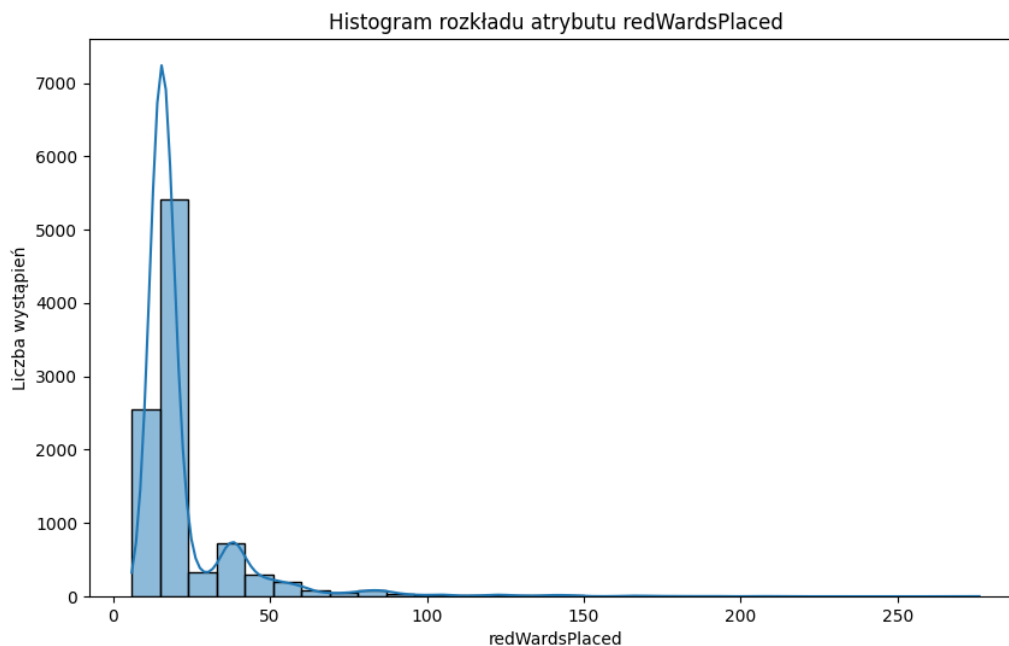
Rozkład całkowitej ilości złota graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, całkowita ilość złota najczęściej oscylowała w okolicy 16 tysięcy.



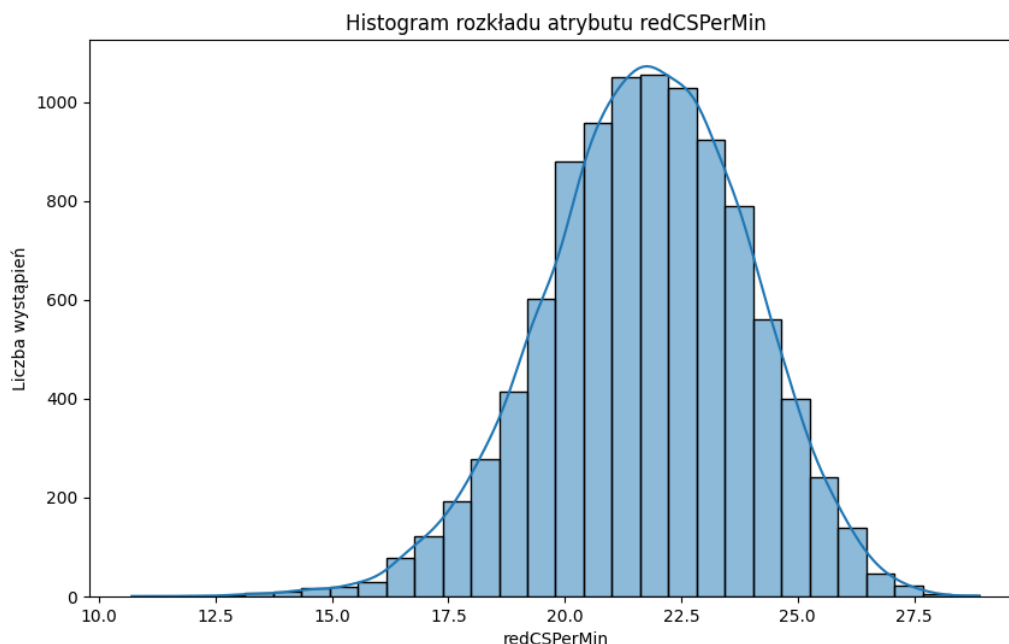
Rozkład złota graczy po stronie czerwonej na minutę po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość złota na minutę najczęściej oscylowała w okolicy 1.6 tysiąca.



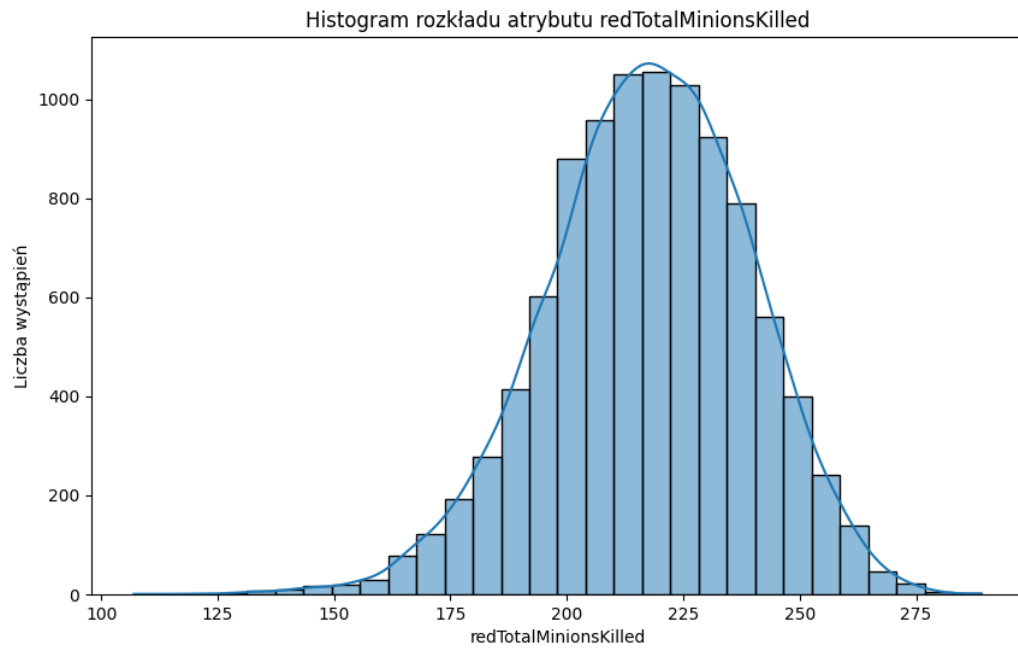
Rozkład ilości położonych totemów przez graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość położonych totemów najczęściej oscylowała w okolicy 15.



Rozkład ilości zabitych stworów na minutę przez graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, ilość zabitych stworów na minutę najczęściej oscylowała w okolicy 22.



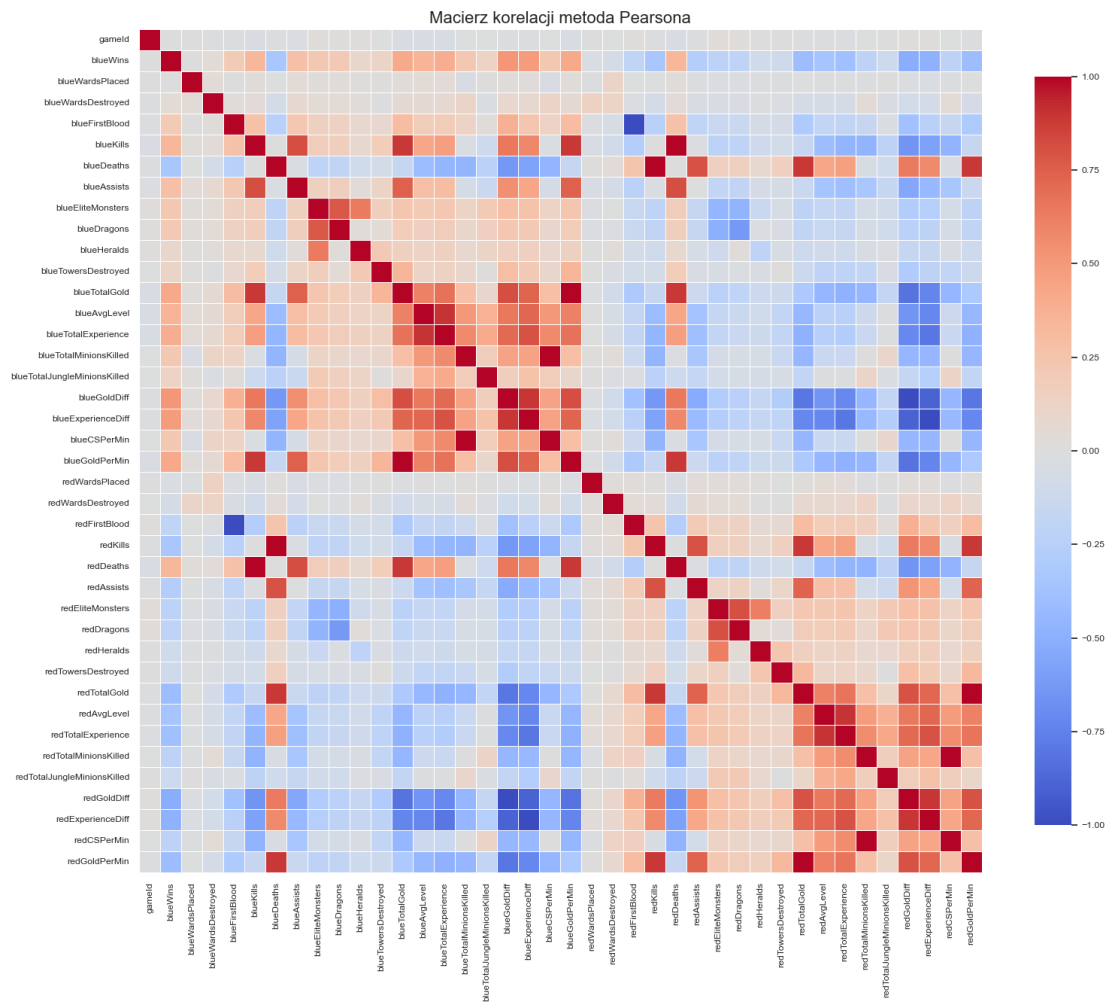
Rozkład całkowitej ilości zabitych stworów przez graczy po stronie czerwonej po pierwszych 10 minutach rozgrywki. Według danych z podanego zbioru, całkowita ilość zabitych stworów najczęściej oscylowała w okolicy 224.

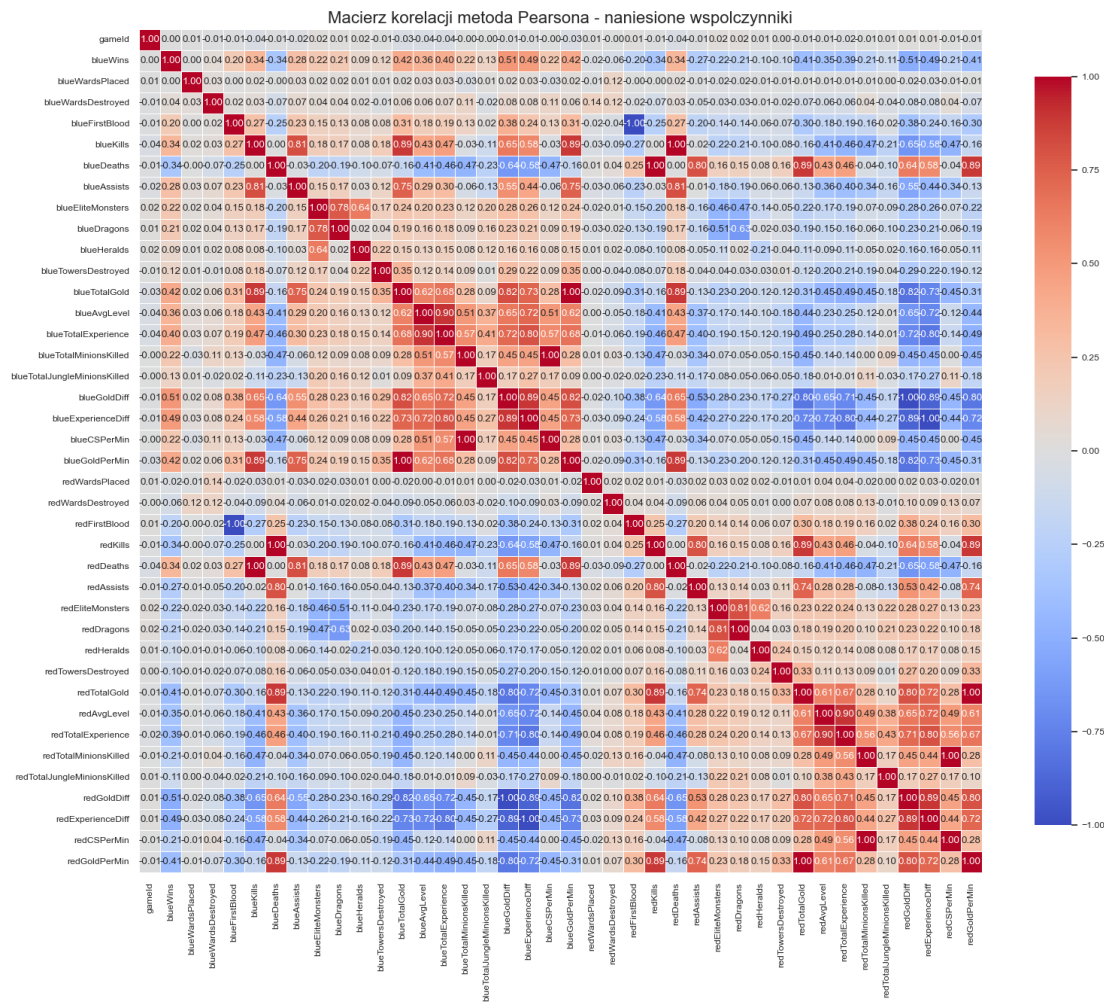


Analiza Korelacji

Macierz korelacji

Korelacja **mierzona współczynnikiem Pearsona** (zakres od -1 do 1). Poniżej otrzymana macierz korelacji dla wszystkich zmiennych liczbowych :





Najsilniejsze dodatnie korelacje z blueWins:

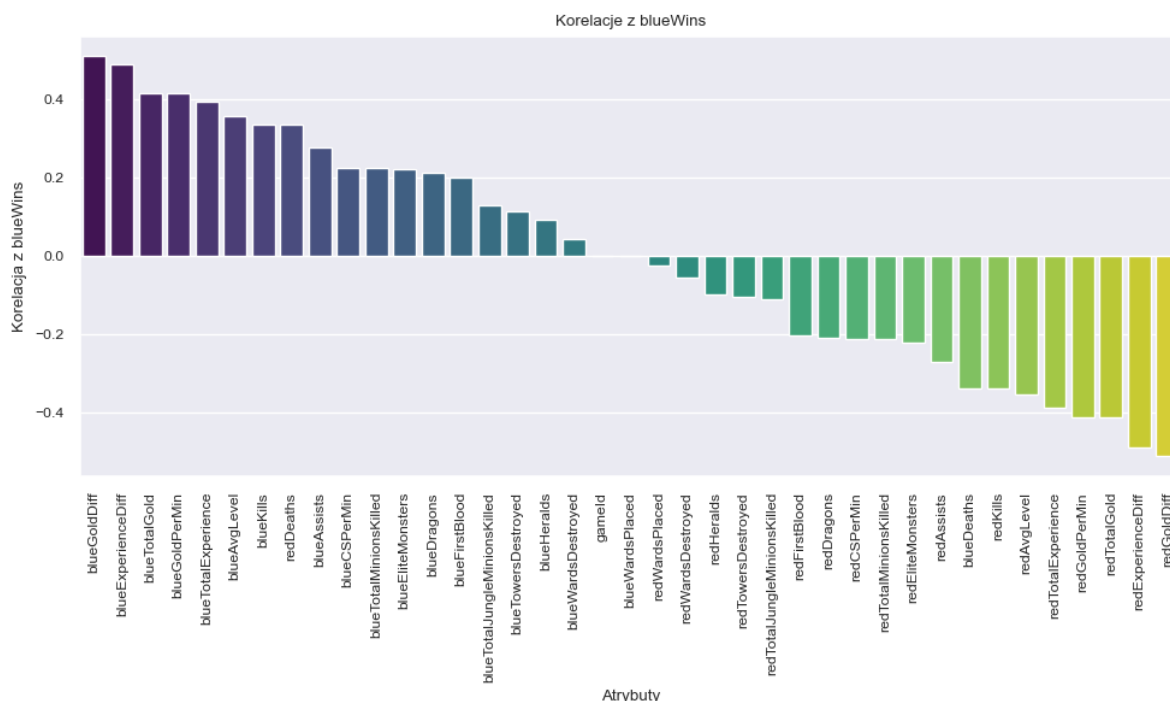
1. blueGoldDiff: 0.92
2. blueExperienceDiff: 0.87
3. blueTowersDestroyed: 0.75
4. blueEliteMonsters: 0.66
5. blueKills: 0.62
6. blueTotalGold: 0.60

Najsilniejsze ujemne korelacje z blueWins:

7. redGoldDiff: -0.92
8. redExperienceDiff: -0.87
9. redTowersDestroyed: -0.75
10. redEliteMonsters: -0.66
11. redKills: -0.62

Nie dziwi nas również jakoby odbicie lustrzane korelacji. **Ujemne korelacje są odbiciem dodatnich dla przeciwnej drużyny.** Na przykład, `blueGoldDiff` ma korelację 0.511 z `blueWins`, podczas gdy `redGoldDiff` (która jest równa - `blueGoldDiff`) ma korelację -0.511. **To naturalne w grze o sumie zerowej** pod względem przewagi.

Korelacje ze zmienną docelową

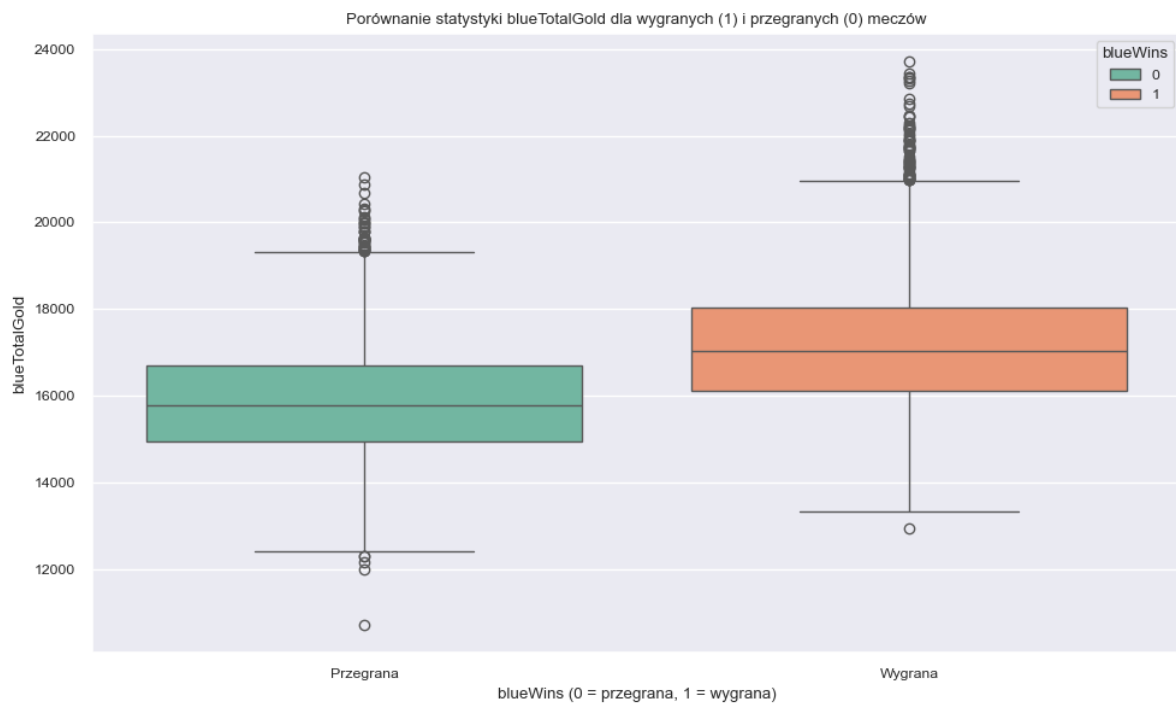


Najsilniejsze korelacje (zarówno dodatnie dla statystyk drużyny niebieskiej, jak i ujemne dla statystyk drużyny czerwonej) ze zmienną `blueWins` **wykazują atrybuty związane z różnicą w zdobytych złocie (`blueGoldDiff`) oraz doświadczeniu (`blueExperienceDiff`).** Potwierdza to intuicyjne założenie, że zdobycie przewagi ekonomicznej i w doświadczeniu w pierwszych 10 minutach gry jest silnym prognostykiem zwycięstwa.

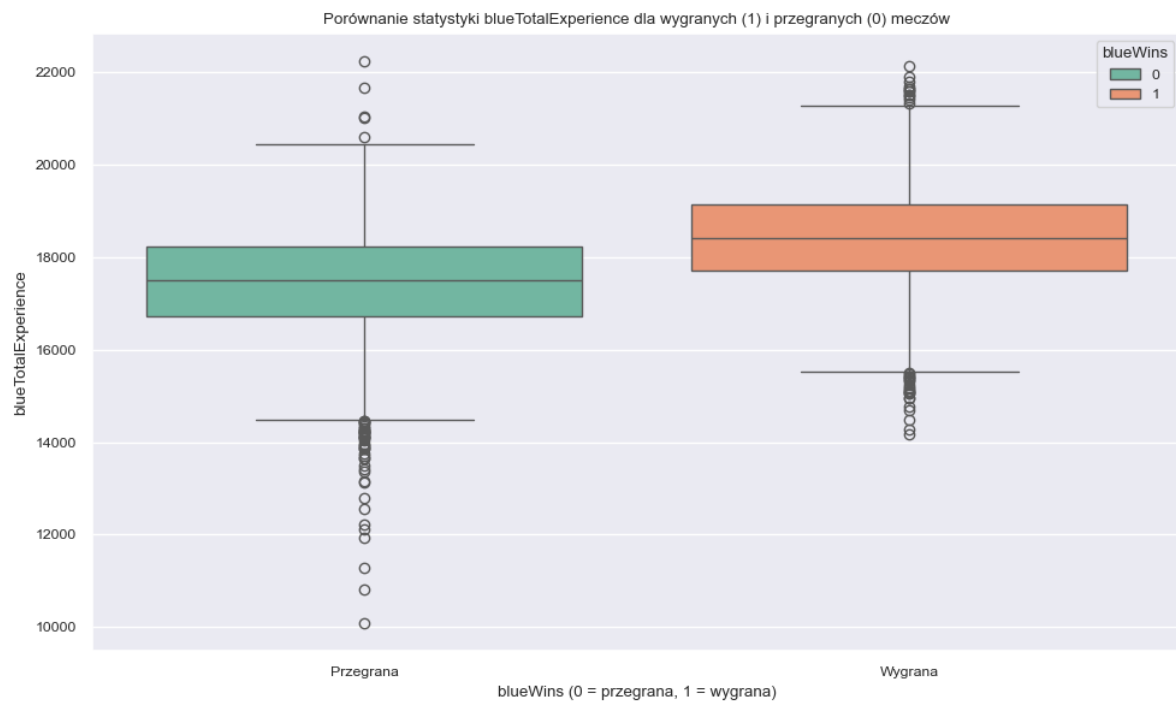
Opis wyników EDA w odniesieniu do celów eksploracji

Porównanie statystyk dla meczy wygranych i przegranych

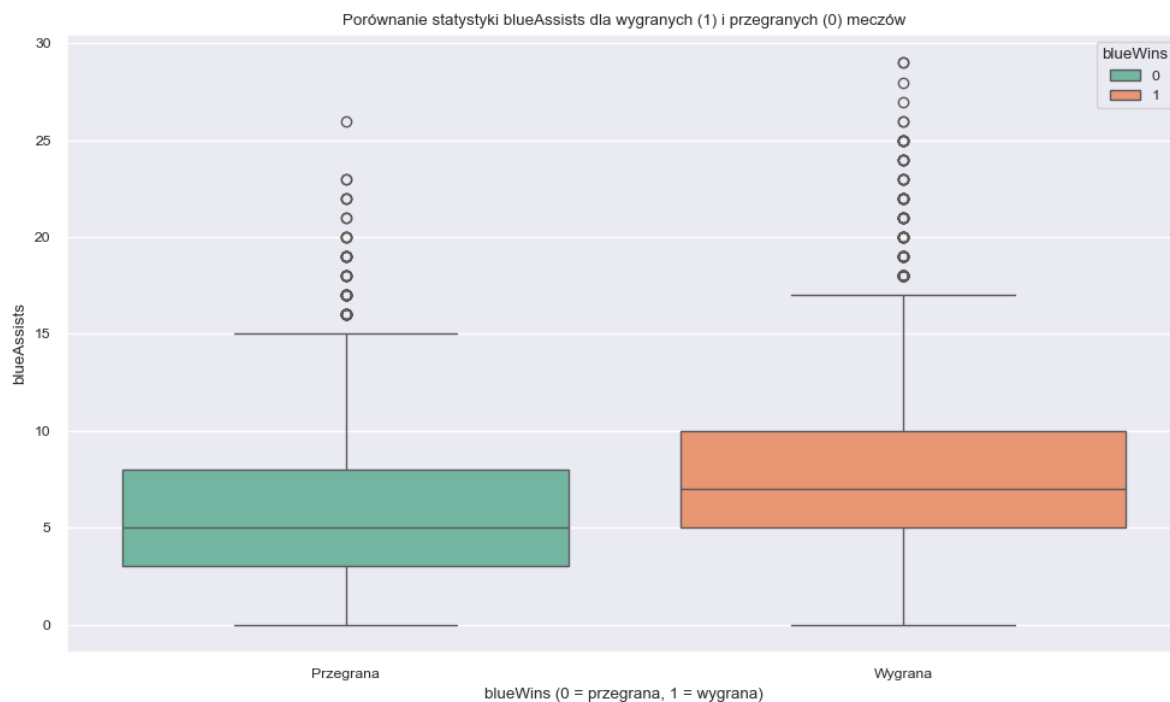
Im więcej złota tym częściej wygrywa się mecze.



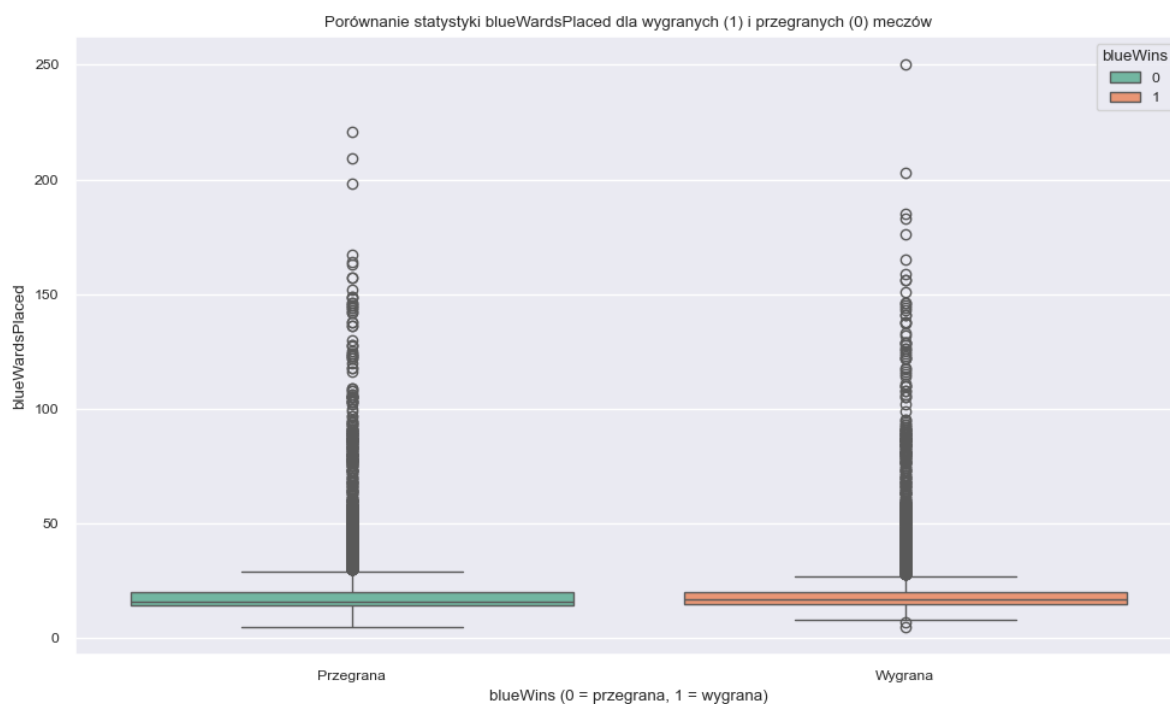
Im więcej doświadczenia tym częściej wygrywa się mecze.



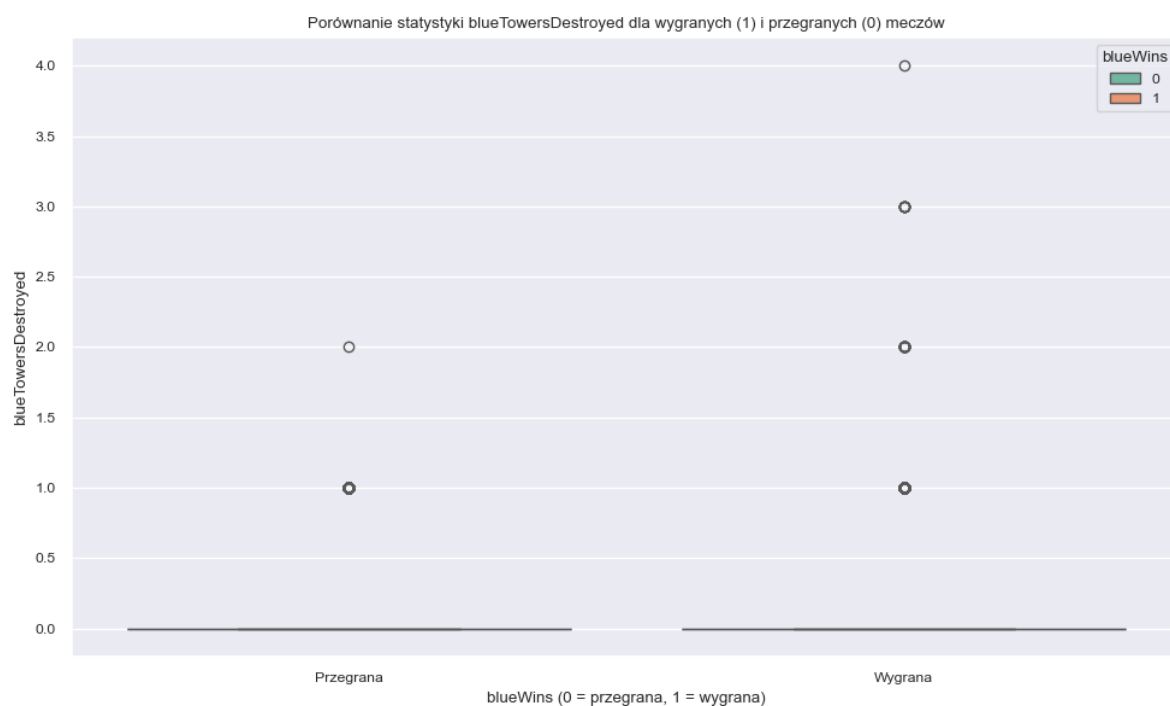
Im więcej asyst tym częściej wygrywa się mecze.



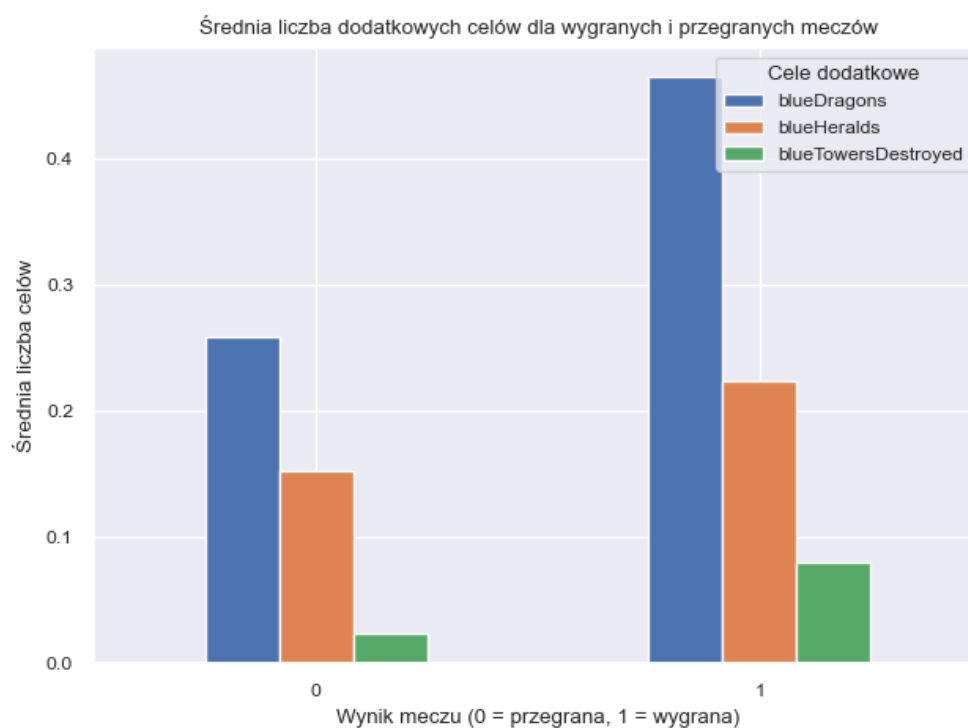
Liczba postawionych totemów do 10 minuty meczu nie przekłada się aż tak na wygraną.

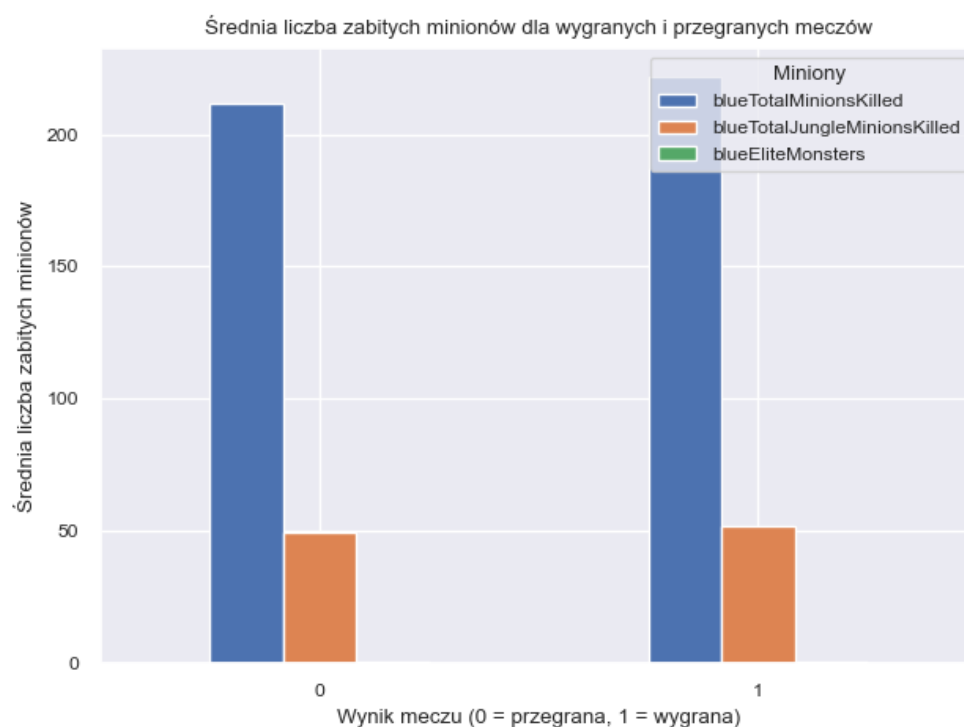
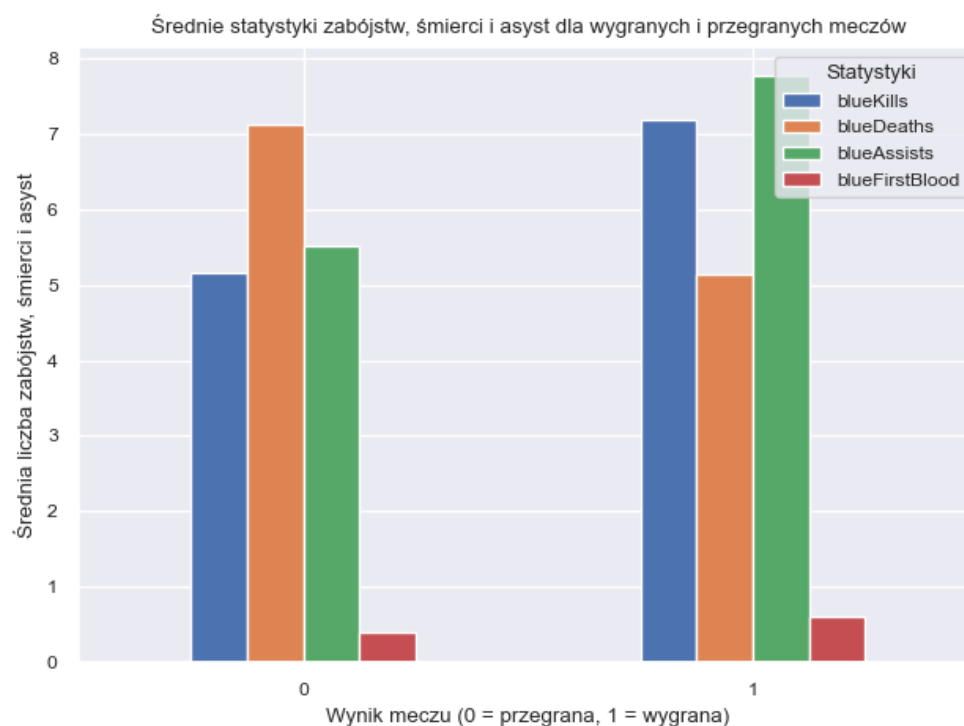


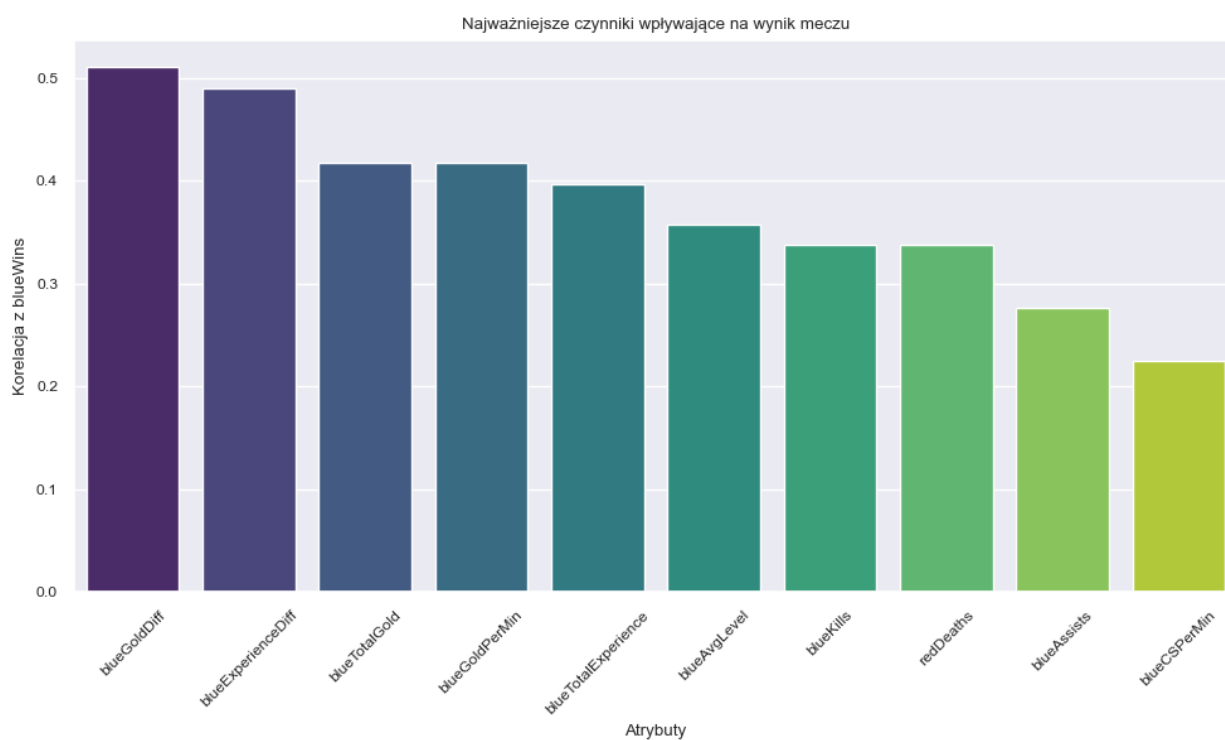
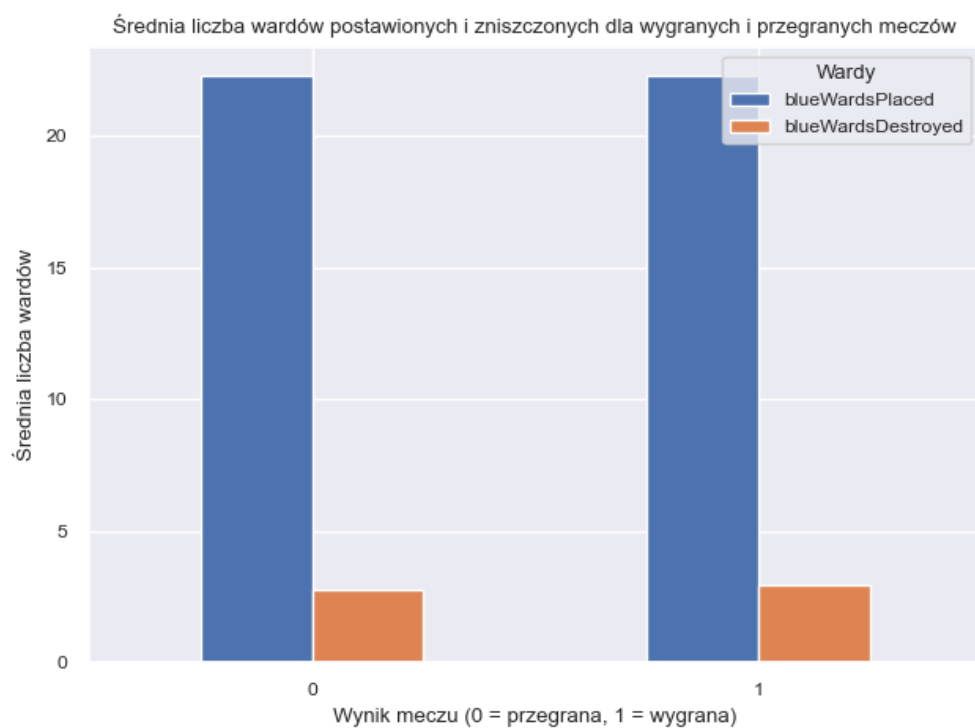
Liczba zniszczonych totemów do 10 minuty meczu nie przekłada się aż tak na wygraną.



Praktyczne wskazówki dla graczy







Możemy zatem na bazie analizy danych przedstawić kilka wartych uwagi reguł:

1. **Bezpieczeństwo przede wszystkim:** Nawet agresywna gra musi uwzględniać bezpieczeństwo. Unikanie niepotrzebnych śmierci ("feeding") jest priorytetem, ponieważ oddanie złota i doświadczenia przeciwnikowi wzmacnia go.
2. **Last Hitting (Dobijanie Stworów):** To podstawowy sposób na zdobywanie złota. Uderzanie w stwora w ostatniej chwili, gdy ma mało zdrowia, jest kluczowe dla efektywnego farmienia.
3. **Reagowanie na cele:** Smoki (Dragons) i Heroldowie (Heralds) są pierwszymi dużymi celami na mapie. Zdolność do ich zabezpieczenia może dać dużą przewagę - jak widzimy ponad 4 na 10 drużyn które zdobędą pierwszego smoka są na dobrej drodze do zwycięstwa.

Z drugiej jednak strony, na podstawie danych kontrola wizji we wczesnych fazach gry przekłada się na dominację w mniejszym stopniu. Poza tym są pewne aspekty które ciężko ocenić takie jak ocenienie swoich szans przeciwko oponentowi. Dla opisanych zagadnień opisałem mimo wszystko jeszcze dwie zasady:

1. **Kontrola wizji (Warding):** Odpowiednie rozmieszczenie totemów (wardów) jest absolutnie krytyczne. Pozwala uniknąć zasadzek (ganków), śledzić ruchy junglera przeciwnika i planować własne zagrania.
2. **Zrozumienie Matchupu:** Każdy bohater ma swoje silne i słabe strony w porównaniu do bohatera przeciwnego. Zrozumienie, czy masz przewagę (jesteś silniejszy) czy musisz grać defensywnie, jest fundamentalne.

Uwagi nt. jakości danych:

Dane brakujące

Brakujących danych **nie wykryto**.

Dane niespójne

Dane są wewnętrznie spójne – statystyki obu drużyn równoważą się, a zależności logiczne (np. suma zabójstw i śmierci, liczba wardów, zdobycie obiektów) są zachowane.

Dane niezrozumiałe

Dzięki wczytaniu się w to, że statystyki są zbierane w 10 minucie meczu oraz znajomości gry League of Legends **dane były zrozumiałe** dla zespołu.

Punkty oddalone

Średnia liczba punktów oddalonych dla analizowanych kolumn numerycznych **wynosi około 96**. Patrząc na specyfikę gry jest to zjawisko zupełnie normalne. W niektórych meczach występują rzadkie zjawiska gdzie na przykład jedna z drużyn zdobywa ogromną przewagę w złocie czy ilości zabójstw. Nie planuje się więc ich usuwania, ponieważ **mogą zawierać istotną informację dla klasyfikatora**.

Atrybutem z największą liczbą punktów oddalonych jest blueTowersDestroyed (464 punkty), co stanowi około 4.7% danych dla tej kolumny. Wysoką liczbę outlierów obserwuje się również dla redTowersDestroyed (396 punktów, ok. 4.0%). Kolumny związane z totemami (blueWardsPlaced - 281, redWardsPlaced - 244) również wykazują znaczną liczbę wartości odstających.

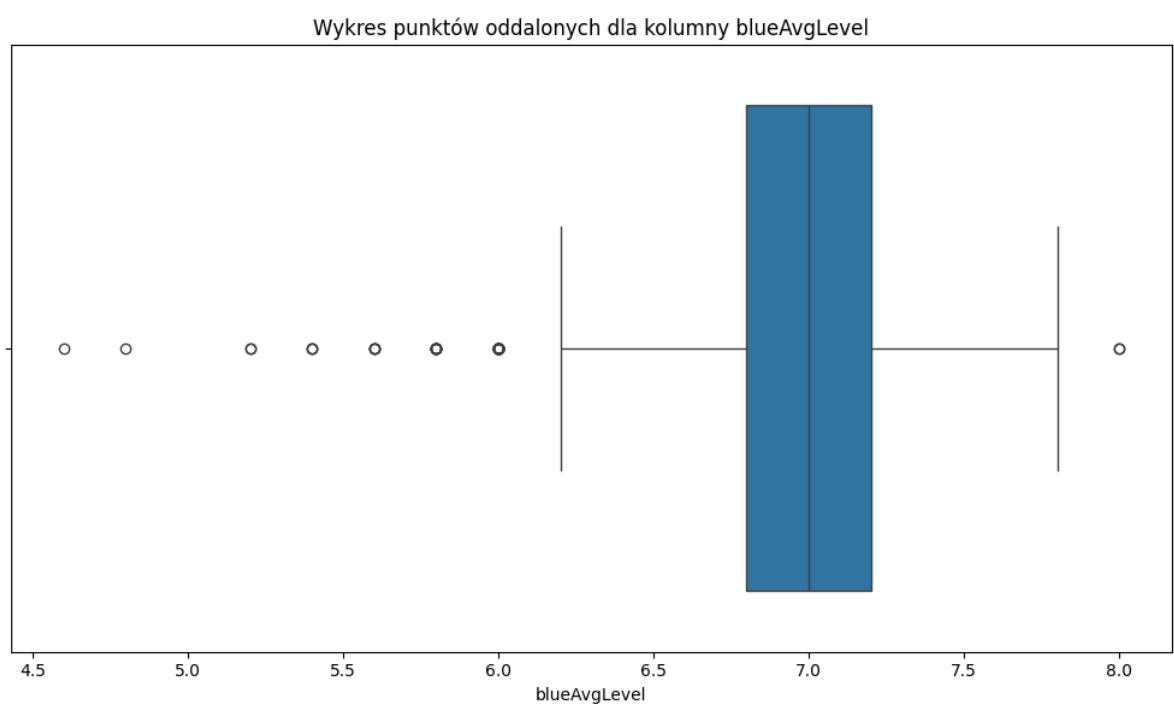
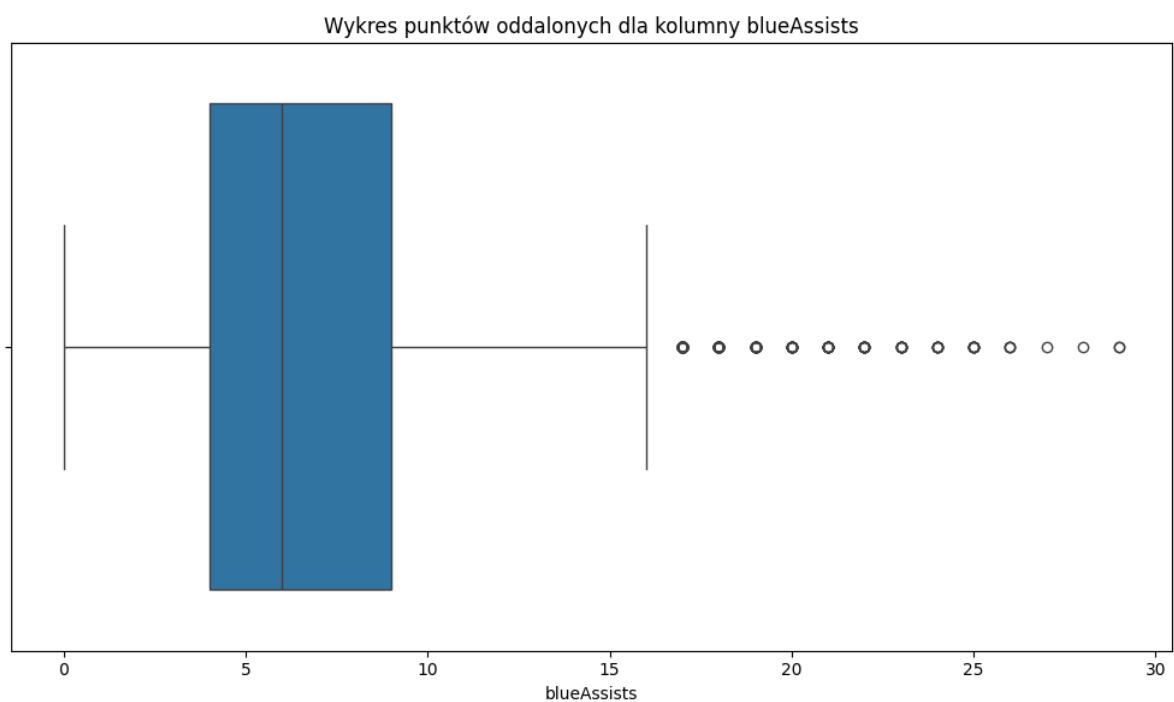
Błędy Standardowe i Punkty Oddalone dla Atrybutów Numerycznych:

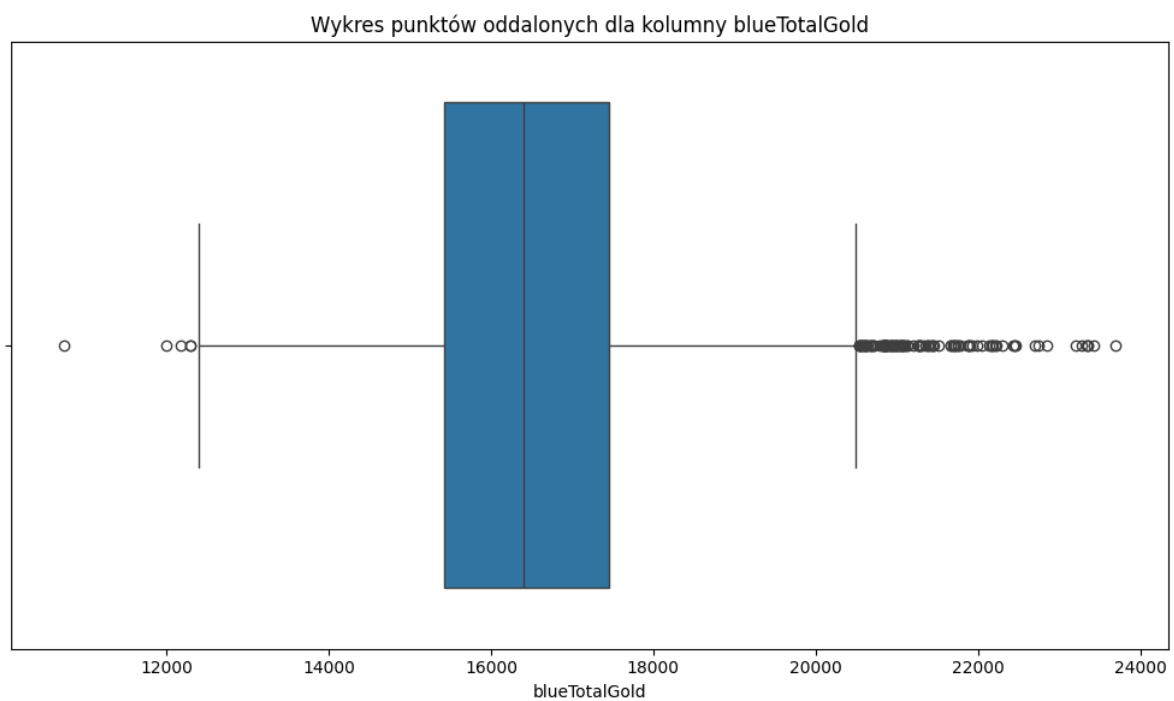
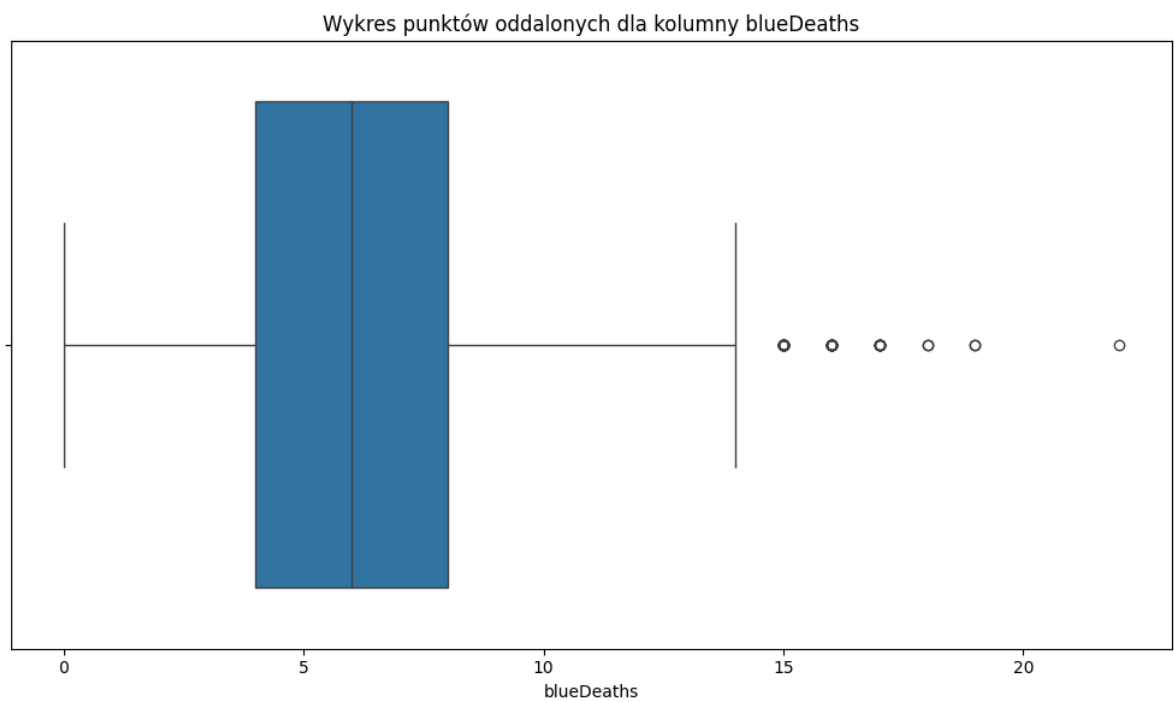
Atrybut	Błąd Standardowy	Liczba Punktów Oddalonych	Procent Danych Stanowiących Punkty Oddalone (%)
gameld	277416.261305	79	0.7997
blueWins	0.005031	N/A	N/A
blueWardsPlaced	0.181292	281	2.8444
blueWardsDestroyed	0.021883	114	1.1540
blueFirstBlood	0.005031	N/A	N/A
blueKills	0.030294	50	0.5061
blueDeaths	0.029517	65	0.6580
blueAssists	0.040893	92	0.9313
blueEliteMonsters	0.006293	N/A	N/A
blueDragons	0.004835	N/A	N/A
blueHeralds	0.003931	N/A	N/A
blueTowersDestroyed	0.002459	464	4.6968
blueTotalGold	15.448212	54	0.5466
blueAvgLevel	0.003070	69	0.6985
blueTotalExperience	12.078535	51	0.5162
blueTotalMinionsKilled	0.219919	31	0.3138

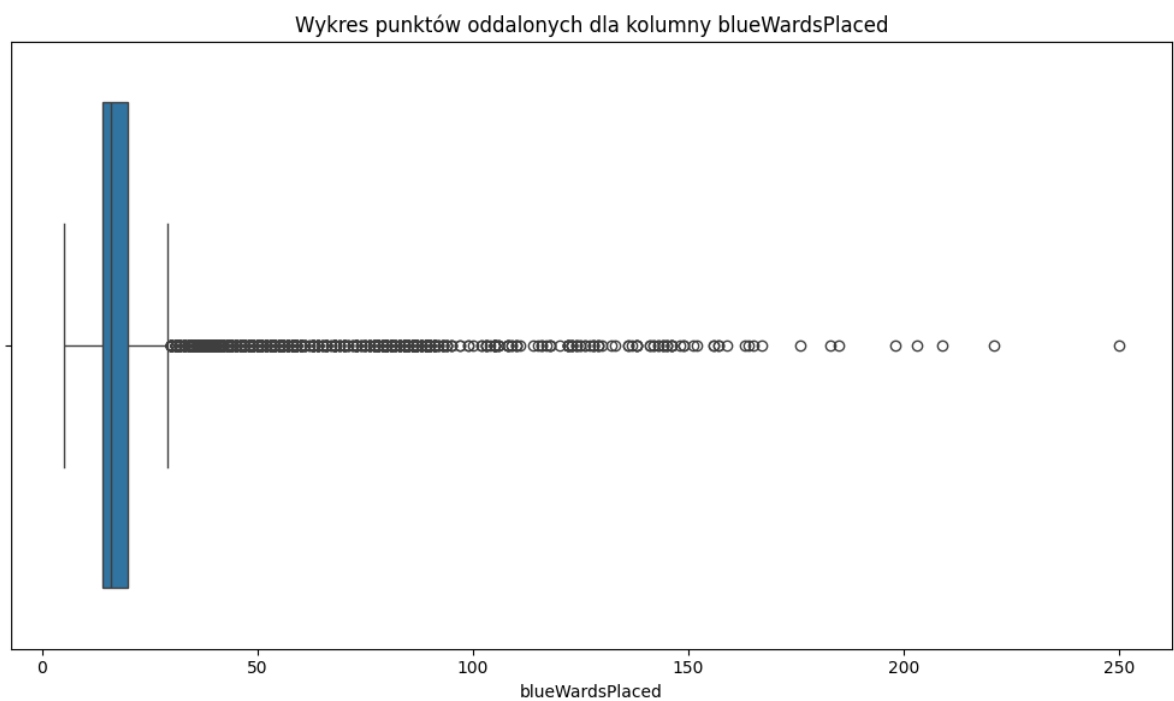
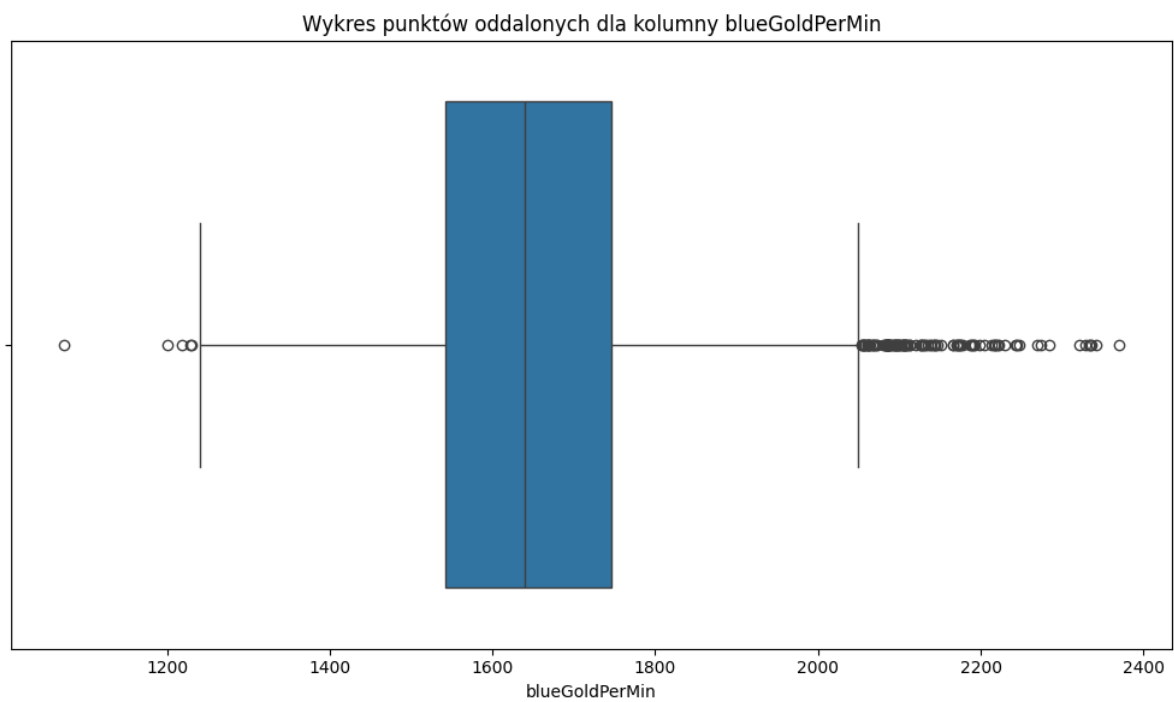
blueTotalJungleMinionsKilled	0.099587	40	0.4049
blueGoldDiff	24.683280	47	0.4758
blueExperienceDiff	19.320952	45	0.4555
blueCSPerMin	0.021992	31	0.3138
blueGoldPerMin	1.544821	54	0.5466
redWardsPlaced	0.185701	244	2.4699
redWardsDestroyed	0.021514	110	1.1135
redFirstBlood	0.005031	N/A	N/A
redKills	0.029517	65	0.6580
redDeaths	0.030294	50	0.5061
redAssists	0.040854	98	0.9920
redEliteMonsters	0.006303	N/A	N/A
redDragons	0.004954	N/A	N/A
redHeralds	0.003689	N/A	N/A
redTowersDestroyed	0.002182	396	4.0085
redTotalGold	14.999910	49	0.4960
redAvgLevel	0.003072	86	0.8705
redTotalExperience	12.059018	54	0.5466
redTotalMinionsKilled	0.220454	42	0.4251
redTotalJungleMinionsKilled	0.100891	38	0.3847
redGoldDiff	24.683280	47	0.4758
redExperienceDiff	19.320952	45	0.4555
redCSPerMin	0.022045	42	0.4251
redGoldPerMin	1.499991	49	0.4960

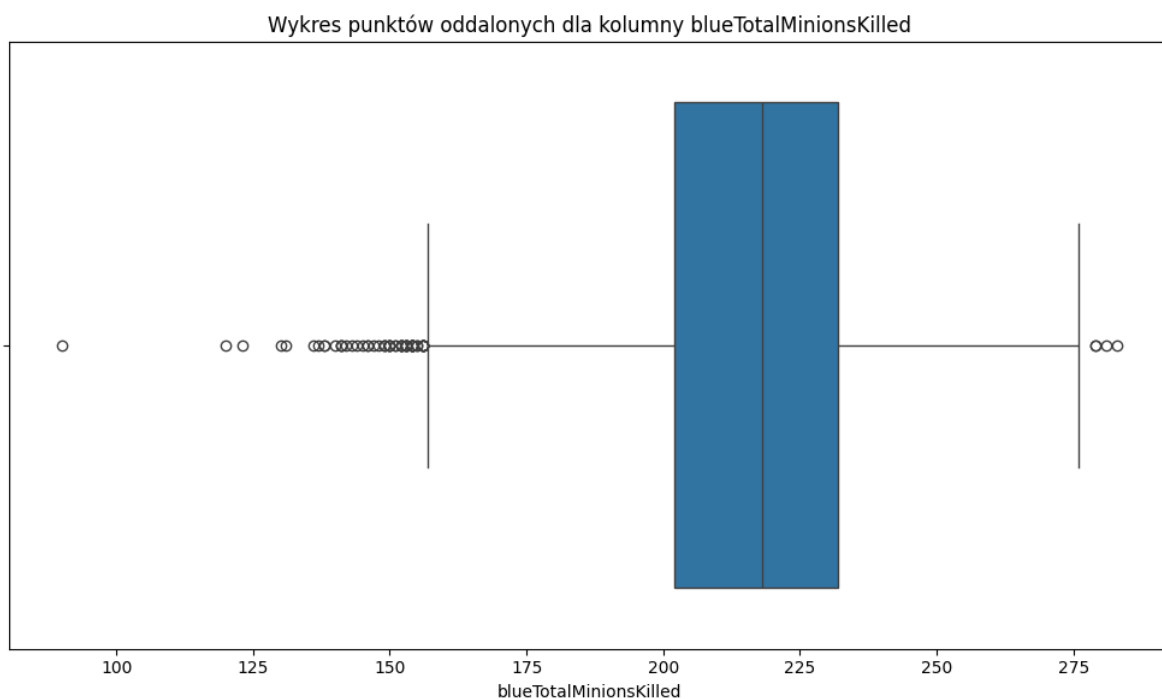
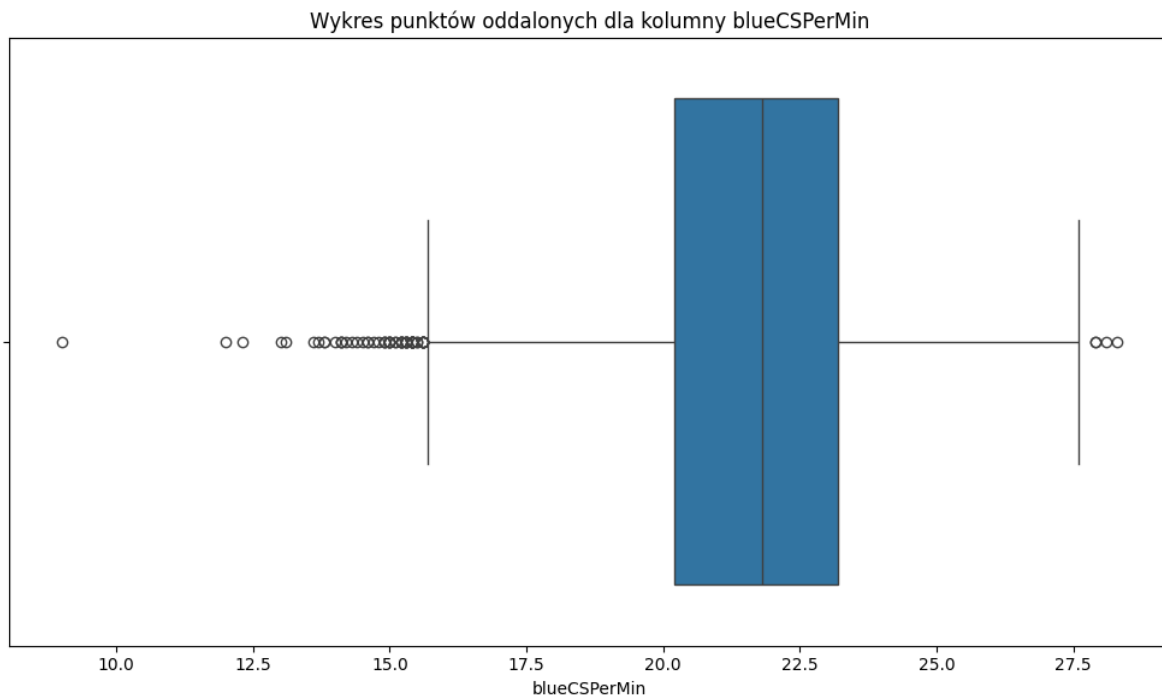
Uwaga: "N/A" oznacza, że szczegółowe dane o liczbie punktów oddalonych dla danego atrybutu nie zostały dostarczone (dotyczy to zmiennych binarnych).

Wybrane wykresy punktów oddalonych









Ocena jakości danych wskazuje, że **zestaw jest bardzo dobrze przygotowany** pod względem czystości, braku duplikatów i spójności typów danych. Szczegółowa analiza punktów oddalonych potwierdza obecność wartości ekstremalnych w danych, które są interpretowane jako odzwierciedlenie naturalnej zmienności w grze i nie będą usuwane na tym etapie.

Ewentualna rewizja celów

Dane są na tyle bogate i dobrze przygotowane, że umożliwiają nie tylko realizację pierwotnego celu klasyfikacji zwycięstwa, ale również rozwój projektu w kierunku bardziej zaawansowanej analizy strategicznej.

Kluczowe wnioski z analizy

Analiza korelacji ujawniła silne zależności wewnątrzgrupowe między zmiennymi dotyczącymi statystyk drużyn. Przykładowo:

- Atrybuty `blueKills`, `blueAssists` i `blueTotalGold` są mocno skorelowane, ponieważ większa liczba zabójstw zazwyczaj przekłada się na większe zyski złota.
- Podobnie, zmienne takie jak `blueTotalExperience`, `blueAvgLevel` i `blueCSPerMin` również wykazują wysoką współzależność, co może świadczyć o podobnym wpływie na przebieg meczu.

Przewaga w złocie i doświadczeniu to najbardziej decydujące czynniki wpływające na wygraną. Aby tą przewagę uzyskać konieczne jest jednak eliminowanie przeciwników oraz zabijanie stworów.

Kluczowa jest ponadto nie tyle ilość zdobytego złota lecz **przewaga nad przeciwnikiem**. Co oznacza, że opłacałoby się być ostrożnym tak, aby nie oddać przewagi w ekonomii przeciwnikowi już na starcie.

Widzimy zatem potrzebę zachowania przez graczy synergii między możliwością maksymalizacji własnych zysków a postępowaniem w ostrożny, wywarzony sposób. Taktyką wydaje się zatem być oczekiwanie na potknięcie graczy drużyny przeciwnej i karanie ich za nieostrożność poprzez np. zaatakowanie przeciwnika we dwójkę.

Jak widzimy wczesna faza gry jest absolutnie kluczowa dla zbudowania przewagi, która może zadecydować o wyniku całej rozgrywki.

Dalsze kroki

Zbiór danych spełnia wymogi jakościowe i ilościowe do budowy modeli predykcyjnych. Celem eksploracji, jakim jest przewidywanie wygranej niebieskiej drużyny na podstawie dostępnych statystyk, można uznać za **realny i możliwy do osiągnięcia**.

Biorąc pod uwagę charakter problemu (klasyfikacja binarna), zbalansowane klasy i numeryczne cechy, można rozważyć następujące modele:

- **Regresja Logistyczna:** Dobry jako model bazowy, prosty i interpretowalny.
- **Drzewa Decyzyjne i Lasy Losowe (Random Forest):** Dobrze radzą sobie z nieliniowościami, są odporne na outliery (do pewnego stopnia) i dostarczają miar ważności cech.
- **Modele wzmacniane gradientowo:** Często osiągają najwyższą skuteczność dla danych tabelarycznych.
- **Maszyny Wektorów Nośnych (SVM):** Mogą być skuteczne, ale mogą wymagać starannego strojenia hiperparametrów i skalowania danych.
- **Proste Sieci Neuronowe (MLP):** Mogą być rozważone, ale mogą wymagać większej ilości danych i bardziej złożonego strojenia. Zaleca się rozpoczęcie od prostszych modeli i stopniowe przechodzenie do bardziej złożonych, porównując ich wyniki. Metryki Oceny Modelu:

W dalszej kolejności po zbadaniu wstępnych rezultatów i **dostosowaniu hiperparametrów** będzie można rozważyć połączenie (tzw. **Metody ensemble**) najbardziej obiecujących rozwiązań w celu dalszej poprawy wydajności i redukcji wariancji.