

Data Analysis Report – Ekstraklasa 2024/25

Autor: Dawid Jasiński

Data: 13.02.2026

Narzędzia: Python (Pandas, Matplotlib, Seaborn, Scikit-Learn), Excel

Źródło danych: FBref.com

Executive Summary

Niniejszy raport przedstawia analizę danych drużyn Ekstraklasy w sezonie 2024/2025, koncentrując się na identyfikacji czynników determinujących liczbę zdobytych punktów oraz charakterystyce profili statystycznych zespołów. Analiza pokazuje wyraźny podział pomiędzy drużynami dominującymi, zespołami środka tabeli oraz drużynami mierzącymi się z problemami strukturalnymi. Kluczowym mechanizmem wyjaśniającym różnice punktowe okazuje się efektywność ofensywna, w szczególności jakość strzałów oraz zdolność do generowania realnych sytuacji bramkowych. Drużyny z czołówki, takie jak Lech czy Raków, łączą wysoką liczbę strzałów celnych z ponadprzeciętną skutecznością i stabilnością defensywną. Zespoły ze strefy spadkowej charakteryzuje natomiast ograniczona intensywność ofensywna oraz wysoka liczba błędów popełnianych w obronie, co znacząco utrudnia im regularne punktowanie.

Przeprowadzona eksploracja danych potwierdza, że posiadanie piłki ma znaczenie pomocnicze i nie stanowi decydującego czynnika sukcesu. Zależności statystyczne wskazują, że większe znaczenie mają te elementy gry, które bezpośrednio przekładają się na gole, takie jak liczba celnych strzałów, gole na strzał oraz gole na celny strzał. Wysokie współczynniki korelacji pomiędzy tymi zmiennymi a liczbą zdobytych punktów potwierdzają, że sukces w Ekstraklasie jest przede wszystkim funkcją efektywności, a nie dominacji w posiadaniu. Równie ważna jest stabilność defensywna, co szczególnie dobrze obrazuje przykład Rakowa, który mimo przeciętnego wolumenu działań ofensywnych osiągnął jedną z najwyższych lokat dzięki doskonale funkcjonującej obronie.

Model regresji liniowej zastosowany w końcowej części analizy dostarczył dodatkowego, ilościowego potwierdzenia obserwacji z części eksploracyjnej. Zmienne związane z jakością strzałów oraz liczba celnych prób na mecz okazały się najsilniejszymi predyktorami liczby zdobytych punktów, podczas gdy liczba straconych bramek wykazała jeden z najsilniejszych negatywnych wpływów. Model osiągnął wysoki poziom dopasowania, co pozwoliło oszacować oczekiwaną liczbę punktów dla każdej drużyny. Analiza porównawcza wykazała, że zespoły

takie jak Korona i Motor punktowały powyżej potencjału wynikającego ze swoich statystyk, natomiast Śląsk, Radomiak i Puszcza uzyskały wyniki istotnie niższe niż przewidywane. Te różnice wskazują na wpływ elementów niewyłapywanych przez podstawowe statystyki, takich jak skuteczność w kluczowych fragmentach meczu, organizacja taktyczna czy stabilność mentalna.

Podsumowując, analiza Ekstraklasy 2024/2025 potwierdza, że o końcowych wynikach drużyn decyduje przede wszystkim zdolność do generowania i wykorzystywania wysokiej jakości sytuacji bramkowych oraz ograniczania błędów defensywnych. Najlepsze zespoły łączą intensywność ofensywną, skuteczność oraz stabilność obronną, podczas gdy drużyny z dolnej części tabeli zmagają się zarówno z brakiem kreacji, jak i problemami w defensywie. Rezultaty modelowania podkreślają wagę efektywności i spójności strukturalnej, a także wskazują, że znacząca część różnic punktowych wynika z jakości decyzji podejmowanych w najważniejszych momentach meczu. Te wnioski wyznaczają kierunki, w których drużyny mogą rozwijać swoje strategie, a analiza statystyczna stanowi użyteczne narzędzie do oceny i optymalizacji ich potencjału na kolejnych etapach rywalizacji.

INTRODUCTION

Celem niniejszego raportu jest przeprowadzenie kompleksowej analizy statystycznej drużyn Ekstraklasy w sezonie 2024/2025 w oparciu o dane pochodzące z serwisu FBref.com. Analiza ta pozwala zrozumieć, które aspekty gry, zarówno ofensywne, jak i defensywne, mają największy wpływ na liczbę zdobytych punktów, a tym samym na końcowy układ tabeli. Projekt obejmuje pełny proces pracy z danymi, charakterystyczny dla stanowisk analitycznych, począwszy od pozyskania i przygotowania danych, poprzez eksplorację i wizualizację, aż po modelowanie predykcyjne i formułowanie wniosków.

Podstawą raportu jest zestaw łączonych danych obejmujących klasyczne statystyki meczowe (gole zdobyte, gole stracone, bilans bramkowy), wskaźniki opisujące styl gry (posiadanie piłki, liczba strzałów, liczba strzałów celnych na 90 minut), a także parametry opisujące jakość i efektywność działań ofensywnych, takie jak gole na strzał (G/Sh) czy gole na celny strzał (G/SoT). Dane pochodzą z kilku arkuszy udostępnionych na FBref i wymagają uprzedniego scalania oraz ujednolicenia struktury. W ramach przygotowania zbioru wyczyszczono nazwy drużyn, przetworzono kolumny tekstowe na wartości liczbowe, a także zadbane o spójność i prawidłowość danych.

Analiza eksploracyjna (EDA) stanowi centralną część pracy. W jej ramach przedstawiono obraz ligi z różnych perspektyw: najpierw przez statystyki podstawowe (posiadanie, strzały, gole), następnie przez zależności między zmiennymi, aż po szczegółowe profile drużyn. Kluczowe wykresy, takie jak „posiadanie vs punkty”, „skuteczność strzałów vs punkty” czy mapa korelacji, pozwalają zidentyfikować, które cechy gry są naprawdę istotne, a które – pomimo częstego podkreślania w przekazie medialnym – mają ograniczony wpływ na wynik końcowy. W analizie wyraźnie widoczny jest fakt, że Ekstraklasa jest ligą, w której najważniejsza okazuje się jakość strzałów oraz skuteczność wykończenia, a nie koniecznie dominacja w posiadaniu piłki.

W bardziej szczegółowej perspektywie raport prezentuje porównanie profili statystycznych drużyn z TOP4, środka tabeli oraz zespołów z dolnych rejonów tabeli. Podejście to umożliwia wskazanie najważniejszych różnic pomiędzy najlepszymi i najgorszymi drużynami w lidze. Analiza radarowa oraz szczegółowe zestawienia pokazują, że drużyny z czołówki nie tylko częściej oddają strzały, ale przede wszystkim częściej trafiają do bramki i znacząco przewyższają rywali pod względem różnicy bramek. Z kolei zespoły ze strefy spadkowej charakteryzują się niską skutecznością oraz dużymi problemami defensywnymi, co bezpośrednio przekłada się na ich pozycję w tabeli.

Istotnym elementem raportu jest również zastosowanie modelu regresji liniowej, którego celem było określenie, w jakim stopniu wybrane zmienne statystyczne wyjaśniają liczbę zdobytych punktów. Model został zbudowany na wystandaryzowanych danych, a jego współczynniki pozwalają ocenić względną siłę poszczególnych zmiennych. Co istotne, model osiągnął bardzo wysoką wartość współczynnika determinacji R^2 , co wskazuje, że zmienne opisujące efektywność strzałów i parametry defensywne bardzo dobrze tłumaczą rzeczywiste

wyniki. W modelu wyznaczono również oczekiwaną liczbę punktów (expected points), co umożliwiło identyfikację drużyn, które osiągnęły wyniki powyżej lub poniżej swojego statystycznego potencjału.

Raport korzysta z narzędzi analitycznych typowych dla współczesnej pracy w obszarze data analytics. Kluczową rolę odgrywa Python (biblioteki pandas, numpy, matplotlib, seaborn, scikit-learn), który został wykorzystany zarówno w procesie czyszczenia i transformacji danych, jak i w analizie statystycznej oraz modelowaniu regresyjnym. Całość została opracowana w środowisku Jupyter Notebook, co umożliwiło przejrzysty podział prac na kolejne etapy oraz łatwą replikowalność wyników.

Choć analiza ma charakter kompleksowy, jej wyniki należy traktować jako analizę opisową, a nie predykcyjną. Dane obejmują wyłącznie sezon 2024/2025, a próba 18 drużyn uniemożliwia tworzenie modeli o charakterze prognostycznym w sensie ścisłym. Niemniej, przedstawione wnioski dają bardzo klarowny obraz tego, jakie elementy gry prowadzą do sukcesu w Ekstraklasie oraz które zespoły wykorzystują swój potencjał lepiej lub gorzej, niż sugerują ich statystyki.

DATA CLEANING & PREPARATION

Proces przygotowania danych stanowił kluczowy etap całego projektu, ponieważ surowe dane pobrane z portalu FBref były rozproszone w kilku arkuszach, posiadały niejednolity format, a część zmiennych wymagała konwersji i dodatkowego przetworzenia. Dane obejmowały sześć różnych zestawów statystyk drużynowych, takich jak tabela ligowa, statystyki standardowe, zestawienia strzałów, statystyki bramkarzy oraz dane dotyczące czasu gry i dodatkowych parametrów meczowych. Aby umożliwić ich wspólną analizę, konieczne było stworzenie jednego, spójnego zbioru danych zawierającego wszystkie kluczowe informacje.

Pierwszym etapem przygotowania danych było wczytanie wszystkich arkuszy i przegląd ich struktury. Dane pochodzące z tabeli ligowej wymagały podstawowego oczyszczenia, w tym usunięcia pustych rekordów oraz konwersji kolumn z typów tekstowych na liczbowe—dotyczyło to między innymi wartości takich jak liczba punktów, bramek, meczów czy średnia punktów na mecz. Najważniejszym elementem wstępnego czyszczenia było ujednolicenie nazw drużyn. W wielu arkuszach nazwy zespołów zawierały odnośniki HTML w postaci fragmentów zapisanych w nawiasach kwadratowych, co uniemożliwiało bezpośrednie łączenie danych. Zastosowano więc ekstrakcję właściwej nazwy za pomocą wyrażeń regularnych, tak aby we wszystkich arkuszach występował identyczny format nazwy drużyny.

Kolejnym etapem było zredukowanie zbędnych kolumn i wybór tylko tych zmiennych, które były istotne z punktu widzenia dalszej analizy. Po wstępnym przeglądzie danych wybrano kluczowe metryki takie jak liczba zawodników, średni wiek drużyny, posiadanie piłki, liczba goli, liczba asyst oraz sumaryczna produkcja bramkowa (G+A). Analogicznie, z arkusza dotyczącego strzałów wybrano zmienne opisujące wolumen (strzały na 90 minut, celne strzały na 90 minut), jak również parametry jakościowe, w tym gole na strzał (G/Sh) i gole na celny strzał (G/SoT). Pozwoliło to ograniczyć rozmiar zbioru i skoncentrować analizę na tych wskaźnikach, które rzeczywiście mogą tłumaczyć efektywność drużyn.

Wszystkie wybrane arkusze zostały połączone w jeden zbiór danych poprzez operację łączenia (merge) po oczyszczonej nazwie zespołu. Dzięki temu udało się uzyskać kompletny zestaw obejmujący 18 drużyn oraz 28 ujednoliconych zmiennych opisujących ich profil statystyczny. Poniższy fragment kodu prezentuje zapis procesu łączenia danych:

```
df = df_table.merge(std_small, on='Squad', how='left', suffixes=('', '_std'))
df = df.merge(shoot_small, on='Squad', how='left', suffixes=('', '_shoot'))
```

Po scaleniu arkuszy wykonano standardową walidację jakości danych, sprawdzając brakujące wartości, rozkłady statystyczne oraz zgodność podstawowych miar, takich jak suma strzelonych i straconych bramek. Na przykład, porównano liczbę goli z tabeli ligowej z liczbą goli zanotowaną w arkuszach statystyk standardowych, aby upewnić się, że dane opisują ten sam zakres meczów. Wynik tych sprawdzeń potwierdził spójność i poprawność danych, co umożliwiło ich dalsze wykorzystanie bez konieczności uzupełniania braków lub korekty wartości.

Przygotowany zbiór danych stał się podstawą dla analiz eksploracyjnych i modelowych przedstawionych w kolejnych częściach raportu. Jego struktura pozwalała nie tylko na analizę pojedynczych wskaźników, lecz również na badanie zależności między nimi, porównania profili drużyn oraz budowę modelu statystycznego przewidującego liczbę punktów. Dzięki wykonanym zabiegom czyszczenia i transformacji danych możliwe było przeprowadzenie analizy w sposób rzetelny, pełny i zgodny z metodyką pracy analitycznej.

EXPLORATORY DATA ANALYSIS

Analiza eksploracyjna stanowi centralną część raportu i ma na celu zrozumienie struktury ligi, identyfikację najważniejszych zależności statystycznych oraz ocenę tego, w jaki sposób różne aspekty gry wpływały na wyniki drużyn Ekstraklasy w sezonie 2024/2025. W ramach EDA przeanalizowano zarówno klasyczne miary meczowe, jak i bardziej zaawansowane wskaźniki efektywności, co pozwoliło wyjść poza proste obserwacje wynikające z tabeli ligowej. Analiza obejmuje zarówno przegląd rankingów podstawowych statystyk, jak również badanie zależności pomiędzy kluczowymi zmiennymi i porównanie profili drużyn o odmiennych wynikach sportowych.

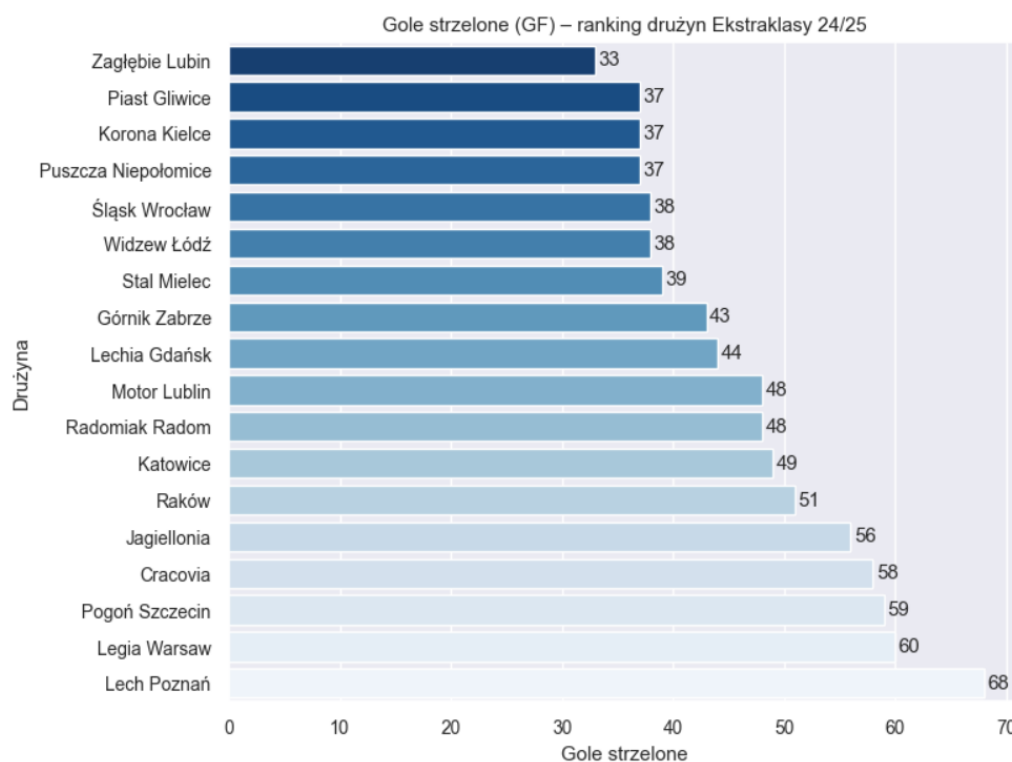
League Overview

W pierwszej kolejności przeanalizowano podstawowe statystyki ofensywne i defensywne drużyn Ekstraklasy, takie jak liczba zdobytych i straconych bramek, różnica bramek oraz średni poziom posiadania piłki. Już na tym etapie widoczne są wyraźne różnice między zespołami dominującymi a drużynami mającymi trudności z utrzymaniem stabilności w grze. Lech Poznań wyróżnia się najbardziej kompletnym profilem, łącząc wysokie posiadanie z dużą liczbą oddawanych strzałów oraz ponadprzeciętną skutecznością, co przekłada się na najwyższą liczbę zdobytych bramek w lidze i jeden z najlepszych bilansów bramkowych. Raków natomiast prezentuje inny model osiągania wyników – mimo niższego posiadania piłki niż Lech, dysponuje najlepiej funkcjonującą defensywą w lidze, co potwierdza najniższa liczba straconych bramek. Ich efektywność wynika więc bardziej ze stabilności i organizacji gry, szczególnie w formacji obronnej.

Na drugim biegunie znajdują się zespoły, które miały trudności przez cały sezon. Puszcza, Stal Mielec czy Zagłębie Lubin charakteryzują się nie tylko niską liczbą zdobytych bramek, ale również słabą skutecznością ofensywną i bardzo niestabilną grą w obronie. W tych drużynach widoczne są jednocześnie braki w kreowaniu sytuacji strzeleckich i wysoka podatność na utratę bramek, co powoduje, że nie są w stanie rywalizować na równi z bardziej zorganizowanymi przeciwnikami. Już na poziomie podstawowych statystyk widać, że różnice pomiędzy drużynami z czołówki a zespołami zajmującymi dolne miejsca tabeli nie ograniczają się wyłącznie do samego bilansu bramkowego, ale obejmują ogólną jakość gry – zarówno w zakresie utrzymywania się przy piłce, jak i w sposobie budowania akcji oraz reagowania na zagrożenia w defensywie.

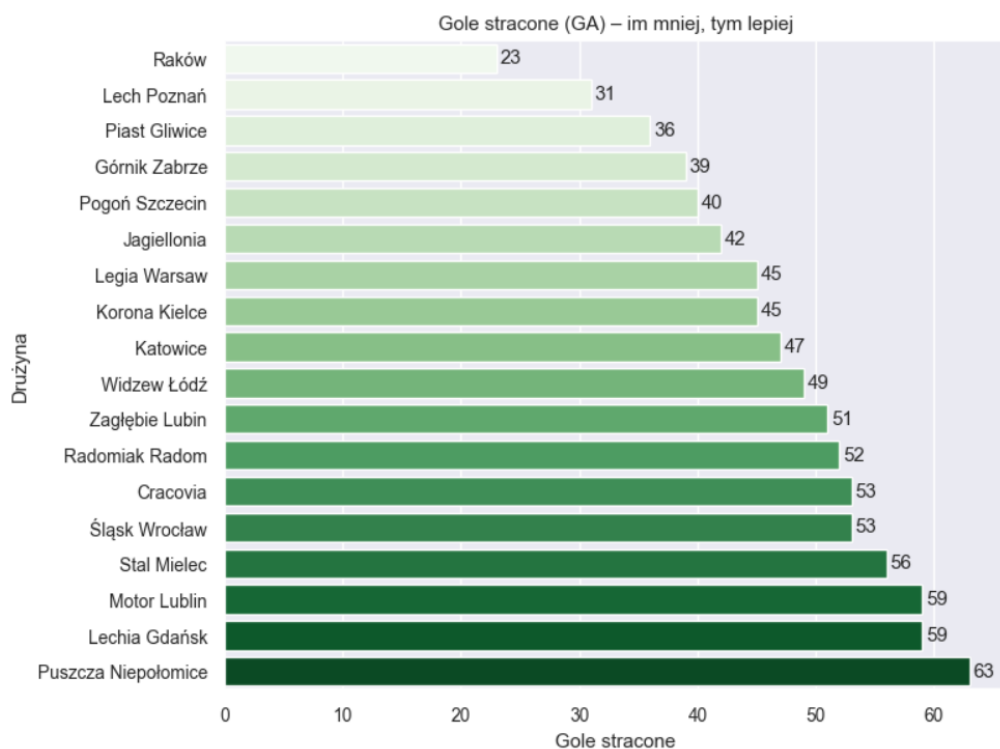
Ta początkowa analiza pozwala zrozumieć, że Ekstraklasa 2024/2025 charakteryzuje się wyraźną dychotomią między drużynami dominującymi i zespołami walczącymi o utrzymanie. Różnice w efektywności, intensywności ofensywnej oraz organizacji gry defensywnej są widoczne już na najprostszych danych. W kolejnych częściach analizy te obserwacje zostają rozwinięte poprzez bardziej szczegółowe zależności statystyczne, które pozwalają uchwycić, dlaczego jedne drużyny były w stanie stale budować przewagę, podczas gdy inne zmagwały się z problemami na większości płaszczyzn gry.

Rysunek 1 Gole strzelone (GF)



Źródło: Opracowanie własne na podstawie danych FBref.

Rysunek 2 Gole stracone (GA)

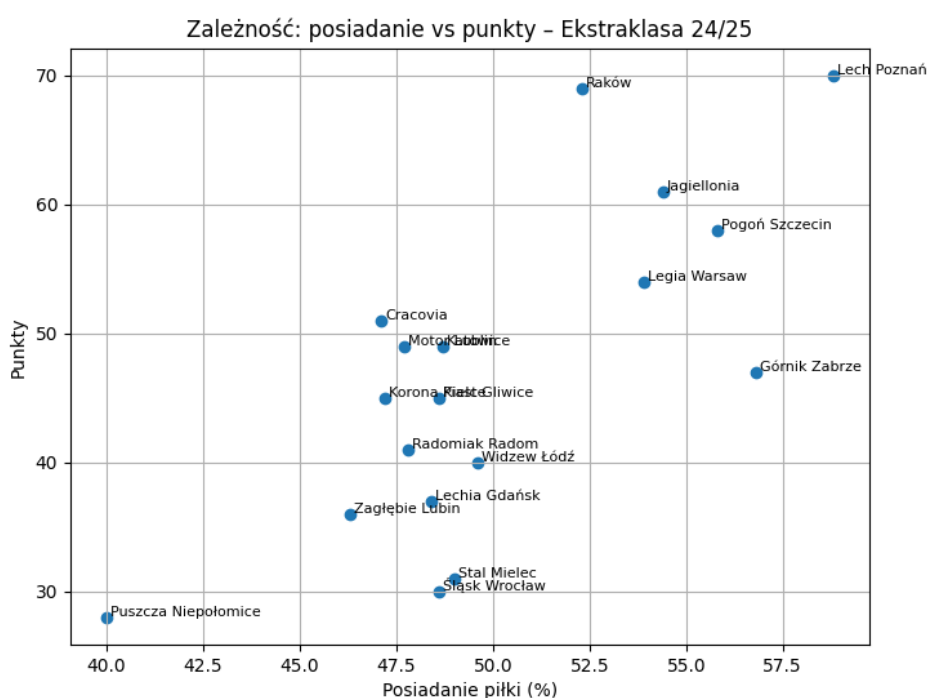


Źródło: Opracowanie własne na podstawie danych FBref.

Key Statistical Relationships

Kolejnym krokiem było zbadanie zależności pomiędzy posiadaniem piłki a liczbą zdobytych punktów. Wykres rozrzutu pokazał, że choć drużyny z górnej części tabeli częściej utrzymywały się przy piłce, zależność ta nie jest silnie deterministyczna. Lech i Raków, znajdujące się w prawym-górnym obszarze wykresu, łączą wysokie posiadanie piłki z dużą liczbą punktów, jednak Górnik — mimo jednego z najwyższych poziomów posiadania — plasuje się znacznie niżej. Z drugiej strony Puszcza, utrzymująca się przy piłce najrzadziej, zdobyła zdecydowanie najmniej punktów. Wskazuje to na to, że posiadanie piłki może pełnić funkcję wspierającą, lecz nie jest kluczowym czynnikiem decydującym o wynikach.

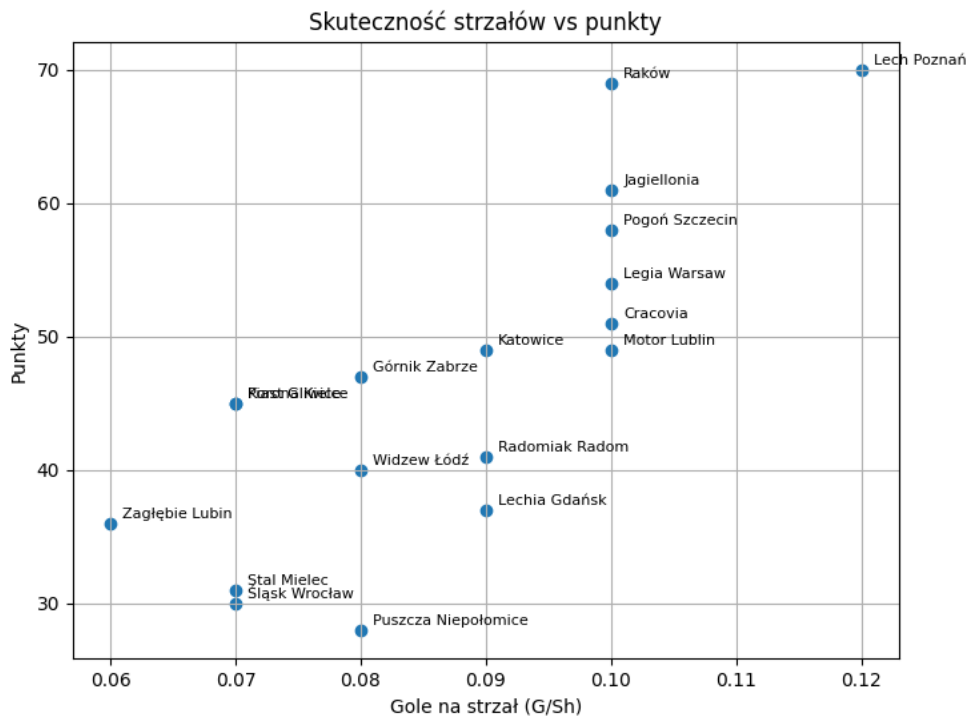
Rysunek 3 Zależność: posiadanie vs punkty



Źródło: Opracowanie własne na podstawie danych FBref.

Znacznie silniejsze zależności ujawniają się natomiast w przypadku skuteczności strzałów. Analiza zależności między golami na strzał (G/Sh) a liczbą punktów pokazuje niemal liniową relację: drużyny, które potrzebują mniejszej liczby strzałów do zdobycia gola, zajmują wyższe miejsca w tabeli. Lech, jako jedyny z wartością G/Sh na poziomie ok. 0,12, znajduje się znacznie powyżej reszty stawki zarówno pod względem skuteczności, jak i punktów. Po przeciwnej stronie zestawienia znajduje się Zagłębie, z najniższą skutecznością strzałów w lidze, co dobrze tłumaczy jego wynik końcowy. Wykresy te pokazują, że w Ekstraklasie o sukcesie decyduje przede wszystkim efektywność ofensywy, a nie jej wolumen.

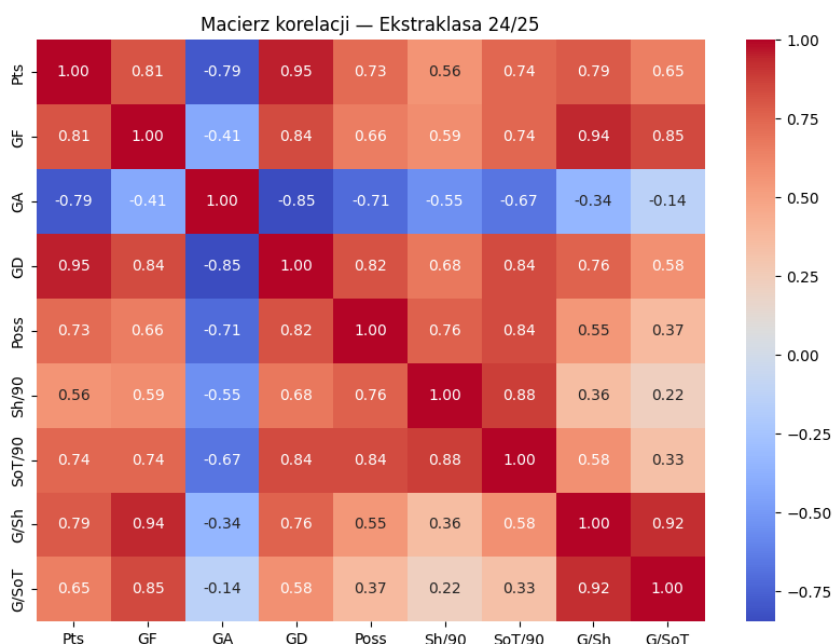
Rysunek 4 Skuteczność strzałów vs punkty



Źródło: Opracowanie własne na podstawie danych FBref.

Dodatkowe potwierdzenie tego faktu dostarcza analiza korelacji pomiędzy zmiennymi. Macierz korelacji ujawnia, że najsilniejszym „statystycznym wyjaśnieniem” liczby punktów jest różnica bramek ($r = 0,95$), liczba goli strzelonych ($r = 0,81$) oraz wskaźniki skuteczności strzałów ($G/Sh = 0,79$ oraz $G/SoT = 0,65$). Z kolei posiadanie piłki, choć często podkreślane w analizach medialnych, wykazuje jedynie umiarkowaną korelację ($r \approx 0,73$). W praktyce oznacza to, że drużyny nie wygrywają meczów poprzez długie utrzymywanie się przy piłce, lecz dzięki umiejętności tworzenia i wykorzystywania klarownych sytuacji strzeleckich. Równie istotnym czynnikiem jest defensywa — liczba goli straconych wykazuje silną korelację ujemną z liczbą punktów ($r = -0,79$), co potwierdza, że ograniczenie strat bramkowych stanowi fundament sukcesu.

Rysunek 5 Macierz korelacji



Źródło: Opracowanie własne na podstawie danych FBref.

Team Profiles

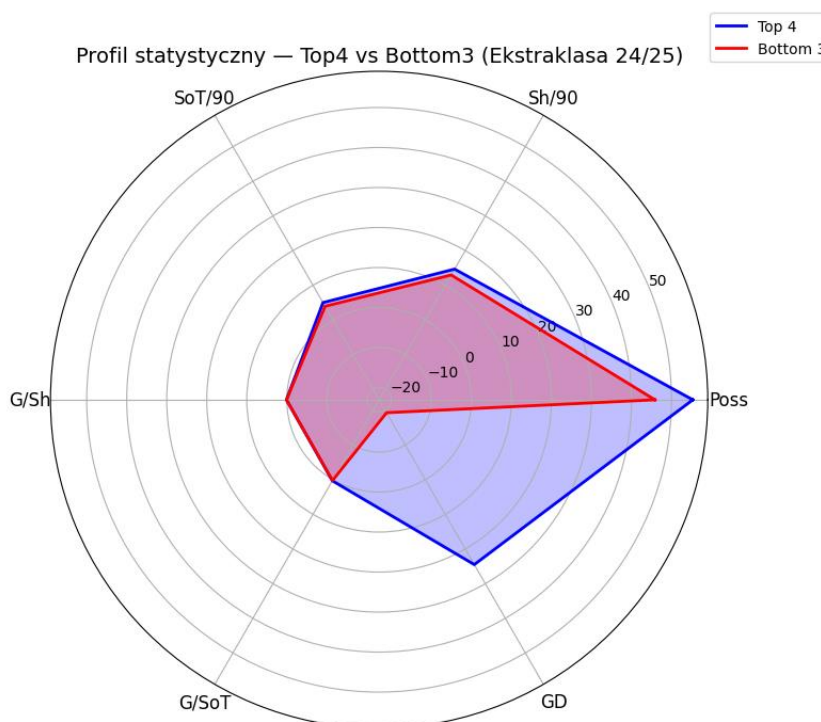
Porównanie profili statystycznych drużyn z TOP4 oraz zespołów ze strefy spadkowej pozwala bardzo wyraźnie zobaczyć różnice w sposobie gry oraz w jakości działań ofensywnych i defensywnych. Zespoły z czołówki cechują się zdecydowanie większą intensywnością ataku: częściej utrzymują się przy piłce, dłużej kontrolują przebieg meczu i regularnie znajdują się na połowie przeciwnika. Przekłada się to na większą liczbę oddawanych strzałów, większą liczbę strzałów celnych oraz większą liczbę sytuacji bramkowych, co w konsekwencji skutkuje wyższą liczbą goli i lepszym bilansem bramkowym.

Drużyny ze strefy spadkowej odstają nie tylko pod względem liczby kreowanych okazji, ale również pod względem stabilności defensywnej. Strata dużej liczby bramek oraz częste błędy w obronie sprawiają, że zespoły te rzadko mają możliwość narzucenia własnego stylu gry, a wiele ich akcji kończy się zanim zdążą stworzyć realne zagrożenie. Co istotne, różnice pomiędzy grupami dotyczą nie tylko wolumenu działań, ale również ich konsekwencji — TOP4 potrafi utrzymać strukturę gry zarówno w ataku, jak i w obronie, podczas gdy drużyny z dołu tabeli tracą bramki po prostych błędach i nie są w stanie akumulować przewagi w dłuższym okresie meczu.

Interesujące jest natomiast to, że sama skuteczność strzałów — mierzona jako gole na strzał czy gole na celny strzał — nie różni się aż tak wyraźnie pomiędzy obiema grupami. Sugeruje to, że kluczowy problem drużyn ze strefy spadkowej nie leży w ich umiejętnościach wykończeniowych, lecz w zbyt małej liczbie kreowanych sytuacji oraz w wysokiej liczbie popełnianych błędów defensywnych. Ostatecznie to właśnie te różnice strukturalne — brak intensywności ofensywnej i niestabilna defensywa — prowadzą do powstania przepaści

punktowej widocznej na koniec sezonu i stanowią główny czynnik odróżniający zespoły walczące o mistrzostwo od drużyn walczących o utrzymanie.

Rysunek 6 Profil statystyczny — Top4 vs Bottom3



Źródło: Opracowanie własne na podstawie danych FBref.

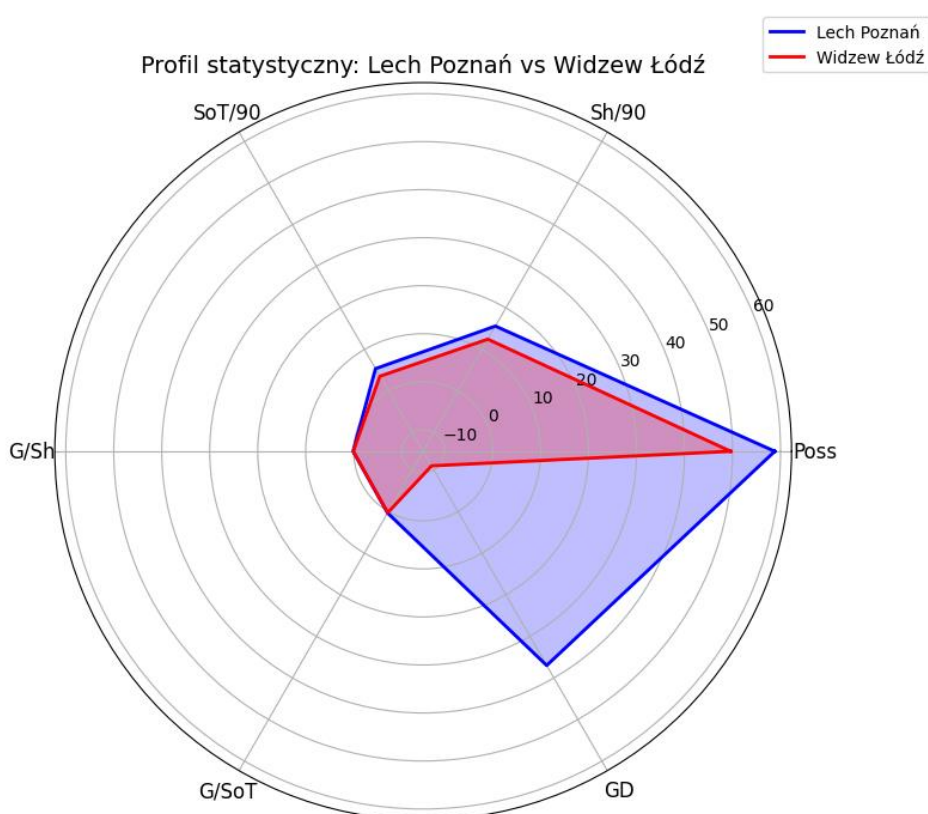
Uzupełnieniem analizy jest bardziej szczegółowe studium przypadku Lecha Poznań i Widzewa Łódź, które doskonale obrazuje różnice pomiędzy drużyną walczącą o najwyższe cele a zespołem zajmującym pozycję środkową. Lech prezentuje profil drużyny dominującej zarówno pod względem stylu gry, jak i efektywności. Utrzymywał jedno z najwyższych w lidze posiadanie piłki, co przekładało się na większą liczbę akcji ofensywnych oraz znacznie wyższy wolumen strzałów i strzałów celnych. Dzięki temu był w stanie konsekwentnie kontrolować tempo spotkań, a wysoka jakość kreowanych sytuacji w połączeniu z bardzo dobrą skutecznością przełożyła się na przewagę bramkową i ostatecznie na wynik punktowy.

Profil Widzewa Łódź różni się od profilu Lecha w prawie każdym aspekcie gry. Drużyna ta generowała mniejszą liczbę strzałów, rzadziej dochodziła do czystych pozycji strzeleckich, a jej posiadanie piłki oscylowało wokół wartości typowych dla zespołów środka tabeli. Mimo że skuteczność strzałów Widzewa nie odbiegała drastycznie od średniej ligowej, niska intensywność ofensywna powodowała, że zespół miał ograniczone możliwości budowania przewagi bramkowej. Istotnym problemem była również defensywa, która traciła bramki znacznie częściej niż czołowe drużyny ligi, co dodatkowo komplikowało zdobywanie punktów.

Zestawienie tych dwóch drużyn w wykresie radarowym uwidacznia wyraźną asymetrię: Lech przewyższa Widzew w niemal każdym wymiarze – od poziomu posiadania, przez liczbę strzałów i ich celność, aż po bilans bramkowy. Widzew natomiast prezentuje profil bardziej

zachowawczy i mało efektywny w kluczowych aspektach gry. Studium przypadku jasno pokazuje, że różnica punktowa pomiędzy tymi zespołami nie wynika z jednego, pojedynczego wskaźnika. Jest efektem skumulowanego wpływu wielu elementów, w których Lech konsekwentnie przewyższał swojego rywala. To porównanie stanowi praktykę w pigułce dla wniosków płynących z całej analizy EDA - drużyny z czołówki łączą wysoki wolumen sytuacji, dobrą ich jakość oraz stabilną defensywę, natomiast zespoły środka tabeli zwykle brakuje jednego lub kilku kluczowych komponentów.

Rysunek 7 Profil statystyczny: Lech Poznań vs Widzew Łódź



Źródło: Opracowanie własne na podstawie danych FBref.

Całość przeprowadzonej analizy eksploracyjnej pokazuje, że Ekstraklasa 2024/2025 jest ligą, w której o sukcesie decyduje przede wszystkim efektywność gry ofensywnej oraz stabilność w obronie, a nie sama liczba oddawanych strzałów czy dominacja w posiadaniu piłki. Wnioski te stanowią fundament do kolejnego etapu raportu, czyli modelowania regresyjnego, które pozwoliło precyzyjnie określić wagę poszczególnych zmiennych w kontekście liczby zdobytych punktów.

Predictive Modeling

W końcowym etapie analizy zastosowano model regresji liniowej, którego celem było określenie, w jakim stopniu wybrane zmienne statystyczne wyjaśniają liczbę zdobytych punktów w sezonie 2024/2025. Model ten nie miał pełnić funkcji prognostycznej, lecz interpretacyjną — jego zadaniem było wskazanie, które aspekty gry, spośród analizowanych wcześniej parametrów, są statystycznie najważniejsze w kontekście finalnych wyników. Zastosowanie regresji pozwala spojrzeć na Ekstraklasę z perspektywy ilościowej: nie tylko zauważyć różnice między drużynami, ale także zrozumieć, jak silny jest wpływ poszczególnych zmiennych na zdobycz punktową, gdy pozostałe czynniki są brane pod uwagę jednocześnie.

Przed zbudowaniem modelu dane zostały wystandaryzowane, aby zapewnić porównywalność wszystkich zmiennych niezależnie od ich jednostek i skali. Standaryzacja pozwala także interpretować współczynniki regresji jako względną siłę wpływu danego czynnika na zmienną zależną, co w tym przypadku jest kluczowe dla wyciągnięcia wartościowych wniosków. Do modelu wprowadzono zestaw zmiennych opisujących zarówno wolumen działań ofensywnych, takich jak liczba strzałów i strzałów celnych na 90 minut, jak również parametry jakościowe, skuteczność strzałów ogółem oraz skuteczność strzałów celnych, posiadanie piłki oraz podstawowe mierniki bramkowe.

Model regresji wykazał bardzo wysoki stopień dopasowania do danych, o czym świadczy wartość współczynnika R^2 na poziomie około 0,95. Oznacza to, że wybrane zmienne wyjaśniają aż 95% zróżnicowania punktów pomiędzy drużynami Ekstraklasy, co jest wyjątkowo wysokim wynikiem w kontekście danych sportowych, często obarczonych dużą wariancją i trudnych do modelowania. Jednocześnie tak wysoka wartość R^2 powinna być interpretowana z ostrożnością, ponieważ liczba obserwacji jest niewielka, a model opisuje sezon zakończony, co sprawia, że dopasowanie jest naturalnie wyższe. Niemniej jego wyniki pozwalają niezwykle trafnie zinterpretować zależności między statystykami a miejscem drużyny w tabeli.

Najsilniejszym predyktorem liczby zdobytych punktów okazała się skuteczność strzałów celnych (G/SoT). Wysoka wartość tego parametru oznacza, że zespół nie tylko oddaje strzały, lecz robi to w sposób przemyślany i efektywny, zamieniając znaczną część strzałów celnych na gole. Kolejnymi kluczowymi zmiennymi okazały się liczba celnych strzałów na 90 minut oraz ogólna skuteczność strzałów (G/Sh). Ich istotność potwierdza tezę, że w Ekstraklasie o sukcesie decyduje jakość działań ofensywnych i umiejętność finalizowania akcji. To nie sama liczba strzałów ma największe znaczenie, ale ich celność i efektywność. Co interesujące, tradycyjne miary takie jak liczba strzelonych bramek uzyskały niższe współczynniki, ponieważ ich wpływ jest „przechwytywany” przez zmienne wyższego rzędu, bardziej opisujące jakość procesu ofensywnego niż sam jego wynik końcowy.

Tabela 1 Wyniki regresji liniowej

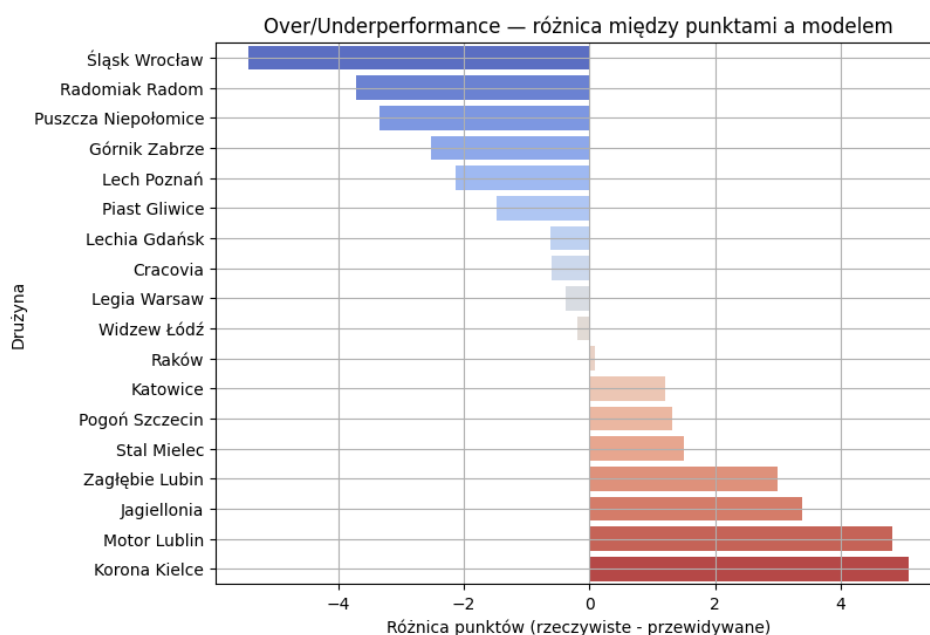
Feature	Coefficient
G/SoT	5.561082
SoT/90	3.439537
G/Sh	1.573352
Sh/90	-0.668302
GF	-0.869304
Poss	-1.316131
GA	-7.632773

Źródło: Opracowanie własne na podstawie danych FBref.

Równie silny, lecz ujemny wpływ na liczbę punktów wykazała liczba straconych bramek. Wysoka wartość tej zmiennej obniża wynik drużyny w sposób przewidywalny i konsekwentny, co potwierdza, że solidna defensywa stanowi fundament sukcesu. W kontekście analizy regresji oznacza to, że nawet drużyny ofensywnie zorganizowane i skuteczne mają ograniczoną możliwość nadrobienia słabości w obronie — jeśli tracą dużo bramek, model natychmiast odzwierciedla to niższą wartością przewidywanych punktów.

Po zbudowaniu modelu wyznaczono także oczekiwaną liczbę punktów dla każdej drużyny, co pozwoliło zidentyfikować zespoły grające powyżej i poniżej swojego statystycznego potencjału. Wyniki wskazały, że drużyny takie jak Korona czy Motor Lublin zdobyły więcej punktów, niż przewidywał model na podstawie ich statystyk, co może świadczyć o umiejętności wykorzystywania kluczowych momentów meczu, indywidualnych błyskach zawodników lub dobrze funkcjonujących stałych fragmentach gry. Z drugiej strony Śląsk, Radomiak oraz Puszcza osiągnęły wyniki wyraźnie niższe od oczekiwanych, co sugeruje problemy z efektywnością w fazach decydujących o punktach — takich jak końcówki spotkań, sytuacje zero-jedynkowe czy obrona własnej bramki po stracie prowadzenia.

Rysunek 8 Over/Underperformance



Źródło: Opracowanie własne na podstawie danych FBref.

Różnica pomiędzy liczbą punktów rzeczywistych a oczekiwanych stanowi niezwykle interesujący element analizy. Pokazuje, które zespoły były bardziej efektywne od swojej statystycznej charakterystyki oraz które grały poniżej potencjału. Zespoły z dodatnią różnicą wykazały się wyższą skutecznością w kluczowych momentach sezonu, podczas gdy drużyny z różnicą ujemną mogą wskazywać na problemy taktyczne, mentalne lub organizacyjne, których nie oddają surowe statystyki.

Model regresji oraz analiza oczekiwanych punktów stanowią istotny element raportu, ponieważ zamykają część analityczną w sposób liczbowy i precyzyjny. Stanowią one logiczne dopełnienie obserwacji płynących z EDA, potwierdzając, że fundamentami sukcesu w Ekstraklasie są skuteczność ofensywna, intensywność generowania sytuacji oraz stabilność defensywna, a nie sama liczba strzałów czy dominacja w posiadaniu piłki. Pozwoliło to przygotować klarowny obraz tego, które drużyny zbudowały swoją pozycję na solidnych podstawach statystycznych, a które odstawały od swoich wyników, pokazując w praktyce, jak wartościowym narzędziem może być analiza predykcyjna w ocenie piłkarskiej rzeczywistości.

Key Findings

Przeprowadzona analiza danych Ekstraklasy 2024/2025 pozwala sformułować kilka kluczowych wniosków dotyczących mechanizmów decydujących o wynikach ligowych. Najważniejszym z nich jest dominująca rola efektywności ofensywnej. Drużyny osiągające wysokie miejsca w tabeli charakteryzowały się nie tylko większą liczbą oddawanych strzałów, ale przede wszystkim wyższą jakością i skutecznością tych prób. Zależności ujawnione w analizie korelacji oraz wykresach relacyjnych jednoznacznie wskazują, że gole na strzał oraz gole na celny strzał należą do najmocniej skorelowanych zmiennych z liczbą zdobytych punktów. Wynika z tego, że sukces opiera się nie tyle na intensywności ofensywy, ile na zdolności do jej efektywnego wykorzystywania. W przypadku drużyn z dolnej części tabeli widoczny był natomiast deficyt zarówno pod względem wolumenu działań ofensywnych, jak i stabilności w grze obronnej, co skutkowało większą podatnością na utratę bramek i trudnościami w akumulowaniu przewagi w przebiegu sezonu.

Drugim kluczowym wnioskiem jest wyraźna rola organizacji gry defensywnej, czego najlepszym przykładem pozostaje Raków. Jego wysoka pozycja w tabeli była wynikiem przede wszystkim wzorowo funkcjonującej defensywy, która redukowałą liczbę strat bramkowych i pozwalała na zdobywanie punktów nawet w meczach o ograniczonej liczbie sytuacji ofensywnych. Model regresji potwierdził wagę tego czynnika, wskazując liczbę straconych bramek jako jedną z najsilniej, negatywnie skorelowanych zmiennych w odniesieniu do punktów. Wreszcie analiza expected points pokazała, że część drużyn uzyskiwała wyniki powyżej lub poniżej potencjału wynikającego ze statystyk. Zespoły takie jak Korona czy Motor Lublin wyróżniały się efektywnością w kluczowych momentach spotkań, podczas gdy Śląsk, Radomiak czy Puszcza osiągały wyniki słabsze, niż wynikałoby to z ich parametrów gry. Zestawienie tych obserwacji pozwala stwierdzić, że wyniki końcowe drużyn Ekstraklasy są rezultatem kombinacji skuteczności, organizacji defensywnej oraz zdolności do wykorzystywania najważniejszych fragmentów meczu, a nie jedynie prostą funkcją liczby oddawanych strzałów czy poziomu posiadania piłki.

Recommendations

Wnioski płynące z przeprowadzonych analiz pozwalają sformułować kilka praktycznych rekomendacji dotyczących kierunków rozwoju drużyn Ekstraklasy, zwłaszcza zespołów, które znajdują się w dolnej części tabeli. Najważniejszym obszarem wymagającym poprawy jest jakość kreowania sytuacji ofensywnych. Zespoły znajdujące się w strefie spadkowej nie odstają drastycznie pod względem samej skuteczności strzałów, jednak generują zdecydowanie zbyt małą liczbę okazji, aby móc regularnie zdobywać punkty. Wskazane jest zatem zwiększenie intensywności działań ofensywnych zarówno poprzez przyspieszenie faz przejściowych, jak i poprawę jakości podań w finalnej tercji boiska. W praktyce może to oznaczać konieczność zmiany tempa gry, wprowadzenie bardziej dynamicznych skrzydłowych lub większy nacisk na schematy, które pozwalają częściej dochodzić do pozycji strzeleckich o wyższej jakości.

Drugim kluczowym obszarem wymagającym poprawy jest organizacja defensywy. Dane jednoznacznie pokazują, że liczba straconych bramek ma bardzo silny wpływ na liczbę zdobywanych punktów, a w wielu przypadkach problemy drużyn nie wynikały z samej liczby dopuszczanych strzałów, lecz z braku stabilności i koncentracji w newralgicznych momentach spotkań. Usprawnienie struktury obronnej oraz zwiększenie dyscypliny taktycznej może pozwolić drużynom nie tylko ograniczyć liczbę strat, ale także lepiej kontrolować przebieg meczu. Zespoły, które miały problemy w defensywie, powinny przeanalizować swoje mecze pod kątem organizacji podczas faz przejściowych oraz jakości ustawienia w polu karnym, ponieważ wiele utraconych bramek wynikało z błędów indywidualnych lub braku asekuracji.

Odrębną rekomendacją jest większe wykorzystanie analityki w przygotowaniu do meczów i w podejmowaniu decyzji kadrowych. Analiza expected points pokazuje, że niektóre zespoły punktowały znacznie poniżej swojego potencjału statystycznego, co może wskazywać na niewłaściwe zarządzanie kluczowymi fragmentami gry, brak stabilności psychicznej lub problemy z finalizacją akcji w sytuacjach wysokiego napięcia. Drużyny o ujemnym bilansie różnicy między punktami rzeczywistymi a oczekiwanymi powinny skupić się na szczegółowej analizie faz meczu, w których dochodzi do utraty kontroli, a także na pracy nad mentalnością i precyzją w kluczowych momentach, tak aby minimalizować koszt pojedynczych błędów.

Podsumowując, aby skutecznie rywalizować w Ekstraklasie, drużyny powinny dążyć do zwiększenia intensywności ofensywnej, jednocześnie dbając o stabilność w obronie oraz optymalizując proces wykorzystywania sytuacji strzeleckich. Kluczem do poprawy wyników nie jest jedynie zwiększenie liczby strzałów czy dominacja w posiadaniu piłki, ale spójność pomiędzy wszystkimi fazami gry oraz umiejętność podejmowania jakościowych decyzji w momentach decydujących o wyniku meczu. Zastosowanie tych rekomendacji może pozwolić drużynom na poprawę efektywności punktowej i lepsze wykorzystanie swojego potencjału w kolejnych sezonach.