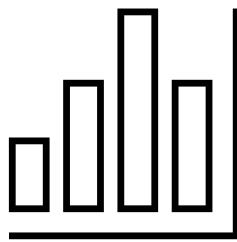


Projekt zaliczeniowy

Piotr Niemira



4 stycznia 2024

—

Prezentacja i wizualizacja danych

—

Dr Jarosław Olejniczak

Wprowadzenie

Informacje ogólne

Wszystkie wizualizacje w niniejszym raporcie zostały utworzone przy pomocy języka programowania R. Tematem przewodnim projektu jest piłka nożna. Dane zostały pobrane ze strony, do której link znajduje się poniżej:
<https://fbref.com/en/comps/9/2022-2023/stats/2022-2023-Premier-League-Stats>

Wykresy przedstawiają dane na temat piłkarzy zarejestrowanych we wszystkich klubach, które w sezonie 2022-2023 grały w najwyższej klasie rozgrywkowej w Anglii - Premier League. Zestaw danych obejmuje obserwacje z jednego, wyżej wspomnianego, sezonu oraz zawiera szereg zmiennych charakteryzujących cechy zawodników.

Przekształcenia danych

Dane wymagały modyfikacji w przypadku wielu zmiennych. Najistotniejszą zmianą w zestawie danych było odrzucenie wszystkich zawodników, którzy mieli rozgrane mniej niż 900 minut ze względu na to, że wszelkie statystyki jakie zostały dla nich wyliczone, mogły nie być wystarczająco wiarygodne. Korekcji podległy również niektóre zmienne. Wszystkie niezbędne informacje na temat nazw zmiennych, ich znaczenia oraz wszelkich modyfikacji, którym zostały poddane, znajdują się w poniższej tabeli:

Nazwa zmiennej po modyfikacji	Nazwa zmiennej przed modyfikacją	Pełna nazwa	Opis modyfikacji
NA	Rk	NA	usunięta
Player	Player	Imię i nazwisko gracza	NA
Nation	Nation	Narodowość	usunięto pierwsze 3 znaki
Pos	Pos	Najczęściej grana pozycja	usunięto drugą pozycję
Squad	Squad	Nazwa drużyny	NA
Age	Age	Wiek	brak
Born	Born	Rok urodzenia	brak
MP	MP	Rozegrane mecze	brak
Starts	Starts	Rozpoczęte mecze	brak
Min	Min	Rozegrane minuty	usunięto przecinek
full_90s	90s	Pełne mecze	zmieniono nazwę
Gls	Gls	Gole	brak
Ast	Ast	Asysty	brak
G_pl_A	G+A	Gole + asysty	zmieniono nazwę
G_mi_PK	G-PK	Gole - rzuty karne	zmieniono nazwę
PK	PK	Rzuty karne wykorzystane	brak
PKatt	PKatt	Podjęte próby rzutów karnych	brak
CrdY	CrdY	Żółte kartki	brak
CrdR	CrdR	Czerwone kartki	brak
xG	xG	Oczekiwane bramki (xG)	brak
npxG	npxG	Oczekiwane bramki bez karnych (npxG)	brak
xAG	xAG	Oczekiwane asysty (xAx)	brak
npxG_pl_xAG	npxG+xAG	Oczekiwane bramki bez rzutów karnych + oczekiwane asysty	zmieniono nazwę
PrgC	PrgC	Rajdy z piłką	brak
PrgP	PrgP	Podania progresywne	brak
PrgR	PrgR	Otrzymane podania progresywne	brak
Gls_p90	Gls	Gole na 90 minut	zmieniono nazwę
Ast_p90	Ast	Asysty na 90 minut	zmieniono nazwę
G_pl_A_p90	G+A	Gole + asysty na 90 minut	zmieniono nazwę
G_mi_PK_p90	G-PK	Gole - rzuty karne na 90 minut	zmieniono nazwę
G_pl_A_mi_PK_p90	G+A-PK	Gole + asysty - rzuty karne na 90 minut	zmieniono nazwę
xG_p90	xG	Oczekiwane bramki (xG) na 90 minut	zmieniono nazwę

xAG_p90	xAG	Oczekiwane asysty (xAG) na 90 minut	zmieniono nazwę
xG_pl_xAG_p90	xG+xAG	Oczekiwane bramki + oczekiwane asysty na 90 minut	zmieniono nazwę
npxG_p90	npxG	Oczekiwane bramki bez rzutów karnych (npxG) na 90 minut	zmieniono nazwę
npxG_pl_xAG_p90	npxG+xAG	Oczekiwane bramki bez rzutów karnych + oczekiwane asysty na 90 minut	zmieniono nazwę
PrgC_p90	NA	Rajdy z piłką na 90 minut	stworzono zmienną na podstawie istniejącej
PrgP_p90	NA	Podania progresywne na 90 minut	stworzono zmienną na podstawie istniejącej
PrgR_p90	NA	Otrzymane podania progresywne na 90 minut	stworzono zmienną na podstawie istniejącej
Gls_p90_perc	NA	Ranga procentowa bramek na 90 minut	stworzono zmienną na podstawie istniejącej
Ast_p90_perc	NA	Ranga procentowa asyst na 90 minut	stworzono zmienną na podstawie istniejącej
xG_p90_perc	NA	Ranga procentowa oczekiwanych bramek (xG) na 90 minut	stworzono zmienną na podstawie istniejącej
xAG_p90_perc	NA	Ranga procentowa oczekiwanych asyst (xAG) na 90 minut	stworzono zmienną na podstawie istniejącej
PrgC_p90_perc	NA	Ranga procentowa progresywnych podań na 90 minut	stworzono zmienną na podstawie istniejącej
PrgP_p90_perc	NA	Ranga procentowa podań progresywnych na 90 minut	stworzono zmienną na podstawie istniejącej
PrgR_p90_perc	NA	Ranga procentowa otrzymanych podań progresywnych na 90 minut	stworzono zmienną na podstawie istniejącej
NA	matches	NA	usunięta

Wizualizacje

Pierwsza wizualizacja jest charakterystyką piłkarza. Dane na nim wyświetcone przedstawiają jak dany zawodnik wypada na tle ligi pod względem siedmiu najważniejszych statystyk. Każda z nich przedstawiona jest w centylach, które informują od jakiego odsetka piłkarzy ligi, badany zawodnik, wypada lepiej w danej statystyce.

Pierwszy wykres został poświęcony Taiwo Awoniyi. Od razu rzuca się w oczy, że wyróżnia się on w dwóch miarach – Gole na 90min i Gole oczekiwane na 90min. Jest to spowodowane faktem, że piłkarz ten gra na środku ataku, gdzie ma najczęściej okazji do zdobywania bramek oraz jest bardzo dobry w wykorzystywaniu nadarzających się okazji. Z tego samego powodu Taiwo wypada słabo w pozostałych statystykach, które obrazują jak dobrze piłkarz radzi sobie z budowaniem akcji i stwarzaniem szans kolegom z drużyny co po prostu nie jest jego zadaniem.

Statistics of Taiwo Awoniyi per 90 minutes

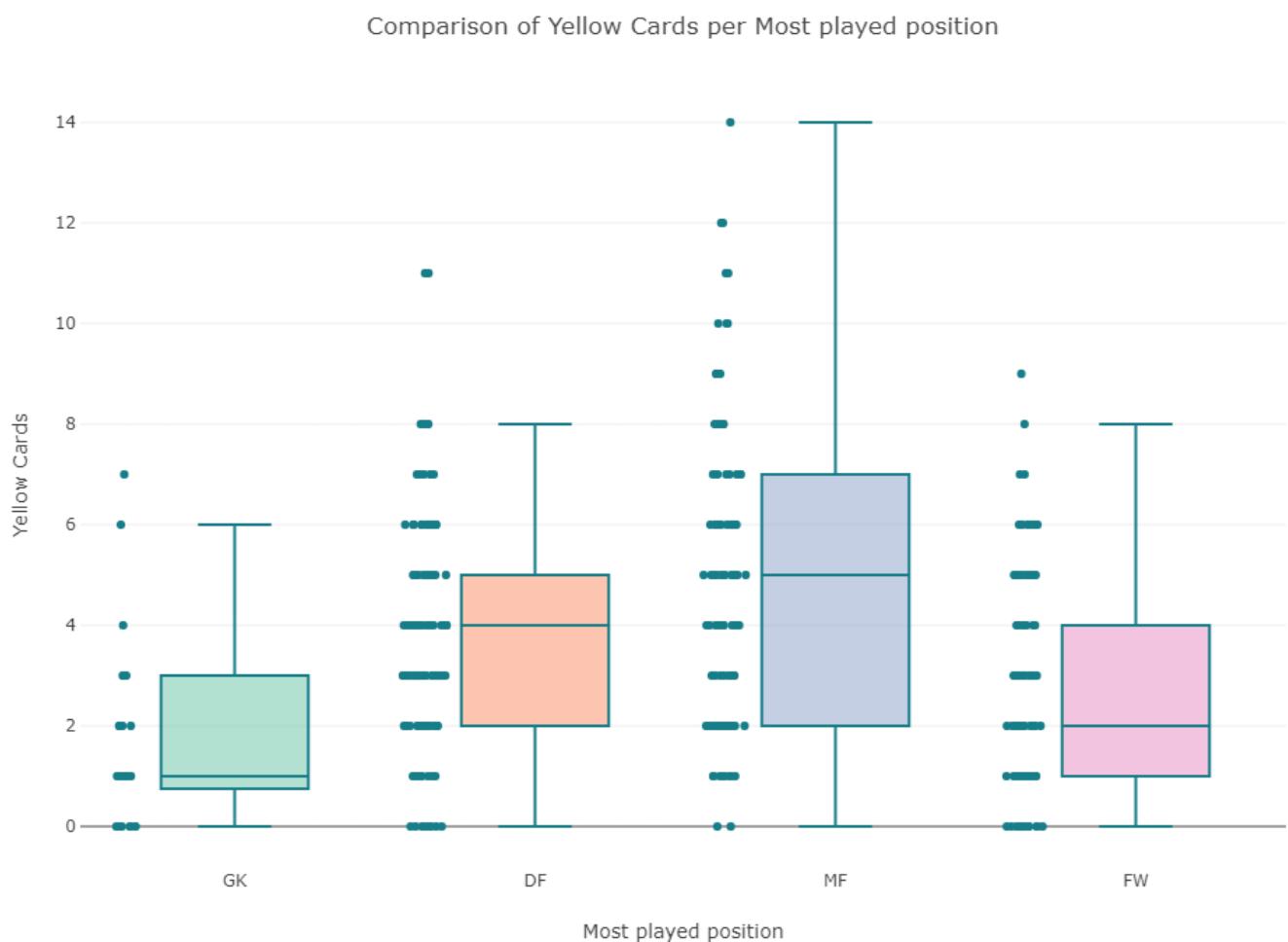


Drugi wykres prezentuje informacje na temat Enzo Fernández, który gra jako pomocnik. Gołym okiem widać różnice między pierwszym a drugim wykresem. Enzo wyróżnia się w statystykach odpowiadających za dostarczanie piłki innym piłkarzom, a przede wszystkim w podaniach progresywnych na 90min. Jest to także efekt gry w głębi pola która umożliwia mu zagrywanie piłki do przodu co jest utrudnione dla najbardziej wysuniętego piłkarza takiego jak Taiwo.

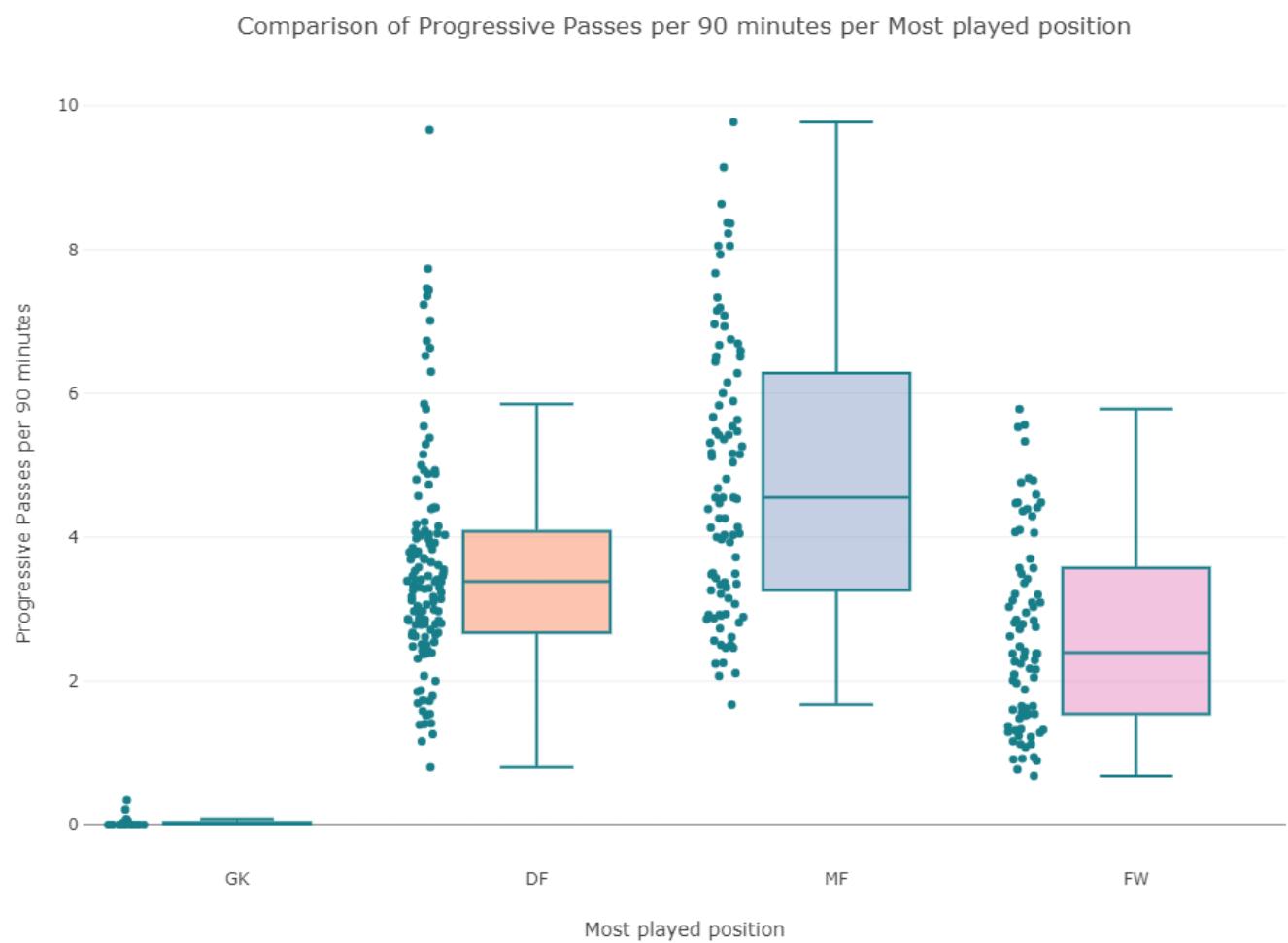
Statistics of Enzo Fernández per 90 minutes



Następna wizualizacja pokazuje zagregowane dane w postaci wykresów pudelkowych. Pierwszy z nich pokazuje jaki rozkład ma liczba żółtych kartek otrzymanych przez zawodnika w przeciągu sezonu z podziałem na pozycje. Jak widać najwięcej żółtych kartek dostają pomocnicy co prawdopodobnie wynika z faktu, iż oni najczęściej są zmuszeni do przerywania akcji przeciwnika, gdy ten jest w fazie przejścia z obrony do ataku. Na tej pozycji obserwujemy również największą wariancję, która może być wyjaśniona różnorodnością piłkarzy grających na tej pozycji. Najniższą medianę i wariancję reprezentują bramkarze, którzy nie mają zbyt wiele okazji do kontaktu z przeciwnikiem to też rzadziej zdarza im się obejrzeć żółty kartonik.

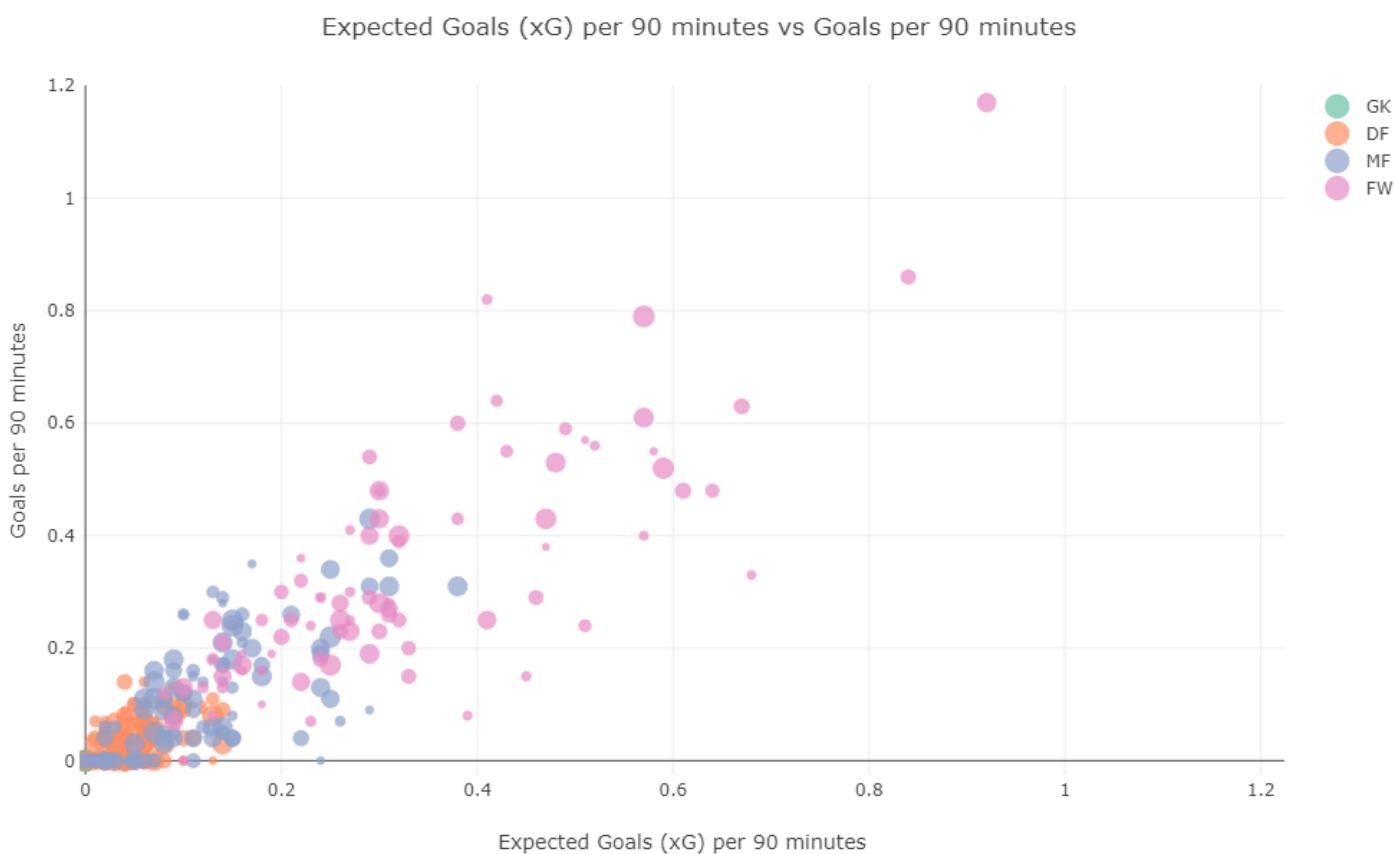


Na drugim wykresie tego typu widać porównanie liczby zagranych podań progresywnych względem pozycji. Podobnie jak w poprzednim przykładzie, najlepiej prezentują się pomocnicy co niewątpliwie wynika z faktu, że bardzo często znajdują się przy piłce i ich zadaniem jest dostarczyć piłkę do napastnika z głębi pola gry. Istotnie gorzej wypadają napastnicy, ponieważ ich zadaniem jest odbierać podania i wykorzystywać sytuacje bramkowe. Jednakże, ciekawą sytuację obserwujemy w przypadku obrońców, gdzie wariancja jest bardzo duża. Jest to efektem tego, że zwykle boczni obrońcy grają znacznie bliżej bramki rywala i mają więcej okazji do kreowania gry. Natomiast środkowi obrońcy mają przede wszystkim za zadanie bronić dostępu do własnej bramki i rzadziej znajdują się przy piłce. Skrajnie niskimi wartościami charakteryzują się bramkarze, którzy nie mają zbyt wielu okazji do zagrywania piłek na połowę rywala.



Ostatnia wizualizacja to wykres punktowy, którego celem jest zobrazowanie jak poszczególni piłkarze wypadają względem siebie w dwóch wybranych statystykach. Pozycja w układzie współrzędnych wskazuje na wartości statystyk piłkarzy, wielkość punktów odpowiada za liczbę minut zagranych w sezonie, a kolor punktu wskazuje na pozycję na boisku.

Na pierwszym wykresie widać zestawienie goli na 90min z oczekiwanyimi golami na 90min. Najciekawszą cechą tego wykresu jest możliwość dostrzeżenia, którzy piłkarze wyróżniają się pod względem liczby strzelonych goli w stosunku do goli oczekiwanych. Można również dostrzec, że piłkarze, których punkty odstają od ogólnego trendu, mają zwykle mniej rozegranych minut. Na podstawie tego można wnioskować, że mniejsza próba generuje więcej odstępstw od normy.



Drugim wykresem punktowym jest zestawienie wyników piłkarzy pod względem rajdów z piłką oraz progresywnych podań. Dzięki temu widzimy, którzy piłkarze mają większą tendencję do zdobywania terenu za pomocą prowadzenia piłki, a którzy za pomocą podania. Na wykresie widać również, że pomocnicy mają większą tendencję do podań, a napastnicy do rajdów. Natomiast obrońcy są bardzo rozproszeni.

Progressive Passes per 90 minutes vs Progressive carries per 90 minutes

