

Analiza podstawowych statystyk meczowych zawodników ligi NBA

Marcin Miśkiewicz, Adrian Sobczak

12 grudnia 2021

1 Wprowadzenie

Odkąd profesjonalne organizacje sportowe zaczęły zatrudniać na pełen etat analityków i statystyków, charakter całych dyscyplin uległ pewnego rodzaju rewolucji. Specjaliści od danych z miejsca stali się bezpośrednimi doradcami kadry trenerskiej, a wyniki ich prac wprowadziły nowe trendy w sposobach treningu i samej gry. Dzisiaj system szkolenia oparty jedynie na solidnym przygotowaniu fizycznym ustępuje miejsca podejściu strategicznemu, mającemu w arsenale analizy wideo, pomiary wydolnościowe czy taktyki wyliczone pod konkretnych rywali. Jako kolebkę wdrażania zaawansowanej statystyki do sportu powszechnie uznaje się najsilniejszą ligę koszykarską na świecie — NBA (*National Basketball Association*). Dopiero amerykańscy pionierzy zainspirowali nowym sposobem myślenia działaczy z innych dyscyplin, takich jak piłka nożna czy siatkówka.

2 Cel pracy

Mimo że wewnętrzne statystyki klubów NBA są ściśle chronione, istnieją również powszechne bazy danych gromadzące szereg informacji na temat konkretnych zawodników. W niniejszej pracy wykorzystamy jeden z takich zbiorów i na jego podstawie wysuniemy garść wniosków m.in. na temat panujących trendów w lidze, czy zależności pomiędzy poszczególnymi wynikami i parametrami fizycznymi graczy. Głównym celem analizy będzie uzyskanie charakterystyki najskuteczniejszych zawodników, czyli innymi słowy, odpowiedź na pytanie co łączy ze sobą najlepszych graczy. Przy okazji wyróżnimy zawodników o profilach unikalnych — grających inaczej niż większość teoretycznie podobnych do nich koszykarzy.

	min.	max.	średnia	mediana	braków
wzrost [m]	1.62	2.28	1.95	1.95	97
punkty	0	31.8	6.36	5.1	15
asysty	0	11.2	1.44	1	15
zbiórki	0	22.9	2.98	2.4	307
liczba sezonów	0	22	4.3	3	0

Tabela 1: Podsumowanie zmiennych ilościowych.

3 Opis danych

Do analizy wykorzystamy tabelę „Player_Attributes” z bazy danych „Basketball dataset (version 211)” udostępnionej przez Wyatta Walsh na platformie Kaggle [1] (na licencji CC BY-SA 4.0 [2]). Zaznaczymy, że na moment pobrania danych, ich ostatnia aktualizacja odbyła się 15 stycznia 2021 roku. W tabeli znajdziemy aż 37 zmiennych określonych dla 4500 zawodników (zarówno aktywnych, jak i tych na sportowej emeryturze). Na potrzeby tej analizy wykorzystamy jedynie 9 interesujących nas cech: imię i nazwisko zawodnika, wzrost, narodowość, nazwę klubu, średnią liczbę punktów/asyst/zbiórek na mecz, nominalną pozycję oraz liczbę pełnych sezonów rozegranych w lidze.

3.1 Przygotowanie zmiennych ilościowych

Część danych zawiera braki — nie każdy zawodnik ma przypisany wzrost, czy wszystkie statystyki meczowe. W tabeli [1] umieściliśmy krótkie podsumowanie dla zmiennych ilościowych. Widzimy, że braków danych jest stosunkowo niewiele jak na 4500 rekordów. Dla zbiórek¹ mamy znacząco więcej braków niż dla pozostałych statystyk meczowych. Być może wpływ ma na to fakt, że zbiórki w meczach zostały oficjalnie zliczane dopiero od 1950 roku, w piątym sezonie ligi NBA.

Warto zaznaczyć, że wzrost domyślnie zapisany był w calach, jednak mając na uwadze czytelność pracy, konwertujemy wszystkie wartości na metry (korzystamy z przelicznika: 1 cal = 2,54 cm).

3.2 Przygotowanie zmiennych kategoriycznych

W rozpatrywanym zbiorze danych możemy wyróżnić aż 51 różnych drużyn. Spośród nich 30 funkcjonuje do dzisiaj. Bez przypisanego klubu figuruje 664

¹Na przestrzeni całej pracy będziemy posługiwać się wygodnym skrótem myślowym — pisać „zbiórki”, „asysty” lub „punkty” zawodnika, mamy oczywiście na myśli ich średnią liczbę w przeliczeniu na jeden mecz.

zawodników. Chcielibyśmy zaznaczyć, że nie znamy dokładnej metody kojarzenia zawodnika z danym klubem. Autor bazy danych nie podał informacji na jakiej zasadzie graczom przypisywany jest zespół, ale wiemy, że nie jest to ostatni klub w którym emerytowany już zawodnik kończył karierę (gdyby tak było Michael Jordan byłby skojarzony z Washington Wizards, a jest z Chicago Bulls). Na potrzeby tej analizy zakładamy, że graczom przypisywany jest klub w którym zaliczyli najwięcej występów meczowych, jednak nie dysponujemy danymi by móc jednoznacznie potwierdzić zasadność tej interpretacji.

W analizowanych danych możemy wyróżnić 3 podstawowe pozycje koszykarskie: *Guard* (rozgrywający i rzucający obrońcy), *Forward* (skrzydłowi) oraz *Center* (podkoszowi). Dodatkowo niektórym zawodnikom przypisana jest kombinacja dwóch z wymienionych pozycji np. *Center-Forward*, co w tym przypadku oznacza podkoszowego, który może grać również na skrzydle. Dla uproszczenia analiz przyjmujemy, że w przypadku takich graczy, jako nominalną pozycję na parkiecie, będziemy brać pierwszą z wymienionych. Zatem wszystkich *Center-Forward* klasyfikujemy jako *Center*, a wszystkich *Forward-Center* jako *Forward*. Zawodników bez przypisanej jakiegokolwiek pozycji mamy w zbiorze danych zaledwie 60.

Dodajmy jeszcze, że każdy zawodnik ma przypisaną narodowość oraz nie ma zawodników bez wpisanego imienia i nazwiska.

4 Analiza

4.1 Wzrost

Koszykówka z założenia jest sportem faworyzującym wysokich zawodników. Chociaż liga NBA gościła i niższych graczy, którzy byli w stanie nadrobić szybkością i sprytem na boisku, to dla ogółu graczy wzrost daje nieocenioną przewagę. Dlatego w poszukiwaniu najskuteczniejszych zawodników zaczniemy od sprawdzenia wyników na podstawie wzrostu.

Na wykresach widzimy, że biorąc za punkt odniesienia wzrost, ciężko nam będzie wyznaczyć najskuteczniejszych graczy. Chociaż można podejrzewać, że wzrost ma istotny wpływ na skuteczność w zbiórkach, to z jakichś przyczyn wyżsi gracze zdają się gorzej asystować. W zdobywaniu punktów wzrost zaś zdaje się w ogóle nie mieć znaczenia.

4.2 Wpływ doświadczenia na wyniki meczowe

Czy przeciętny ligowy weteran to lepszy wybór dla trenera niż przeciętny *rookie*²? A może to ambitna młodzież rządzi na parkiecie? Odpowiedź na te pytania nie jest jednoznaczna. Jednak na podstawie rozpatrywanych danych możemy wyciągnąć szereg ciekawych wniosków. Na rysunku 1 widzimy łańcuch wykresów pudełkowych średniej liczby punktów na mecz, skonstruowanych dla zawodników o konkretnym doświadczeniu w lidze. Od razu możemy zauważyć, że wraz z doświadczeniem, średnia liczba punktów na mecz zawodnika w większości przypadków rośnie. Jest to zaskakująco wyraźny trend — przeciętnie, doświadczeni gracze zdobywają znacznie więcej punktów. Jednak zawodnikiem punkującym najlepiej wcale nie jest weteran. Mający 8 lat doświadczenia w lidze Bradley Beal rzuca średnio 31.8 punktów na mecz. Wyniki pojedynczych drużyn w meczach NBA oscylują często wokół liczby 90, zatem można powiedzieć, że Bradley odpowiada zwykle za jedną trzecią zdobyczy punktowej swojego zespołu. Na drugim miejscu, z wynikiem 30.1 punktów, uplasowali się *ex aequo* Michael Jordan (14 sezonów w NBA), Wilt Chamberlain (13 sezonów w NBA) oraz Damian Lillard (8 sezonów w NBA).

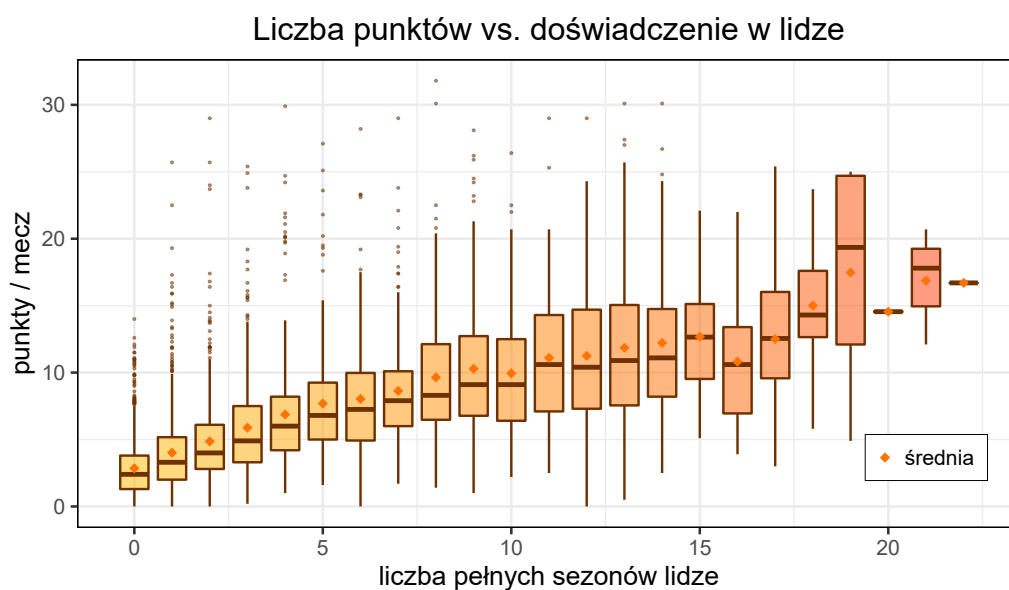
Analizując dalej rysunek 1, możemy zauważyć, że wraz ze wzorstem doświadczenia, w większości przypadków zwiększa się rozstęp międzykwartyłowy próbki. Jednocześnie przy tym maleje liczba obserwacji nietypowych. Dla zawodników, którzy w NBA rozegrali 0, 1, 2, 3, 4 i 5 pełnych sezonów mamy odpowiednio 41, 32, 23, 15, 15 i 9 obserwacji nietypowych. Warto także zaznaczyć, że w żadnym przypadku nie otrzymujemy obserwacji nietypowo małych — nieszablonowi zawodnicy to wyłącznie ci którzy zaskakują w pozytywnym sensie.

Dodajmy, że tylko dwóch koszykarzy rozegrało dokładnie 20 pełnych sezonów w lidze (Jamal Crawford i Robert Parish) i tylko jeden rozegrał dokładnie 22 pełne sezony (Vince Carter).

Dla średniej asyst i zbiórek również obserwujemy tendencję wzrostową przy rosnącym doświadczeniu, jednak efekt ten nie jest aż tak wyraźny jak w przypadku punktów. Rozstępy międzykwartyłowe także się zwiększają, a liczba obserwacji nietypowych stopniowo maleje.

Na podstawie powyższych analiz, możemy stwierdzić, że opcją „bezpieczniejszą” dla trenera jest zakontraktowanie zawodnika grającego dłużej w lidze. Oczywiście mówimy tu o czysto teoretycznej sytuacji, w której trener musiałby podjąć decyzję dysponując jedynie informacją o doświadczeniu ligowym kandydatów do składu.

²*rookie* (ang.) — debiutant, nowicjusz, żółtodziób. W koszykarskim żargonie jest to powszechnie stosowane określenie na niedoświadczonych zawodników.



Rysunek 1: Wykresy pudełkowe średniej liczby punktów na mecz dla zawodników o konkretnym doświadczeniu w lidze.

Źródła

- [1] www.kaggle.com/wyattowalsh/basketball
- [2] creativecommons.org/licenses/by-sa/4.0/