

Analiza podstawowych statystyk meczowych zawodników ligi NBA

Marcin Miśkiewicz, Adrian Sobczak

16 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]	1.65	2.31	1.98	1.98	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 8 interesujących nas cech: imię i nazwisko zawodnika, wzrost, 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 nie ma zawodników bez wpisanego imienia i nazwiska.

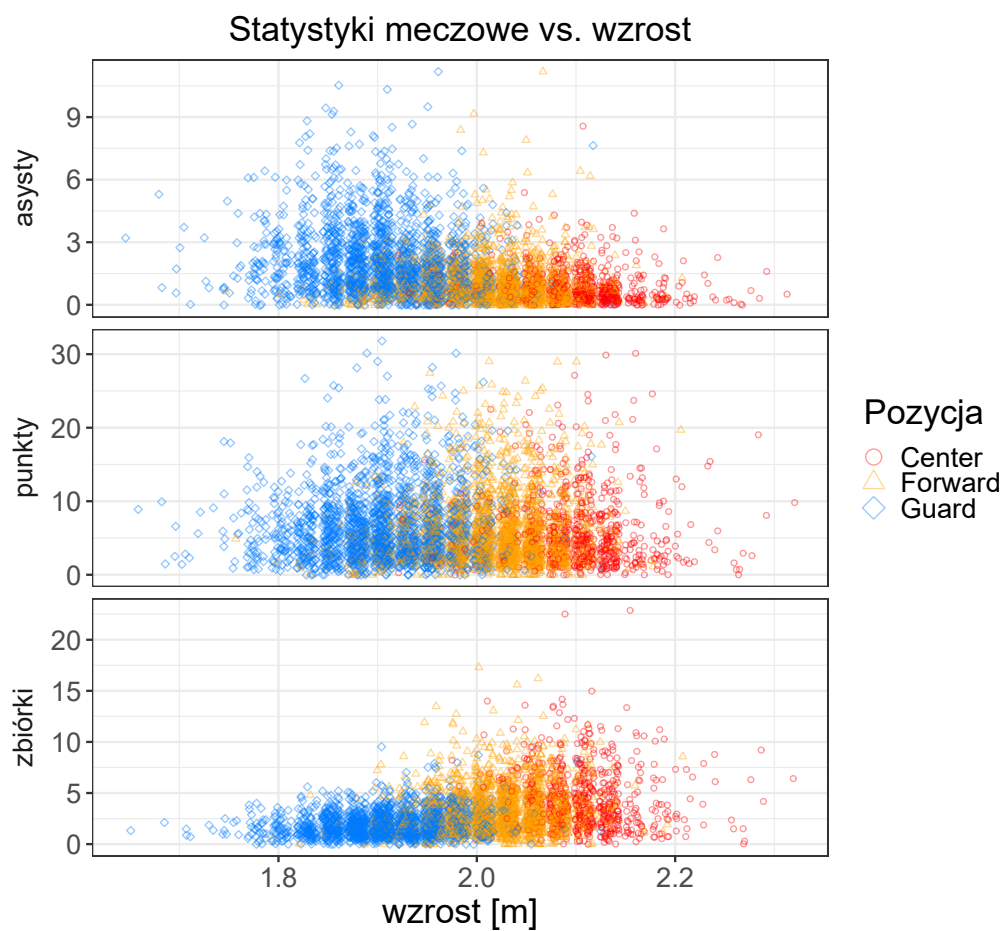
4 Analiza

Analizę danych podzielimy na kilka pomniejszych, wybranych przez nas problemów, dla których wnioski będziemy formułować na bieżąco. Ogół wyników i przemyśleń zbierzemy w sekcji 5, gdzie postaramy się także odpowiedzieć na postawione wcześniej pytania badawcze.

4.1 Czy w NBA wyższy znaczy lepszy?

Koszykówka z założenia jest sportem faworyzującym wysokich zawodników. Chociaż liga NBA gościła i niższych graczy, powszechnie wiadomo, że wzrost daje nieocenioną przewagę. Dlatego w poszukiwaniu najskuteczniejszych zawodników zaczniemy od sprawdzenia wyników na podstawie wzrostu.

W rozpatrywanych danych wzrost został podany z dokładnością do jednego cala (czyli po konwersji 2,54 cm). Jest to zaskakująco niska dokładność, dlatego, do celów graficznych, dodamy do wzrostu każdego z graczy niewielki szum losowy. Nie wpłynie on znacząco na charakter zmiennej, ale za to zapobiegnie nakładaniu się punktów podczas rysowania wykresu i poprawi ogólną czytelność.



Rysunek 1: Wykresy rozrzutu dla asyst, punktów i zbiórek, względem wzrostu. Rysunek nie uwzględnia zawodników dla których wzrost lub rozpatrywane statystyki meczowe są nieznane.

Rysunek 1 przedstawia zależność podstawowych statystyk meczowych od wzrostu zawodnika. Pierwszym co rzuca się w oczy jest wyraźne rozgraniczenie pozycji koszykarzy względem osi poziomej — *Guard* są zwykle najniżsi, *Center* najwyżsi, a *Forward* gdzieś pomiędzy. Na podstawie wykresu możemy stwierdzić, że statystyki meczowe nie są tak mocno skorelowane ze wzrostem gracza jak możnaby przypuszczać. Co więcej, okazuje się nawet, że pod względem asyst zawodnicy niżsi zdają się być skuteczniejsi. Należy tu jednak być ostrożnym. Stwierdzenie, że wyżsi są gorsi w podawaniu piłki byłoby pewnym nadużyciem, gdyż domyślamy się, że podkoszowi z założenia mają inne zadania na parkiecie. Dlatego rysunek 1 wskazuje nam na to kto asystuje, punktuje i zbiera mniej lub więcej, a nie gorzej czy lepiej.

Możemy zauważyć, że w przypadku punktów, wzrost nie ma aż tak wielkiego znaczenia. Jednak przy zbieraniu piłki, nawet kilka centymetrów przewagi może zrobić różnicę w meczu.

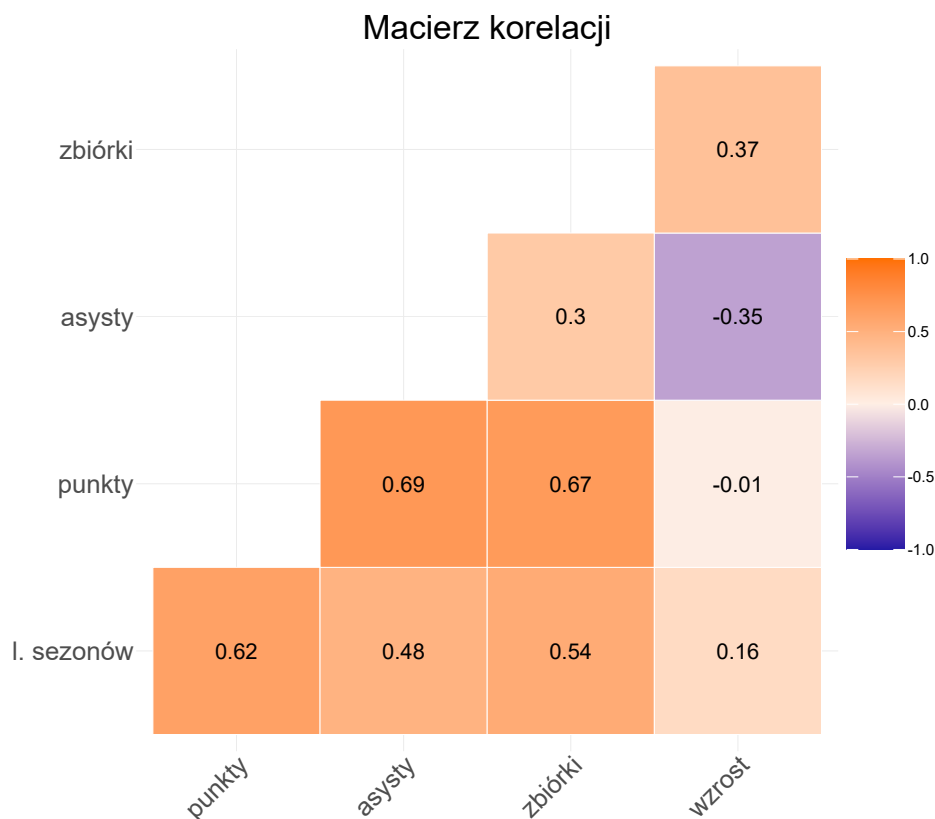
Naszym głównym wnioskiem, który wyciągamy z powyższej analizy, jest to, że w realiach NBA ponadprzeciętny wzrost nie gwarantuje lepszych wyników. Wybitnych graczy nie wyłoniśmy wybierając jedynie tych najwyższych, gdyż z różną budową ciała związane są różne pozycje i role w drużynie. Tak więc złożenie zespołu z samych podkoszowych byłoby prawdopodobnie taktycznym strzałem w kolano.

4.2 Korelacje między statystykami meczowymi

W poszukiwaniu cech wyróżniających najlepszych graczy sensownie będzie rozważyć zależności między zmiennymi ilościowymi. Mając dostęp do statystyk takich jak zdobyte punkty, które są bezpośrednim wyznacznikiem umiejętności gracza, możemy sprawdzić czy wraz z nimi występują inne cechy. Rysunek 2 przedstawia macierz korelacji rozpatrywanych zmiennych ilościowych, sporządzoną przy użyciu współczynnika korelacji Pearsona.

Rzeczą która od razu rzuca się w oczy jest praktycznie zerowy współczynnik korelacji pomiędzy wzrostem oraz zdobytymi punktami. Potwierdza to nasze obserwacje z sekcji 4.1, w której wykres rozrzutu 1 dokładnie przedstawia to zjawisko. Współczynnik korelacji wzrostu i asyst jest ujemny, co również jest zgodne z poprzednią analizą. Najsilniejsze zależności liniowe występują zaś pomiędzy punktami i pozostałymi statystykami meczowymi.

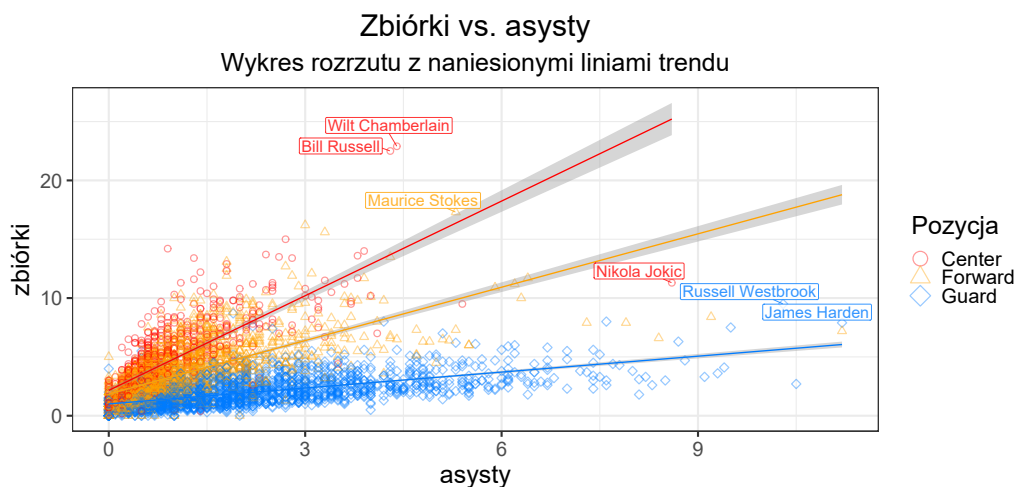
Założmy że umiejętności gracza w lidze NBA mają stałą sumę. Oznaczałoby to, że inwestując w swoje wyniki w zbiórkach musimy przykładowo zrezygnować nieco ze zdobywania punktów. Choć intuicyjnie mogłoby to mieć sens, macierz korelacji pokazuje nam coś zupełnie odwrotnego. Każdorazowo między osiąganymi przez graczy występuje dodatni współczynnik korelacji. Mówi nam to o tym, że koszykarz w NBA, będąc efektywny w jednym aspekcie gry,



Rysunek 2: Macierz korelacji zmiennych ilościowych.

prawdopodobnie jest też efektywny w pozostałych.

Spośród rozpatrywanych korelacji, chcemy zwrócić uwagę na pewien, na pozór mało interesujący przykład. Asysty i zbiórki nie wykazują względem siebie silnej korelacji, jednak jeśli graczy podzielimy na pozycje, sytuacja ta ulegnie zmianie. Na wykresie 3 dobieramy trzy różne proste regresji, każdą dla innej pozycji na parkiecie. Widzimy, że pozycja ma istotny wpływ na siłę korelacji. Dla *Guard* współczynnik korelacji wynosi 0.61, dla *Forward* 0.66, a dla *Center* aż 0.76. Na wykresie 3 dodatkowo wyróżniliśmy graczy, których iloczyn asyst i zbiórek jest większy niż 85. Warto zwrócić uwagę na dwóch geniuszy zbiórek — Wilta Chamberlaina i Billa Russela. Ich wyniki są ponad dwa razy wyższe od zdecydowanej większości pozostałych zawodników.



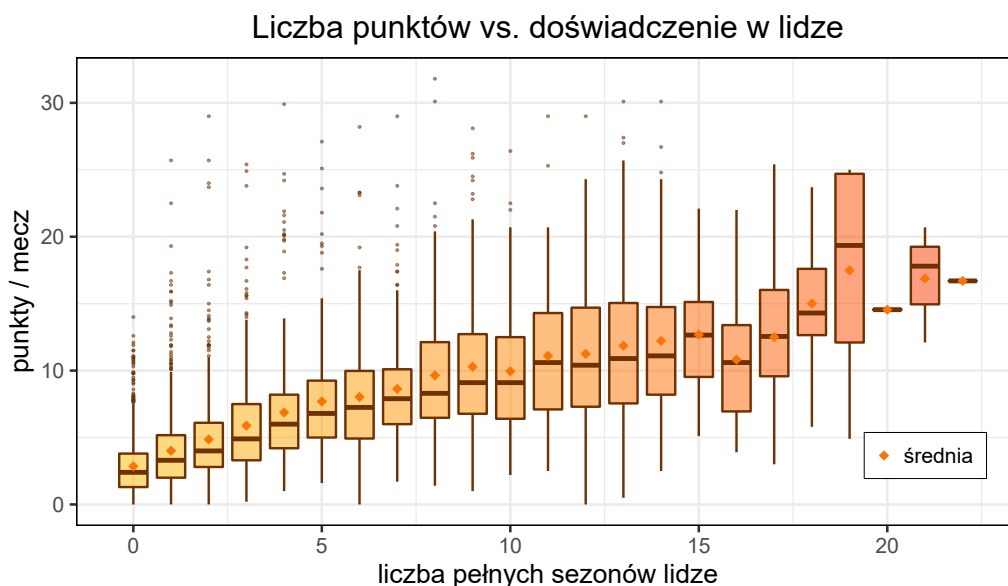
Rysunek 3: Wykres rozrzutu dla zależności asyst i zbiórek. Pogrupowany względem pozycji, z czego dla każdej dopasowuje prostą regresji. Dodatkowo podpisani zostali gracze z najlepszymi wynikami.

4.3 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 4 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 4, możemy zauważyć, że wraz ze wzrostem doświadczenia, w większości przypadków zwiększa się rozstęp międzykwartylowy próbki. Jednocześnie przy tym maleje liczba obserwacji nietypowych.

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



Rysunek 4: Wykresy pudełkowe średniej liczby punktów na mecz dla zawodników o konkretnym doświadczeniu w lidze.

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 międzykwartylowe 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.

4.4 Drużyna a wyjątkowość zawodników

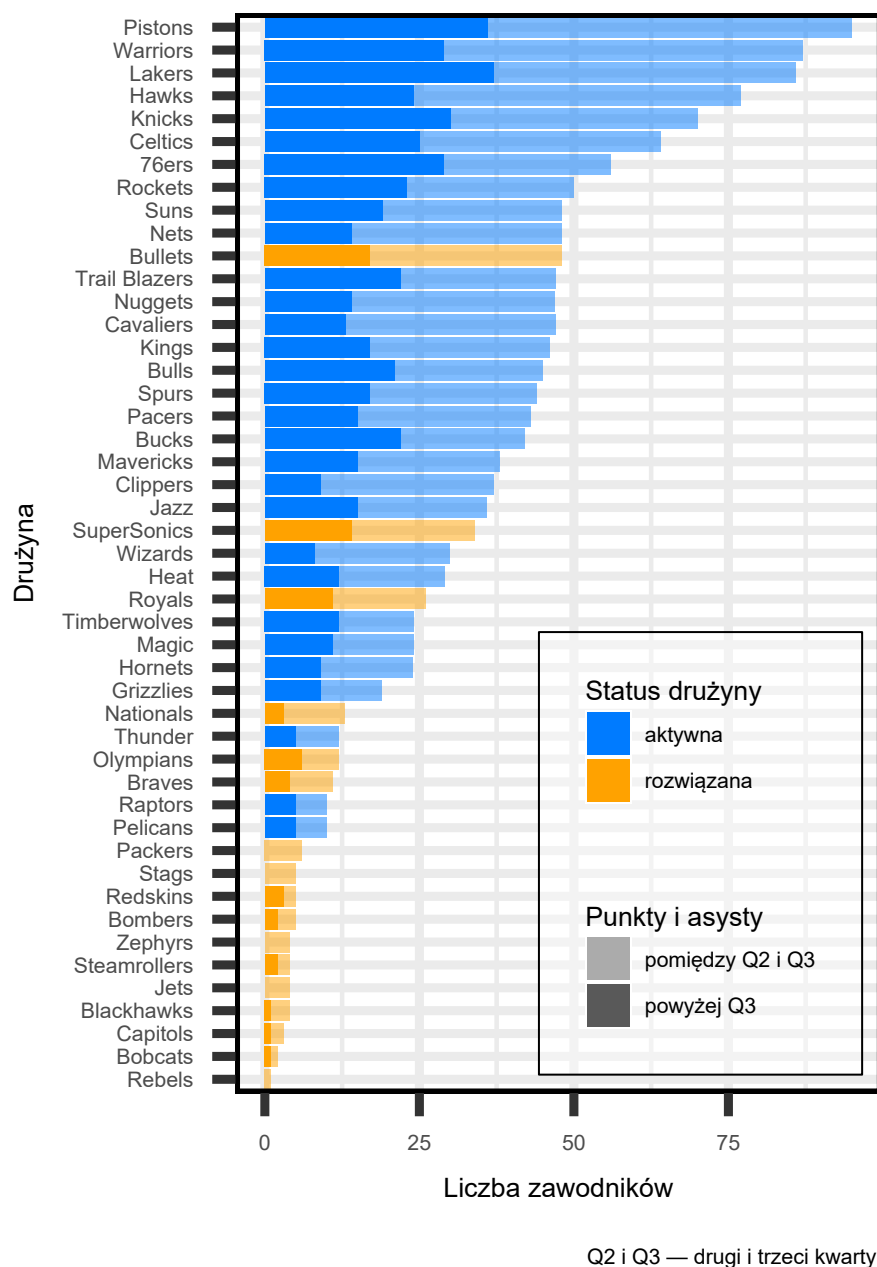
Czy istnieją drużyny w NBA specjalizujące się w zrzeszaniu zawodników o ponadprzeciętnych umiejętnościach? Aby móc odpowiedzieć na to pytanie, należy zdefiniować umowną granicę pomiędzy graczem zwykłym, a wyjątkowym. Przyjmijmy, że gracze wyjątkowi to ci, dla których dwie najważniejsze statystyki meczowe — punkty i asysty, są jednocześnie większe niż odpowiadające im mediany wyliczone dla danych z całej ligi (wartości median znajdują się w tabeli 1). Jeśli zawodnik spełnia ten warunek również dla trzecich kwartyli statystyk, to mówimy, że jest unikalny. Nie uwzględniamy zbiorów ze względu na to, że nie znamy ich wartości aż dla 307 zawodników, a pominięcie tak wielu obserwacji wpłynęłoby negatywnie na wyniki. Zawodników z brakami danych dla punktów i asyst jest stosunkowo niewiele, dlatego decydujemy się by nie uwzględniać ich w analizie.

Wykorzystując dostępny zbiór danych, możemy sporządzić wykres liczby wyjątkowych i unikalnych zawodników w historii poszczególnych drużyn NBA. Na rysunku 5 kolorem niebieskim wyróżnione zostały zespoły aktywne, a kolorem żółtym zespoły już nieistniejące. Odcień jaśniejszy wskazuje na liczbę graczy wyjątkowych (ale nie unikalnych), a odcień ciemniejszy na liczbę graczy unikalnych.

Widzimy, że Detroit Pistons to klub który zrzesza najwięcej wyjątkowych zawodników, a Los Angeles Lakers wygrywa pod względem koszykarzy unikalnych. Na pierwszy rzut oka zaskakujące może być to, że Thunder, Raptors i Pelicans są tak nisko w rankingu, będąc drużynami aktywnymi. Jednak spoglądając na oficjalne dane NBA dotyczące lat działalności klubów [3], możemy szybko zauważyć, że organizacje te są stosunkowo młode w porównaniu z resztą ligi. Przykładowo, Oklahoma City Thunder została założona w 2008 roku, po rozwiązaniu funkcjonujących od 40 lat Seattle SuperSonics. Na uwagę zasługuje także Philadelphia 76ers — organizacja ta zrzesza 27 koszykarzy wyjątkowych i 29 unikalnych, co czyni ją jedyną drużyną w czołówce naszego rankingu, w której zawodników bardzo dobrych jest więcej niż po prostu dobrych.

Na wykresie 5 nie uwzględniamy koszykarzy bez przypisanego klubu, dlatego chcemy dodatkowo skomentować ich przypadek. Przypomnijmy, że takich graczy jest aż 664. Wśród nich 184 jest według naszej definicji wyjątkowych.

Liczba wyjątkowych zawodników w historii klubu



Rysunek 5: Wykres liczby wyjątkowych zawodników w historii klubów NBA. Zawodnicy wyjątkowi to ci, których punkty i asysty są jednocześnie większe od opowiadających im median obliczonych dla danych z całej ligi.

5 Podsumowanie

Zbierzmy zatem wnioski z analiz. W pierwszej podsekcji pokazaliśmy, że mimo intuicji sugerującej, że wzrost może mieć bezpośredni wpływ na całkowitą wartość gracza, w lidze NBA takie zjawisko zdaje się nie występować. Dalej okazuje się, że jedyną zmienną wykazującą zdecydowany brak korelacji jest wzrost i zdobyte punkty, a wszystkie inne zmienne ilościowe oprócz wzrostu i asyst są mniej lub bardziej dodatnio skorelowane. Obserwacja ta sugeruje, że gracze lepsi w konkretnych elementach gry są też przeciętnie lepsi w pozostałych.

Źródła

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