

Raport z projektu

Analiza i wizualizacja danych

Gracjan Popiółkowski (grapop@st.amu.edu.pl)
Tomasz Sewastynowicz (tomsew@st.amu.edu.pl)
Stanisław Miedziejko (stamie@st.amu.edu.pl)

Kognitywistyka II rok.

1. Źródło danych.	1
2. Opis zawartości danych.	1
3. Proces czyszczenia danych.	3
4. Cel i kierunek przeprowadzonych analiz.	3
5. Analiza.	4
6. Wnioski.	10

1. Źródło danych.

W projekcie zastosowaliśmy zbiór danych NBA Players Performance z strony kaggle.com. Zestaw danych zawiera informacje o wynikach graczy National Basketball Association (NBA) w sezonie 2019-2020. Dane są typu open-source.

2. Opis zawartości danych.

Zbiór danych obejmuje wszystkie kluczowe wskaźniki wydajności, niezbędne do oceny jakości gry gracza. Zestaw danych zawiera informacje na temat takich rzeczy jak liczba rozegranych meczy, nazwa gracza, pozycja, współczynnik obrotów asyst, bloki na mecz, faule na rozegrane minuty, zbiórki na mecz, nazwę drużyny i więcej. W analizie danych skupimy się przede wszystkim na takich zmiennych jak: pozycja gracza, ilość rzutów wolnych, ilość rzutów za 3pkt, ile czasu gracz przebywał na meczu i ilość zdobytych punktów w polu. Wydaje nam się, że są to kluczowe czynniki mające związek z pozycją gracza.

Łącznie zbiór zawiera 14 zmiennych numerycznych oraz 3 zmienne nominalne.

Znacznie poszczególnych zmiennych naszej bazy danych:

GP Games Played - zmienna odnosi się do ilości gier rozegranych przez danego gracza.

GS Games Started - w przypadku tej zmiennej nie potrafimy określić co autor miał na myśli.

Mpg Minutes Per Game - odnosi się do ilości czasu (podanego w minutach) jaką średnio gracz spędził na boisku na jeden mecz.

Ppg Points Per Game - odnosi się do średniej ilości punktów zdobytych przez danego gracza na jeden mecz.

Fgm Field Goals Made - ta zmienna odnosi się do ilości trafień wszystkich rzutów do kosza, poza rzutami wolnymi, danego gracza.

Fga Field Goals Attempted - ta zmienna odnosi się do ilości prób trafienia wszystkich rzutów do kosza, wykluczając rzuty wolne, danego gracza.

FG % Field Goal Percentage - jest to zmienna opisująca procent trafień wszystkich rzutów do kosza, poza rzutami wolnymi, danego gracza.

3fgm Three Point Field Goals Made - ta zmienna odnosi się do ilości trafień wszystkich rzutów do kosza za 3 punkty, poza rzutami wolnymi, danego gracza.

3fga Three Point Field Goals Attempted - ta zmienna odnosi się do ilości prób trafień wszystkich rzutów do kosza za 3 punkty, poza rzutami wolnymi, danego gracza.

FTM Free Throws Made - ta zmienna odnosi się do ilości wykonanych, trafionych rzutów wolnych przez danego gracza.

FTA Free Throws Attempted - ta zmienna odnosi się do ilości wykonanych rzutów wolnych przez danego gracza.

FT% Free Throw Percentage - to zmienna która opisuje procent trafionych rzutów wolnych przez danego gracza.

Player - to zmienna, pod którą znajdują się: imię i nazwisko danego gracza.

Zmiennymi kategoryjnymi są **pozycje na boisku**, których mamy 5:

- SG - shooting guard (rzucający obrońca),
- SF - small forward (niski skrzydłowy),
- PF - power forward (silny skrzydłowy),
- PG - point guard (rozgrywający),
- C - center (środkowy).

i **nazwy drużyn**, drużyn w lidze NBA mamy 30 (i tyle jest w naszym zbiorze), lecz na potrzeby naszej analizy, wybraliśmy pięć drużyn, które uważamy za najistotniejsze z opisanych niżej powodów:

- **Boston Celtics** (BOS), zdobyli największą liczbę tytułów mistrzowskich ligi NBA, jako pierwsi zdobyli 17 tytułów.
- **Los Angeles Lakers** (LAL), również posiadają 17 tytułów mistrzowskich, drużynę z Bostonu dogonili w 2020 roku.
- **Golden State Warriors** (GS), zwycięzcy poprzedniego sezonu 2021-2022, którzy mają na koncie 7 krotne zwycięstwo, co pod względem liczby tytułów mistrzowskich stawia ich na trzecim miejscu.
- **Chicago Bulls** (CHI), drużyna na 4 miejscu pod względem liczby tytułów, rozpoznawalna dzięki postaci Michaela Jordana, który spędził w tej drużynie 13 sezonów.
- **Washington Wizards** (WAS), zespół mocno kojarzony w Polsce ze względu na Marcina Gortata, który rozegrał w tej drużynie 5 sezonów w swojej szczytowej formie. Również jest to druga drużyna Michaela Jordana, który rozegrał w niej swoje dwa ostatnie sezony.

3. Proces czyszczenia danych.

Do obliczenia średnich w tabelach usunęliśmy za pomocą excela puste dane, które mogły być odebrane za 0. Wszędzie gdzie jest mowa o drużynach zostawiliśmy w analizie tylko pięć wymienionych w poprzednim punkcie (BOS, LAL, GS, CHI, WAS). Tabele i wykresy, w których nic nie wspominamy o drużynach są na bazie wszystkich zawodników z trzydziestu drużyn NBA.

4. Cel i kierunek przeprowadzonych analiz.

Celem naszej analizy jest znalezienie, opisanie i zdefiniowanie zależności pomiędzy różnymi zmiennymi ilościowymi i nominalnymi w zbiorze danych, aby móc wyciągnąć wnioski na temat sezonu 2019-2020 ligi NBA. W raporcie postaramy się odpowiedzieć na takie pytania jak:

- I. Którzy gracze najlepiej rzucają za 3 punkty (mieli najwyższą liczbę trafień)?
- II. Którzy gracze mają największą średnią zdobytych punktów na mecz?
- III. Którzy gracze mają największą celność rzutów wolnych w tym sezonie?
- IV. Którzy gracze rzucili najwięcej rzutów wolnych (są najwięcej faulowani albo umiejętnie wymuszają faule)?
- V. Gracze z jakiej pozycji najczęściej przystępują do rzutów wolnych?
- VI. Jaka z naszych pięciu wybranych drużyn, ma największy skumulowany procent celności rzutów?
- VII. Gracze z jakich pozycji trafiają najwięcej rzutów za 3?
- VIII. Czy w lidze NBA, graczy różnych pozycji jest tyle samo?
- IX. Jak wygląda porównanie naszych pięciu drużyn pod względem rzutów z boiska (bez rzutów wolnych)?
- X. Czy zawsze czas na boisku, ile dany gracz jest na boisku jest skorelowane z ilością zdobytych przez niego punktów?

5. Analiza.

ranga	nazwa gracza	pozycja	ilość trafień za 3 pkt
1	Stephen Curry	PG	103
2	Anfernee Simons	SG	79
3	Malik Beasley	SG	75
4	Donovan Mitchell	SG	71
5	Buddy Hield	SG	71
6	Klay Thompson	SG	67
7	Jayson Tatum	PF	66
8	Jordan Clarkson	PG	63
9	Kevin Huerter	SG	62
10	Jerami Grant	PF	58

Tabela nr 1: przedstawiająca top 10 graczy NBA pod względem ilości trafień za 3 pkt.

W top 10 graczach pod względem średniej ilości punktów na mecz pozycje się rozkładają: 4 rozgrywający (PG), 3 silny skrzydłowy (PF), 3 rzucający obrońca (SG).

ranga	nazwa gracza	pozycja	średnia ilość pkt na mecz
1	Luka Doncic	PG	33.5
2	Stephen Curry	PG	31.4
3	Shai Gilgeous-Alexander	SG	31.1
4	Giannis Antetokounmpo	PF	30.9
5	Jayson Tatum	PF	30.8
6	Kevin Durant	PF	30.0
7	Donovan Mitchell	SG	28.9
8	Ja Morant	PG	28.5
9	Devin Booker	SG	27.9
10	Trae Young	PG	27.7

Tabela nr 2: przedstawiająca top 10 graczy NBA pod względem średniej ilości punktów na mecz.

Co może być lekkim zaskoczeniem dwóch zawodników z największą średnią zdobytych punktów, Luka Doncic i Stephen Curry grają na pozycji rozgrywających (PG), jest to najmniej

skojarzona pozycja z zdobywaniem punktów, co świadczy o dużej uniwersalności tych zawodników.

ranga	nazwa gracza	pozycja	procent trafień rzutów wolnych [%]	ilość oddanych rzutów wolnych	ilość trafionych rzutów wolnych
1	AJ Griffin	SF	100	7	7
2	Troy Brown Jr.	SF	100	11	11
3	Andrew Nembhard	PG	100	7	7
4	Derrick Rose	PG	100	10	10
5	Jake LaRavia	SF	100	5	5
6	Garrison Mathews	SF	95	20	19
7	Reggie Jackson	PG	94.1	34	32
8	Ayo Dosunmu	PG	94.1	17	16
9	Trey Murphy III	SF	93.9	33	31
10	Edmond Sumner	PG	93.9	33	31

Tabela nr 3: przedstawia topowych 10 graczy NBA ze względu na procent trafień rzutów wolnych.

ranga	nazwa gracza	pozycja	procent trafień rzutów wolnych [%]	ilość oddanych rzutów wolnych	ilość trafionych rzutów wolnych
1	Luka Dončić	PG	72.3	213	154
2	Giannis Antetokounmpo	PF	60	200	120
3	Shai Gilgeous-Alexander	SG	91.9	185	170
4	Trae Young	PG	90.2	174	157
5	Kevin Durant	PF	91.9	172	158
6	Jayson Tatum	PF	87	169	147
7	DeMar DeRozan	SF	88	158	139
8	Ja Morant	PG	75.6	156	118
9	Devin Booker	SG	88	133	117
10	Anthony Davis	C	81.5	130	106

Tabela nr 4: przedstawia 10 graczy, którzy wykonali największą ilość rzutów wolnych.

Obserwując wyniki tabeli nr 3 i 4 możemy zauważyć, że wysoki procent trafionych rzutów wolnych ($\geq 95\%$) był uzależniony od małej ilości wykonanych rzutów wolnych. Najwyższe procentowe wyniki trafień osiągnęli zawodnicy z pozycji SF i PG. Patrząc na największą ilość wykonanych rzutów wolnych przez zawodników (Tab. nr 4) nie jesteśmy w stanie wyciągnąć jednoznacznych wniosków co do wpływu ilości wykonanych rzutów wolnych na ilość trafień i wykazać jakiegokolwiek korelacji z pozycją.

pozycja	średnia ilość rzutów wolnych
C (Środkowy)	94,62
PF (Silny skrzydłowy)	54,98
PG (Rozgrywający)	33,74
SF (Niski skrzydłowy)	29,02
SG (Strzelający obrońca)	24,86

Tabela nr 5: przedstawia średnią ilość rzutów wolnych wykonanych przez graczy na odpowiednich pozycjach (wzięci są pod uwagę gracze z naszych pięciu wybranych zespołów).

Widać tutaj, że najwięcej rzutów wolnych rzucają środkowi i silni skrzydłowi, co wiąże się najpewniej z faktem, iż te pozycje działają najbliżej kosza i często są narażone na faule pod koszem, które są najczęstszym powodem rzutów wolnych.

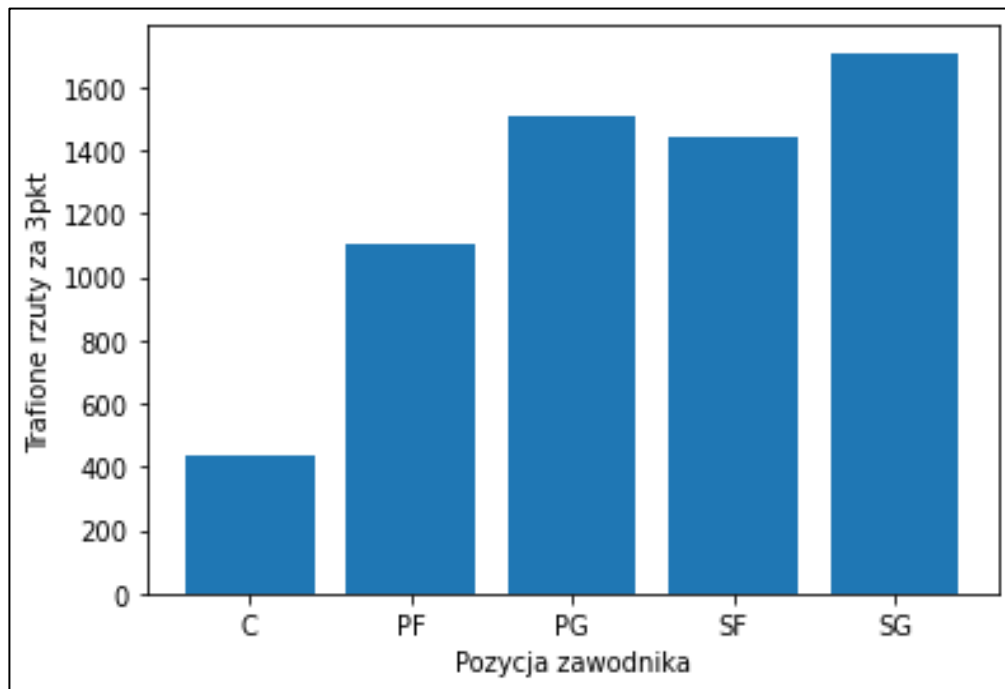
drużyna	średni procent trafień z boiska (nie licząc wolnych)	średni procent za 3 punkty	średni procent trafień rzutów wolnych	uśredniona skuteczność rzutów drużyny
BOS	52,12	42,13	83,26	59,17
CHI	47,54	31,95	81,57	53,69
GS	49,00	33,15	77,67	53,27
LAL	48,31	31,26	85,79	55,12

Tabela nr 6: przedstawia procenty skuteczności rzutów członków danych drużyn.

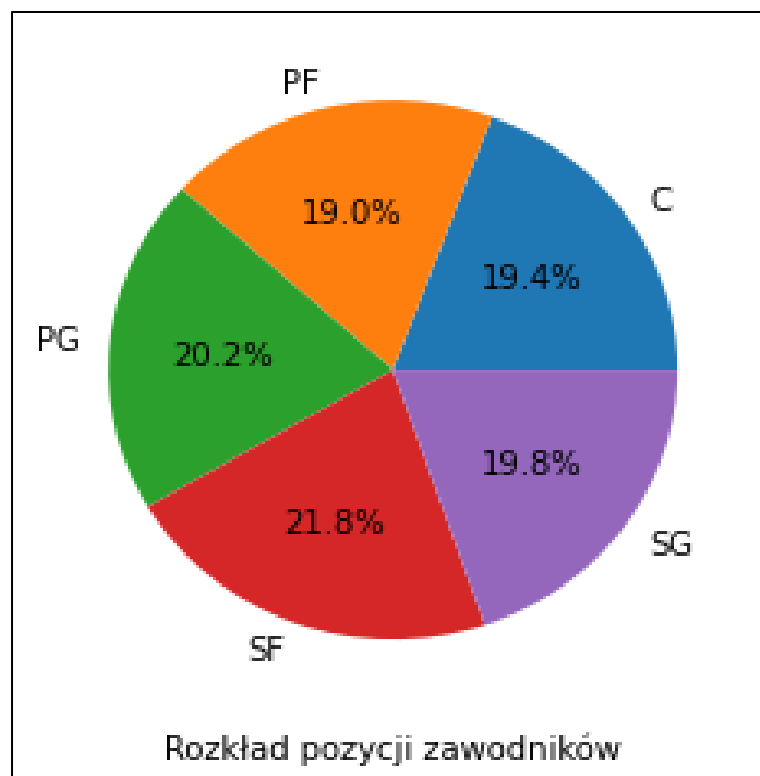
Widzimy tutaj, iż średnia celność trafień zawodników z różnych zespołów jest bardzo zbliżona. Tylko drużyna z Bostonu (BOS) wyróżnia się wyjątkowo wysoką celnością rzutów za trzy punkty (aż o około 10 punktów procentowych więcej niż pozostałe drużyny). Co zaważy również średnią trafień w ogóle przez Bostońską drużynę.

Podejrzewaliśmy, że tego przyczyną może być jakiś znakomity strzelający obrońca tego zespołu, jednak co się okazało w danych, wszyscy gracze z Bostońskiego zespołu mają zbliżoną celność rzutów za trzy punkty, właśnie około 40%.

Z pierwszego wykresu możemy odczytać związek między zmienną pozycja zawodnika, a ilością trafień rzutów za 3pkt. Dominantą jest pozycja rzucającego obrońcy (SG), Również pozycje niskiego skrzydłowego (SF) oraz rozgrywający (PG) wyróżniają się ilością trafionych rzutów za 3pkt na tle pozostałych pozycji.



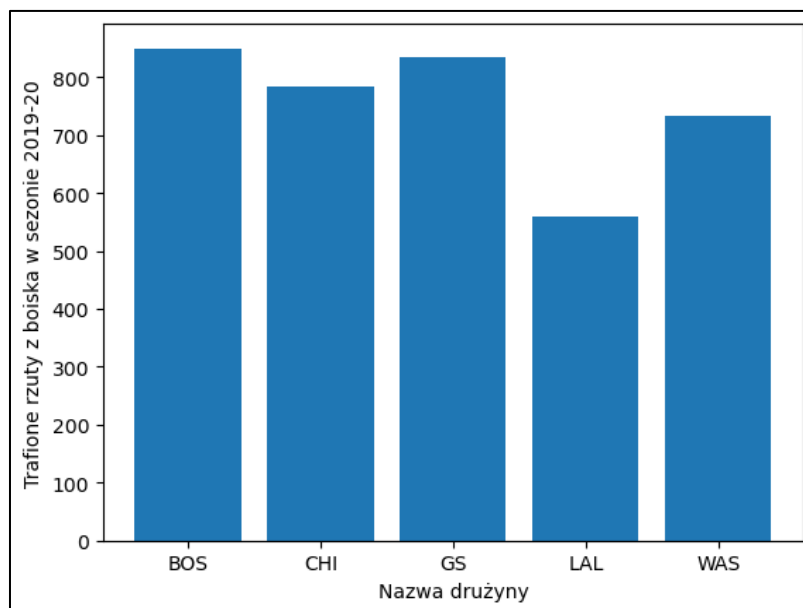
Rys. nr 1: wykres słupkowy przedstawiający stosunek pozycji gracza do trafionych rzutów za 3 punkty.



Rys. nr. 2: wykres kołowy przedstawiający rozkład pozycji.

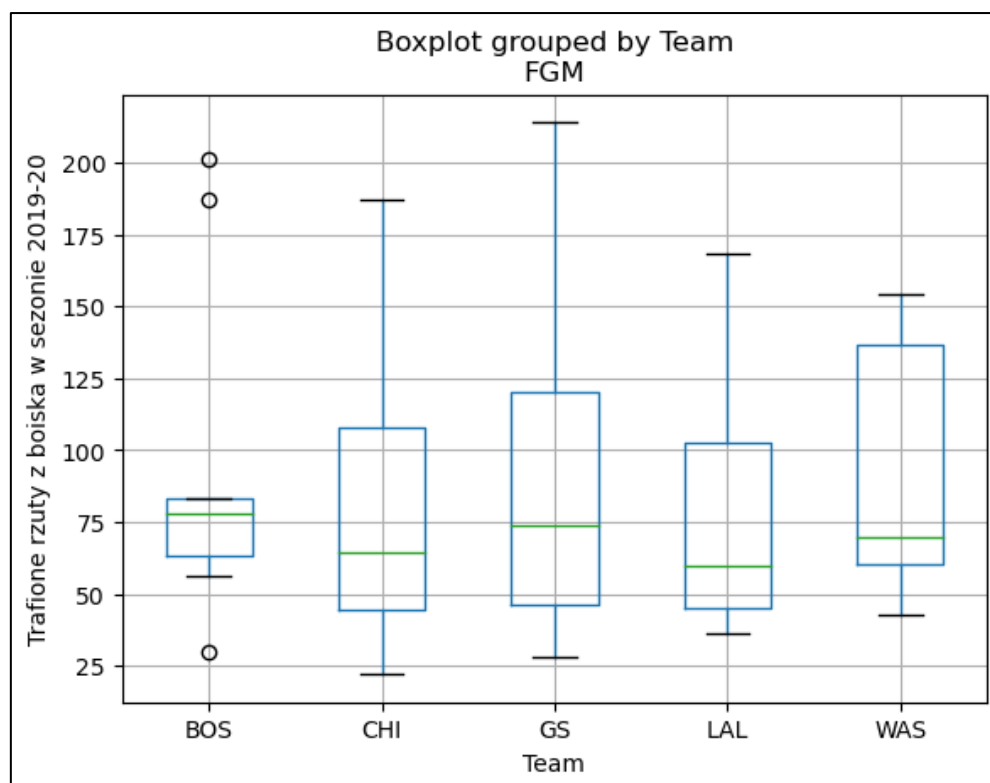
Zmienna nominalna "Player position" posiada 5 kategorii.

Występuje jedna dominanta - pozycja SF (niski skrzydłowy). Na tej pozycji w lidze NBA jest 21.8% graczy. Natomiast drugą najczęściej spotykaną pozycją jest PG (rozgrywający) 20.2%.



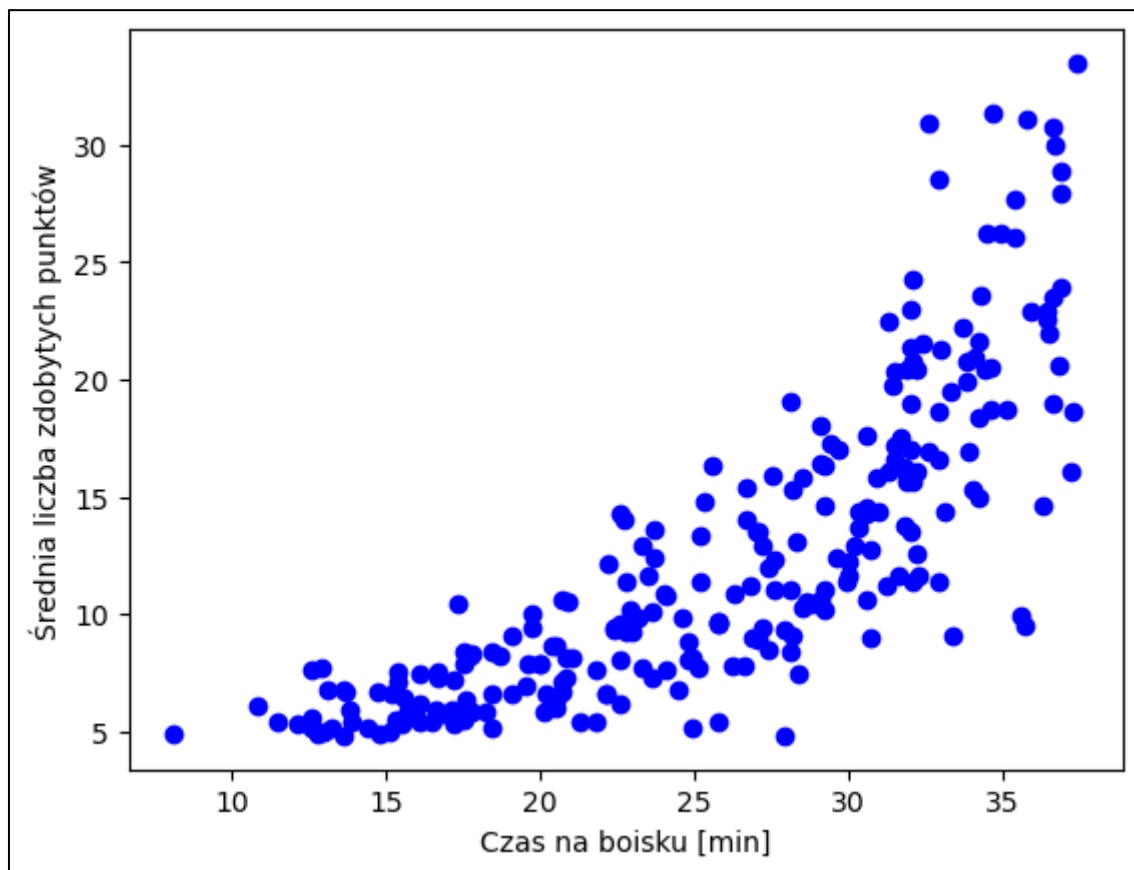
Rys. nr 3: wykres słupkowy przedstawiający, ile dane drużyny trafiły rzutów do kosza w sezonie 2019-2020.

Wykres porównujący liczbę trafionych rzutów z boiska, naszych pięciu wyszczególnionych drużyn. Widać tutaj, że drużyna z Los Angeles (LAL), miała zdecydowanie słabszy sezon. A największą liczbą trafień może się pochwalić drużyna z Bostonu.



Rys. nr 4: wykres pudełkowy, przedstawia wartości minimalne, pierwszy kwartył, medianę, trzeci kwartył i wartość maksymalną dla trafionych rzutów, zawodników z naszych pięciu drużyn.

Zamieszczamy również wykres pudełkowy, na którym można zobaczyć rozkład trafień zawodników z tych drużyn. Widzimy tutaj, że mediana trafień na zawodnika również jest największa w drużynie Bostonu (BOS), dwóch zawodników z tej drużyny miało wyjątkowo dużo trafnych rzutów, widoczni jako niezamalowane kropki Jayson Tatum, który ma na swoim koncie 201 trafień i Jaylen Brown z 187 trafieniami. Widzimy też, że drużyna Golden State Warriors posiada zawodnika z największą liczbą trafień, jest to Stephen Curry z 214 trafieniami z boiska.

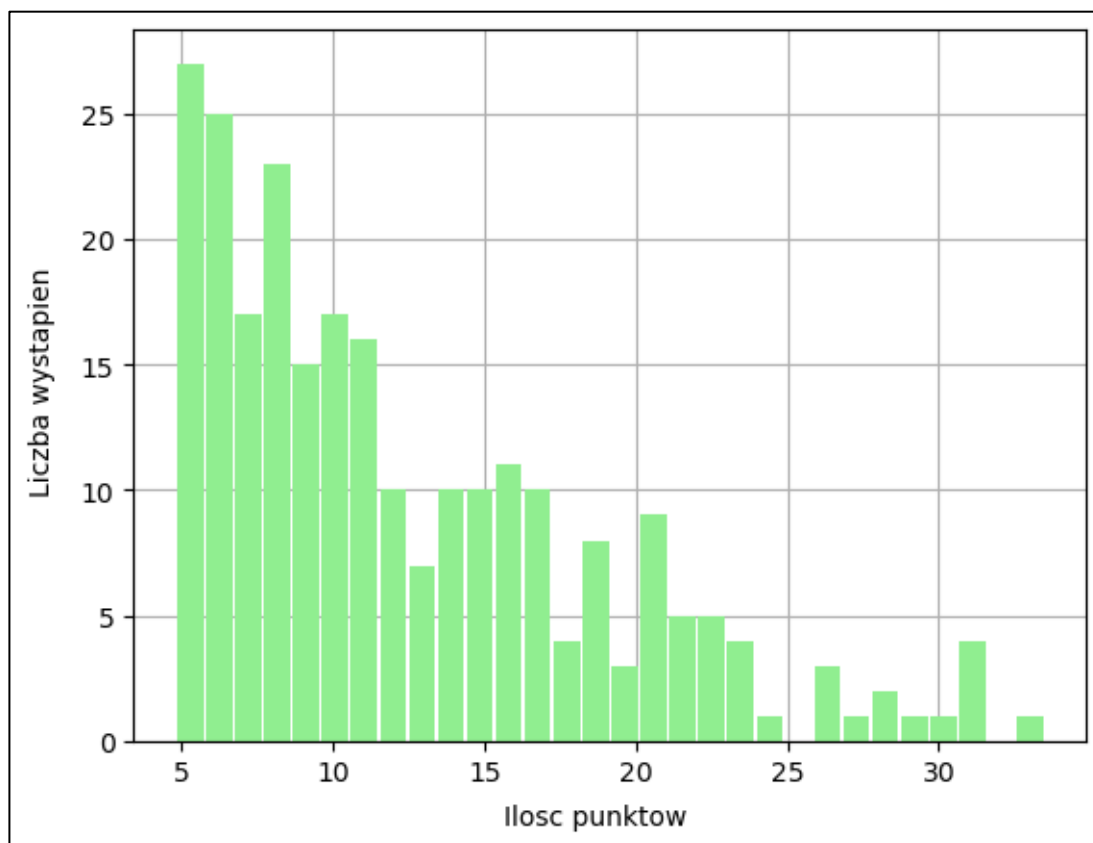


Rys. nr 5: wykres przedstawia rozstrzał graczy pod względem zdobytych punktów i czasu na boisku.

Jak możemy zauważyć, wraz z czasem spędzonym na boisku rośnie ilość punktów zdobyta przez zawodnika. Nie rośnie ona jednak jednostajnie liniowo, już w interwale czasowym 20-25 min możemy zauważyć spadek ilości punktów, jaką gracz zdobywa spędzając tyle czasu na boisku.

Na podstawie tych danych sprawdziliśmy hipotezę, zachodzenia korelacji między czasem spędzonym na boisku, a liczbą zdobytych punktów.

Wyliczyliśmy za pomocą biblioteki „scipy.stats” współczynnik korelacji liniowej Spearmana, którego siła wyszła równa 0.87, co przy dostatecznej istotności może świadczyć o bardzo silnej korelacji.



Rys. nr 6: Przedstawiający ilość średniej zdobytej ilości punktów przez zawodników.

Histogram przedstawia jak często gracze średnio zdobywają daną ilość punktów. Jak możemy zauważyć, w miarę jak średnia punktów na mecz jest wyższa, maleje liczba wystąpień. Mediana dla wartości z wykresu z rys. nr 6 jest równa 10.6, średnia 12.45, dominanta, 5.4.

6. Wnioski.

Nasz raport poruszył najróżniejsze aspekty sezonu 2019-2020 ligi NBA, z każdej tabeli i wykresu można wyciągnąć interesujące dla fanów sportu dane, a również być może przydatną analizę dla samych zespołów i zawodników.

Przeanalizowaliśmy wyniki poszczególnych graczy i również wyniki drużyn, porównaliśmy cztery drużyny z największą liczbą wygranych tytułów mistrzowskich i jako dodatek dla drużyny Washington Wizards, by zobaczyć jak sobie poradziła pierwszy sezon bez naszej polskiej gwiazdy koszykówki Marcina Gortata.

Uważamy, że takie przeglądy danych z poszczególnych sezonów są bardzo ważnym elementem analizy sportowych wyników i nie powinno się wyciągać wniosków na samych wygranych czy przegranych meczach.