

Analiza danych z bazy "Curling Masters"

Joanna Matuszak, Joanna Wojciechowicz, Wiktoria Marzec, Tymoteusz Kempa,
Krystian Walewski

20 czerwca 2022

Spis treści

1	Połączenie z bazą	1
2	Analiza danych	2
2.1	Liczba spotkań każdego miesiąca	2
2.2	Bilans wygranych i przegranych drużyny	2
2.3	Lista najlepszych zawodników	2
2.4	Liczba kibiców w zależności od momentu sezonu	2
2.5	Budżet klubu na przestrzeni czasu	3
2.6	Lista najbardziej aktywnych zawodników	3
2.7	Najbardziej hojni sponsorzy	3
2.8	Korelacja między średnią efektywnością a wynikiem meczu	3
2.9	Drużyna zdobywająca średnio najwięcej punktów	4
3	Wykresy i tabele	4

1 Połączenie z bazą

Aby połączyć się z bazą danych skorzystamy z bibliotek DBI oraz RMariaDB. Następnie możemy pobrać potrzebne dane.

```
connection <- dbConnect(RMariaDB::MariaDB(),
                        host='giniewicz.it',
                        dbname='team9',
                        user=USER,
                        password=PASSWORD)

matches <- tbl(connection, "Matches") %>% data.frame()
teams <- tbl(connection, "Teams") %>% data.frame()
player_match_performance <- tbl(connection, "PlayerMatchPerformance") %>% data.frame()
players <- tbl(connection, "Players") %>% data.frame()
info <- tbl(connection, "Info") %>% data.frame()
cash_flow <- tbl(connection, "CashFlow") %>% data.frame()
outside_income <- tbl(connection, "OutsideIncome") %>% data.frame()

dbDisconnect(connection)
```

Przed przystąpieniem do analizy musimy przygotować odpowiednio nasze dane.

```

matches$Date <- as.Date(matches$Date, "%Y-%m-%d")
matches$Month <- as.Date(cut(matches$Date, breaks = "month"))

matches_played <- matches %>% drop_na(ScoreA)

info$BirthDate <- as.Date(info$BirthDate, "%Y-%m-%d")

```

W dalszej części sprawozdania prezentowane są jedynie wyniki w postaci tabel i wykresów. Pełny kod generujący raport znajduje się w pliku `report.Rnw`.

2 Analiza danych

2.1 Liczba spotkań każdego miesiąca

Na wykresie 1 możemy zobaczyć liczbę spotkań, które odbyły się każdego miesiąca. Widzimy, że mecze odbywają się z regularną okresowością. Sezon rozpoczyna się w pierwszym tygodniu listopada i trwa 20 tygodni, a w każdym tygodniu mamy po jednym meczu w danej kategorii (FEMALE JUNIOR, FEMALE SENIOR, MALE JUNIOR, MALE SENIOR). Stąd co tydzień mamy po cztery spotkania – dwa w sobotę i dwa w niedzielę. Na wykresie 2 możemy zobaczyć, jak będzie wyglądać liczba meczów w miesiącu, włączając także spotkania zaplanowane, które się jeszcze nie odbyły.

2.2 Bilans wygranych i przegranych drużyny

Z racji tego, że każda drużyna do tej pory zagrała taką samą liczbę meczów, możemy łatwo porównać, które z nich prezentują się najlepiej pod kątem łącznej liczby wygranych spotkań. Wyniki możemy zobaczyć na wykresie 3 oraz w tabeli 1. Możemy też wskazać najlepsze drużyny w każdej kategorii – tabela 2.

Przy okazji porównywania bilansu wygranych i przegranych możemy też wyłonić zwycięzców każdej kategorii w sezonie 2020/2021 oraz w sezonie 2021/2022 (tabele 3 i tabela 4). Gdy pojawią się remisy, wtedy zwycięzcą zostaje ta drużyna, która ma łącznie najwięcej zdobytych punktów.

2.3 Lista najlepszych zawodników

Jako kryterium wyboru najlepszych zawodników przyjęliśmy ich średnią efektywność. Czołówka zawodników została przedstawiona w tabeli 5. Możemy też wskazać najlepszego zawodnika w każdej drużynie. Wyniki zostały przedstawione w tabelach 6 - 9.

2.4 Liczba kibiców w zależności od momentu sezonu

Na wykresach 4 oraz 5 możemy zobaczyć, jak zmieniało się zainteresowanie meczami na przestrzeni czasu, odpowiednio podczas sezonu 2020/2021 oraz 2021/2022. W obu przypadkach widzimy pewien schemat – na początku sezonu zainteresowanie jest duże, w jego trakcie liczba kibiców danego dnia spada, a w końcówce sezonu pojawia się trend rosnący – zainteresowanie meczami wzrasta.

Możemy formalnie sprawdzić istnienie zależności między liczbą widzów a momentem sezonu (a konkretnie oddaleniem momentu sezonu od jego środka) wykonując test na korelację. Otrzymaliśmy $p\text{-value} < 0$, zatem odrzucamy hipotezę zerową, mówiącą o nieskorelowaniu zmiennych. Wnioskujemy, że zainteresowanie meczami jest powiązane z czasem od środka sezonu.

```

##
## Pearson's product-moment correlation
##

```

```
## data: spectators$SelledToday and as.numeric(abs(dates$Date - mid_date$Date))
## t = 8.5245, df = 38, p-value = 2.369e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.6672470 0.8957317
## sample estimates:
## cor
## 0.8103269
```

2.5 Budżet klubu na przestrzeni czasu

Ważną kwestią jest także budżet klubu. Wykres 6 przedstawia jak zmieniał się on na przestrzeni czasu. W momencie założenia klubu (1 września 2020 roku) budżet był dosyć wysoki, wyniósł około milion złotych. Widzimy, że generalnie wydatki klubu przewyższają przychody - trend jest ewidentnie malejący. Możemy też zauważyć, że zmiany w budżecie są mniej więcej cykliczne, stan konta zmienia się jedynie kilka razy w miesiącu.

2.6 Lista najbardziej aktywnych zawodników

W tabeli 10 przedstawiona została lista zawodników, którzy zagraли w największej liczbie spotkań. W tabeli znajdziemy dane dla graczy, którzy nie brali czynnego udziału maksymalnie w jednym meczu. Możemy też sprawdzić, którzy zawodnicy rozegrali najmniejszą liczbę spotkań – wyniki przedstawione są w tabeli 11.

2.7 Najbardziej hojni sponsorzy

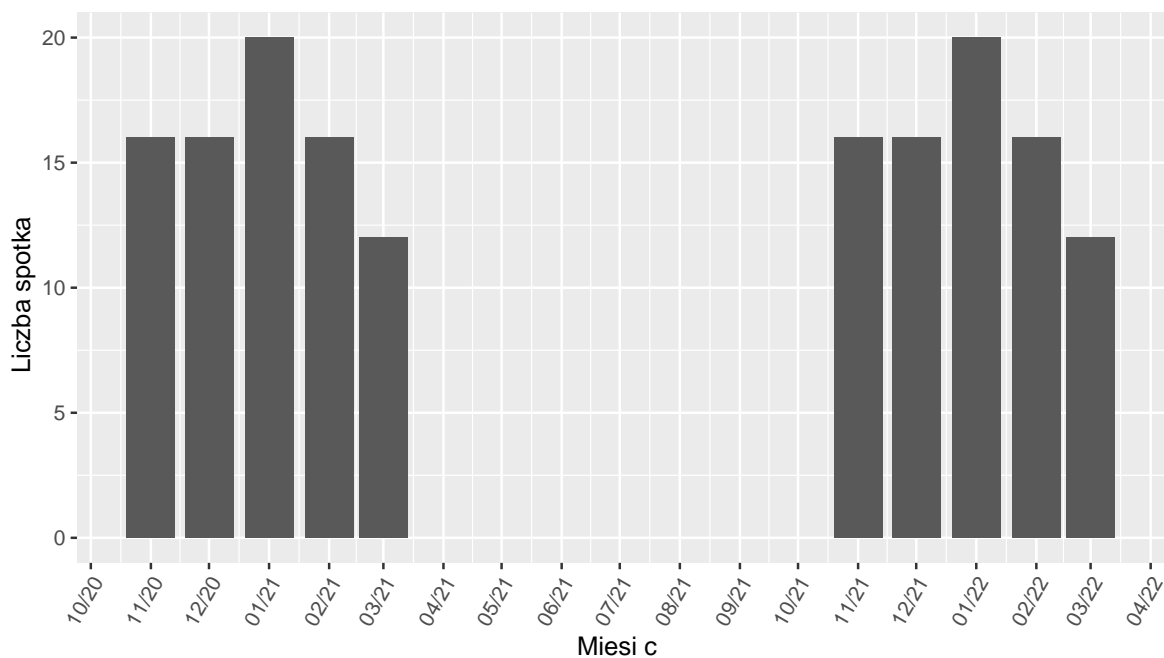
Znaczącym źródłem dochodów klubu są dotacje od sponsorów. W tabeli 12 możemy zobaczyć zestawienie, z którego możemy odczytać, które współprace są najbardziej korzystne finansowo dla klubu.

2.8 Korelacja między średnią efektywnością a wynikiem meczu

Chcemy sprawdzić, czy istnieje związek między efektywnością drużyny a wynikiem meczu – wygrana czy przegrana. Jako efektywność drużyny przyjmujemy przeciętną efektywność zawodnika biorącego udział w meczu. W tym celu przeprowadzimy test korelacji pomiędzy zmiennymi, gdzie hipoteza zerowa oznacza korelację równą 0.

```
##
## Pearson's product-moment correlation
##
## data: TeamID_effectiveness and if_TeamID_wins
## t = 7.5799, df = 318, p-value = 3.818e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.2941661 0.4802322
## sample estimates:
## cor
## 0.3911893
```

W naszym przypadku $p\text{-value} < 0.05$, zatem odrzucamy hipotezę zerową. Wnioskujemy, że efektywność zawodników oraz rezultat meczu są ze sobą powiązane.

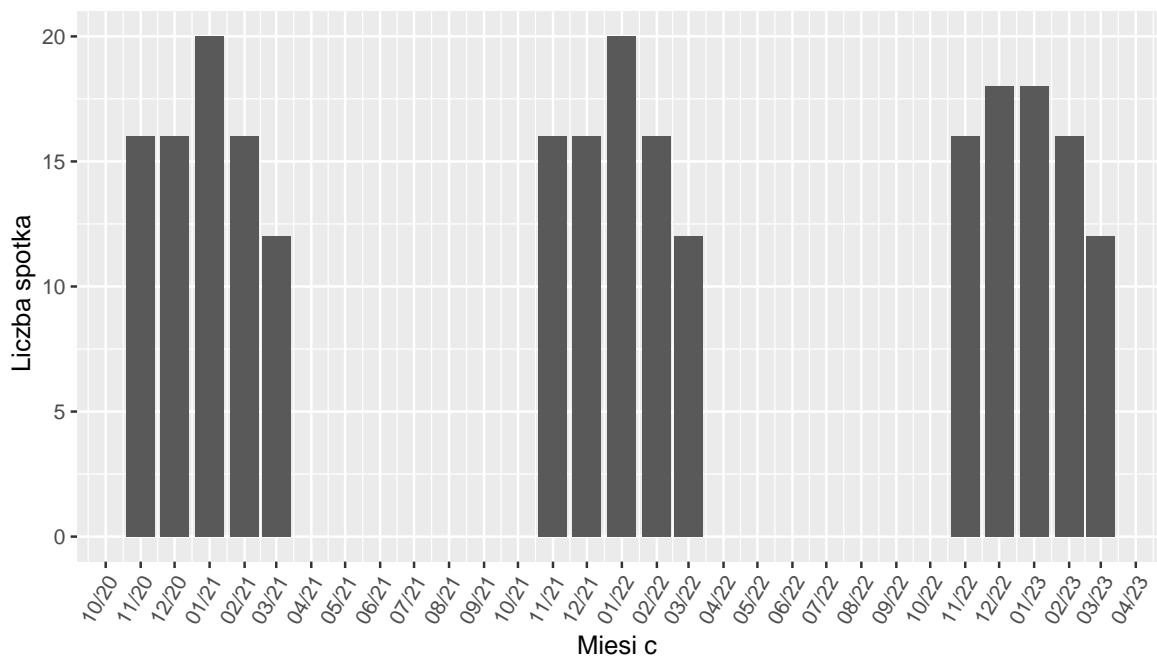


Rysunek 1: Liczba spotkań każdego miesiąca

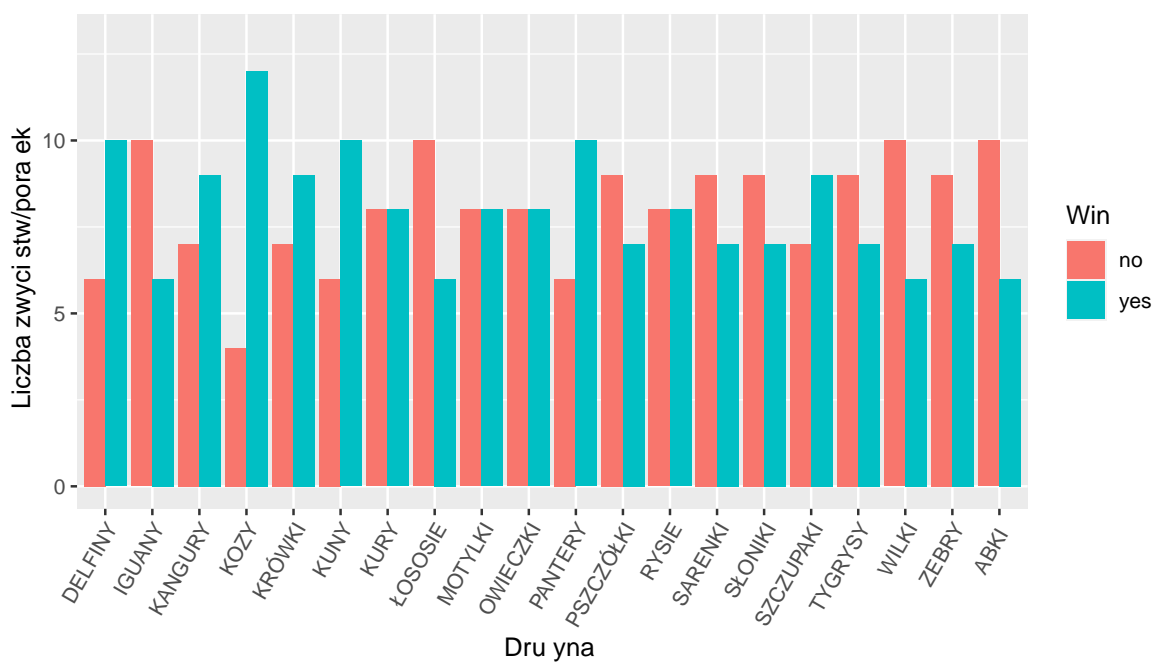
2.9 Drużyna zdobywająca średnio najwięcej punktów

Chcemy porównać drużyny pod kątem średniej liczby zdobytych punktów. Wyniki możemy zobaczyć w tabeli 13. Najlepsze drużyny w każdej kategorii zostały przedstawione w tabeli 14.

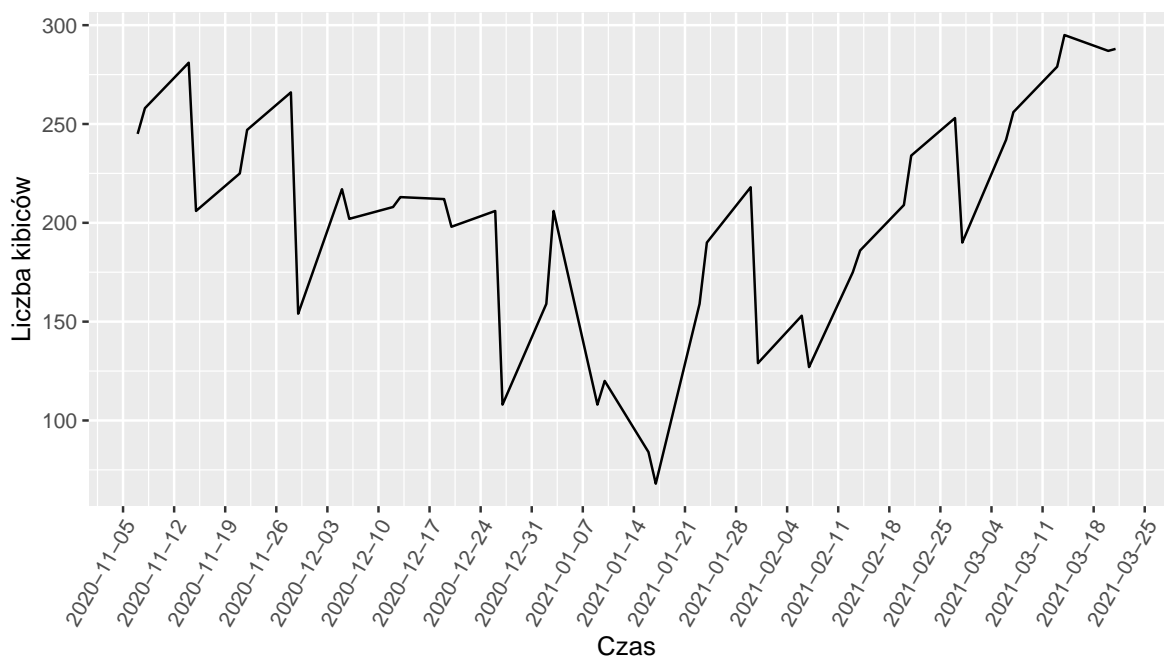
3 Wykresy i tabele



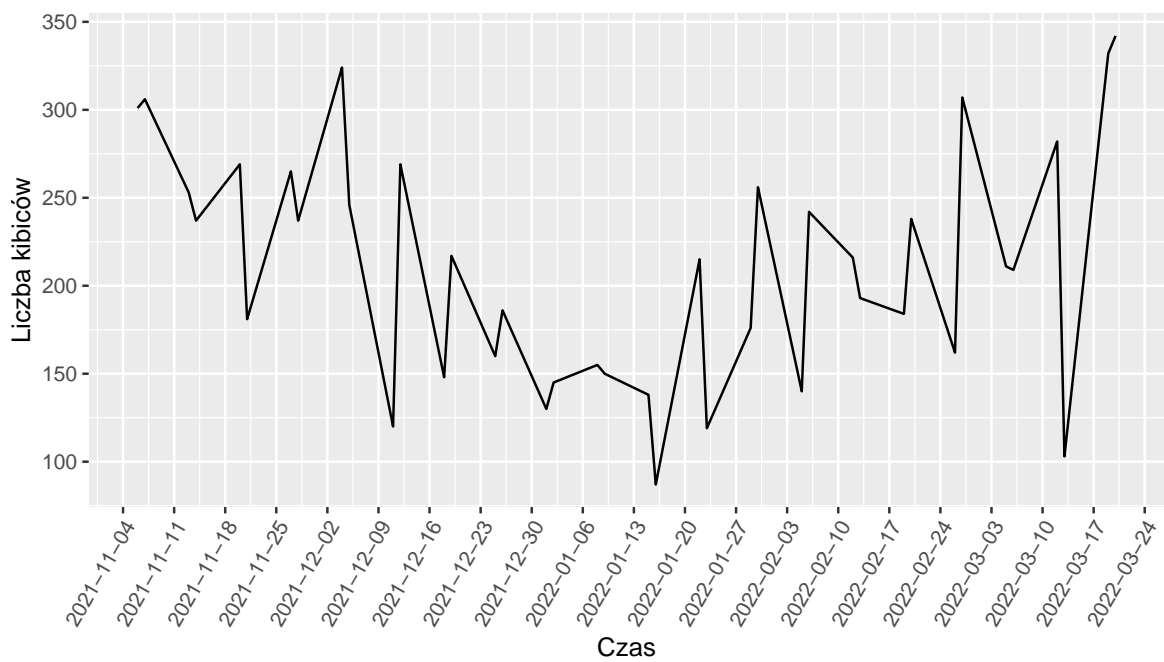
Rysunek 2: Liczba spotkań każdego miesiąca - spotkania zakończone oraz zaplanowane



Rysunek 3: Bilans wygranych i przegranych drużyny



Rysunek 4: Liczba kibiców w sezonie 2020/2021



Rysunek 5: Liczba kibiców w sezonie 2021/2022

	TeamID	TeamName	Type	AgeCategory	NumberOfWins	NumberOfLoses
1	20	KOZY	MALE	SENIORS	12	4
2	7	PANTERY	FEMALE	SENIORS	10	6
3	8	DELFINY	FEMALE	SENIORS	10	6
4	15	KUNY	MALE	JUNIORS	10	6
5	5	KRÓWKI	FEMALE	JUNIORS	9	7
6	13	KANGURY	MALE	JUNIORS	9	7
7	14	SZCZUPAKI	MALE	JUNIORS	9	7
8	1	MOTYLKI	FEMALE	JUNIORS	8	8
9	2	OWIECZKI	FEMALE	JUNIORS	8	8
10	4	KURY	FEMALE	JUNIORS	8	8
11	17	RYSIE	MALE	SENIORS	8	8
12	3	TYGRYSY	FEMALE	JUNIORS	7	9
13	6	ZEBRY	FEMALE	SENIORS	7	9
14	10	SARENKI	FEMALE	SENIORS	7	9
15	16	SŁONIKI	MALE	SENIORS	7	9
16	19	PSZCZÓŁKI	MALE	SENIORS	7	9
17	9	WILKI	FEMALE	SENIORS	6	10
18	11	ŁOSOSIE	MALE	JUNIORS	6	10
19	12	ŻABKI	MALE	JUNIORS	6	10
20	18	IGUANY	MALE	SENIORS	6	10

Tabela 1: Bilans sukcesów i porażek

	TeamID	TeamName	Type	AgeCategory	NumberOfWins	NumberOfLoses
1	5	KRÓWKI	FEMALE	JUNIORS	9	7
2	7	PANTERY	FEMALE	SENIORS	10	6
3	8	DELFINY	FEMALE	SENIORS	10	6
4	15	KUNY	MALE	JUNIORS	10	6
5	20	KOZY	MALE	SENIORS	12	4

Tabela 2: Najlepsze drużyny w każdej kategorii

	TeamID	TotalPoints	TeamName	Type	AgeCategory	NumberOfWins	NumberOfLoses
1	5	65	KRÓWKI	FEMALE	JUNIORS	5	3
2	8	77	DELFINY	FEMALE	SENIORS	8	0
3	13	67	KANGURY	MALE	JUNIORS	5	3
4	17	68	RYSIE	MALE	SENIORS	5	3

Tabela 3: Zwycięzcy Ligi - sezon 2020/2021

	TeamID	TotalPoints	TeamName	Type	AgeCategory	NumberOfWins	NumberOfLoses
1	2	64	OWIECZKI	FEMALE	JUNIORS	6	2
2	7	67	PANTERY	FEMALE	SENIORS	6	2
3	15	65	KUNY	MALE	JUNIORS	6	2
4	20	68	KOZY	MALE	SENIORS	7	1

Tabela 4: Zwycięzcy Ligi - sezon 2021/2022

	PlayerID	MeanEffectiveness	FirstName	LastName	TeamName
1	83	0.80	ALAN	JAMRÓZ	RYŚIE
2	80	0.80	ARTUR	MORAWSKI	SŁONIKI
3	72	0.80	ZENON	GŁĄB	KUNY
4	37	0.80	NATALIA	SZKUDLAREK	DELFINY
5	74	0.79	MATEUSZ	MIŚ	KUNY
6	57	0.79	WOŁODYMYR	KRUSZEWSKI	ŻABKI
7	26	0.79	EWA	MROCZKOWSKA	ZEBRY
8	69	0.78	KONRAD	WOJTOWICZ	SZCZUPAKI
9	85	0.78	PAWEŁ	ROSZAK	RYŚIE
10	55	0.78	JANUSZ	SKIBA	ŁOSOSIE

Tabela 5: Lista 10 najlepszych zawodników

	TeamID	PlayerID	MeanEffectiveness	FirstName	LastName	TeamName
1	1	3	0.78	PAULINA	WOLNA	MOTYLKI
2	5	25	0.78	IZABELA	MIERZEJEWSKA	KRÓWKI
3	2	10	0.77	KINGA	MROWIEC	OWIECZKI
4	3	13	0.74	MARIOŁA	PILARSKA	TYGRYSY
5	4	19	0.74	BARBARA	BILIŃSKA	KURY

Tabela 6: Lista najlepszych zawodniczek w danej drużynie z kategorii FEMALE JUNIORS

	TeamID	PlayerID	MeanEffectiveness	FirstName	LastName	TeamName
1	8	37	0.80	NATALIA	SZKUDLAREK	DELFINY
2	6	26	0.79	EWA	MROCZKOWSKA	ZEBRY
3	7	32	0.77	HELENA	MODZELEWSKA	PANTERY
4	9	43	0.76	ANITA	PIWOWARCZYK	WILKI
5	10	46	0.74	WŁADYSŁAWA	MORAWSKA	SARENKI

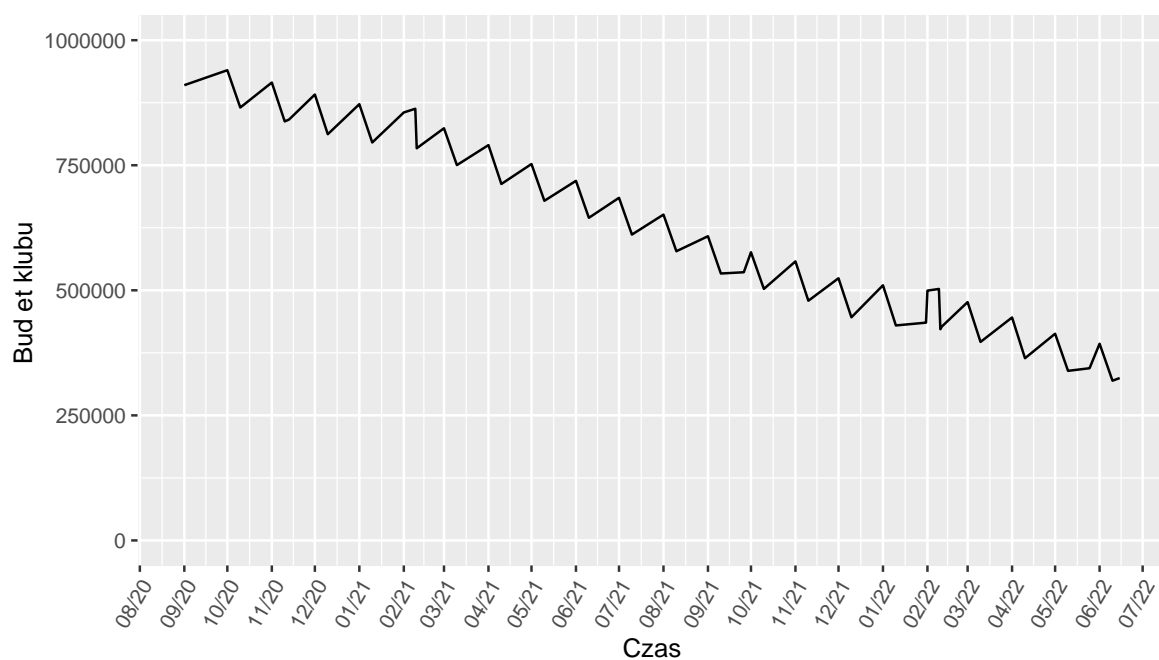
Tabela 7: Lista najlepszych zawodniczek w danej drużynie z kategorii FEMALE SENIORS

	TeamID	PlayerID	MeanEffectiveness	FirstName	LastName	TeamName
1	15	72	0.80	ZENON	GŁĄB	KUNY
2	12	57	0.79	WOŁODYMYR	KRUSZEWSKI	ŻABKI
3	14	69	0.78	KONRAD	WOJTOWICZ	SZCZUPAKI
4	11	55	0.78	JANUSZ	SKIBA	ŁOSOSIE
5	13	62	0.77	IGNACY	SZYSZKA	KANGURY

Tabela 8: Lista najlepszych zawodników w danej drużynie z kategorii MALE JUNIORS

	TeamID	PlayerID	MeanEffectiveness	FirstName	LastName	TeamName
1	17	83	0.80	ALAN	JAMRÓZ	RYŚIE
2	16	80	0.80	ARTUR	MORAWSKI	SŁONIKI
3	20	97	0.76	ZBIGNIEW	BUGAJ	KOZY
4	18	89	0.75	IGOR	CICHOŃ	IGUANY
5	19	95	0.73	JAROSŁAW	MIELCZAREK	PSZCZÓŁKI

Tabela 9: Lista najlepszych zawodników w danej drużynie z kategorii MALE SENIORS



Rysunek 6: Budżet klubu

	PlayerID	FirstName	LastName	TeamName	MatchesPlayed
1	14	IZABELA	WOŚ	TYGRYSY	16
2	47	TERESA	KRUSZYŃSKA	SARENKI	16
3	4	EWA	SOJKA	MOTYLKI	15
4	28	AGATA	PYTEL	ZEBRY	15
5	34	KAZIMIERA	CZAJKA	PANTERY	15
6	36	JADWIGA	OLCZYK	DELFINY	15
7	51	BOGUSŁAW	STOKŁOSA	ŁOSOSIE	15
8	72	ZENON	GŁĄB	KUNY	15
9	75	ZENON	BEDNARCZYK	KUNY	15
10	76	SYLWESTER	GWÓŹDŹ	SŁONIKI	15
11	81	FABIAN	WEBER	RYSIE	15
12	94	TADEUSZ	ŻUK	PSZCZÓŁKI	15

Tabela 10: Zawodnicy z największą liczbą zagranych meczów

	PlayerID	FirstName	LastName	TeamName	MatchesPlayed
1	35	EDYTA	WASIAK	PANTERY	9
2	3	PAULINA	WOLNA	MOTYLKI	10
3	13	MARIOLA	PILARSKA	TYGRYSY	10
4	52	JANUSZ	OLEJNICZAK	ŁOSOSIE	10
5	56	STANISŁAW	KRUSZEWSKI	ŻABKI	10
6	62	IGNACY	SZYSZKA	KANGURY	10
7	71	MARIAN	SOBCZYK	KUNY	10
8	80	ARTUR	MORAWSKI	SŁONIKI	10
9	83	ALAN	JAMRÓZ	RYSIE	10
10	96	GABRIEL	DZIURA	KOZY	10

Tabela 11: Zawodnicy z najmniejszą liczbą zagranych meczów

	SponsorID	TotalAmount	SponsorName
1	1	1000000.00	EUROPEAN UNION
2	3	440000.00	TARCZYŃSKI
3	2	220000.00	CITY COUNCIL
4	7	114000.00	BIEDRONKA
5	5	110000.00	ZAKŁAD POGRZEBOWY AS BYTOM
6	4	80000.00	VOLTAREN
7	6	60000.00	STOPERAN

Tabela 12: Porównanie łącznych przychodów od sponsorów

	TeamID	AveragePoints	TeamName	Type	AgeCategory
1	6	8.44	ZEBRY	FEMALE	SENIORS
2	8	8.31	DELFINY	FEMALE	SENIORS
3	13	8.25	KANGURY	MALE	JUNIORS
4	20	8.06	KOZY	MALE	SENIORS
5	5	8.00	KRÓWKI	FEMALE	JUNIORS
6	18	7.94	IGUANY	MALE	SENIORS
7	14	7.88	SZCZUPAKI	MALE	JUNIORS
8	15	7.88	KUNY	MALE	JUNIORS
9	12	7.69	ŻABKI	MALE	JUNIORS
10	1	7.56	MOTYLKI	FEMALE	JUNIORS
11	16	7.50	SŁONIKI	MALE	SENIORS
12	2	7.44	OWIECZKI	FEMALE	JUNIORS
13	11	7.44	ŁOSOSIE	MALE	JUNIORS
14	17	7.38	RYSIE	MALE	SENIORS
15	19	7.38	PSZCZÓŁKI	MALE	SENIORS
16	7	7.25	PANTERY	FEMALE	SENIORS
17	4	7.00	KURY	FEMALE	JUNIORS
18	9	6.94	WILKI	FEMALE	SENIORS
19	10	6.69	SARENKI	FEMALE	SENIORS
20	3	6.62	TYGRYSY	FEMALE	JUNIORS

Tabela 13: Średnia liczba zdobytych punktów

	TeamID	AveragePoints	TeamName	Type	AgeCategory
1	5	8.00	KRÓWKI	FEMALE	JUNIORS
2	6	8.44	ZEBRY	FEMALE	SENIORS
3	13	8.25	KANGURY	MALE	JUNIORS
4	20	8.06	KOZY	MALE	SENIORS

Tabela 14: Najlepsze drużyny w każdej kategorii