

Analiza wyników meczów piłkarskich w klubowych rozgrywkach UEFA

*opr. Tomasz Modzelewski **

** Wydział Matematyki i Nauk Informacyjnych, Inżynieria i analiza danych, I rok*

Opis zbioru danych

Będziemy analizować mecze z Ligi Mistrzów, Ligi Europy oraz nieistniejącego już Pucharu Zdobywców Pucharów. W rundach rozgrywanych systemem pucharowym drużyny dobierane są w pary i grają na następujących zasadach (poniższy system stosowany jest na pełną skalę od sezonu 1970/71):

- Dwie drużyny rozgrywają w danej rundzie między sobą dwa mecze. Każda z drużyn rozgrywa dokładnie jeden mecz na własnym boisku. Do kolejnej rundy kwalifikuje się drużyna, która łącznie w obu meczach zdobyła więcej bramek.
- Jeżeli obie drużyny zdobyły tyle samo bramek, o awansie decyduje wyższa liczba bramek zdobytych w meczu wyjazdowym.
- Jeżeli nie jest możliwe rozstrzygnięcie kolejności, bezpośrednio po drugim meczu zarządzana jest dogrywka, a przy dalszym braku rozstrzygnięcia – seria rzutów karnych.

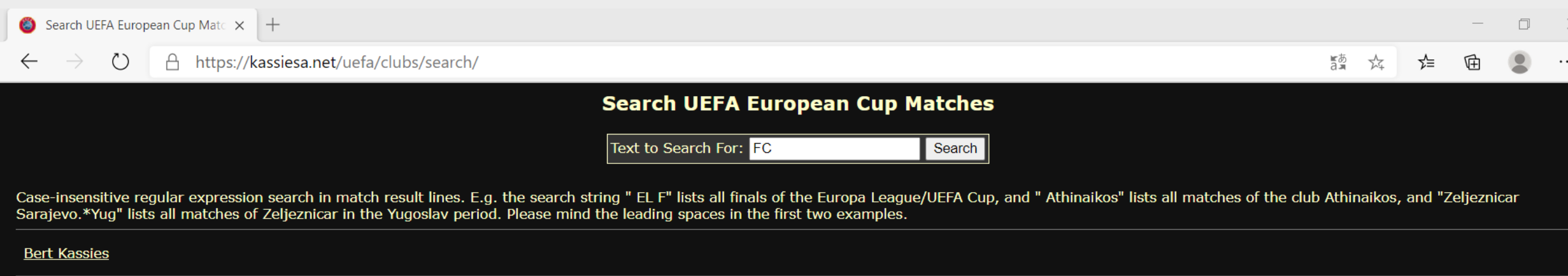
Opis zbioru danych

Ponadto niektóre rundy rozgrywane są systemem grupowym. Każda grupa składa się z czterech drużyn. Każda drużyna rozgrywa z każdą inną dwa mecze: jeden jako gospodarz, drugi jako gość. Za zwycięstwo przyznaje się 3 punkty, za remis 1 punkt, za porażkę 0 punktów; o wyższym miejscu w grupie decyduje wyższa suma zdobytych punktów.

To, kto okazał się lepszy w danym dwumeczu, nie ma zatem bezpośredniego przełożenia na awans do dalszej fazy rozgrywek.

Opis metody

Skorzystamy z wyszukiwarki znajdującej się na stronie internetowej <https://kassiesa.net/uefa/clubs/search/>



Z wspomnianych wcześniej przyczyn analizie poddajemy mecze od sezonu 1970/71 do dziś. W pewnych sytuacjach przeanalizujemy również oddzielnie okres od 2005/06 do dziś.

Opis metody

Wyszukiwarka ta wyświetla dla zadanego ciągu tekstowego rok (przyjmujemy, że rok n jest równoważny sezonowi $(n-1)/n$), nazwę rozgrywek (CL=Liga Mistrzów, EL=Liga Europy, CW=Puchar Zdobywców Pucharów), rundę, nazwy drużyn (najpierw gospodarz pierwszego meczu), kraje oraz wyniki w następującej kolejności: wynik pierwszego meczu, wynik drugiego meczu (jeśli się odbył) oraz rezultat serii rzutów karnych (jeśli się odbyły).

[back](#)[Print](#)

Search UEFA European Cup Matches

Search string: FC

year	cup	rnd							
2021	EL	Q4	Standard Liège	Bel	FC Fehérvár	Hun	3-1		
2021	EL	Q4	FC København	Den	NK Rijeka	Cro	0-1		
2021	EL	Q4	FC Basel	Sui	CSKA Sofia	Bul	1-3		
2021	EL	Q3	Legia Warsaw	Pol	FC Drita	Kos	2-0		
2021	EL	Q3	Riga FC	Lat	Celtic	Sco	0-1		
2021	EL	Q3	FC Basel	Sui	Anorthosis Famagusta	Cyp	3-2		
2021	EL	Q3	FC København	Den	Piast Gliwice	Pol	3-0		

Opis metody

W/w formularz przeszukuje wszystkie pola tabeli. W praktyce więc wprowadzenie zapytania składającego się z samego myślnika spowoduje wyświetlenie zawartości całej bazy, gdyż wynik każdego meczu zapisywany jest w postaci napisu $a-b$, gdzie a i b to liczby bramek zdobyte odpowiednio przez gospodarza i gościa pierwszego meczu danej rywalizacji.

Ponadto wpisanie do formularza określonego roku (np. 2018) spowoduje wypisanie nadzbioru wszystkich meczów, które odbyły się w danym sezonie.

Dużym udogodnieniem będzie fakt, iż formularz wyszukujący jest obsługiwany za pomocą metody GET, więc do każdego wyszukiwania można odwołać się za pomocą adresu WWW:

<https://kassiesa.net/uefa/clubs/search/fm.php?search=zapytanie>

Opis metody

Do eksploracji danych użyjemy języka Python. Zdefiniujemy funkcję *ramka* z opcjonalnym parametrem *rok*, która zwraca ramkę z danymi z wszystkich meczów od sezonu 1970/1971, a jeśli został podany rok – z zadanego sezonu.

```
import numpy as np
import pandas as pd

def konwersja(wynik):
    if(type(wynik) != type("_")):
        return [-1,-1]
    return list(map(int, wynik.split("-")))

def ramka(rok_rozgrywek=0):
    if(rok_rozgrywek):
        www = "https://kassiesa.net/uefa/clubs/search/fm.php?search="+str(rok_rozgrywek)
    else:
        www = "https://kassiesa.net/uefa/clubs/search/fm.php?search=-"
    r = pd.read_html(www)
    r = r[1]
    r = r.drop([0, 1, 2], axis=0)

    r.columns = ["rok", "rozgrywki", "runda", "druzyna1", "kraj1", "druzyna2", "kraj2", "wynik1A", "wynik2A",
                 "wynik3A"]
    r = r.set_index(np.r_[0:len(r)])

    for j in range(len(r)):
        r["rok"][j] = int(r["rok"][j])

    for i in range(1, 4):
        for j in range(len(r)):
            r["wynik" + str(i) + "A"][j] = konwersja(r["wynik" + str(i) + "A"][j])
            r["wynik" + str(i) + "B"] = r["wynik" + str(i) + "A"]
        for j in range(len(r)):
            r["wynik" + str(i) + "A"][j] = (r["wynik" + str(i) + "A"][j])[0]
            r["wynik" + str(i) + "B"][j] = (r["wynik" + str(i) + "B"][j])[1]

    if(rok_rozgrywek):
        return r[r.rok == rok_rozgrywek]
    else:
        return r[r.rok > 1970]
```

W kodzie HTML występują dwie tabele, nasze dane znajdują się w drugiej z nich, ponadto trzy pierwsze wiersze tej drugiej tabeli są zbędne, więc pozbywamy się ich. Następnie rozbijamy ciągi tekstowe zawierające wyniki na liczby. Kolumny zostały odpowiednio nazwane, przy czym *wynik1A* i *wynik1B* odpowiadają za wynik pierwszego meczu, *wynik2A* i *wynik2B* za wynik drugiego meczu, a *wynik3A* i *wynik3B* za rzuty karne. Na koniec niezbędne jest zadbanie, by zostały zwrócone tylko informacje o meczach z konkretnego sezonu (jeżeli podano parametr *rok_rozgrywek*), bowiem rok może również występować m.in. w nazwie klubu.

Opis metody

Teraz możemy swobodnie wyszukiwać interesujące nas informacje. Poniżej przykład funkcji *wyniki*, która później wykorzystamy do sprawdzenia, ile dwumeczów zakończyło się wygranymi gospodarzy pierwszego meczu, a ile wygranymi gospodarzy rewanżu. Wszystkie pozostałe funkcje użyte na potrzeby tego zadania również można napisać, wykorzystując zaprezentowane niżej pakiety i elementy składni języka.

```
from ramka import ramka

def wyniki(rodzaj=0,r=ramka()):
    #0 - dwumecze pucharowe, 1 - dwumecze grupowe, 2 - wszystkie
    wygrany = (r.wynik1A + r.wynik2A > r.wynik1B + r.wynik2B) \
        | ((r.wynik1A + r.wynik2A == r.wynik1B + r.wynik2B) & (r.wynik2A > r.wynik1B)) \
        | ((r.wynik1A + r.wynik2A == r.wynik1B + r.wynik2B) & (r.wynik2A == r.wynik1B) & (r.wynik3A > r.wynik3B))
    przegrany = (r.wynik1A + r.wynik2A < r.wynik1B + r.wynik2B) \
        | ((r.wynik1A + r.wynik2A == r.wynik1B + r.wynik2B) & (r.wynik2A < r.wynik1B)) \
        | ((r.wynik1A + r.wynik2A == r.wynik1B + r.wynik2B) & (r.wynik2A == r.wynik1B) & (r.wynik3A < r.wynik3B))
    remis = ( (r.wynik1A == r.wynik2B) & (r.wynik2A == r.wynik1B) & (r.wynik3A == -1) )

    dwumecz_pucharowy = ((r.runda != "G") & (r.runda != "F") & (r.wynik2A != -1))
    dwumecz_grupowy = ((r.runda == "G") & (r.wynik2A != -1))

    if(rodzaj == 0):
        return [len(r[dwumecz_pucharowy & przegrany]), len(r[dwumecz_pucharowy & wygrany])]
    if(rodzaj == 1):
        return [len(r[dwumecz_grupowy & przegrany]), len(r[dwumecz_grupowy & remis]), len(r[dwumecz_grupowy & wygrany])]
    if(rodzaj == 2):
        return [len(r[przegrany]), len(r[remis]), len(r[wyrany])]
```


Przewaga własnego boiska

Sprawdźmy, czy rozgrywanie meczu na własnym boisku faktycznie zwiększa szanse wygranej. Poniżej procent **poszczególnych wyników** (nie bierzemy pod uwagę meczów finałowych, które odbywają się na neutralnych boiskach, nie bierzemy pod uwagę serii rzutów karnych):

Wygrana gospodarzy	Remis	Wygrana gości
10931	4705	5607
51,5%	22,1%	26,4%

dane z sezonów 1970/71-2020/21

Wygrana gospodarzy	Remis	Wygrana gości
4806	2412	2982
47,1%	23,6%	29,2%

dane z sezonów 2005/06-2020/21

Procent zwycięstw w serii rzutów karnych:

Wygrana gospodarzy	Wygrana gości
145	114
56,0%	44,0%

Siła własnego boiska

Sprawdźmy teraz, jaki procent wszystkich wygranych meczów przez kluby z danego kraju stanowią mecze rozegrane na boisku klubu z tego kraju. Brałem pod uwagę dane od sezonu 2005/06, uwzględniłem tylko te kraje, z których kluby wygrały łącznie minimum 30 meczów. Poniżej pięć najwyższych wskaźników:

Kraj	Procent
Kaz	74,4%
Alb	72,5%
Nir	71,9%
Fin	70%
Est	70%

Własne boisko jest więc największym atutem dla klubów z Kazachstanu, Albanii, Irlandii Północnej, Finlandii i Estonii. Nie jest to zaskoczeniem, zważywszy, iż większość z tych krajów stanowią te uznawane za „egzotyczne” i nieprzyjemne miejsce do wyjazdu.

Miejsce rozegrania rewanżu w dwumeczu pucharowym a prawdopodobieństwo awansu

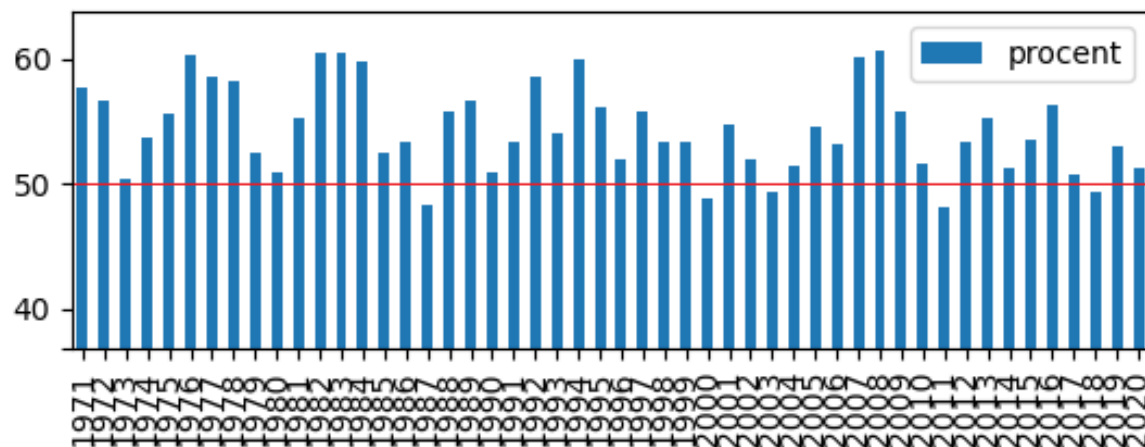
Zweryfikujemy obiegową opinię, że bardziej opłaca się grać rewanż na własnym boisku. Opinię tę wyznaje sama UEFA, np. zgodnie z regulaminem Ligi Mistrzów w 1/8 finału zwycięzcy grup mają przywilej grania rewanżu u siebie.

Uwaga. Nie będziemy brali pod uwagę dwumeczów rozegranych w ćwierćfinale Ligi Mistrzów w latach 1995-2003 (symbol rundy QF), w 1/8 finału Ligi Mistrzów od roku 2004 (symbol rundy $R2$), a także w 1/16 finału Ligi Europy od roku 2005 (symbol rundy $R3$ w latach 2005-2009 oraz $R2$ od roku 2010). W tych rundach o miejscu rozegrania rewanżu decydowały wyniki z poprzednich rund. W pozostałych przypadkach gospodarza wyznaczało losowanie.

Miejsce rozegrania rewanżu w dwumeczu pucharowym a prawdopodobieństwo awansu

Łącznie: **8024** rywalizacje, **4330** awanse gospodarzy rewanżów (ok. **54,0%**).

Przy uwzględnieniu danych tylko od sezonu 2005/06 mamy **2913** rywalizacje, **1555** awansów gospodarzy rewanżów (ok. **53,4%**).



Zdarzyły się jednak sezony, w których wskaźnik ten był niższy niż 50% (po lewej wykres przedstawiający procentową liczbę awansów gospodarzy rewanżów w każdym sezonie pucharowym od 1970/1971).

Miejsce rozegrania rewanżu w dwumeczu grupowym a rezultat rywalizacji

„Wygrana” gospodarza drugiego meczu	Remis	„Wygrana” gospodarza pierwszego meczu
834	104	850
46,6%	5,8%	47,5%

Jak widać, liczby wygranych gospodarzy rewanżu i gospodarzy pierwszego meczu są niemal identyczne.

Można zatem postawić tezę, że w piłce nożnej duży wpływ na ostateczny wynik mają morale drużyny w końcowej fazie rywalizacji. W dwumeczach pucharowych, gdzie w drugim meczu rozstrzyga się sprawa awansu, lepsze rezultaty uzyskują drużyny, na których boiskach odbywa się drugi, decydujący mecz. W fazie grupowej, gdzie nie gra się o lepszy bilans dwumeczu, a jedynie o punkty do tabeli, wyniki są porównywalne, nawet nieznacznie na niekorzyść gospodarzy drugiego meczu.

Wynik pierwszego meczu a potencjalny rezultat dwumeczu

Poniżej przedstawiamy, jaki procent dwumeczów pucharowych (bez fazy grupowej!), w których pierwszą rywalizację wygrał gospodarz pierwszego spotkania, zakończył się jego awansem. Przeanalizowałem dziewięć najczęstszych wyników pierwszych meczów.

wynik I meczu	1-0	2-0	2-1	3-0	3-1	3-2	4-0	4-1	5-0	← dane od sezonu 1970/71
Liczba wystąpień	1042	727	632	477	351	174	201	137	126	
Liczba awansów gospodarza I m.	630	613	325	451	274	76	199	122	126	
Procent	60,5%	84,3%	51,4%	94,5%	78,1%	43,7%	99,0%	89,1%	100,0%	

wynik I meczu	1-0	2-0	2-1	3-0	3-1	3-2	4-0	4-1	5-0	← dane od sezonu 2005/06
Liczba wystąpień	386	247	251	160	115	51	61	31	41	
Liczba awansów gospodarza I m.	261	228	140	157	94	23	59	29	41	
Procent	67,6%	92,3%	55,8%	98,1%	81,7%	45,1%	96,7%	93,5%	100,0%	

Wynik pierwszego meczu a potencjalny rezultat dwumeczu

Teraz analogiczne podsumowanie dwumeczów, których pierwszy mecz zakończył się remisem lub wygraną drużyny gości. Tym razem podajemy, jaki procent tych dwumeczów zakończył się awansem gościa pierwszego spotkania.

wynik I meczu	0-0	0-1	0-2	0-3	1-1	1-2	1-3	2-2	2-3
Liczba wystąpień	775	646	378	209	859	382	205	295	103
Liczba awansów gościa I m.	496	563	368	208	613	355	203	227	99
Procent	64,0%	87,2%	97,4%	99,5%	71,4%	92,9%	99,0%	76,9%	96,1%

←
dane od
sezonu
1970/71

wynik I meczu	0-0	0-1	0-2	0-3	1-1	1-2	1-3	2-2	2-3
Liczba wystąpień	307	285	176	106	396	175	88	124	39
Liczba awansów gościa I m.	179	241	169	105	259	157	88	92	36
Procent	58,3%	84,6%	96,0%	99,1%	65,4%	89,7%	100,0%	74,2%	92,3%

←
dane od
sezonu
2005/06

Wynik pierwszego meczu a potencjalny rezultat dwumeczu

Po przeanalizowaniu tych zestawień nasuwają się m.in. następujące spostrzeżenia:

- 1) Mimo że wynik 3-2 w pierwszym meczu rozgrywanym na własnym boisku jest teoretycznie korzystnym rezultatem (drużyna przegrana w rewanżu u siebie musi odrabiać straty), to w praktyce jest to rezultat niekorzystny dla wygranych, którzy w większości przypadków ostatecznie przegrywają cały dwumecz.
- 2) Remis 0-0 w pierwszym meczu nikogo nie promuje (nikt nie musi w rewanżu odrabiać strat), praktyka pokazuje jednak, że częściej w takich sytuacjach awansuje gospodarz rewanżu.

Wynik pierwszego meczu a potencjalny rezultat dwumeczu

3) Konfrontując dane z sezonów 1970/71-2020/21 z danymi z sezonów 2005/06-2020/21, można stwierdzić, że w ostatnich latach gospodarzom pierwszych meczów częściej udaje się w drugim meczu na wyjeździe obronić zaliczkę albo odrobić straty z pierwszego meczu. Oznacza to, że z roku na rok to, na którym stadionie rozgrywany jest mecz, ma nieco mniejsze znaczenie dla jego wyniku. Pokrywa się to również z innymi statystykami (np. slajd 9 – przewaga własnego boiska).

Przyczyn takiej sytuacji należy upatrywać w szeroko pojętej globalizacji. Obecnie wyjazd do innego kraju nie jest tak uciążliwy i skomplikowany logistycznie jak np. 30 lat temu.

KONIEC

Dziękuję za uwagę 😊