

Analiza danych klubu FC Barcelona w sezonie 21/22

Autor: Wiktoria Cwała

23 listopada 2025



1 Wprowadzenie

Celem niniejszego projektu jest proste przeanalizowanie danych meczowych klubu piłkarskiego FC Barcelona w sezonie 21/22. Dane pochodzą z platformy Kaggle <https://www.kaggle.com/datasets/yashsrivastava51213/barcelona2021-2022>.

2 Instalacja bibliotek i wyświetlenie danych

Aby zacząć kodowanie na początku zaimportujemy potrzebne biblioteki:

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Pobieramy dane i je wyświetlamy, w tym celu posłużymy się następującą komendą:

```
df = pd.read_csv("/kaggle/input/barcelona2021-2022/match_report_2021-22.csv")
df
```

W rezultacie otrzymujemy tabelę danych (pozostała część kolumn i wierszy nie wyświetla się, gdyż cała tabela zawiera dużo danych, które nie mieszczą się na zdjęciu):

	Date	Time	Comp	Round	Day	Venue	Result	GF	GA	Opponent	xG	xGA	Poss	Attendance	Captain	Formation	Referee	Match Report	Notes
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2021-08-15	20:00 (23:30)	La Liga	Matchweek 1	Sun	Home	W	4.0	2.0	Real Sociedad	3.1	1.1	62.0	20384.0	Sergio Busquets	4-3-3	Alejandro Hernández	Match Report	NaN
2	2021-08-21	22:00 (01:30)	La Liga	Matchweek 2	Sat	Away	D	1.0	1.0	Athletic Club	1.4	1.5	67.0	9394.0	Sergio Busquets	4-3-3	Juan Martínez	Match Report	NaN
3	2021-08-29	17:00 (20:30)	La Liga	Matchweek 3	Sun	Home	W	2.0	1.0	Getafe	1.2	0.4	66.0	26543.0	Sergio Busquets	4-3-3	Pablo González	Match Report	NaN
4	2021-09-14	21:00 (00:30)	Champions Lg	Group stage	Tue	Home	L	0.0	3.0	de Bayern Munich	0.2	2.7	49.0	39737.0	Sergio Busquets	3-1-4-2	Michael Oliver	Match Report	NaN
5	2021-09-20	21:00 (00:30)	La Liga	Matchweek 5	Mon	Home	D	1.0	1.0	Granada	1.8	0.7	76.0	27097.0	Sergio Busquets	4-3-3	Santiago Jaime	Match Report	NaN
6	2021-09-23	22:00 (01:30)	La Liga	Matchweek 6	Thu	Away	D	0.0	0.0	Cádiz	0.6	1.4	67.0	12180.0	Sergio Busquets	4-3-3	Carlos del Cerro	Match Report	NaN
7	2021-09-26	16:15 (19:45)	La Liga	Matchweek 7	Sun	Home	W	3.0	0.0	Levante	3.1	0.7	70.0	35334.0	Sergio Busquets	4-2-3-1	Isidro Díaz de Mera	Match Report	NaN
8	2021-09-29	20:00 (00:30)	Champions Lg	Group stage	Wed	Away	L	0.0	3.0	pt Benfica	1.2	1.9	59.0	29454.0	Sergio Busquets	3-4-1-2	Daniele Orsato	Match Report	NaN

Jeżeli tabela zawiera bardzo dużo wierszy, dla ułatwienia możemy wyświetlić tylko kilka pierwszych wierszy wykorzystując następującą komendę:

```
df.head()
```

3 Informacje o danych

Możemy bliżej przyjrzeć się naszej tabelce danych. Na początek, możemy uzyskać informacje ile wierszy oraz kolumn zawiera nasza tabela danych. Wykorzystujemy do tego poniższą komendę:

```
print(df.shape)
```

W rezultacie otrzymujemy informacje, że nasza tabela zawiera 54 wiersze i 19 kolumn, co wyświetla się w następującej formie:

(54, 19)

Przy użyciu komendy **df.info()** otrzymujemy dokładne informacje, takie jak nazwy wszystkich kolumn, ilość niepustych wartości oraz typ danych zawartych w danej kolumnie.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54 entries, 0 to 53
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date                  53 non-null    object
1   Time                  53 non-null    object
2   Comp                  53 non-null    object
3   Round                 53 non-null    object
4   Day                   53 non-null    object
5   Venue                 53 non-null    object
6   Result                53 non-null    object
7   GF                    53 non-null    float64
8   GA                    53 non-null    float64
9   Opponent              53 non-null    object
10  xG                     50 non-null    float64
11  xGA                    50 non-null    float64
12  Poss                   53 non-null    float64
13  Attendance             51 non-null    float64
14  Captain               53 non-null    object
15  Formation              53 non-null    object
16  Referee                53 non-null    object
17  Match Report          53 non-null    object
18  Notes                  8 non-null     object
dtypes: float64(6), object(13)
```

4 Czyszczenie i analiza danych

Następnie, przechodzimy do czyszczenia danych. Przy pomocy komendy **df.dropna()** usuwamy wiersze lub kolumny zawierające brakujące wartości. W rezultacie nasza tabela danych zawiera już tylko 6 wierszy i prezentuje się następująco:

Specjalnie dane zostały tak dobrane (i były wystarczająco małe), by można było naocznie sprawdzić wyniki naszej analizy. Na początek możemy sprawdzić bilans wygranych, remisów i przegranych. W tym celu użyjemy komendy:

```
df['Result'].value_counts()
```

	Date	Time	Comp	Round	Day	Venue	Result	GF	GA	Opponent	xG	xGA	Poss	Attendance	Captain	Formation	Referee	Match Report	Notes
33	2022-02-17	18:45 (23:15)	Europa Lg	Knockout round play-offs	Thu	Home	D	1.0	1.0	it Napoli	2.4	0.8	66.0	73525.0	Gerard Piqué	4-3-3	István Kovács	Match Report	Leg 1 of 2
35	2022-02-24	21:00 (01:30)	Europa Lg	Knockout round play-offs	Thu	Away	W	4.0	2.0	it Napoli	1.9	1.3	56.0	37858.0	Sergio Busquets	4-3-3	Sergey Karasev	Match Report	Leg 2 of 2; Barcelona won
38	2022-03-10	21:00 (01:30)	Europa Lg	Round of 16	Thu	Home	D	0.0	0.0	tr Galatasaray	1.6	0.2	68.0	61740.0	Jordi Alba	4-3-3	Benoit Bastien	Match Report	Leg 1 of 2
40	2022-03-17	20:45 (23:15)	Europa Lg	Round of 16	Thu	Away	W	2.0	1.0	tr Galatasaray	2.4	0.7	67.0	50110.0	Sergio Busquets	4-3-3	Daniele Orsato	Match Report	Leg 2 of 2; Barcelona won
43	2022-04-07	21:00 (00:30)	Europa Lg	Quarter-finals	Thu	Away	D	1.0	1.0	de Eint Frankfurt	0.9	1.4	65.0	48000.0	Sergio Busquets	4-3-3	Srdan Jovanović	Match Report	Leg 1 of 2
45	2022-04-14	21:00 (00:30)	Europa Lg	Quarter-finals	Thu	Home	L	2.0	3.0	de Eint Frankfurt	2.6	2.0	74.0	79468.0	Sergio Busquets	4-3-3	Artur Dias	Match Report	Leg 2 of 2; Eint Frankfurt won

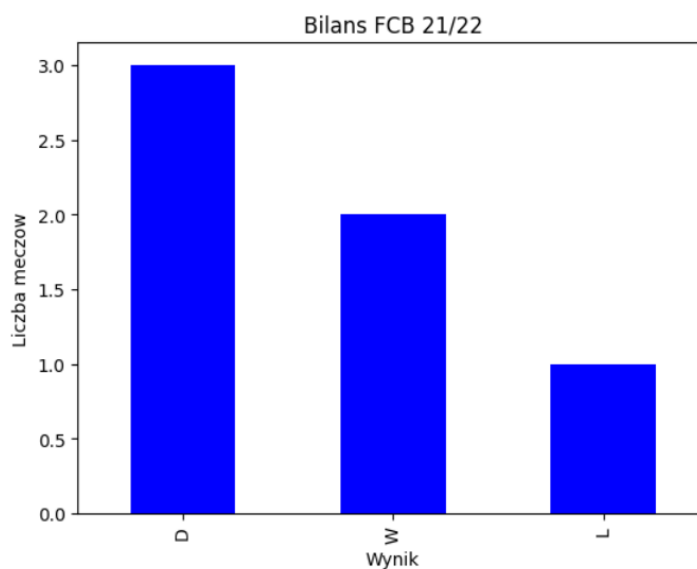
W rezultacie otrzymujemy:

```
Result
D      3
W      2
L      1
Name: count, dtype: int64
```

Możemy wyświetlić powyższy bilans na wykresie. Używając poniższej komendy w rezultacie otrzymujemy wykres słupkowy obrazujący ilość wygranych, przegranych oraz remisów klubu FC Barcelona w sezonie 21/22:

```
df['Result'].value_counts().plot(kind='bar', color='blue')
plt.title("Bilans FCB 21/22")
plt.xlabel("Wynik")
plt.ylabel("Liczba meczow")
plt.show()
```

Wówczas wykres prezentuje się następująco:



Możemy także posortować nasze dane. Przypuśćmy, że chcemy posortować nasze dane według frekwencji na stadionie (od najmniejszej do największej). Skupiamy się zatem na kolumnie **Attendance**. Wykorzystujemy komendę

```
df.sort_values(by = "Attendance", ascending = True).head()
```

W rezultacie otrzymujemy poniższą tabelę posortowaną względem frekwencji na meczu:

	Date	Time	Comp	Round	Day	Venue	Result	GF	GA	Opponent	xG	xGA	Poss	Attendance	Captain	Formation	Referee	Match Report	Notes
35	2022-02-24	21:00 (01:30)	Europa Lg	Knockout round play-offs	Thu	Away	W	4.0	2.0	it Napoli	1.9	1.3	56.0	37858.0	Sergio Busquets	4-3-3	Sergey Karasev	Match Report	Leg 2 of 2; Barcelona won
43	2022-04-07	21:00 (00:30)	Europa Lg	Quarter-finals	Thu	Away	D	1.0	1.0	de Eint Frankfurt	0.9	1.4	65.0	48000.0	Sergio Busquets	4-3-3	Srdan Jovanović	Match Report	Leg 1 of 2
40	2022-03-17	20:45 (23:15)	Europa Lg	Round of 16	Thu	Away	W	2.0	1.0	tr Galatasaray	2.4	0.7	67.0	50110.0	Sergio Busquets	4-3-3	Daniele Orsato	Match Report	Leg 2 of 2; Barcelona won
38	2022-03-10	21:00 (01:30)	Europa Lg	Round of 16	Thu	Home	D	0.0	0.0	tr Galatasaray	1.6	0.2	68.0	61740.0	Jordi Alba	4-3-3	Benoit Bastien	Match Report	Leg 1 of 2
33	2022-02-17	18:45 (23:15)	Europa Lg	Knockout round play-offs	Thu	Home	D	1.0	1.0	it Napoli	2.4	0.8	66.0	73525.0	Gerard Piqué	4-3-3	István Kovács	Match Report	Leg 1 of 2

5 Zamiana danych

Wcześniej z naszych danych usunięte zostały wszystkie kolumny i wiersze zawierające puste pola. W konsekwencji czego pozostało tylko 6 wierszy, które mogliśmy analizować. Teraz, w puste pola niezawierające żadnych danych wstawimy pewne wartości. Zauważmy, że kolumna **"Notes"** w większości miejsc ma pola wypełnione symbolem **NaN**. To są właśnie te braki w danych, które teraz uzupełnimy. W te miejsca wpiszemy wartość **"Empty"**.

Wczytajmy ponownie nasze dane.

```
data = pd.read_csv("/kaggle/input/barcelona2021-2022/match_report_2021-22.csv")
data
```

Dalej, wykonujemy wcześniej opisaną procedurę:

```
values = "Notes": 'Empty'
data = data.fillna(value=values)
data
```

Jak możemy zauważyć na poniższym obrazku, w kolumnie ”Notes” puste wartości zostały wypełnione wartością ”Empty”.

	Date	Time	Comp	Round	Day	Venue	Result	GF	GA	Opponent	xG	xGA	Poss	Attendance	Captain	Formation	Referee	Match Report	Notes
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Empty
1	2021-08-15	20:00 (23:30)	La Liga	Matchweek 1	Sun	Home	W	4.0	2.0	Real Sociedad	3.1	1.1	62.0	20384.0	Sergio Busquets	4-3-3	Alejandro Hernández	Match Report	Empty
2	2021-08-21	22:00 (01:30)	La Liga	Matchweek 2	Sat	Away	D	1.0	1.0	Athletic Club	1.4	1.5	67.0	9394.0	Sergio Busquets	4-3-3	Juan Martínez	Match Report	Empty
3	2021-08-29	17:00 (20:30)	La Liga	Matchweek 3	Sun	Home	W	2.0	1.0	Getafe	1.2	0.4	66.0	26543.0	Sergio Busquets	4-3-3	Pablo González	Match Report	Empty
4	2021-09-14	21:00 (00:30)	Champions Lg	Group stage	Tue	Home	L	0.0	3.0	de Bayern Munich	0.2	2.7	49.0	39737.0	Sergio Busquets	3-1-4-2	Michael Oliver	Match Report	Empty
5	2021-09-20	21:00 (00:30)	La Liga	Matchweek 5	Mon	Home	D	1.0	1.0	Granada	1.8	0.7	76.0	27097.0	Sergio Busquets	4-3-3	Santiago Jaime	Match Report	Empty
6	2021-09-23	22:00 (01:30)	La Liga	Matchweek 6	Thu	Away	D	0.0	0.0	Cádiz	0.6	1.4	67.0	12180.0	Sergio Busquets	4-3-3	Carlos del Cerro	Match Report	Empty
7	2021-09-26	16:15 (19:45)	La Liga	Matchweek 7	Sun	Home	W	3.0	0.0	Levante	3.1	0.7	70.0	35334.0	Sergio Busquets	4-2-3-1	Isidro Díaz de Mera	Match Report	Empty
8	2021-09-29	20:00 (00:30)	Champions Lg	Group stage	Wed	Away	L	0.0	3.0	pt Benfica	1.2	1.9	59.0	29454.0	Sergio Busquets	3-4-1-2	Daniele Orsato	Match Report	Empty
9	2021-10-02	21:00 (00:30)	La Liga	Matchweek 8	Sat	Away	L	0.0	2.0	Atlético Madrid	0.9	1.3	70.0	60594.0	Sergio Busquets	4-2-3-1	César Soto	Match Report	Empty
10	2021-10-17	21:00 (00:30)	La Liga	Matchweek 9	Sun	Home	W	3.0	1.0	Valencia	2.1	0.5	60.0	47317.0	Sergio Busquets	4-3-3	Jesús Gil	Match Report	Empty
11	2021-10-20	18:45 (22:15)	Champions Lg	Group stage	Wed	Home	W	1.0	0.0	ua Dynamo Kyiv	2.2	0.1	62.0	45968.0	Sergio Busquets	4-3-3	Clément Turpin	Match Report	Empty

6 Analiza danych - ciąg dalszy

Możemy przeanalizować, ile meczów sędziował każdy sędzia. Na początek wyświetlimy imiona i nazwiska wszystkich sędziów, którzy sędziowali co najmniej jeden mecz w sezonie 21/22.

```
data['Referee'].unique()
```

Rezultat możemy zaobserwować w poniższej grafice:

```
array([nan, 'Alejandro Hernández', 'Juan Martínez', 'Pablo González',
       'Michael Oliver', 'Santiago Jaime', 'Carlos del Cerro',
       'Isidro Díaz de Mera', 'Daniele Orsato', 'César Soto', 'Jesús Gil',
       'Clément Turpin', 'José Sánchez', 'Antonio Matéu Lahoz',
       'Jorge Figueroa', 'Ovidiu Hategan', 'Sergey Karasev',
       'Alberola Rojas', 'Guillermo Cuadra', 'José Luis Munuera',
       'István Kovács', 'Benoît Bastien', 'Ricardo de Burgos',
       'Srđan Jovanović', 'Artur Dias', 'Miguel Ángel Ortiz Arias'],
      dtype=object)
```

W celu przeanalizowania ilości sędziowanych meczów przez każdego sędziego, użyjemy w tym celu polecenia:

```
ref = data['Referee'].value_counts()
print(ref)
```

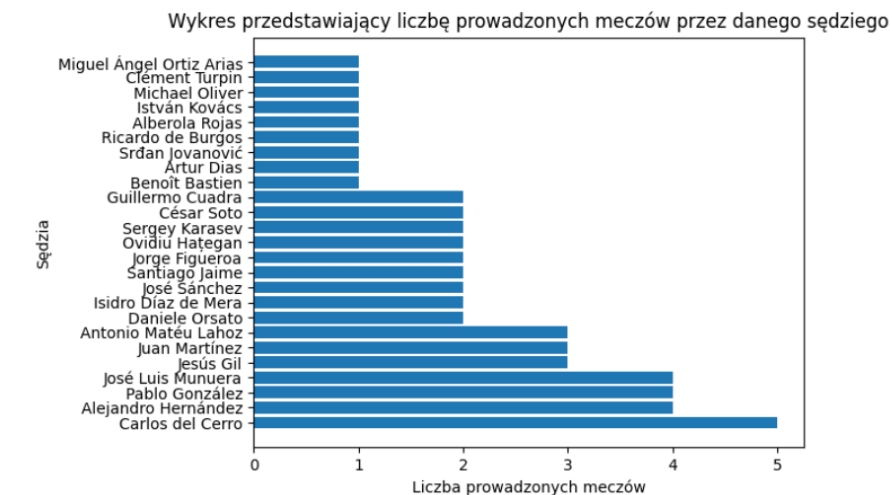
W rezultacie otrzymujemy następujący wynik:

```
Referee
Carlos del Cerro      5
Alejandro Hernández  4
Pablo González        4
José Luis Munuera     4
Jesús Gil             3
Juan Martínez         3
Antonio Matéu Lahoz   3
Daniele Orsato        2
Isidro Díaz de Mera   2
José Sánchez          2
Santiago Jaime        2
Jorge Figueroa        2
Ovidiu Hategan        2
Sergey Karasev        2
César Soto            2
Guillermo Cuadra      2
Benoît Bastien        1
Artur Dias            1
Srđan Jovanović       1
Ricardo de Burgos     1
Alberola Rojas        1
István Kovács         1
Michael Oliver        1
Clément Turpin        1
Miguel Ángel Ortiz Arias 1
Name: count, dtype: int64
```

Uwzględnijmy powyższe rezultaty na wykresie. Wykonując poniższą komendę mamy:

```
ref=data['Referee'].value_counts()
plt.barh(ref.index, ref.values)
plt.xlabel("Liczba prowadzonych meczów")
plt.ylabel("Sędzia")
plt.title("Wykres przedstawiający liczbę prowadzonych meczów przez danego sędziego")
plt.show()
```

Wówczas wykres prezentuje się następująco:



Możemy także sprawdzić, który z zawodników Barcelony najczęściej pełnił rolę kapitana. Uzyskamy to przy uruchomieniu następującej komendy:

```
captain = data['Captain'].value_counts()  
print(captain)
```

Widzimy, że najczęściej kapitanem drużyny był Sergio Busquets:

```
Captain  
Sergio Busquets    49  
Gerard Piqué       2  
Jordi Alba         2  
Name: count, dtype: int64
```

Z ciekawości, sprawdźmy ile meczy piłkarze rozegrali u siebie, a ile na wyjeździe. Informacje na temat miejsca rozgrywania meczu zawiera kolumna 'Venue'. Przy użyciu komendy `data['Venue'].value_counts()` dostajemy informacje meczowe:

```
Venue  
Away    27  
Home    26  
Name: count, dtype: int64
```