

Analiza Danych Sportowych

Czy osiągnięte wyniki wpływają na obecność kibiców na stadionie + Pogoda





Źródła Danych (Mecze w Ekstraklasie)

Dane o frekwencji na meczu: <https://ekstrastats.pl/>

Dane o meczach: <https://www.flashscore.pl/>

Dane o pogodzie: <https://www.wolframalpha.com>

W załączniku znajduje się przykładowy kod w Jupyterze obliczający wszystkie wyniki dla Cracovii



Wybrane drużyny i Sezony

| Drużyny | | Sezony | |
|---------|----------|--------|-------|
| 1. | Cracovia | 1. | 14/15 |
| 2. | Legia | 2. | 15/16 |
| 3. | Lechia | 3. | 16/17 |
| | | 4. | 17/18 |



Drużyna 1. Cracovia (Wersja pełna)

- > Pojemność stadionu: 15 016
- > Posiada podgrzewaną murawę
- > W pełni zadaszony





Odstające wyniki

2014.09.28 -> Derby Krakowa

2016.03.12 -> Mecz ze śmiertelnym rywalem (Legia)

2015.07.24 -> Derby Krakowa

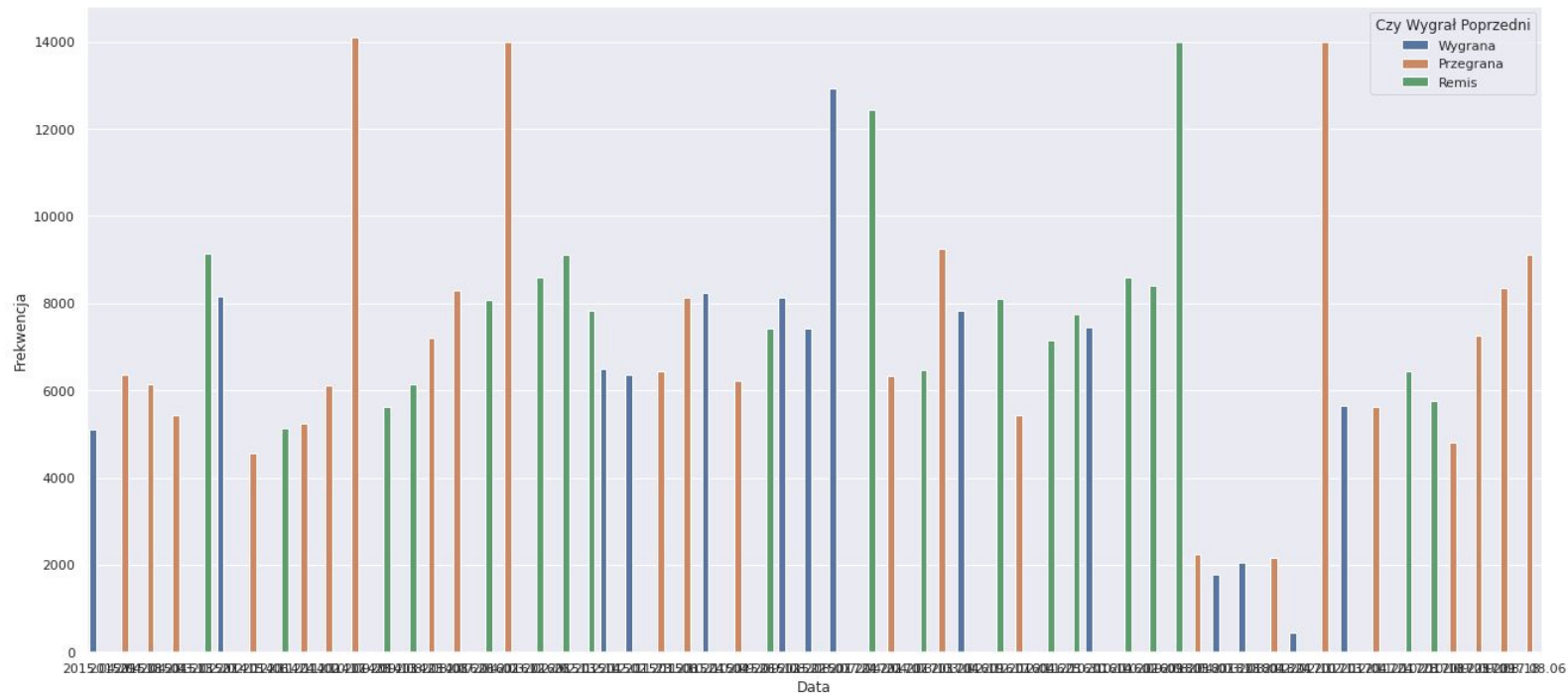
2017.04.22 -> Mecz ze śmiertelnym rywalem (Legia)

2016.08.05 -> Derby Krakowa

2017.12.13 -> Derby Krakowa

2018.02.10 -> 450 kibiców (Nie znalazłem informacji dlaczego, ale jest to bardzo mało)

Frekwencja na stadionie przy 1 zwycięstwie





Frekwencja na stadionie przy 1 zwycięstwie

(Z usunięciem odstających wyników)

50 Meczów

Średnia ilość Kibiców: 6624.180000

Mediana Kibiców: 6479.500000

12 Meczów

Średnia ilość Kibiców po jednej wygranej: 6222.500000

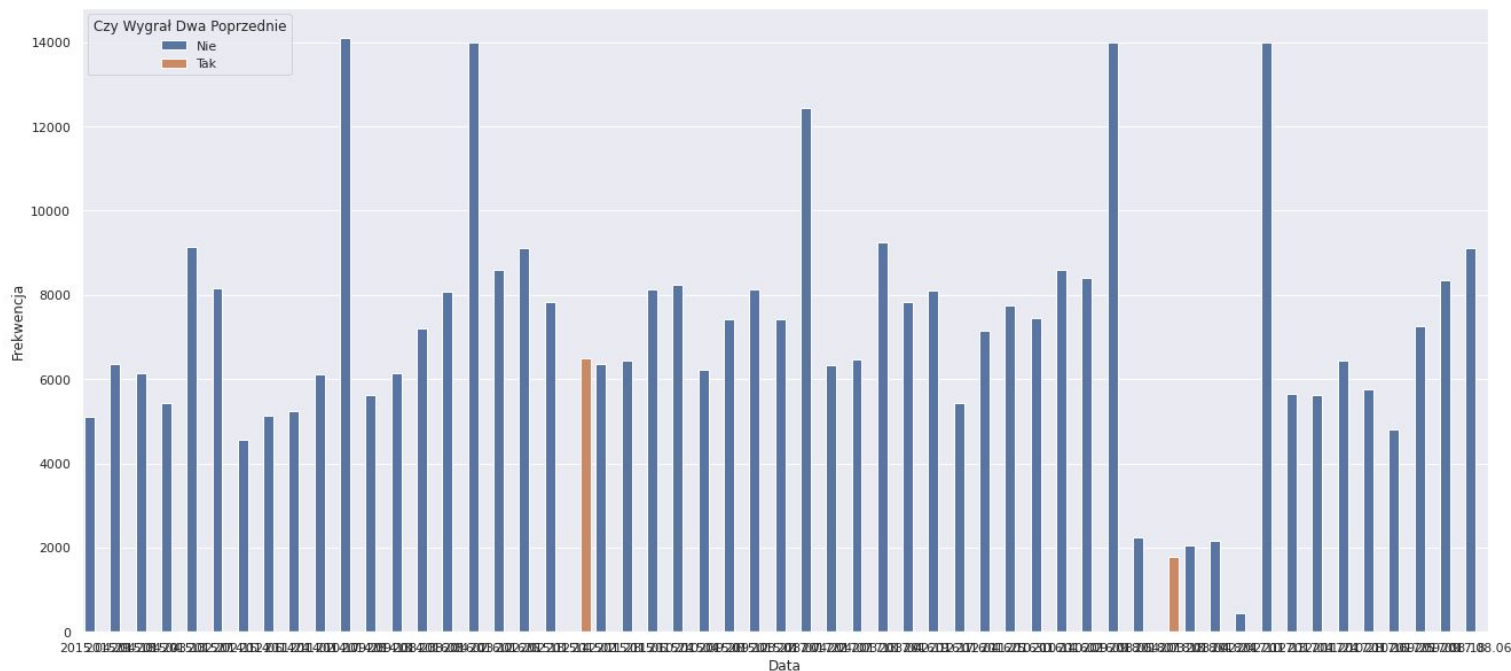
Mediana Kibiców po jednej wygranej: 6951.000000

TEZA: W przypadku Cracovii wzrost kibiców na stadionie po jednej wygranej, nie jest zaobserwowany/jednoznaczny.

Niższa średnia, ale większa mediana (Ale przy o wiele mniejszej ilości danych)



Frekwencja na stadionie przy 2 zwycięstwach





Frekwencja na stadionie przy 2 zwycięstwach

(Z usunięciem odstających wyników)

49 Meczów

Średnia ilość Kibiców: 6589.857143

Mediana Kibiców: 6473.000000

2 Mecze

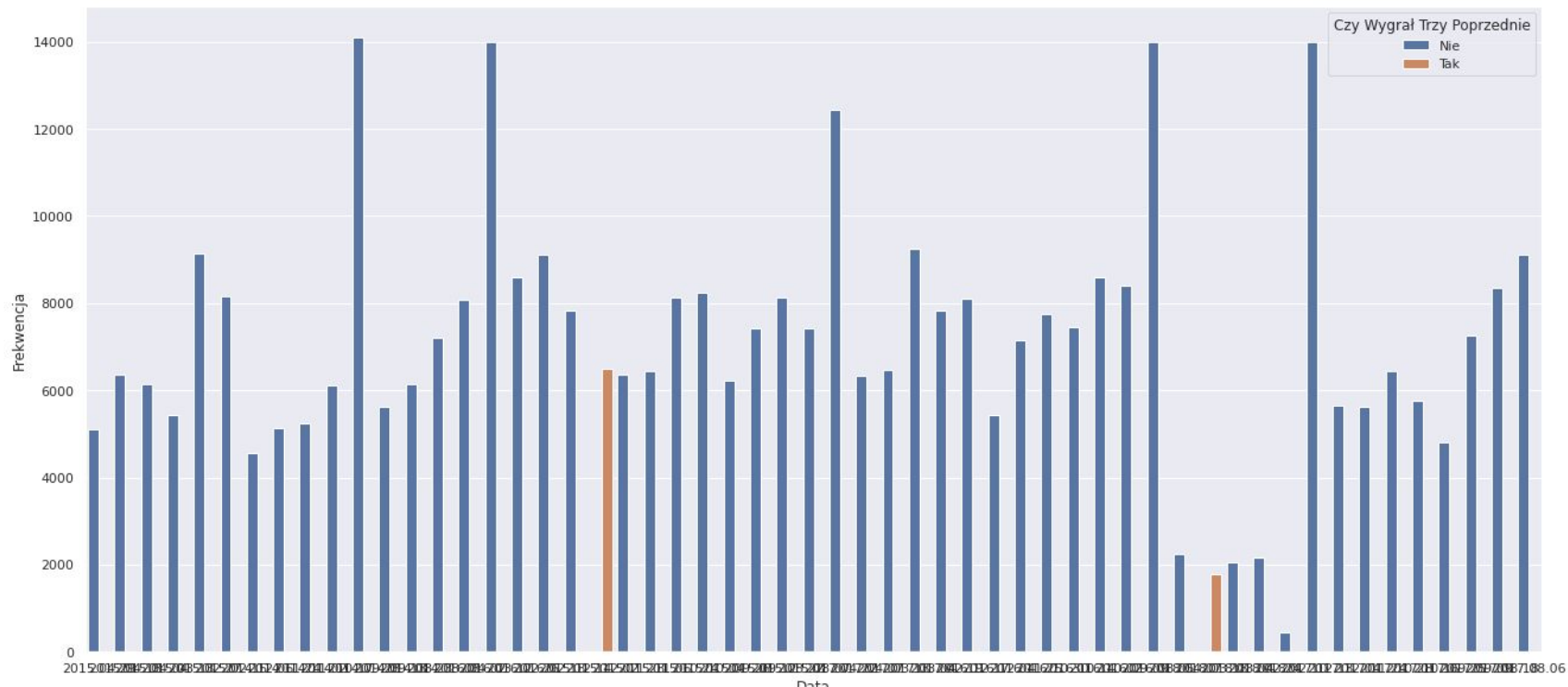
Średnia ilość Kibiców po jednej wygranej: 4134.500000

Mediana Kibiców po jednej wygranej: 4134.500000

TEZA: W przypadku Cracovii wzrost kibiców na stadionie po dwóch wygranych, zdecydowanie nie jest zaobserwowany.



Frekwencja na stadionie przy 3 zwycięstwach





Frekwencja na stadionie przy 3 zwycięstwach

(Z usunięciem odstających wyników)

49 Meczów

Średnia ilość Kibiców: 6589.857143

Mediana Kibiców: 6473.000000

2 Mecze

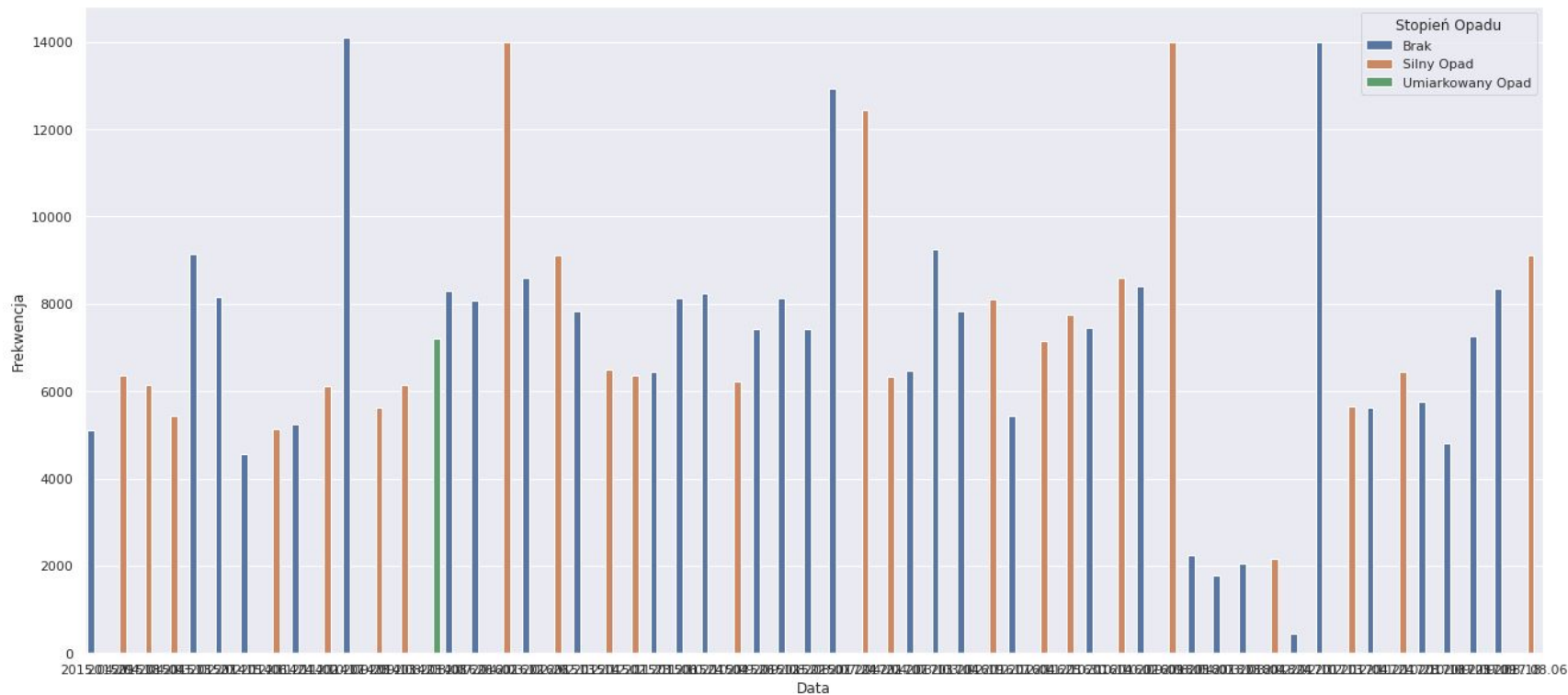
Średnia ilość Kibiców po jednej wygranej: 4134.500000

Mediana Kibiców po jednej wygranej: 4134.500000

TEZA: W przypadku Cracovii wzrost kibiców na stadionie po trzech wygranych, zdecydowanie nie jest zaobserwowany.



Frekwencja W zależności od deszczu





Frekwencja W zależności od deszczu

LEGENDA

Brak Deszczu -> 0 mm

Lekki Opad -> 0-2.5 mm

Umiarkowany Opad -> 2.5-7.5 mm

Silny Opad -> więcej niż 7.5mm



Frekwencja W zależności od deszczu

(Z usunięciem odstających wyników)

50 Meczów

Średnia ilość Kibiców: 6624.180000

Mediana Kibiców: 6479.500000

29 Meczów -> Brak Opadów

Średnia ilość Kibiców: 6673.413793

Mediana Kibiców: 7425.000000

1 Mecz -> Umiarkowany Opad

Średnia ilość Kibiców: 7192.0

Mediana Kibiców: 7192.0

20 Meczów -> Silny Opad

Średnia ilość Kibiców: 6524.400000

Mediana Kibiców: 6348.500000

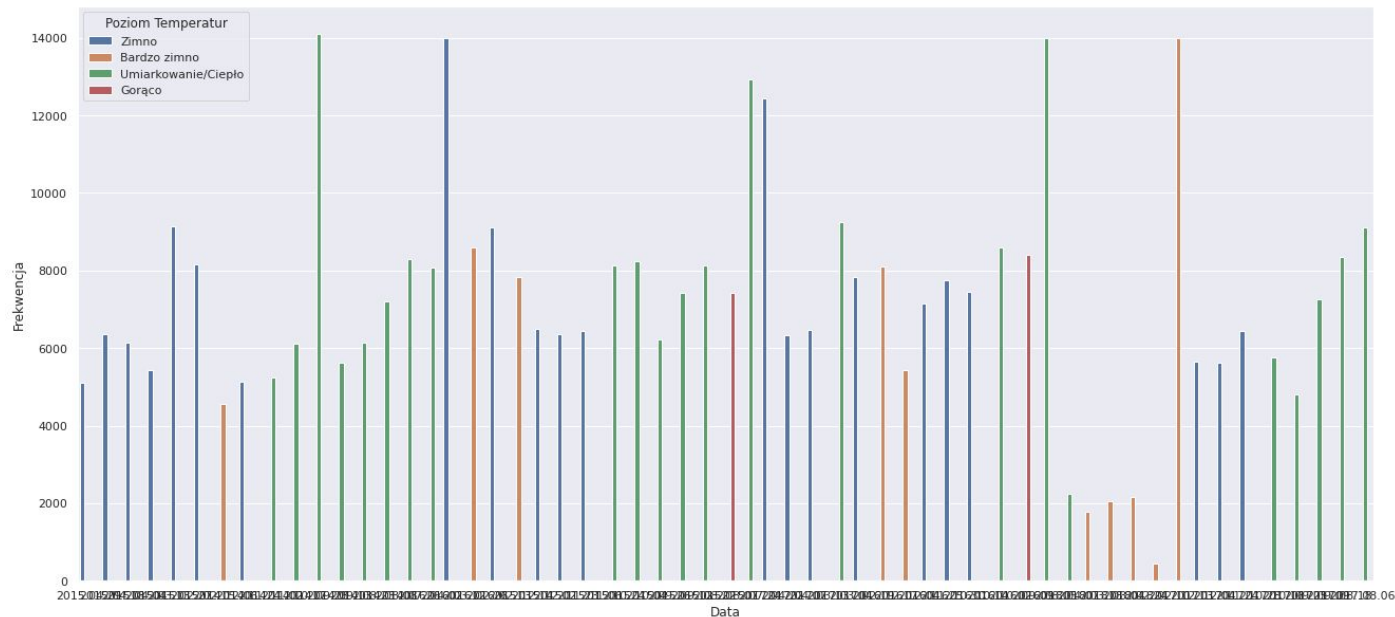


Frekwencja W zależności od deszczu

TEZA: Deszcz nie ma żadnego wpływu na frekwencję na stadionie. Głównym powodem jest zapewne całkowita ochrona przed deszczem na stadionie



Frekwencja W zależności od Temperatury





Frekwencja W zależności od Temperatury

LEGENDA

Bardzo Zimno -> Mniej niż 0 stopni celsjusza

Zimno -> 0-10 stopni celsjusza

Umiarkowanie/Ciepło -> 10-25 stopni celsjusza

Gorąco -> Więcej niż 25 stopni celsjusza



Frekwencja W zależności od Temperatury

(Z usunięciem odstających wyników)

50 Meczów

Średnia ilość Kibiców: 6624.180000

Mediana Kibiców: 6479.500000

8 Meczów -> Bardzo Zimno

Średnia ilość Kibiców: 5066.375000

Mediana Kibiców: 4993.000000

20 Meczów -> Zimno

Średnia ilość Kibiców: 6729.950000

Mediana Kibiców: 6434.500000

20 Meczów -> Umiarkowanie/Ciepło

Średnia ilość Kibiców: 7012.500000

Mediana Kibiców: 7343.500000

1 Mecze -> Gorąco

Średnia ilość Kibiców: 7914.500000

Mediana Kibiców: 7914.500000



Frekwencja W zależności od Temperatury

TEZA: Wraz ze wzrostem temperatury rośnie ilość kibiców na stadionie. Nawet pomimo upałów (25 stopni celsjusza) Na stadionie pojawia się bardzo duża liczba kibiców.



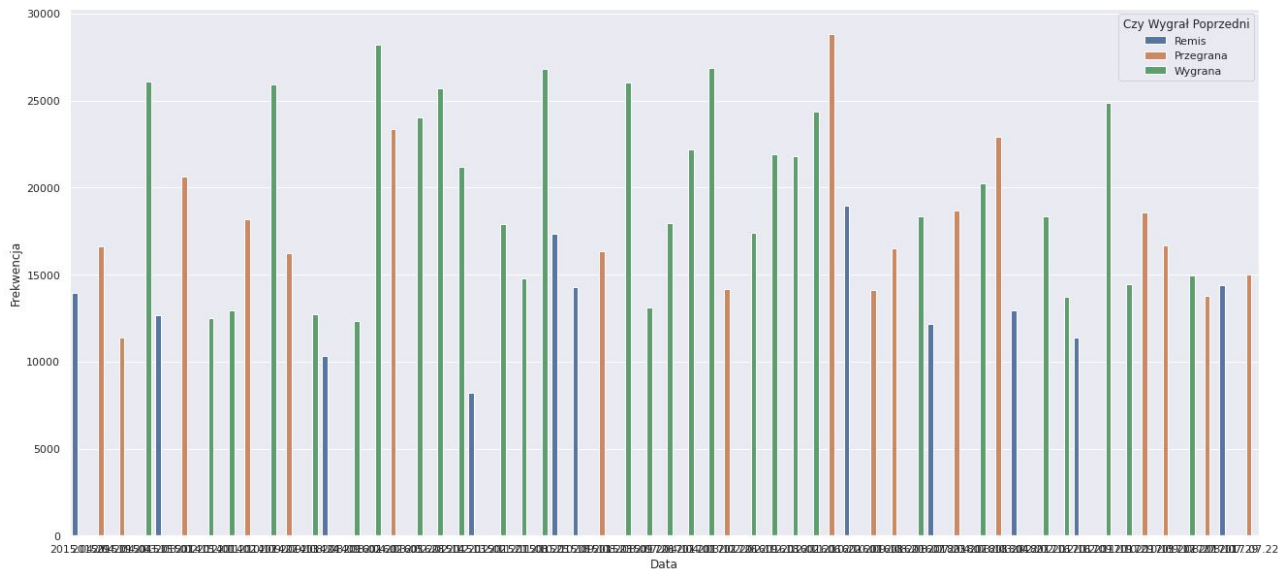
Drużyna 2. Legia Warszawa (Wersja skrócona)

- > Pojemność stadionu: 30 967
- > Posiada podgrzewaną murawę
- > W pełni zadaszony
- > Brak odstających wyników





Frekwencja na stadionie przy 1 zwycięstwie





Frekwencja na stadionie przy 1 zwycięstwie

```
[64] print(final_df_1['Frekwencja'].describe())
```

| | |
|-------|--------------|
| count | 57.000000 |
| mean | 18003.315789 |
| std | 5195.898438 |
| min | 8181.000000 |
| 25% | 13958.000000 |
| 50% | 17343.000000 |
| 75% | 21901.000000 |
| max | 28842.000000 |

Name: Frekwencja, dtype: float64

```
[65] final_df_1['Frekwencja'].loc[final_df_1['Czy Wygrał Poprzedni'] == "Wygrana"].describe()
```

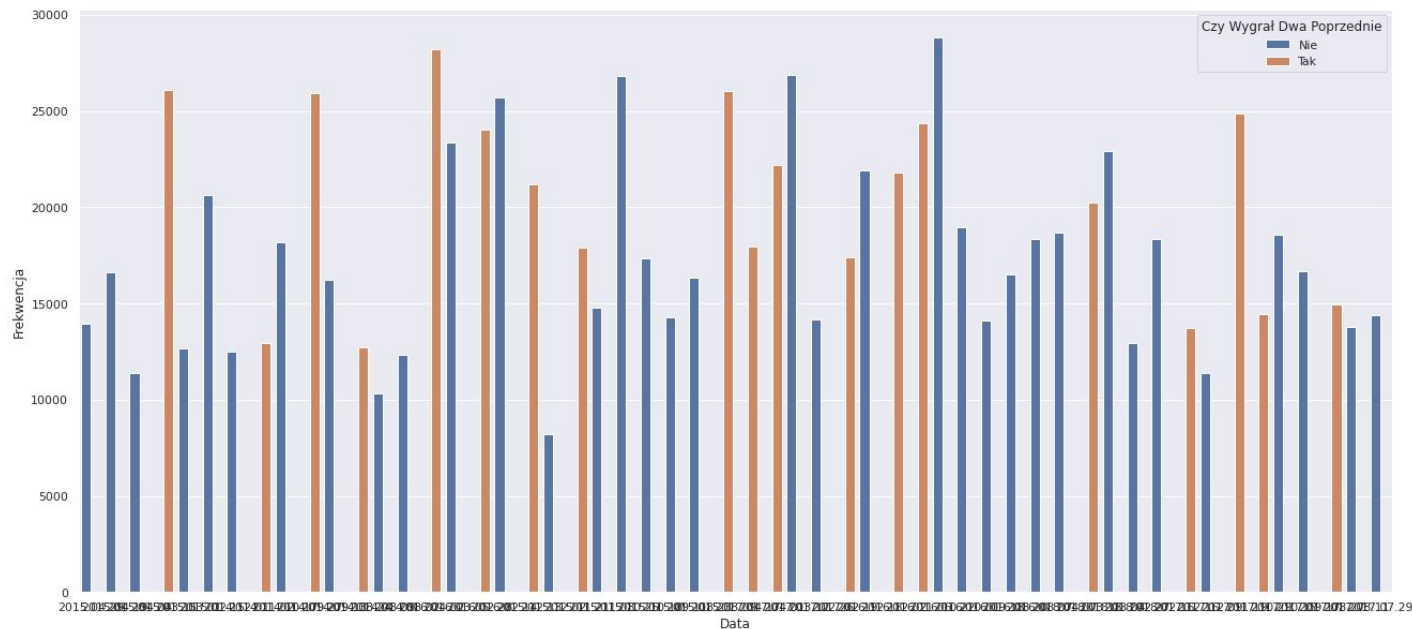
| | |
|-------|--------------|
| count | 29.000000 |
| mean | 19918.620690 |
| std | 5289.076083 |
| min | 12351.000000 |
| 25% | 14780.000000 |
| 50% | 20236.000000 |
| 75% | 24890.000000 |
| max | 28207.000000 |

Name: Frekwencja, dtype: float64

TEZA: Po 1 zwycięstwie na stadionie legii pojawia się więcej kibiców



Frekwencja na stadionie przy 2 zwycięstwach





Frekwencja na stadionie przy 2 zwycięstwach

```
print(final_df_2['Frekwencja'].describe())
```

| | |
|----------------------------------|--------------|
| count | 54.000000 |
| mean | 18257.833333 |
| std | 5214.617031 |
| min | 8181.000000 |
| 25% | 14104.500000 |
| 50% | 17642.000000 |
| 75% | 22132.750000 |
| max | 28842.000000 |
| Name: Frekwencja, dtype: float64 | |

```
[69] final_df_2['Frekwencja'].loc[final_df_2['Czy Wygrał Dwa Poprzednie'] == "Tak"].describe()
```

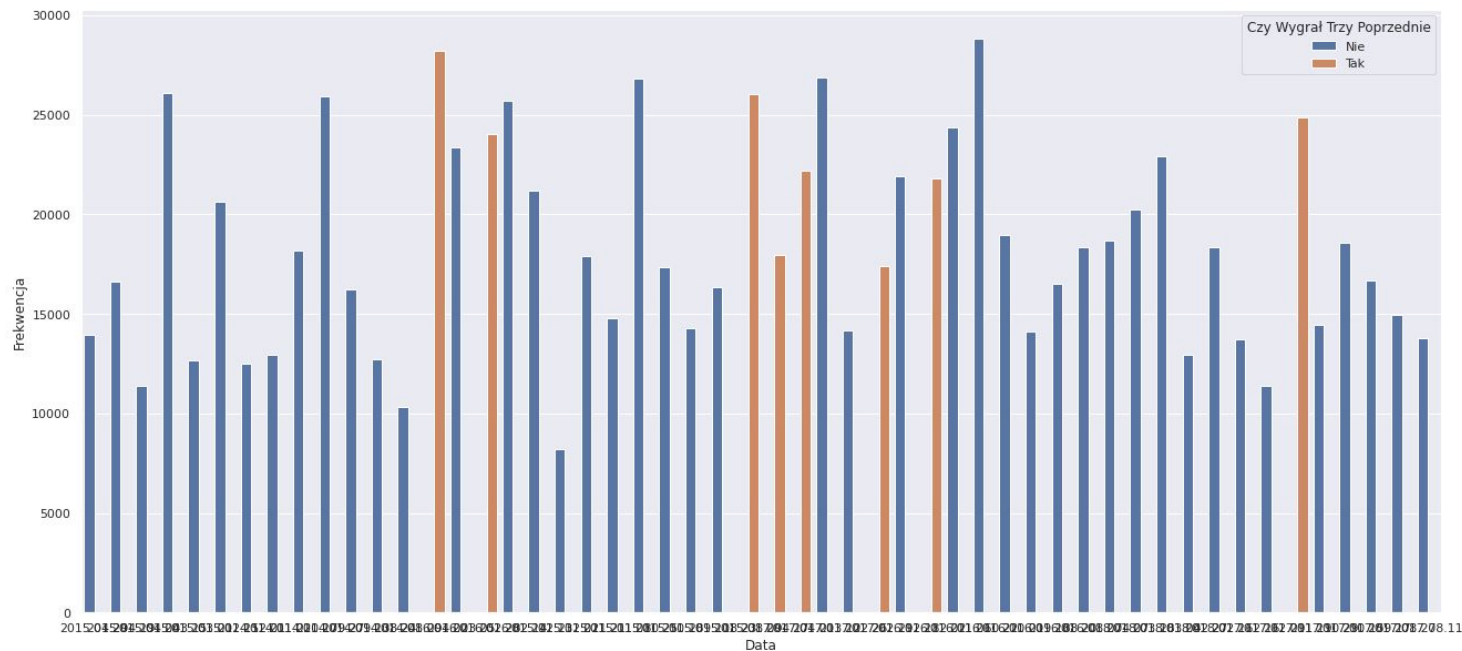
| | |
|----------------------------------|--------------|
| count | 19.000000 |
| mean | 20366.105263 |
| std | 5034.351386 |
| min | 12729.000000 |
| 25% | 16175.000000 |
| 50% | 21197.000000 |
| 75% | 24616.000000 |
| max | 28207.000000 |
| Name: Frekwencja, dtype: float64 | |

TEZA: Po 2 zwycięstwach na stadionie legii pojawia się więcej kibiców

(i więcej niż po 1)



Frekwencja na stadionie przy 3 zwycięstwach





Frekwencja na stadionie przy 3 zwycięstwach

```
print(final_df_3['Frekwencja'].describe())
```

```
count    52.000000
mean     18445.673077
std       5219.790322
min       8181.000000
25%      14143.500000
50%      17918.500000
75%      22382.250000
max      28842.000000
Name: Frekwencja, dtype: float64
```

```
[76] final_df_3['Frekwencja'].loc[final_df_3['Czy Wygrał Trzy Poprzednie'] == "Tak"].describe()
```

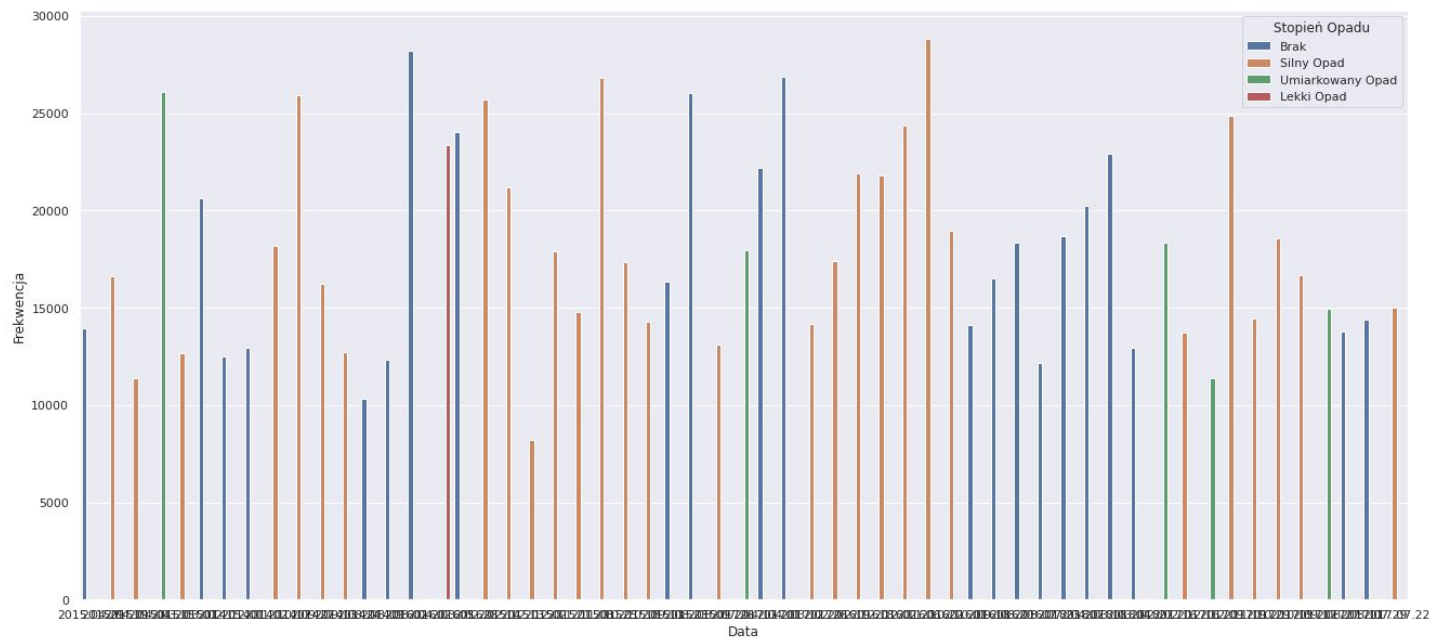
```
count      8.000000
mean     22816.000000
std       3763.651343
min      17409.000000
25%      20856.250000
50%      23116.000000
75%      25169.250000
max      28207.000000
Name: Frekwencja, dtype: float64
```

TEZA: Po 3 zwycięstwach na stadionie legii pojawia się więcej kibiców

(i więcej niż po 1 oraz 2 zwycięstwach)



Frekwencja W zależności od deszczu



Frekwencja W zależności od deszczu

```
[78] print(final_df_1['Frekwencja'].describe())
```

```
count      57.000000
mean     18003.315789
std       5195.898438
min       8181.000000
25%      13958.000000
50%      17343.000000
75%      21901.000000
max       28842.000000
Name: Frekwencja, dtype: float64
```

```
[79] final_df_1['Frekwencja'].loc[final_df_1['Stopień Opadu'] == "Brak"].describe()
```

```
count      22.000000
mean     17746.772727
std       5387.457523
min      10334.000000
25%     13146.250000
50%     16438.500000
75%     21815.750000
max      28207.000000
Name: Frekwencja, dtype: float64
```

```
[80] final_df_1['Frekwencja'].loc[final_df_1['Stopień Opadu'] == "Lekki Opad"].describe()
```

```
count      1.0
mean     23358.0
std       NaN
min     23358.0
25%     23358.0
50%     23358.0
75%     23358.0
max     23358.0
Name: Frekwencja, dtype: float64
```

```
[81] final_df_1['Frekwencja'].loc[final_df_1['Stopień Opadu'] == "Umiarkowany Opad"].describe()
```

```
count      5.000000
mean     17746.800000
std       5445.910456
min      11373.000000
25%     14941.000000
50%     17962.000000
75%     18362.000000
max      26096.000000
Name: Frekwencja, dtype: float64
```

```
[82] final_df_1['Frekwencja'].loc[final_df_1['Stopień Opadu'] == "Silny Opad"].describe()
```

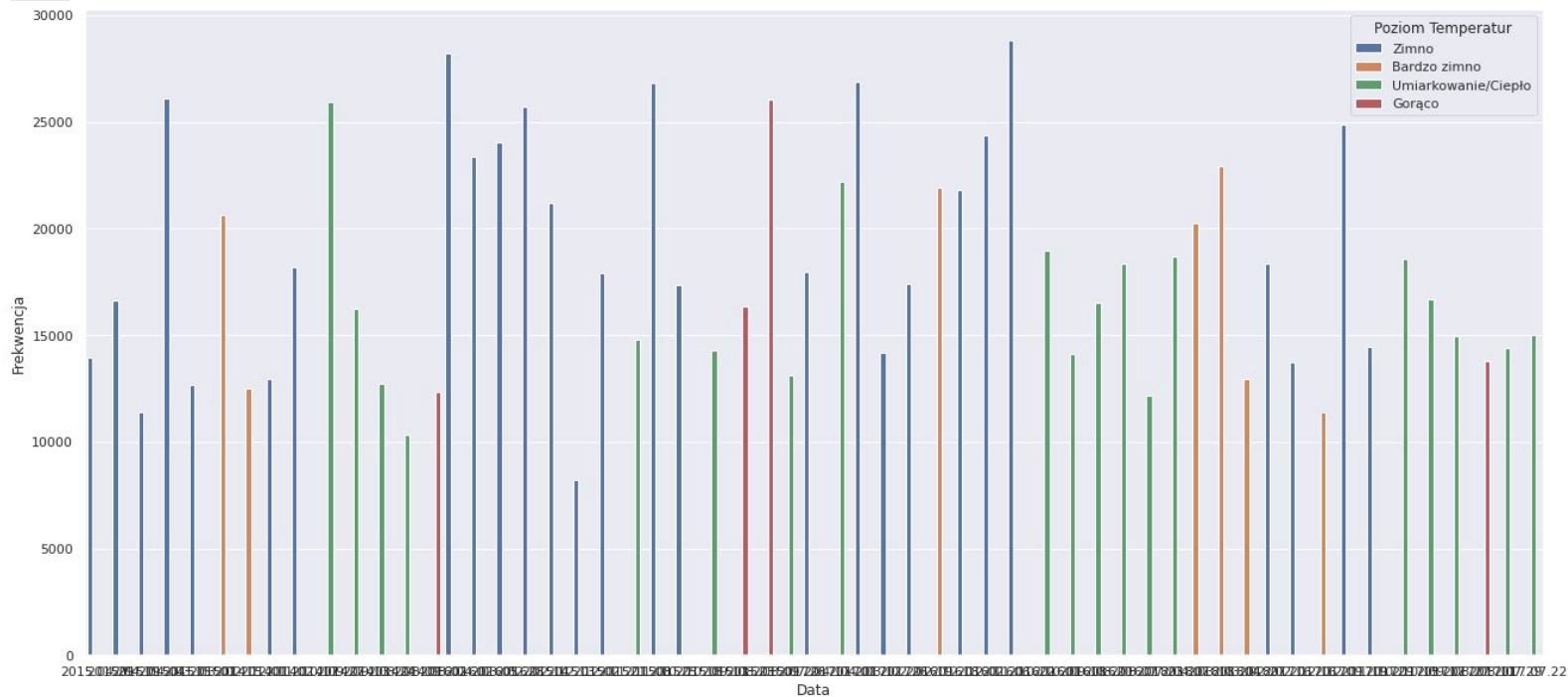
```
count      29.000000
mean     18057.517241
std       5186.407383
min       8181.000000
25%     14281.000000
50%     17343.000000
75%     21821.000000
max      28842.000000
Name: Frekwencja, dtype: float64
```



Frekwencja W zależności od deszczu

TEZA: Nie widać zależności między poziomem opadów a ilością kibiców na stadionie

Frekwencja W zależności od Temperatury



Frekwencja W zależności od Temperatury

```
[84] print(final_df_1['Frekwencja'].describe())
```

```
count      57.000000
mean     18003.315789
std       5195.898438
min       8181.000000
25%      13958.000000
50%      17343.000000
75%      21901.000000
max       28842.000000
Name: Frekwencja, dtype: float64
```

```
[85] final_df_1['Frekwencja'].loc[final_df_1['Poziom Temperatur'] == "Bardzo zimno"].describe()
```

```
count      7.000000
mean     17498.714286
std       4984.347591
min      11373.000000
25%     12724.500000
50%     20236.000000
75%     21267.000000
max      22899.000000
Name: Frekwencja, dtype: float64
```

```
[86] final_df_1['Frekwencja'].loc[final_df_1['Poziom Temperatur'] == "Zimno"].describe()
```

```
count      27.000000
mean     19530.222222
std       5804.885756
min       8181.000000
25%     14298.000000
50%     18171.000000
75%     24616.000000
max       28842.000000
Name: Frekwencja, dtype: float64
```

```
[87] final_df_1['Frekwencja'].loc[final_df_1['Poziom Temperatur'] == "Umiarkowanie/Ciepło"].describe()
```

```
count      19.000000
mean     16205.684211
std       3680.735616
min      10334.000000
25%     14183.000000
50%     15011.000000
75%     18453.000000
max      25934.000000
Name: Frekwencja, dtype: float64
```

```
[88] final_df_1['Frekwencja'].loc[final_df_1['Poziom Temperatur'] == "Gorąco"].describe()
```

```
count      4.000000
mean     17118.500000
std       6152.343971
min      12351.000000
25%     13413.750000
50%     15058.000000
75%     18762.750000
max      26007.000000
Name: Frekwencja, dtype: float64
```



Frekwencja W zależności od Temperatury

TEZA: Na stadionie Legii nie widać korelacji między temperaturą a ilością kibiców



Drużyna 3. Lechia Gdańsk (Wersja skrócona)

- > Pojemność stadionu: 41 620
- > Posiada podgrzewaną murawę
- > W pełni zadaszony
- > Brak odstających wyników





Odstające wyniki

2015.04.11 -> Przedostatni mecz w sezonie z Legią. Zespoły się nie lubią. Ciekawy (dobry) przeciwnik.

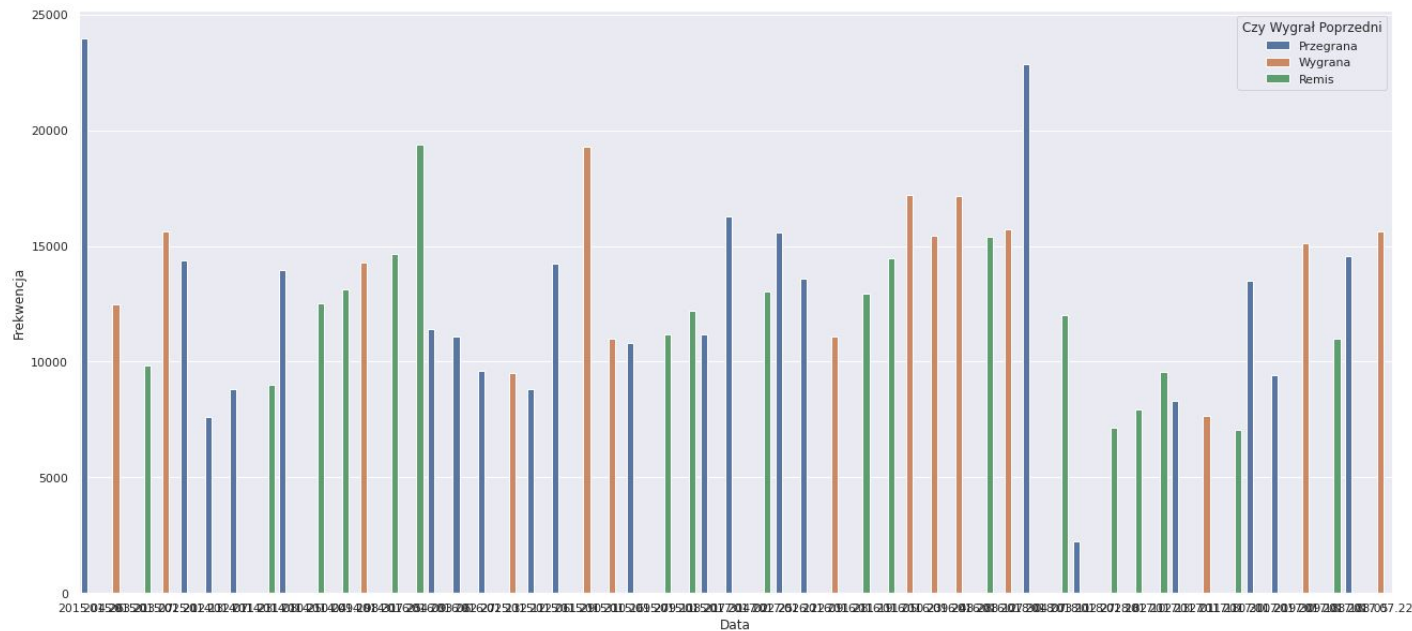
2014.08.10 -> Ciekawy (dobry) przeciwnik, Zespoły się nie lubią.

2017.03.19 -> Ciekawy (dobry) przeciwnik, Zespoły się nie lubią.

2017.04.17 -> Derby Trójmiasta



Frekwencja na stadionie przy 1 zwycięstwie





Frekwencja na stadionie przy 1 zwycięstwie

```
[46] print(final_df_1['Frekwencja'].describe())
```

| | |
|-------|--------------|
| count | 53.000000 |
| mean | 12675.132075 |
| std | 3989.969245 |
| min | 2235.000000 |
| 25% | 9587.000000 |
| 50% | 12510.000000 |
| 75% | 15135.000000 |
| max | 23990.000000 |

Name: Frekwencja, dtype: float64

```
[47] final_df_1['Frekwencja'].loc[final_df_1['Czy Wygrał Poprzedni'] == "Wygrana"].describe()
```

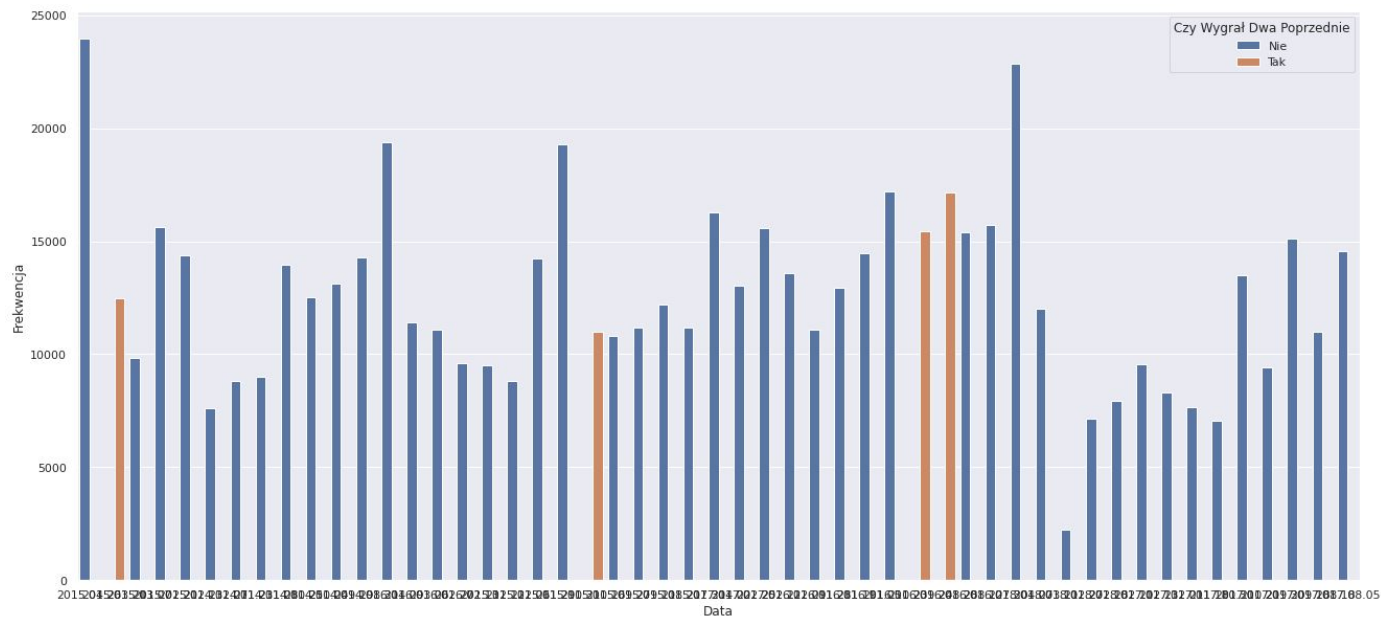
| | |
|-------|--------------|
| count | 14.000000 |
| mean | 14088.000000 |
| std | 3277.806865 |
| min | 7670.000000 |
| 25% | 11429.500000 |
| 50% | 15284.000000 |
| 75% | 15711.000000 |
| max | 19285.000000 |

Name: Frekwencja, dtype: float64

TEZA: Po jednym zwycięstwie pojawia się więcej kibiców



Frekwencja na stadionie przy 2 zwycięstwach





Frekwencja na stadionie przy 2 zwycięstwach

```
print(final_df_2['Frekwencja'].describe())
```

| | |
|-------|--------------|
| count | 51.000000 |
| mean | 12578.764706 |
| std | 4036.826375 |
| min | 2235.000000 |
| 25% | 9567.500000 |
| 50% | 12493.000000 |
| 75% | 14851.000000 |
| max | 23990.000000 |

Name: Frekwencja, dtype: float64

```
[51] final_df_2['Frekwencja'].loc[final_df_2['Czy Wygrał Dwa Poprzednie'] == "Tak"].describe()
```

| | |
|-------|--------------|
| count | 4.000000 |
| mean | 14013.250000 |
| std | 2784.576853 |
| min | 10987.000000 |
| 25% | 12116.500000 |
| 50% | 13963.000000 |
| 75% | 15859.750000 |
| max | 17140.000000 |

Name: Frekwencja, dtype: float64

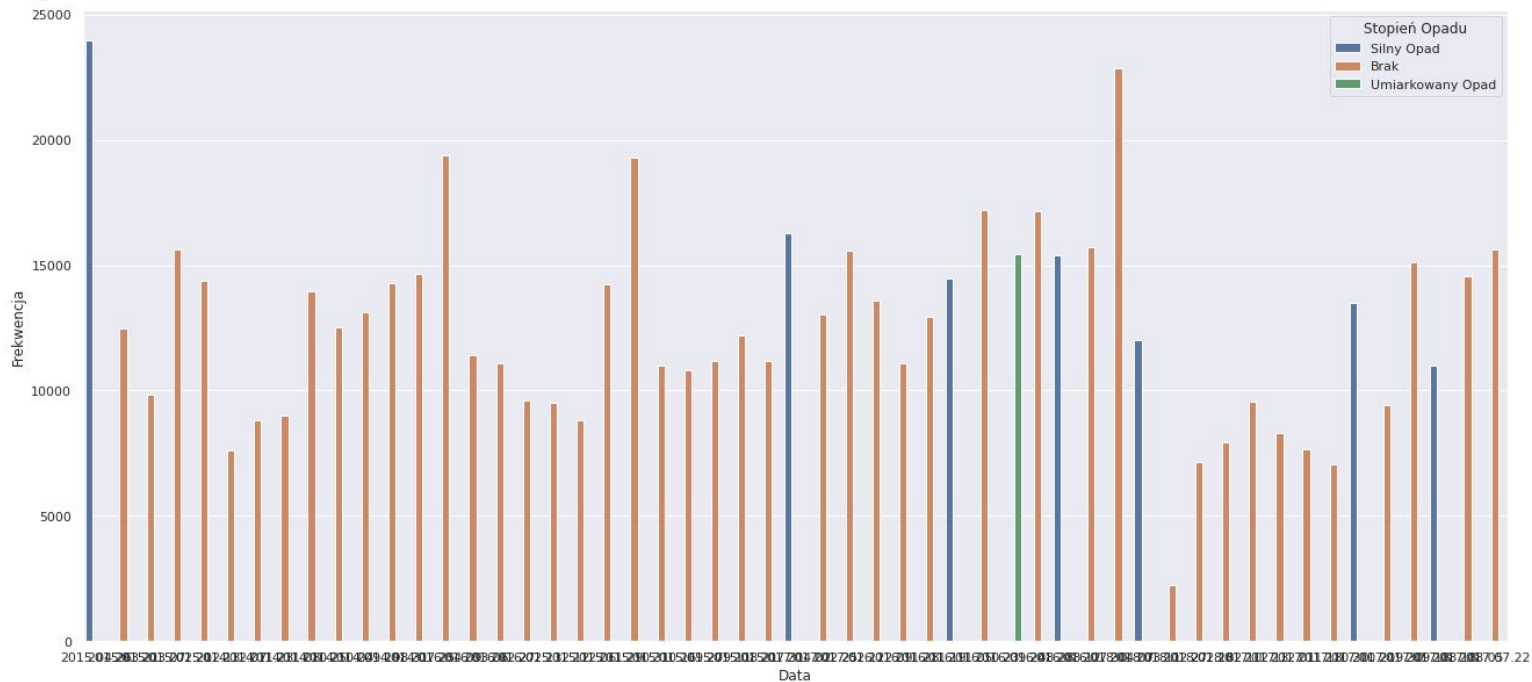
TEZA: Po dwóch zwycięstwach pojawia się więcej kibiców (ale podobna liczba co przy jednym)



Frekwencja na stadionie przy 3 zwycięstwach

Brak Danych. Zespół nigdy nie wygrał 3 meczów z rzędu w tym okresie.

Frekwencja W zależności od deszczu



Frekwencja W zależności od deszczu

```
[59] print(final_df['Frekwencja'].describe())
```

```
count    54.000000
mean     12683.796296
std       3952.661594
min       2235.000000
25%       9653.000000
50%      12715.500000
75%      15011.500000
max       23990.000000
Name: Frekwencja, dtype: float64
```

```
[60] final_df['Frekwencja'].loc[final_df['Stopień Opadu'] == "Brak"].describe()
```

```
count     46.000000
mean     12236.065217
std       3818.961187
min       2235.000000
25%       9532.250000
50%      12343.000000
75%      14516.000000
max       22871.000000
Name: Frekwencja, dtype: float64
```

```
[61] final_df['Frekwencja'].loc[final_df['Stopień Opadu'] == "Lekki Opad"].describe()
```

```
count     0.0
mean      NaN
std       NaN
min       NaN
25%       NaN
50%       NaN
75%       NaN
max       NaN
Name: Frekwencja, dtype: float64
```

```
[62] final_df['Frekwencja'].loc[final_df['Stopień Opadu'] == "Umiarkowany Opad"].describe()
```

```
count      1.0
mean     15433.0
std        NaN
min     15433.0
25%     15433.0
50%     15433.0
75%     15433.0
max     15433.0
Name: Frekwencja, dtype: float64
```

```
[63] final_df['Frekwencja'].loc[final_df['Stopień Opadu'] == "Silny Opad"].describe()
```

```
count      7.000000
mean    15233.285714
std     4271.371392
min    11019.000000
25%    12757.000000
50%    14461.000000
75%    15824.500000
max    23990.000000
Name: Frekwencja, dtype: float64
```

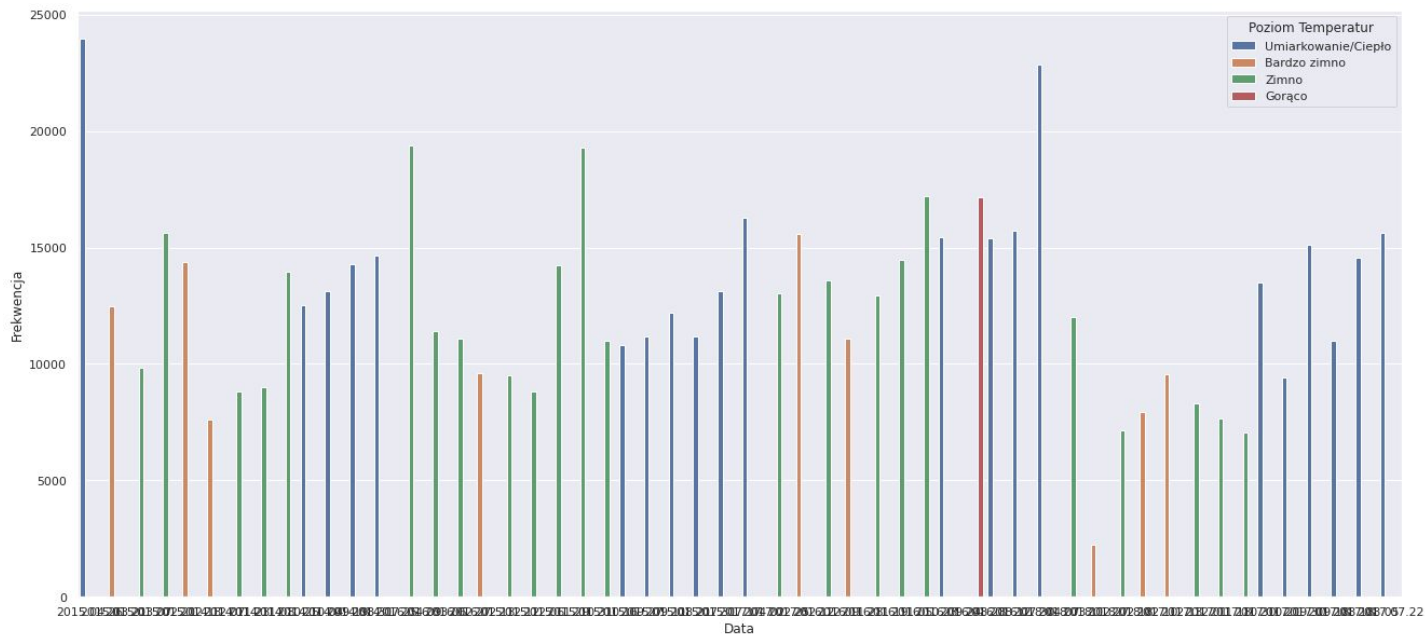


Frekwencja W zależności od deszczu

TEZA: Nie widać korelacji między poziomem opadu deszczu a publiką na stadionie



Frekwencja W zależności od Temperatury



Frekwencja W zależności od Temperatury

```
print(final_df['Frekwencja'].describe())
```

| | |
|-------|--------------|
| count | 54.000000 |
| mean | 12683.796296 |
| std | 3952.661594 |
| min | 2235.000000 |
| 25% | 9653.000000 |
| 50% | 12715.500000 |
| 75% | 15011.500000 |
| max | 23990.000000 |

Name: Frekwencja, dtype: float64

```
[66] final_df['Frekwencja'].loc[final_df['Poziom Temperatur'] == "Bardzo zimno"].describe()
```

| | |
|-------|--------------|
| count | 9.000000 |
| mean | 10045.333333 |
| std | 4002.073837 |
| min | 2235.000000 |
| 25% | 7915.000000 |
| 50% | 9587.000000 |
| 75% | 12493.000000 |
| max | 15573.000000 |

Name: Frekwencja, dtype: float64

```
[67] final_df['Frekwencja'].loc[final_df['Poziom Temperatur'] == "Zimno"].describe()
```

| | |
|-------|--------------|
| count | 23.000000 |
| mean | 11973.782609 |
| std | 3627.691777 |
| min | 7066.000000 |
| 25% | 8909.000000 |
| 50% | 11411.000000 |
| 75% | 14113.000000 |
| max | 19368.000000 |

Name: Frekwencja, dtype: float64

```
final_df['Frekwencja'].loc[final_df['Poziom Temperatur'] == "Umiarkowanie/Ciepło"].describe()
```

| | |
|-------|--------------|
| count | 21.000000 |
| mean | 14380.000000 |
| std | 3576.773336 |
| min | 9399.000000 |
| 25% | 12193.000000 |
| 50% | 14293.000000 |
| 75% | 15433.000000 |
| max | 23990.000000 |

Name: Frekwencja, dtype: float64

```
[69] final_df['Frekwencja'].loc[final_df['Poziom Temperatur'] == "Gorąco"].describe()
```

| | |
|-------|---------|
| count | 1.0 |
| mean | 17140.0 |
| std | NaN |
| min | 17140.0 |
| 25% | 17140.0 |
| 50% | 17140.0 |
| 75% | 17140.0 |
| max | 17140.0 |

Name: Frekwencja, dtype: float64



Frekwencja W zależności od Temperatury

TEZA: W Gdańsku na stadionie, im cieplej, tym więcej kibiców.

Koniec

