EDA

May 4, 2025

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

0.1 Zestaw danych

0.1.1 Liczba uczniów cudzoziemców według statusu - dane SIO według stanu na 30 września 2024 r.

[2]:
```
# url = "https://dane.gov.pl/embed/resource/65513"
# df = pd.read_csv(url)

df = pd.read_csv("./
 ↪liczba_uczniów_nie_są_obywat_polskimi_wg_statusu_typ_podm_30.09.2024.csv")

# Sprawdzenie rozmiaru i kolumn
print(f"Liczba rekordów: {df.shape[0]}, liczba kolumn: {df.shape[1]}")
print("Kolumny:", df.columns.tolist()[:10], "...")

display(df.head())
```

```
Liczba rekordów: 7297, liczba kolumn: 28
Kolumny: ['idTerytGmina', 'idTerytWojewodztwo', 'Wojewodztwo', 'Powiat',
'Gmina', 'Typ obszaru', 'idTypPodmiotu', 'Typ podmiotu',
'lb_ucz_cudzoziem_ogółem', 'obywatel państwa członkowskiego UE, państwa
członkowskiego EFTA lub Konfederacji Szwajcarskiej albo członek rodziny takiej
osoby posiadający prawo pobytu lub prawo stałego pobytu'] …
```

	idTerytGmina	idTerytWojewodztwo	Wojewodztwo	Powiat	Gmina	\
0	201011	2	DOLNOŚLĄSKIE	bolesławiecki	Bolesławiec	
1	201011	2	DOLNOŚLĄSKIE	bolesławiecki	Bolesławiec	
2	201011	2	DOLNOŚLĄSKIE	bolesławiecki	Bolesławiec	
3	201011	2	DOLNOŚLĄSKIE	bolesławiecki	Bolesławiec	
4	201011	2	DOLNOŚLĄSKIE	bolesławiecki	Bolesławiec	

	Typ obszaru	idTypPodmiotu	Typ podmiotu	\
0	obszar miejski	1	Przedszkole	
1	obszar miejski	3	Szkoła podstawowa	

```
2  obszar miejski             14  Liceum ogólnokształcące
3  obszar miejski             16                Technikum
4  obszar miejski             19       Szkoła policealna

   lb_ucz_cudzoziem_ogółem  \
0                      102
1                      317
2                       45
3                       90
4                      177

   obywatel państwa członkowskiego UE, państwa członkowskiego EFTA lub␣
 ↪Konfederacji Szwajcarskiej albo członek rodziny takiej osoby posiadający prawo␣
 ↪pobytu lub prawo stałego pobytu  \
0                                                  2.0
1                                                  NaN
2                                                  1.0
3                                                  1.0
4                                                  NaN

   …  \
0  …
1  …
2  …
3  …
4  …

   osoba, której na terytorium RP udzielono zezwolenia na zamieszkanie na czas␣
 ↪oznaczony w związku z okolicznością, o której mowa w art. 53 ust. 1 pkt 7, 13␣
 ↪i 14 ustawy o cudzoziemcach  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

   członek rodziny osoby ubiegającej się o nadanie statusu uchodźcy  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

   osoba, której uprawnienie do nauki wynika z umowy międzynarodowej.   inny  \
0                                                  NaN                  34.0
1                                                  NaN                 241.0
2                                                  NaN                  35.0
```

```
3                                                   NaN                    62.0
4                                                   NaN                     NaN

   osoba, której udzielono zezwolenia na pobyt stały na terytorium RP  \
0                                                   9
1                                                   5
2                                                 NaN
3                                                   2
4                                                   2

   osoba, której udzielono zgody na pobyt ze względów humanitarnych, albo␣
 ↪członek rodziny takiej osoby  \
0                                                 NaN
1                                                 NaN
2                                                 NaN
3                                                 3.0
4                                                 NaN

   osoba, której na terytorium RP udzielono zezwolenia na pobyt czasowy w␣
 ↪związku z okolicznością, o której mowa w art. 127, art.159 ust. 1, art. 176␣
 ↪lub art. 186 ust. 1 pkt 3 lub 4 ustawy o cudzoziemcach  \
0                                                 3.0
1                                                14.0
2                                                 NaN
3                                                 1.0
4                                                67.0

   osoba, która posiada kartę pobytu z adnotacją "dostęp do rynku pracy", wizę␣
 ↪Schengen lub wizę krajową wydaną w celu wykonywania pracy na terytorium RP  \
0                                                 NaN
1                                                 NaN
2                                                 NaN
3                                                 NaN
4                                                63.0

   członek rodziny osoby ubiegającej się o udzielenie ochrony międzynarodowej  \
0                                                 NaN
1                                                 NaN
2                                                 NaN
3                                                 NaN
4                                                 NaN

   brak statusu w SIO
0                 1.0
1                 2.0
2                 6.0
3                 1.0
```

```
4                 NaN

[5 rows x 28 columns]
```

0.2 Wstępne etapy pipeline'u ML

0.2.1 Wczytanie i podstawowy opis danych

[3]:
```
# Podstawowe statystyki zmiennej celu
target = 'lb_ucz_cudzoziem_ogółem'
print(df[target].describe())

# Przykład: suma uczniów-cudzoziemców wg typu obszaru
print(df.groupby('Typ obszaru')[target].sum())
```

```
count     7297.000000
mean        49.442099
std        403.368199
min          1.000000
25%          2.000000
50%          6.000000
75%         18.000000
max      21355.000000
Name: lb_ucz_cudzoziem_ogółem, dtype: float64
Typ obszaru
obszar miejski    331412
obszar wiejski     29367
Name: lb_ucz_cudzoziem_ogółem, dtype: int64
```

0.2.2 Możliwości zastosowania w uczeniu maszynowym

Dane umożliwiają wykorzystanie uczenia nadzorowanego, ponieważ mamy wyraźny zbiór zmiennych objaśniających i można określić zmienną celu (target). Przykładowo, można próbować przewidzieć liczbę uczniów cudzoziemców w szkole na podstawie cech lokalizacyjnych i typu szkoły. Jeśli ustawimy liczbę cudzoziemców jako zmienną docelową (regresja), mamy typowe zadanie regresyjne. Alternatywnie można by rozważyć klasyfikację (np. czy liczba uczniów obcokrajowców przekroczy pewną wartość progową), ale tu logiczniejsze wydaje się użycie regresji dla zmiennej ciągłej, która sumuje informacje ze statusów pobytowych.

> Krótka refleksja: ponieważ w danych występują zmienne objaśniające (np. Województwo, typ szkoły, etc.) oraz cel, zadanie jest naturalnie nadzorowane. Uczenie nadzorowane z regresją/klasyfikacją jest zasadne – pozwala np. na modelowanie wpływu regionu i rodzaju szkoły na liczbę uczniów-cudzoziemców. Ponadto dane nie posiadają etykiet, które wymagałyby grupowania czy wykrywania struktur ukrytych (co charakteryzuje uczenie nienadzorowane).

0.2.3 Feature Engineering

[4]:
```
# 1) Zmiana nazw kolumn na krótsze (usunięcie polskich znaków i spacji)
df = df.rename(columns={
    'Typ obszaru': 'Typ_obszaru',
    'Typ podmiotu': 'Typ_podmiotu',
    'lb_ucz_cudzoziem_ogółem': 'Liczba_ogolem',
    'brak statusu w SIO': 'Brak_statusu'
})
# Uproszczone nazwy statusów (przytnij długie opisy)
df = df.rename(columns= {
    'obywatel państwa członkowskiego UE, państwa członkowskiego EFTA lub␣
 ↪Konfederacji Szwajcarskiej albo członek rodziny takiej osoby posiadający␣
 ↪prawo pobytu lub prawo stałego pobytu': 'obywatel_UE_EFTA_SZWAJCARIA',
    'osoba pochodzenia polskiego w rozumieniu przepisów o repatriacji':␣
 ↪'pochodzenie_polskie',
    'osoba, której udzielono zezwolenia na osiedlenie się na terytorium RP':␣
 ↪'zezwolenie_osiedlenie',
    'osoba posiadająca ważną Kartę Polaka': 'karta_polaka',
    'osoba, której nadano status uchodźcy, albo członek rodziny takiej osoby':␣
 ↪'status_uchodźcy',
    'osoba posiadająca zgodę na pobyt tolerowany': 'pobyt_tolerowany',
    'osoba, której udzielono ochrony uzupełniającej, albo członek rodziny␣
 ↪takiej osoby': 'ochrona_uzupelniajaca',
    'osoba korzystająca z ochrony czasowej na terytorium RP': 'ochrona_czasowa',
    'osoba, której na terytorium RP udzielono zezwolenia na pobyt rezydenta␣
 ↪długoterminowego UE': 'pobyt_rezydent_UE',
    'osoba, której na terytorium RP udzielono zezwolenia na zamieszkanie na␣
 ↪czas oznaczony w związku z okolicznością, o której mowa w art. 53 ust. 1 pkt␣
 ↪7, 13 i 14 ustawy o cudzoziemcach': 'pobyt_czasowy_art53',
    'członek rodziny osoby ubiegającej się o nadanie statusu uchodźcy':␣
 ↪'rodzina_uchodźcy',
    'osoba, której uprawnienie do nauki wynika z umowy międzynarodowej.':␣
 ↪'nauka_umowa_miedzynarodowa',
    'osoba, której udzielono zezwolenia na pobyt stały na terytorium RP':␣
 ↪'pobyt_stały',
    'osoba, której udzielono zgody na pobyt ze względów humanitarnych, albo␣
 ↪członek rodziny takiej osoby': 'pobyt_humanitarny',
    'osoba, której na terytorium RP udzielono zezwolenia na pobyt czasowy w␣
 ↪związku z okolicznością, o której mowa w art. 127, art.159 ust. 1, art. 176␣
 ↪lub art. 186 ust. 1 pkt 3 lub 4 ustawy o cudzoziemcach':␣
 ↪'pobyt_czasowy_inne',
    'osoba, która posiada kartę pobytu z adnotacją "dostęp do rynku pracy",␣
 ↪wizę Schengen lub wizę krajową wydaną w celu wykonywania pracy na terytorium␣
 ↪RP': 'karta_pobytu_rynek_pracy',
```

```
    'członek rodziny osoby ubiegającej się o udzielenie ochrony␣
  ↪międzynarodowej': 'rodzina_ochrona_miedzynarodowa'
})
```

[5]:
```
display(df.head())
print(df.columns)
```

```
   idTerytGmina  idTerytWojewodztwo   Wojewodztwo         Powiat        Gmina  \
0        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
1        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
2        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
3        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
4        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec

      Typ_obszaru  idTypPodmiotu             Typ_podmiotu  Liczba_ogolem  \
0  obszar miejski              1              Przedszkole            102
1  obszar miejski              3        Szkoła podstawowa            317
2  obszar miejski             14  Liceum ogólnokształcące             45
3  obszar miejski             16                Technikum             90
4  obszar miejski             19        Szkoła policealna            177

   obywatel_UE_EFTA_SZWAJCARIA  …  pobyt_czasowy_art53  rodzina_uchodźcy  \
0                          2.0  …                  NaN               NaN
1                          NaN  …                  NaN               NaN
2                          1.0  …                  NaN               NaN
3                          1.0  …                  NaN               NaN
4                          NaN  …                  NaN               NaN

   nauka_umowa_miedzynarodowa   inny  pobyt_stały  pobyt_humanitarny  \
0                         NaN   34.0            9                NaN
1                         NaN  241.0            5                NaN
2                         NaN   35.0          NaN                NaN
3                         NaN   62.0            2                3.0
4                         NaN    NaN            2                NaN

   pobyt_czasowy_inne  karta_pobytu_rynek_pracy  \
0                 3.0                       NaN
1                14.0                       NaN
2                 NaN                       NaN
3                 1.0                       NaN
4                67.0                      63.0

   rodzina_ochrona_miedzynarodowa  Brak_statusu
0                             NaN           1.0
1                             NaN           2.0
2                             NaN           6.0
3                             NaN           1.0
4                             NaN           NaN
```

```
[5 rows x 28 columns]
```

```
Index(['idTerytGmina', 'idTerytWojewodztwo', 'Wojewodztwo', 'Powiat', 'Gmina',
       'Typ_obszaru', 'idTypPodmiotu', 'Typ_podmiotu', 'Liczba_ogolem',
       'obywatel_UE_EFTA_SZWAJCARIA', 'pochodzenie_polskie',
       'zezwolenie_osiedlenie', 'karta_polaka', 'status_uchodźcy',
       'pobyt_tolerowany', 'ochrona_uzupelniajaca', 'ochrona_czasowa',
       'pobyt_rezydent_UE', 'pobyt_czasowy_art53', 'rodzina_uchodźcy',
       'nauka_umowa_miedzynarodowa', 'inny', 'pobyt_stały',
       'pobyt_humanitarny', 'pobyt_czasowy_inne', 'karta_pobytu_rynek_pracy',
       'rodzina_ochrona_miedzynarodowa', 'Brak_statusu'],
      dtype='object')
```

[6]:
```
# 2. Sprawdzenie brakujących wartości (Missing Values)
missing_values = df.isnull().sum().sort_values(ascending=False).reset_index()
missing_values.columns = ['Column', 'Missing values']
missing_values['Missing percent'] = (df.isnull().mean().
 ↪sort_values(ascending=False)).values.round(4)*100
display(missing_values)
print(missing_values)
```

	Column	Missing values	Missing percent
0	rodzina_uchodźcy	7294	99.96
1	pobyt_czasowy_art53	7286	99.85
2	zezwolenie_osiedlenie	7271	99.64
3	rodzina_ochrona_miedzynarodowa	7212	98.84
4	pochodzenie_polskie	7195	98.60
5	ochrona_uzupelniajaca	7122	97.60
6	nauka_umowa_miedzynarodowa	7078	97.00
7	pobyt_tolerowany	7041	96.49
8	pobyt_rezydent_UE	6989	95.78
9	pobyt_humanitarny	6835	93.67
10	obywatel_UE_EFTA_SZWAJCARIA	6623	90.76
11	ochrona_czasowa	6600	90.45
12	karta_pobytu_rynek_pracy	6507	89.17
13	karta_polaka	6443	88.30
14	Brak_statusu	5855	80.24
15	pobyt_czasowy_inne	5488	75.21
16	pobyt_stały	5223	71.58
17	status_uchodźcy	2972	40.73
18	inny	2155	29.53
19	Liczba_ogolem	0	0.00
20	idTerytGmina	0	0.00
21	Typ_obszaru	0	0.00
22	idTypPodmiotu	0	0.00
23	Typ_podmiotu	0	0.00
24	Gmina	0	0.00

```
25              idTerytWojewodztwo               0             0.00
26                    Wojewodztwo                0             0.00
27                         Powiat                0             0.00

                          Column  Missing values  Missing percent
0                 rodzina_uchodźcy            7294            99.96
1              pobyt_czasowy_art53            7286            99.85
2            zezwolenie_osiedlenie            7271            99.64
3   rodzina_ochrona_miedzynarodowa            7212            98.84
4              pochodzenie_polskie            7195            98.60
5            ochrona_uzupelniajaca            7122            97.60
6       nauka_umowa_miedzynarodowa            7078            97.00
7                  pobyt_tolerowany            7041            96.49
8                pobyt_rezydent_UE            6989            95.78
9                pobyt_humanitarny            6835            93.67
10     obywatel_UE_EFTA_SZWAJCARIA            6623            90.76
11                 ochrona_czasowa            6600            90.45
12         karta_pobytu_rynek_pracy            6507            89.17
13                   karta_polaka              6443            88.30
14                   Brak_statusu              5855            80.24
15              pobyt_czasowy_inne             5488            75.21
16                     pobyt_stały              5223            71.58
17                 status_uchodźcy              2972            40.73
18                           inny               2155            29.53
19                  Liczba_ogolem                 0             0.00
20                   idTerytGmina                 0             0.00
21                    Typ_obszaru                 0             0.00
22                  idTypPodmiotu                 0             0.00
23                   Typ_podmiotu                 0             0.00
24                          Gmina                 0             0.00
25              idTerytWojewodztwo                0             0.00
26                    Wojewodztwo                 0             0.00
27                         Powiat                 0             0.00
```

- Wszystkie statusy pobytowe -> MNAR
- `Brak_statusu` -> Jeśli odnosi się do braku jakiegokolwiek statusu w SIO – też MNAR
- Braki w danych dotyczące statusów pobytu cudzoziemców są typu MNAR, ponieważ brak wartości wynika bezpośrednio z faktu nieposiadania danego statusu.
- Wypełniono brakujące wartości zerem (0), co oznacza brak danego statusu.

[7]:
```
print(df.head(10))
```

```
   idTerytGmina  idTerytWojewodztwo   Wojewodztwo         Powiat        Gmina  \
0        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
1        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
2        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
3        201011                   2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
```

```
4        201011                  2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
5        201011                  2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
6        201011                  2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
7        201011                  2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
8        201022                  2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec
9        201022                  2  DOLNOŚLĄSKIE  bolesławiecki  Bolesławiec

      Typ_obszaru  idTypPodmiotu              Typ_podmiotu  Liczba_ogolem  \
0  obszar miejski              1               Przedszkole            102
1  obszar miejski              3         Szkoła podstawowa            317
2  obszar miejski             14    Liceum ogólnokształcące             45
3  obszar miejski             16                  Technikum             90
4  obszar miejski             19         Szkoła policealna            177
5  obszar miejski             85  Szkoła muzyczna I stopnia              9
6  obszar miejski             93  Branżowa szkoła I stopnia             61
7  obszar miejski             94  Branżowa szkoła II stopnia             2
8  obszar wiejski              1               Przedszkole              4
9  obszar wiejski              3         Szkoła podstawowa             30

   obywatel_UE_EFTA_SZWAJCARIA  …  pobyt_czasowy_art53  rodzina_uchodźcy  \
0                          2.0  …                 NaN               NaN
1                          NaN  …                 NaN               NaN
2                          1.0  …                 NaN               NaN
3                          1.0  …                 NaN               NaN
4                          NaN  …                 NaN               NaN
5                          NaN  …                 NaN               NaN
6                          NaN  …                 NaN               NaN
7                          NaN  …                 NaN               NaN
8                          NaN  …                 NaN               NaN
9                          NaN  …                 NaN               NaN

   nauka_umowa_miedzynarodowa   inny  pobyt_stały  pobyt_humanitarny  \
0                         NaN   34.0            9                NaN
1                         NaN  241.0            5                NaN
2                         NaN   35.0          NaN                NaN
3                         NaN   62.0            2                3.0
4                         NaN    NaN            2                NaN
5                         NaN    NaN          NaN                NaN
6                         NaN   57.0          NaN                NaN
7                         NaN    2.0          NaN                NaN
8                         NaN    NaN          NaN                NaN
9                         NaN    5.0          NaN                NaN

   pobyt_czasowy_inne  karta_pobytu_rynek_pracy  \
0                 3.0                       NaN
1                14.0                       NaN
2                 NaN                       NaN
3                 1.0                       NaN
```

```
4                 67.0                      63.0
5                  2.0                       NaN
6                  2.0                       NaN
7                  NaN                       NaN
8                  1.0                       NaN
9                  NaN                       1.0

   rodzina_ochrona_miedzynarodowa  Brak_statusu
0                             NaN           1.0
1                             NaN           2.0
2                             NaN           6.0
3                             NaN           1.0
4                             NaN           NaN
5                             NaN           NaN
6                             NaN           1.0
7                             NaN           NaN
8                             NaN           NaN
9                             NaN           1.0

[10 rows x 28 columns]
```

[8]:
```
df = df.fillna(0)
```

[12]:
```
df['Typ_obszaru_encoded'] = df['Typ_obszaru'].map({'obszar miejski': 1, 'obszar␣
↪wiejski': 0})
```

[13]:
```
print(df.nunique().sort_values(ascending=False))
```

```
idTerytGmina                      2252
Gmina                             2057
Liczba_ogolem                      394
Powiat                             370
inny                               279
status_uchodźcy                    211
pobyt_czasowy_inne                 108
karta_pobytu_rynek_pracy            91
pobyt_stały                         89
Brak_statusu                        75
ochrona_czasowa                     65
nauka_umowa_miedzynarodowa          47
karta_polaka                        45
pobyt_humanitarny                   33
obywatel_UE_EFTA_SZWAJCARIA         33
pobyt_tolerowany                    28
pobyt_rezydent_UE                   23
idTypPodmiotu                       22
Typ_podmiotu                        22
Wojewodztwo                         16
```

```
idTerytWojewodztwo                  16
rodzina_ochrona_miedzynarodowa      16
ochrona_uzupelniajaca               14
pochodzenie_polskie                  9
zezwolenie_osiedlenie                4
rodzina_uchodźcy                     4
pobyt_czasowy_art53                  4
Typ_obszaru                          2
Typ_obszaru_encoded                  2
dtype: int64
```

0.2.4 Wybór zmiennej celu (TARGET)

Wybrana kolumna: `Liczba_ogolem`

Uzasadnienie:

- Jest to agregowana liczba cudzoziemców w danej placówce, naturalny kandydat do regresji (lub klasyfikacji po skategoryzowaniu).
- Zmienna zawiera zróżnicowane wartości (394 unikalne liczby), co pozwala na bogatą analizę.

0.2.5 6. Wybór cech (FEATURES)

Wybrane kolumny:

- `Typ_obszaru_encoded` – wpływ środowiska miejskiego/wiejskiego.
- `idTerytWojewodztwo` – lokalizacja administracyjna.
- `pobyt_ogolem` – suma zezwoleń pobytowych.
- `ochrona_ogolem` – suma form ochrony.
- `karta_pobytu_rynek_pracy` – może wskazywać na cudzoziemców z rodzinami.
- `inny`, `Brak_statusu` – mogą oznaczać mniej uregulowany status, potencjalnie wpływający na dostęp do edukacji.

Uzasadnienie:

- Wszystkie te zmienne mogą być istotnymi predyktorami liczby uczniów cudzoziemców w szkole – bez potrzeby stosowania metod typu feature importance.

0.2.6 EDA

[17]:
```
display(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7297 entries, 0 to 7296
Data columns (total 29 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   idTerytGmina                    7297 non-null   int64
 1   idTerytWojewodztwo              7297 non-null   int64
 2   Wojewodztwo                     7297 non-null   object
 3   Powiat                          7297 non-null   object
```

```
 4   Gmina                           7297 non-null   object
 5   Typ_obszaru                     7297 non-null   object
 6   idTypPodmiotu                   7297 non-null   int64
 7   Typ_podmiotu                    7297 non-null   object
 8   Liczba_ogolem                   7297 non-null   int64
 9   obywatel_UE_EFTA_SZWAJCARIA     7297 non-null   float64
 10  pochodzenie_polskie             7297 non-null   float64
 11  zezwolenie_osiedlenie           7297 non-null   float64
 12  karta_polaka                    7297 non-null   float64
 13  status_uchodźcy                 7297 non-null   float64
 14  pobyt_tolerowany                7297 non-null   float64
 15  ochrona_uzupelniajaca           7297 non-null   float64
 16  ochrona_czasowa                 7297 non-null   float64
 17  pobyt_rezydent_UE               7297 non-null   float64
 18  pobyt_czasowy_art53             7297 non-null   float64
 19  rodzina_uchodźcy                7297 non-null   float64
 20  nauka_umowa_miedzynarodowa      7297 non-null   float64
 21  inny                            7297 non-null   float64
 22  pobyt_stały                     7297 non-null   object
 23  pobyt_humanitarny               7297 non-null   float64
 24  pobyt_czasowy_inne              7297 non-null   float64
 25  karta_pobytu_rynek_pracy        7297 non-null   float64
 26  rodzina_ochrona_miedzynarodowa  7297 non-null   float64
 27  Brak_statusu                    7297 non-null   float64
 28  Typ_obszaru_encoded             7297 non-null   int64
dtypes: float64(18), int64(5), object(6)
memory usage: 1.6+ MB
```

```
None
```

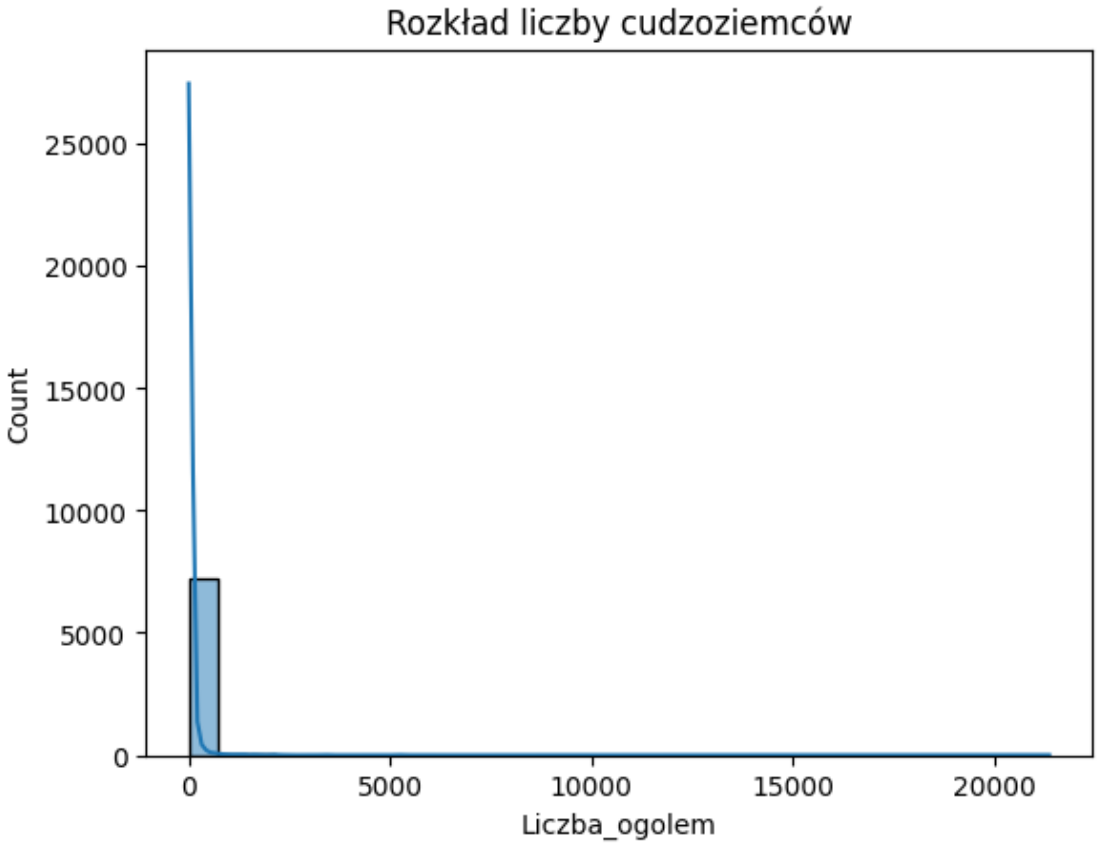
[19]:
```
display(df.describe().round(2))
```

```
       idTerytGmina  idTerytWojewodztwo  idTypPodmiotu  Liczba_ogolem  \
count       7297.00             7297.00        7297.00        7297.00
mean     1699752.25               16.83          19.45          49.44
std       937204.23                9.37          30.87         403.37
min       201011.00                2.00           1.00           1.00
25%      1008032.00               10.00           3.00           2.00
50%      1463011.00               14.00           3.00           6.00
75%      2469011.00               24.00          16.00          18.00
max      3263011.00               32.00          94.00       21355.00

       obywatel_UE_EFTA_SZWAJCARIA  pochodzenie_polskie  \
count                      7297.00              7297.00
mean                          0.34                 0.02
std                           4.62                 0.28
min                           0.00                 0.00
25%                           0.00                 0.00
50%                           0.00                 0.00
```

```
75%                          0.00                0.00
max                        313.00               10.00

       zezwolenie_osiedlenie  karta_polaka  status_uchodźcy  pobyt_tolerowany  \
count                7297.00       7297.00          7297.00           7297.00
mean                    0.00          0.53            11.79              0.16
std                     0.09          5.34            74.71              2.33
min                     0.00          0.00             0.00              0.00
25%                     0.00          0.00             0.00              0.00
50%                     0.00          0.00             1.00              0.00
75%                     0.00          0.00             5.00              0.00
max                     5.00        263.00          2764.00            134.00

       …  pobyt_czasowy_art53  rodzina_uchodźcy  nauka_umowa_miedzynarodowa  \
count  …              7297.00           7297.00                     7297.00
mean   …                 0.00              0.00                        0.46
std    …                 0.07              0.09                        7.62
min    …                 0.00              0.00                        0.00
25%    …                 0.00              0.00                        0.00
50%    …                 0.00              0.00                        0.00
75%    …                 0.00              0.00                        0.00
max    …                 5.00              7.00                      367.00

           inny  pobyt_humanitarny  pobyt_czasowy_inne  \
count   7297.00            7297.00             7297.00
mean      24.00               0.30                3.11
std      221.28               3.47               33.78
min        0.00               0.00                0.00
25%        0.00               0.00                0.00
50%        2.00               0.00                0.00
75%        8.00               0.00                0.00
max    13321.00             223.00             1811.00

       karta_pobytu_rynek_pracy  rodzina_ochrona_miedzynarodowa  Brak_statusu  \
count                   7297.00                         7297.00       7297.00
mean                       3.19                            0.04          1.96
std                       54.79                            0.77         30.62
min                        0.00                            0.00          0.00
25%                        0.00                            0.00          0.00
50%                        0.00                            0.00          0.00
75%                        0.00                            0.00          0.00
max                     2385.00                           34.00       1910.00

       Typ_obszaru_encoded
count              7297.00
mean                  0.59
std                   0.49
min                   0.00
```

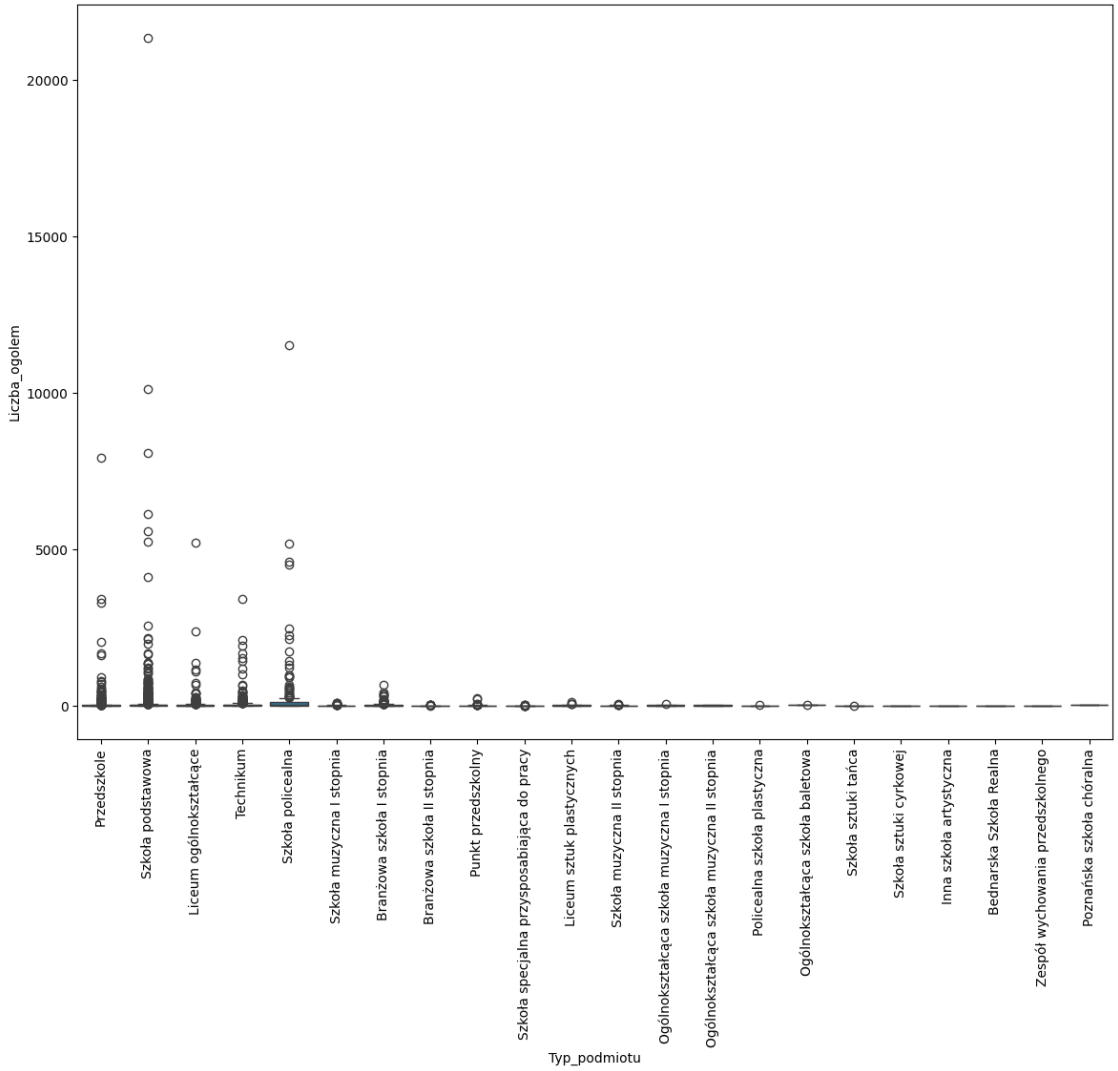
```
25%                   0.00
50%                   1.00
75%                   1.00
max                   1.00

[8 rows x 23 columns]
```

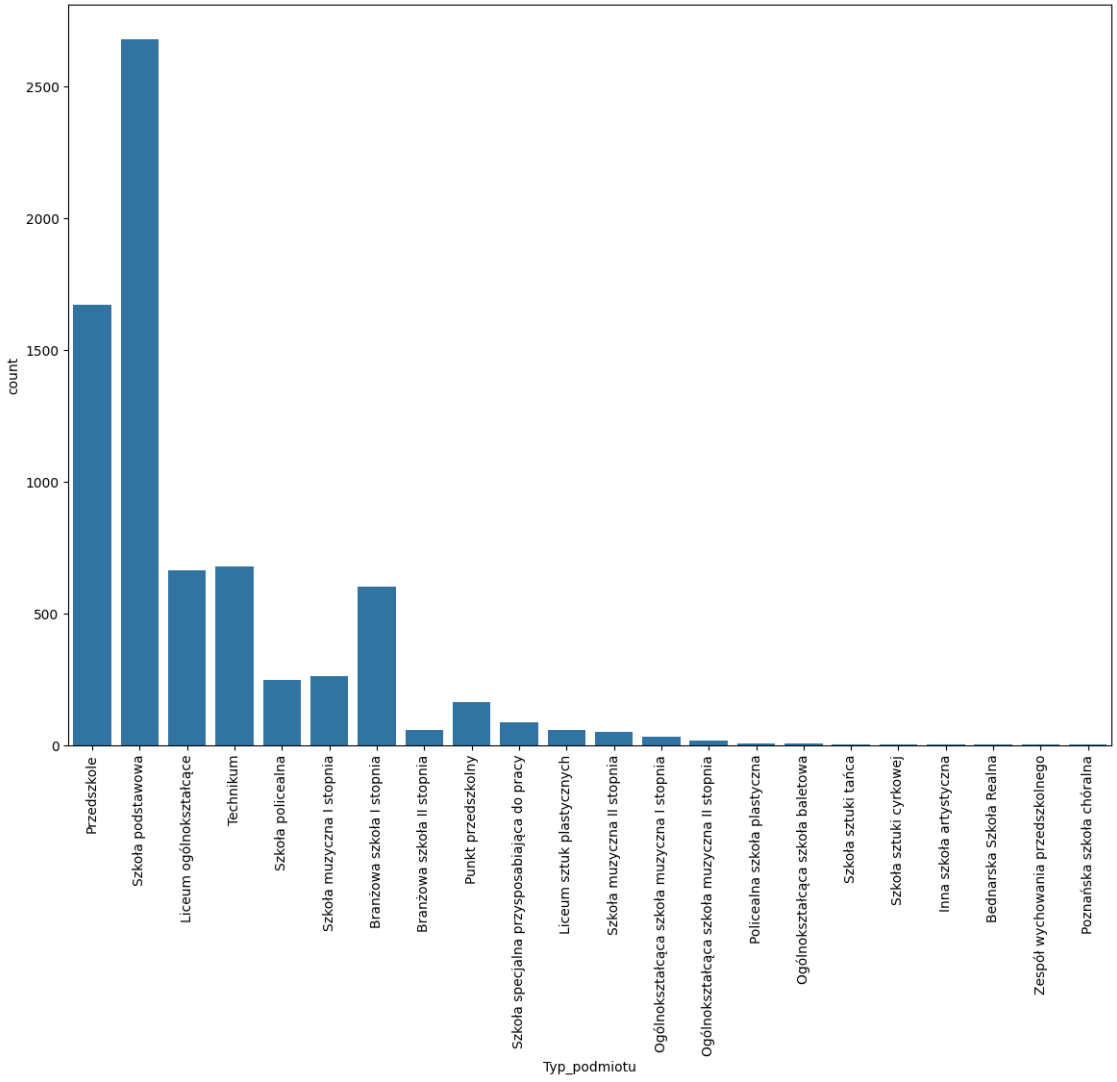
[20]:
```
# 2. Rozkład zmiennej celu
sns.histplot(df['Liczba_ogolem'], bins=30, kde=True)
plt.title("Rozkład liczby cudzoziemców")
plt.show()
```



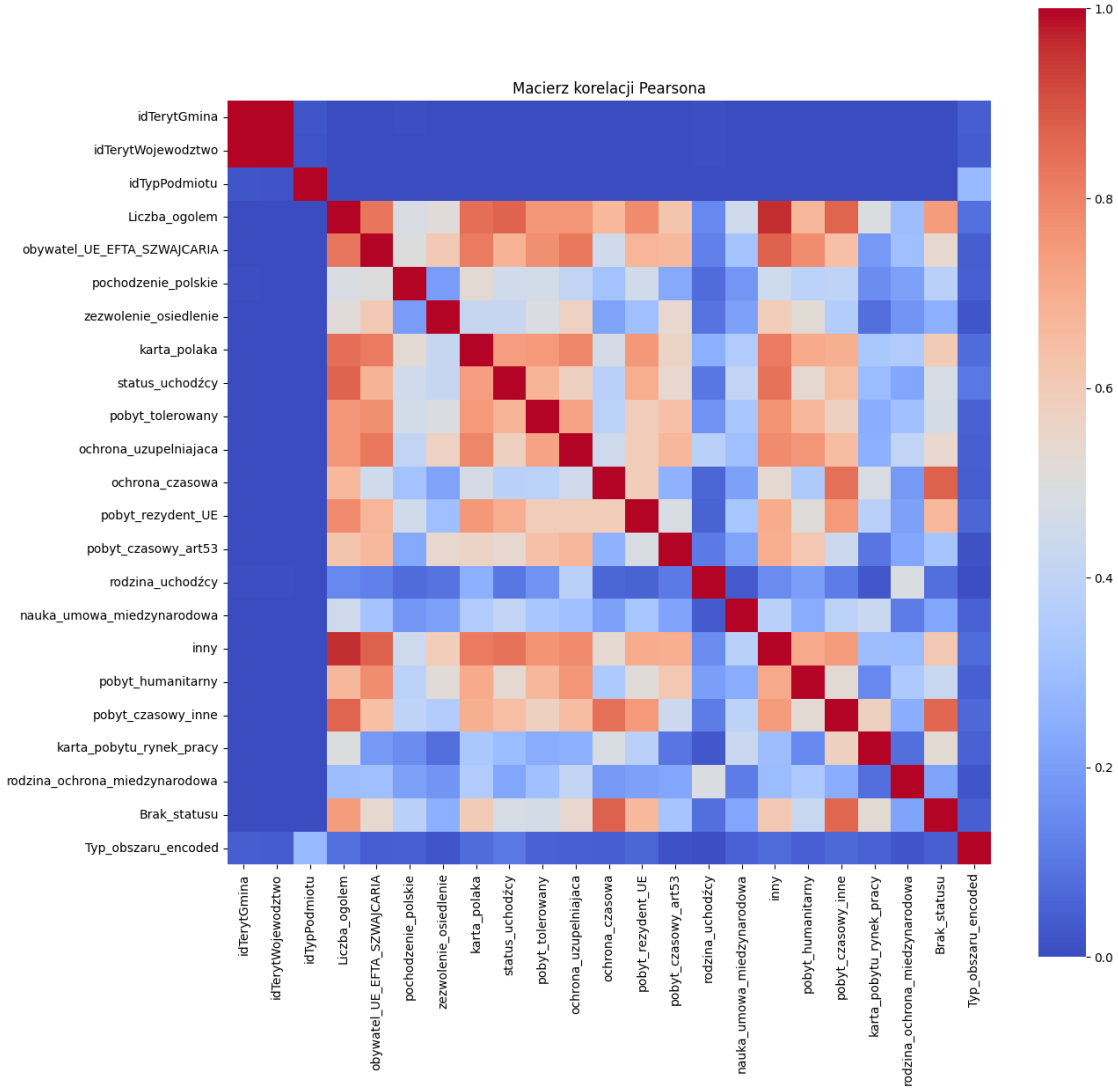
[28]:
```
# 3. Typ podmiotu vs liczba cudzoziemców
plt.figure(figsize=(14, 10))
sns.boxplot(data=df, x='Typ_podmiotu', y='Liczba_ogolem')
plt.xticks(rotation=90)
plt.show()
plt.figure(figsize=(14, 10))
sns.countplot(data=df, x='Typ_podmiotu')
plt.xticks(rotation=90)
```

```
plt.show()
```


Liczba_ogolem
20000
15000
10000
5000
0
Przedszkole
Szkoła podstawowa
Liceum ogólnokształcące
Technikum
Szkoła policealna
Szkoła muzyczna I stopnia
Branżowa szkoła I stopnia
Branżowa szkoła II stopnia
Punkt przedszkolny
Szkoła specjalna przysposabiająca do pracy
Liceum sztuk plastycznych
Szkoła muzyczna II stopnia
Ogólnokształcąca szkoła muzyczna I stopnia
Ogólnokształcąca szkoła muzyczna II stopnia
Policealna szkoła plastyczna
Ogólnokształcąca szkoła baletowa
Szkoła sztuki tańca
Szkoła sztuki cyrkowej
Inna szkoła artystyczna
Bednarska Szkoła Realna
Zespół wychowania przedszkolnego
Poznańska szkoła chóralna
Typ_podmiotu

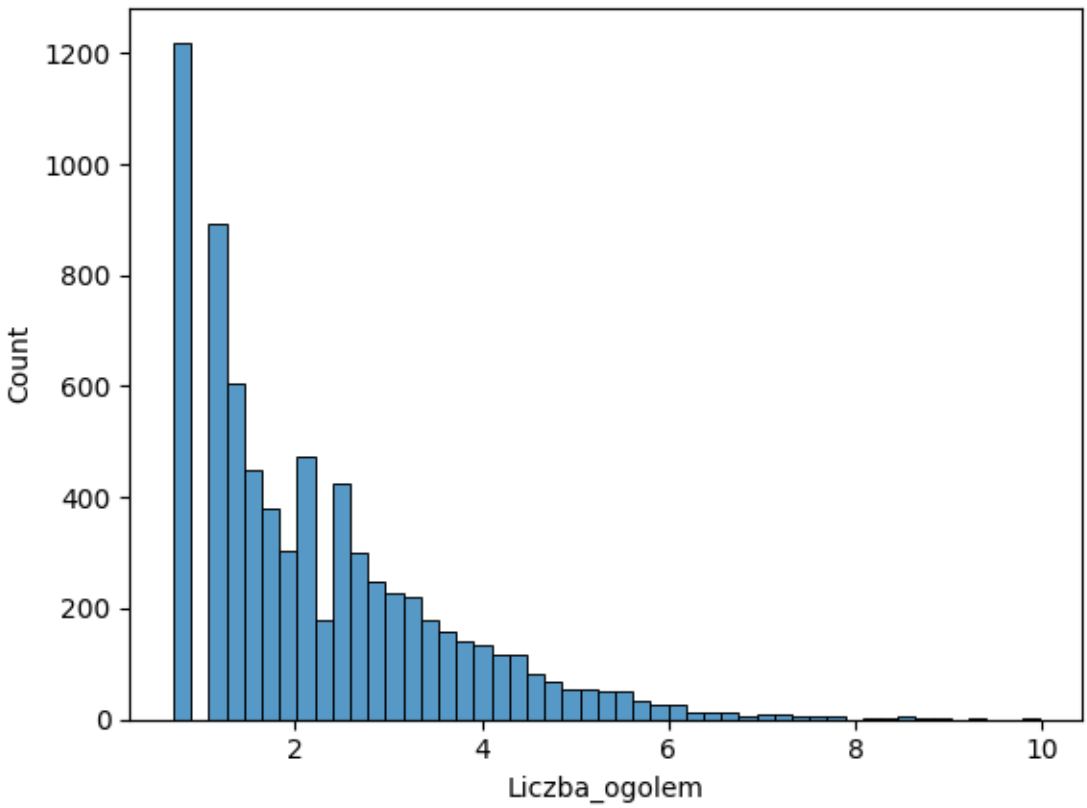


[22]:
```
# 4. Korelacje między kolumnami liczbowymi

plt.figure(figsize=(14, 14))
sns.heatmap(df.corr(numeric_only=True), cmap='coolwarm', annot=False, vmin=0,␣
 ↪vmax=1, square=True)
plt.title("Macierz korelacji Pearsona")
plt.show()
```



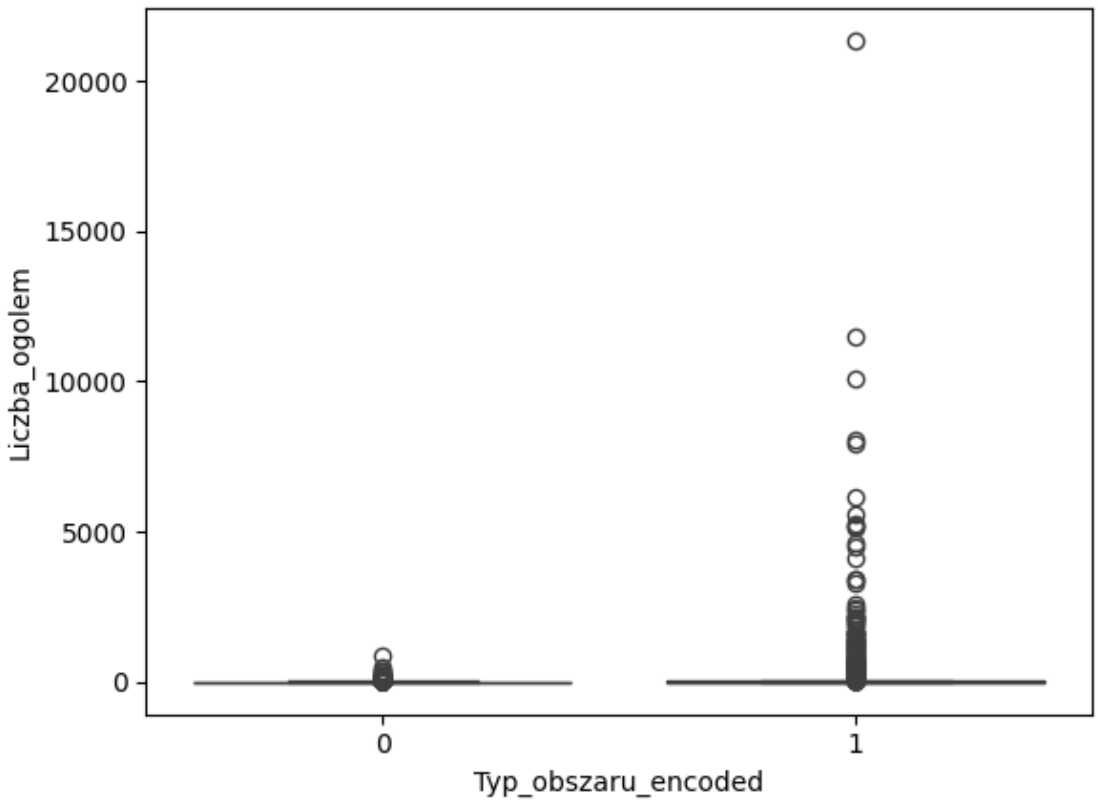
[37]:
```
sns.histplot(np.log1p(df['Liczba_ogolem']))
```

[37]: <Axes: xlabel='Liczba_ogolem', ylabel='Count'>



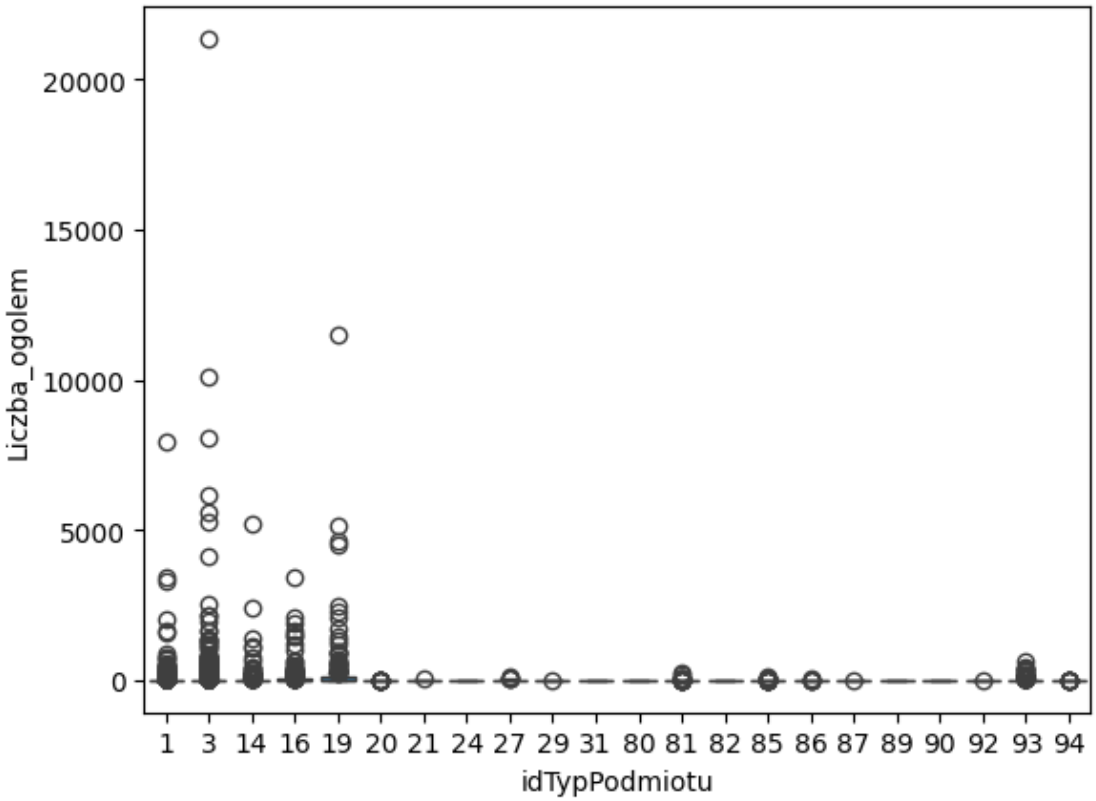
[39]:
```
sns.boxplot(data=df, x='Typ_obszaru_encoded', y='Liczba_ogolem')  # miejski vs␣
↪wiejski
```

[39]: <Axes: xlabel='Typ_obszaru_encoded', ylabel='Liczba_ogolem'>



[40]:
```
sns.boxplot(data=df, x='idTypPodmiotu', y='Liczba_ogolem')  # typy placówek
```

[40]: <Axes: xlabel='idTypPodmiotu', ylabel='Liczba_ogolem'>


20000
15000
10000
5000
0
Liczba_ogolem
1 3 14 16 19 20 21 24 27 29 31 80 81 82 85 86 87 89 90 92 93 94
idTypPodmiotu