

TECHNOLOGIE OBIEKTOWE		
DOKUMENTACJA PROJEKTU „PRZETWARZANIE DANYCH”		
Kierunek: Informatyka	Specjalizacja: Systemy Informacyjne	Nr grupy: 11D21B
Nazwisko i imię: Śpiewak Oskar		

## Wstęp

Otwarte dane to dane instytucji, urzędów, firm, instytutów badawczych, z których każdy może korzystać. Otwarte dane są źródłem oszczędności pieniędzy i czasu dla administracji i obywateli. Obywatele, w tym przedsiębiorcy, mogą korzystać z zasobów danych realizując własne cele, rozwijając swoją działalność gospodarczą lub badania. Otwarte dane charakteryzuje zasada 3R.



## Strony internetowe oferujące otwarte dane

### Dane.gov.pl

Dane.gov.pl oferuje bezpłatny, nie wymagający rejestracji dostęp do informacji publicznej danych z różnych kategorii, takich jak:

- edukacja, kultura i sport,
- energia,
- gospodarka i finanse,
- ludność i społeczeństwo,
- nauka i technologia
- regiony i miasta
- rząd i sektor publiczny,
- środowisko,
- transport.

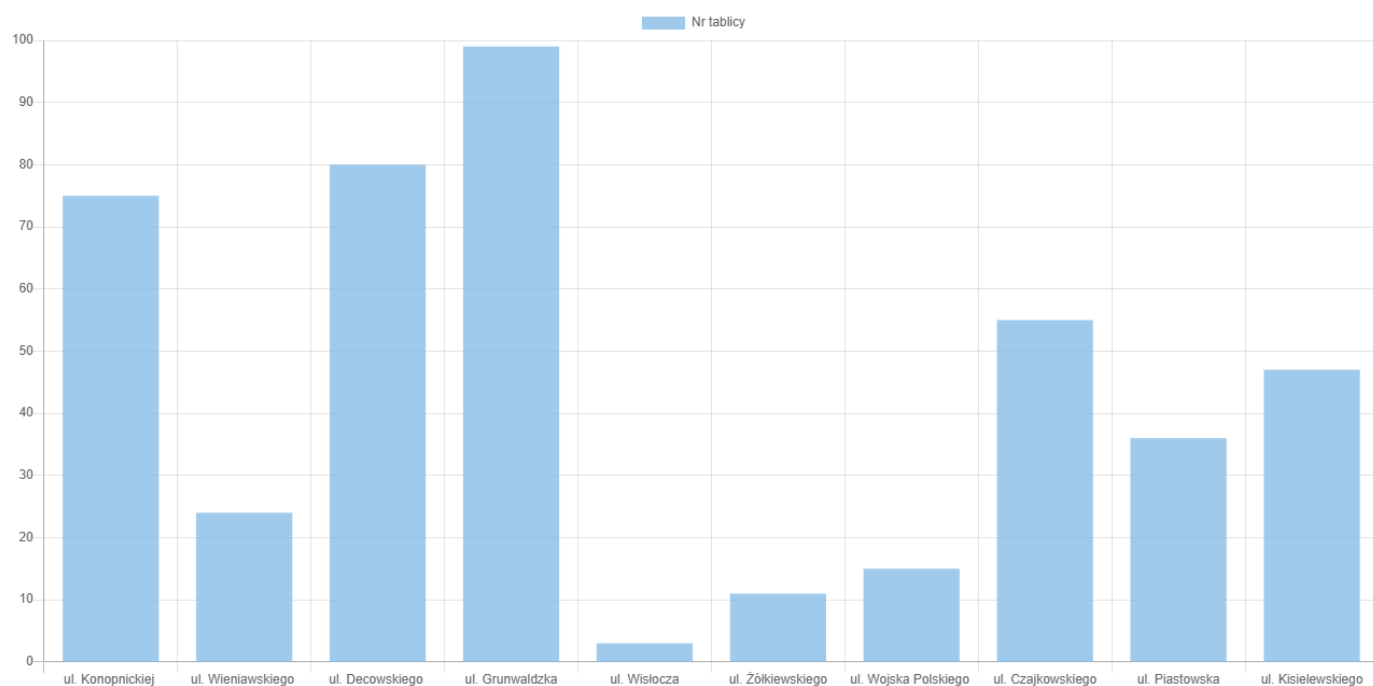
Dane można wyświetlać na różne sposoby, do których należą:

- widok tabelarycznych danych

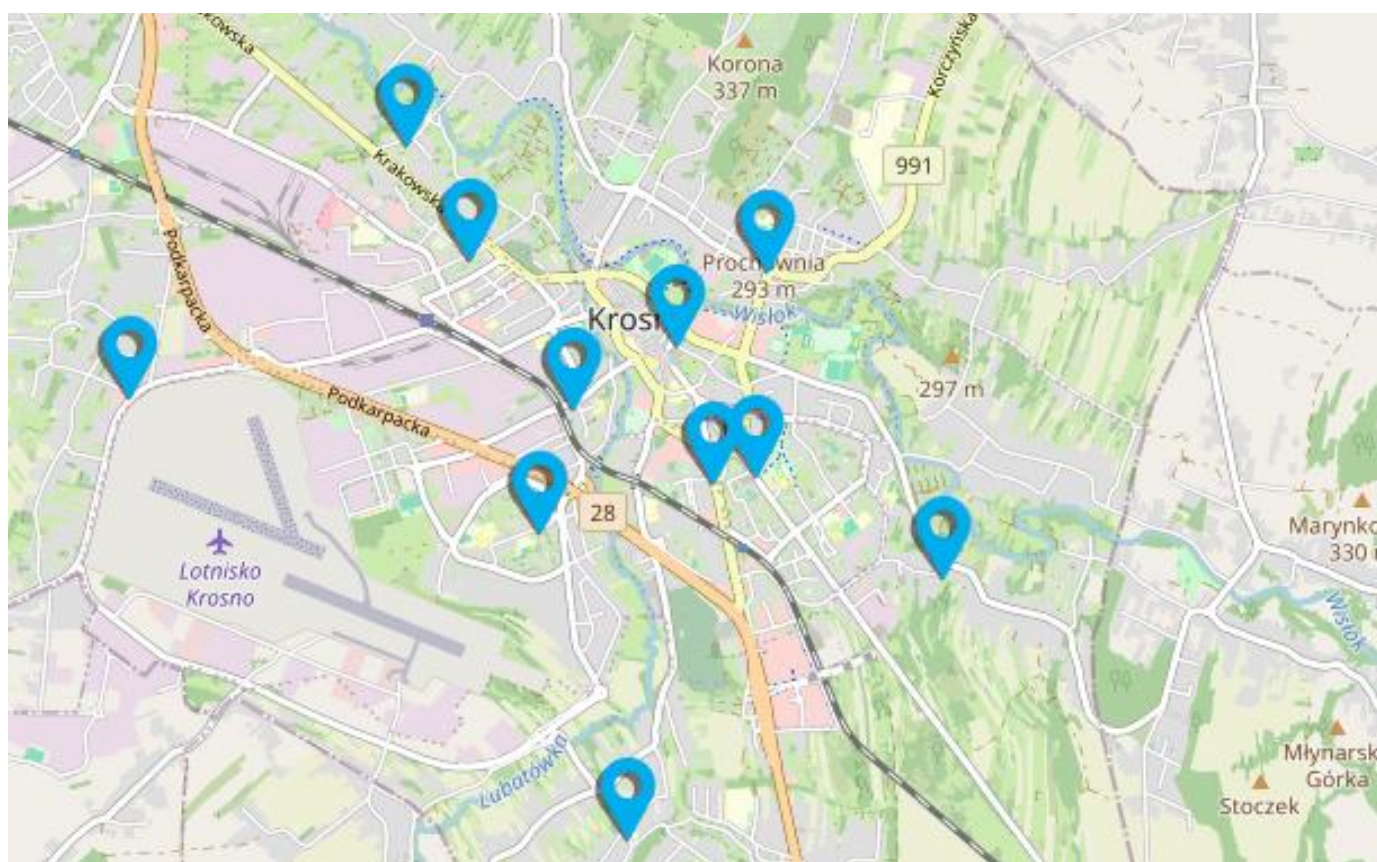
1	Jednostka pomocnicza	Adres	Nr tablicy	Opis	szer.	dł.
2	Dzielnica Białołęka	ul. Konopnickiej	75	przy przystanku MKS, naprzeciw szkoły	49,70258	21,74721
3	Dzielnica Krośnice	ul. Wieniawskiego	24	przy skrzyżowaniu z ul. Chopina	49,68133	21,78805
4	Dzielnica Polanka	ul. Decowskiego	80	przy sklepie Zielony Koszyk	49,69015	21,7261
5	Dzielnica Suchbátów	ul. Grunwaldzka	99	przy Szkole Podstawowej nr 5	49,66846	21,76408
6	Dzielnica Śródmieście	ul. Wiśniewska	3		49,69278	21,76778
7	Dzielnica Zawodzie	ul. Żółkiewskiego	11	przy skrzyżowaniu z ul. Prochownia	49,69634	21,77466
8	Osiedle Ks. B. Markiewicza	ul. Wojska Polskiego	15	przy „Kwadraciaku” i Przychodni Zdrowia	49,68631	21,77381
9	Osiedle Południe	ul. Czajkowskiego	55	przy Zespole Placówek Oświatowych	49,68957	21,75996
10	Osiedle Grota - Roweckiego	ul. Piastowska	36	przy skrzyżowaniu z ul. Lwowską	49,68592	21,77055
11	Osiedle Traugutta	ul. Kisielewskiego	47	przy sklepie Delikatesy Stokrotka	49,68353	21,75733
12	Osiedle Turaszówka	ul. Rzeszowska	87	obok sklepu FRAC	49,71883	21,71575
13	Osiedle Tysiąclecia	ul. Naftowa	68	przy skrzyżowaniu z ul. Mirandoli Pika	49,69698	21,752

Tabela przedstawiająca informacje o wyznaczonych miejscach przeznaczonych na bezpłatne umieszczanie urzędowych obwieszczeń wyborczych i plakatów komitetów wyborczych.

- wizualizacje na wykresach i mapach



Wykres kolumnowy obrazujący adres i przypisany do niego numer tablicy na obwieszczenia i plakaty wyborcze w mieście Krosno.



Mapa wizualizująca, w których miejscach znajdują się tablice na obwieszczenia i plakaty wyborcze w mieście Krosno.

## Stat.gov.pl

Stat.gov.pl – bezpłatne, otwarte dane, które charakteryzują różne dziedziny, takie jak:

- społeczeństwo
  - - małżeństwa, rozwody i separacje
  - – gospodarstwa domowe
  - – stan ludności
- rolnictwo
  - – gospodarstwa rolne
  - – produkcja rolnicza i zwierzęca
  - – uprawy rolnicze
- stan środowiska naturalnego
  - - tereny zieleni
  - - odpady komunalne
  - – składowiska

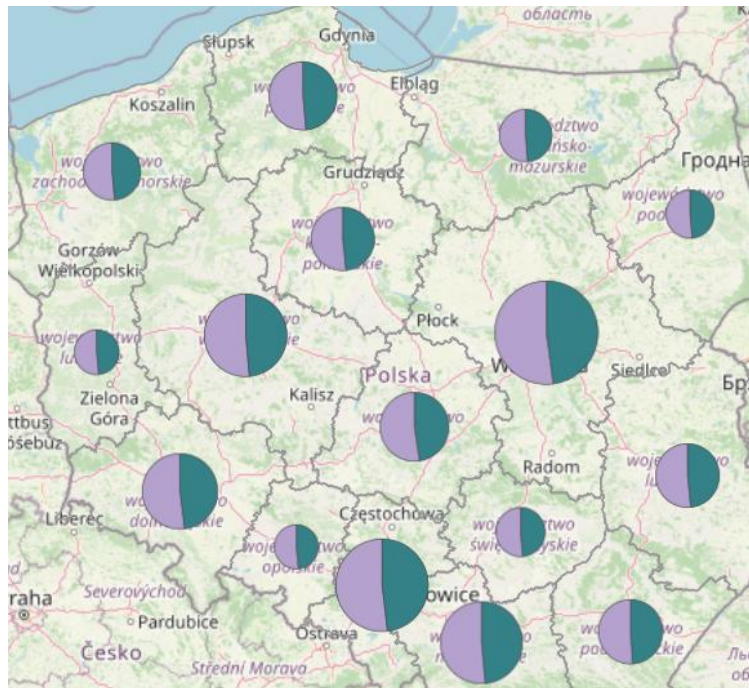
Dane można wyświetlać na takie sposoby jak:

- widok tabelaryczny

Kod	Nazwa	ogółem (w tys.)	
		mężczyźni	kobiety
		2021	2021
		[osoba]	[osoba]
0200000	DOLNOŚLĄSKIE	1393	1505
0400000	KUJAWSKO-POMORSKIE	976	1041
0600000	LUBELSKIE	987	1051
0800000	LUBUSKIE	479	507
1000000	ŁÓDZKIE	1141	1254
1200000	MAŁOPOLSKIE	1664	1766
1400000	MAZOWIECKIE	2639	2873
1600000	OPOLSKIE	459	490
1800000	PODKARPACKIE	1021	1065
2000000	PODLASKIE	559	590
2200000	POMORSKIE	1146	1212
2400000	ŚLĄSKIE	2106	2270
2600000	ŚWIĘTOKRZYSKIE	578	609
2800000	WARMIŃSKO-MAZURSKIE	671	703
3000000	WIELKOPOLSKIE	1701	1799
3200000	ZACHODNIOPOMORSKIE	801	849

Tabela przedstawiająca ludność (w tysiącach) według płci w 2021 roku.

- widok mapy



Mapa obrazująca stosunek liczebności kobiet (kolor fioletowy) do liczebności mężczyzn (kolor zielony).

## **Ec.europa.eu/eurostat**

Ec.europa.eu/eurostat - oficjalny portal statystyczny Unii Europejskiej, czyli urząd statystyczny UE.

Oferuje on dane o krajach UE podzielone na różne kategorie, takie jak:

- Szczegółowe zbiory danych
  - Ekonomia i finanse
  - Populacja i warunki socjalne
  - Przemysł handel i usługi
  - Rolnictwo, leśnictwo i rybołówstwo
  - Handel międzynarodowy
  - Transport
  - Środowisko i energia
  - Nauka, technologia, społeczeństwo cyfrowe
- Polityki UE
  - Wskaźniki procedury nierównowagi makroekonomicznej
  - Wskaźniki gospodarki o obiegu zamkniętym
  - Wskaźniki zatrudnienia i polityki społecznej
- Tematy przekrojowe
  - Jakość życia
  - Wskaźniki globalizacji gospodarczej
  - Jakość zatrudnienia



- Zmiana klimatu

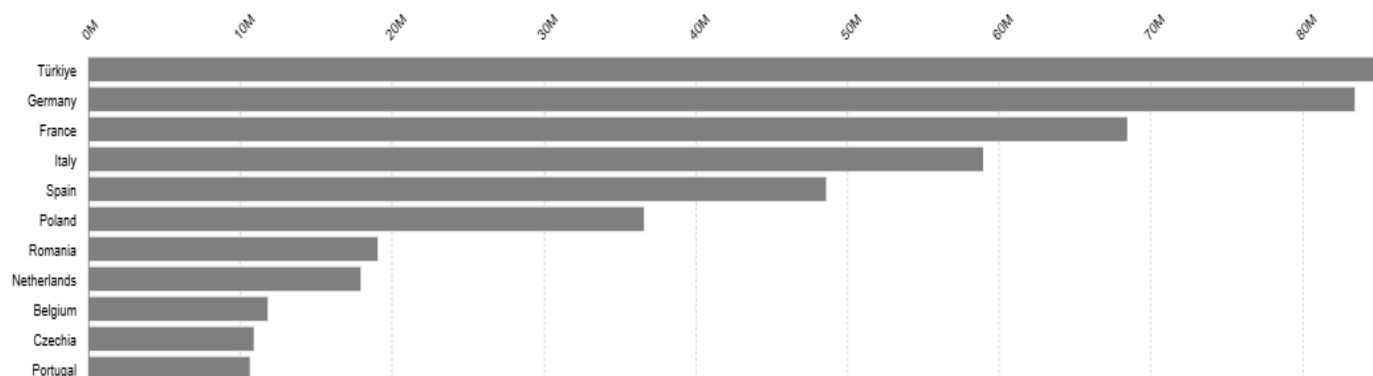
Eurostat pozwala na wyświetlanie danych za pomocą :

- tabeli

1	TIME	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
2	GEO (Labels)										
3	Belgium	11 237 274	11 311 117	11 351 727	11 398 589	11 455 519	11 522 440	11 554 767	11 617 623	11 742 796	11 817 096
4	Bulgaria	7 029 690	6 938 759	6 849 518	6 757 418	6 664 177	6 569 275	6 532 117	6 482 484	6 447 710	6 445 481
5	Czechia	10 538 275	10 553 843	10 578 820	10 610 055	10 649 800	10 693 939	10 494 836	10 516 707	10 827 529	10 900 555
6	Denmark	5 659 715	5 707 251	5 748 769	5 781 190	5 806 081	5 822 763	5 840 045	5 873 420	5 932 654	5 961 249
7	Germany	81 197 537	82 175 684	82 521 653	82 792 351	83 019 213	83 166 711	83 155 031	83 237 124	83 118 501	83 456 045
8	Estonia	1 314 870	1 315 944	1 315 635	1 319 133	1 324 820	1 328 976	1 330 068	1 331 796	1 365 884	1 374 687
9	Ireland	4 677 627	4 726 286	4 799 157	4 855 733	4 940 311	5 012 600	5 066 893	5 154 277	5 271 395	5 351 681
10	Greece	10 858 018	10 783 748	10 768 193	10 741 165	10 724 599	10 718 565	10 678 632	10 459 782	10 413 982	10 400 720
11	Spain	46 425 722	46 418 884	46 497 393	46 645 070	46 918 951	47 318 050	47 400 798	47 486 843	48 085 361	48 619 695
12	France	66 458 153	66 638 391	66 809 816	67 026 224	67 290 471	67 473 651	67 728 568	68 091 703	68 277 210	68 467 362
13	Croatia	4 180 915	4 131 115	4 078 769	4 014 341	3 968 676	3 933 511	3 893 026	3 862 305	3 850 894	3 861 967
14	Italy	60 295 497	60 163 712	60 066 734	59 937 769	59 816 673	59 641 488	59 236 213	59 030 133	58 997 201	58 971 230
15	Cyprus	847 008	848 319	854 802	864 236	875 899	888 005	896 007	904 705	949 084	966 365
16	Latvia	1 986 096	1 968 957	1 950 116	1 934 379	1 919 968	1 907 675	1 893 223	1 875 757	1 883 008	1 871 882

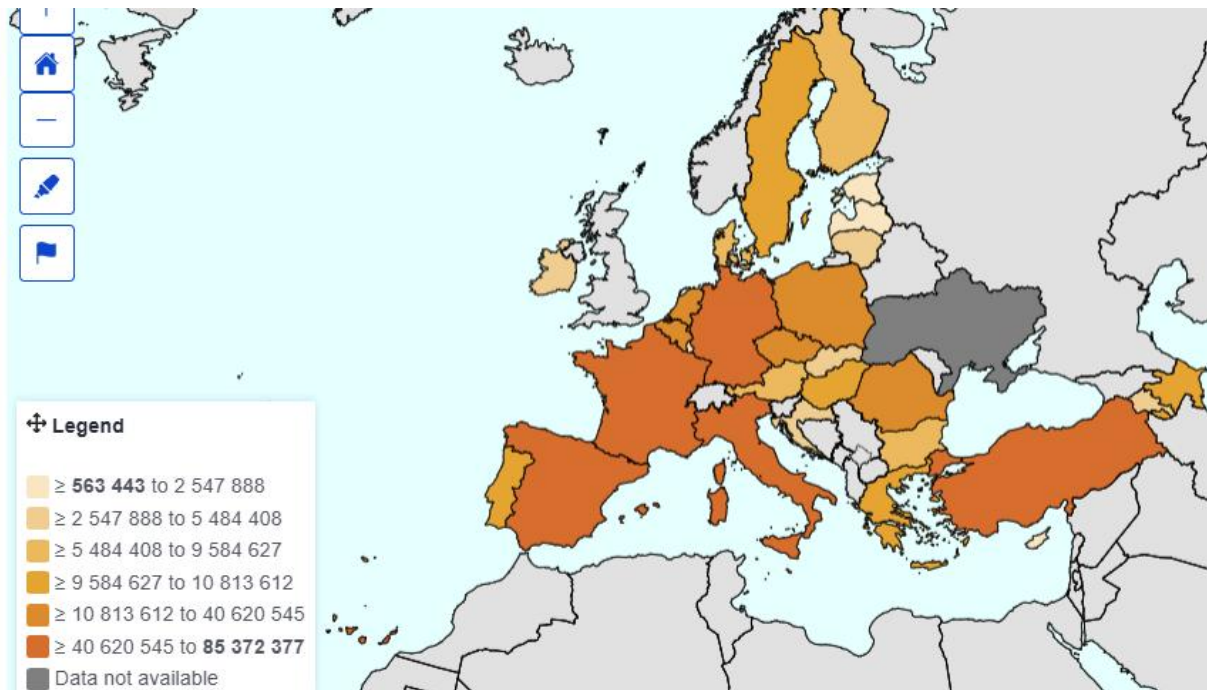
Tabela przedstawiająca liczbę ludności w krajach Europy w latach 2015 – 2024.

- wykresów



Wykres słupkowy wizualizujący zaludnienie w krajach Europy (2024 rok).

- map



Mapa wizualizująca zaludnienie w krajach Europy (2024 rok).

## Zasoby zastosowane w projekcie

W projekcie wykorzystywane są biblioteki języka Python, takie jak:

- requests,
- tkinter,
- ttkbootstrap,
- pandas,
- matplotlib,
- seaborn,
- sklearn,
- scipy,
- numpy,
- re,
- math,
- os.

W celu skorzystania z zasobów bibliotek języka Python należy je zainstalować za pomocą polecenia `pip install <nazwa_biblioteki>` w wierszu poleceń (cmd). Po instalacji trzeba je zaimportować w pliku poleceniem `import np.` `import pandas as pd.`



## Opis działania aplikacji

W celu automatycznego pobierania danych statystycznych aplikacja korzysta z publicznego API Banku Danych Lokalnych (BDL), udostępnianego przez Główny Urząd Statystyczny. Dzięki temu możliwe jest dynamiczne pozyskiwanie danych dla wybranych zmiennych oraz zakresów lat, bez konieczności ręcznego pobierania plików ze strony internetowej. Dane pobierane są stronicowo, co oznacza, że każde zapytanie może zwracać ograniczoną liczbę wyników. W celu pobrania pełnego zbioru danych, aplikacja iteracyjnie przetwarza kolejne strony odpowiedzi. Fragment kodu z aplikacji przedstawiający realizację połączenia z API BDL:

```
API_KEY = "aababab3-da9a-477f-108d-08dd8451a904"
BASE_URL = "https://bdl.stat.gov.pl/api/v1"
HEADERS = {
    "Accept": "application/json",
    "X-ClientId": API_KEY
}
def make_request(url, params=None):
    results = []
    page = 0
    while True:
        if params is None:
            params = {}
        params.update({"page": page, "page-size": 100})
        response = requests.get(url, headers=HEADERS, params=params)
        if response.status_code != 200:
            messagebox.showerror("Błąd API", f"Błąd {response.status_code}: {response.text}")
            return []
        data = response.json()
        batch = data.get("results", [])
        results.extend(batch)
        if not data.get("links", {}).get("next"):
            break
        page += 1
    return results
```

### 1. Wybór zmiennych i zapis danych

Na etapie wstępnym użytkownik ma możliwość wyboru zmiennych, które będą przedmiotem analizy. Dodatkowo określany jest zakres lat, dla których dane mają być pozyskane. Po zatwierdzeniu wyboru, dane są pobierane i zapisywane do pliku CSV w ujednoliconej strukturze: Województwo, Rok, Wartość\_zmiennej

Wybierz temat główny:

CENY

Wybierz podtemat:

CENY W ROLNICTWIE

Wybierz pod-podtemat:

Cena zakupu/sprzedaży użytków rolnych za 1 ha w obrocie prywatnym

Wybierz zmienną:

grunty orne ogółem

Rok od: 2013 Rok do: 2014

Dodaj zmienną Analiza danych

Rys. 1. Aplikacja po uruchomieniu

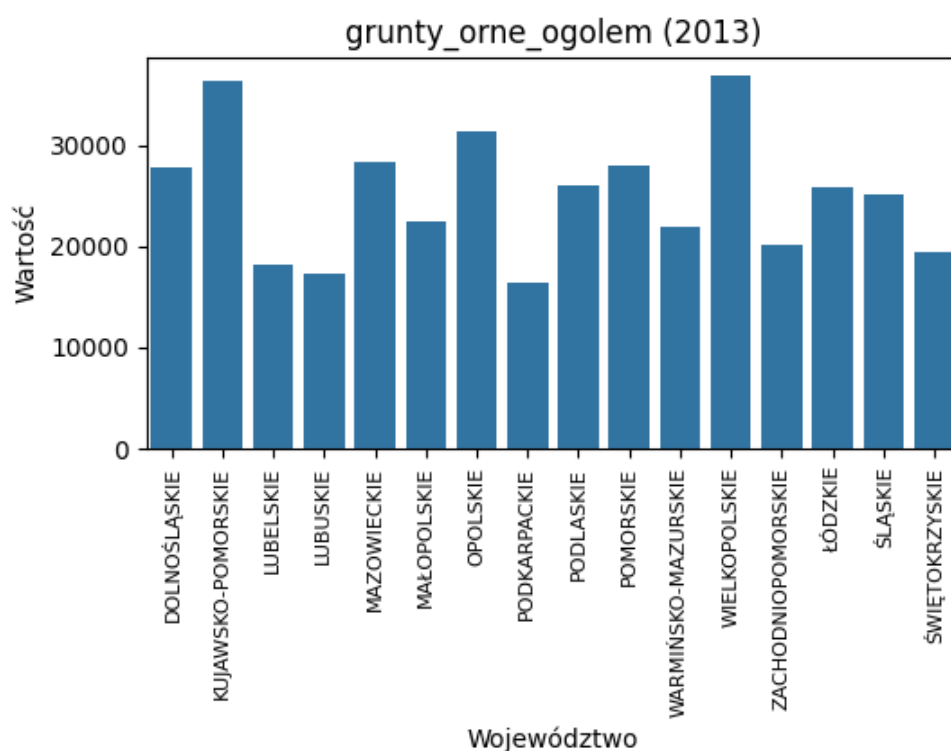
## 2. Sekcja: Informacje podstawowe

Ta sekcja zawiera ogólne informacje o zebranych zbiorze danych. Prezentowane są tam m.in.:

- Liczba rekordów w pliku z danymi,
- Liczba województw objętych analizą,
- Zakres lat, z których pochodzą dane.

Dodatkowo możliwe jest wygenerowanie wykresów ułatwiających wizualną eksplorację danych:

- Wykres słupkowy,
- Wykres liniowy,
- Wykres kołowy,
- Mapa ciepła,
- Wykres pudełkowy,
- Wykres bąbelkowy.



Rys. 2. Wykres słupkowy dla jednej z wybranych zmiennych

### 3. Sekcja: Statystyki i testy normalności

W tej sekcji prezentowane są statystyki opisowe oraz testy normalności rozkładu dla każdej wybranej zmiennej. Dla każdej zmiennej obliczane i prezentowane są następujące miary:

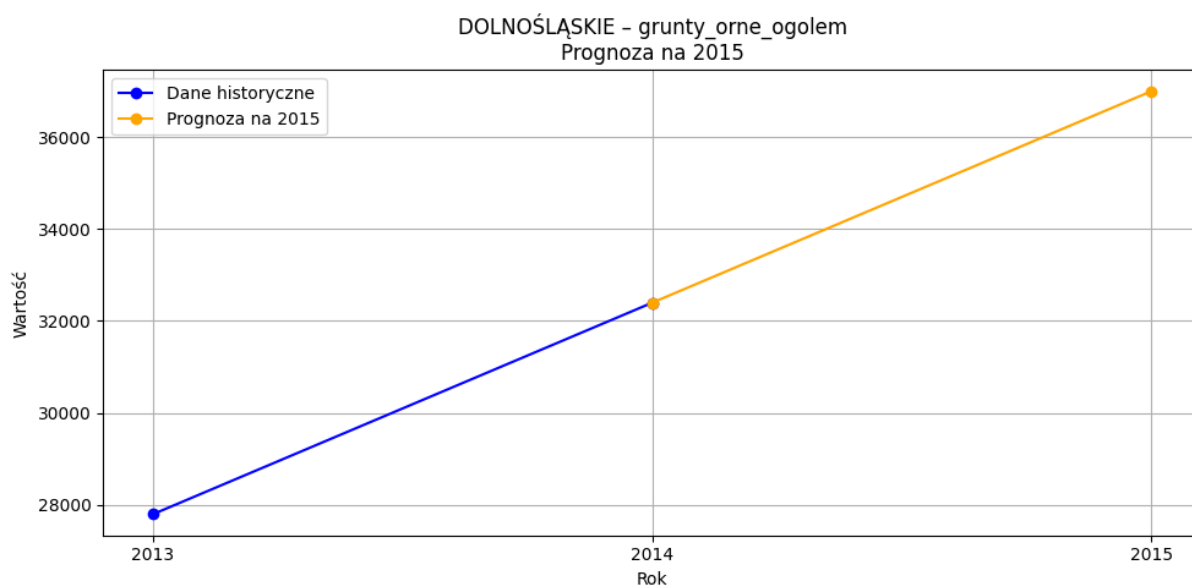
- Średnia arytmetyczna,
- Mediana,
- Odchylenie standardowe,
- Współczynnik zmienności,
- Skośność,
- Kurtoza,
- Wartość minimalna i maksymalna,
- Wynik testu normalności rozkładu (Test Shapiro–Wilka),
- Największy oraz najniższy średni roczny wzrost procentowy - obliczany osobno dla każdego województwa, na podstawie średnich zmian między latami.

#### 4. Sekcja: Trendy i zmiany w czasie

Dla każdej analizowanej zmiennej możliwe jest przeprowadzenie analizy trendu w czasie. Aplikacja oblicza oraz prezentuje:

- Równanie trendu liniowego,
- Wartość współczynnika determinacji ( $R^2$ ),
- Typ wykrytego trendu (rosnący, malejący, brak trendu),
- Istotność statystyczna nachylenia linii trendu,
- Średnia wartość prognozy - obliczona jako średnia prognoz dla wszystkich województw,
- Najwyższa oraz najniższa prognozowana wartość.

Użytkownik może również wyświetlić wykres trendu i prognozy dla wybranego województwa. Aplikacja umożliwia wygenerowanie prognozy wartości zmiennej na kolejny rok, tj. dla zbioru z lat 2013–2014 możliwe jest prognozowanie wartości na rok 2015.

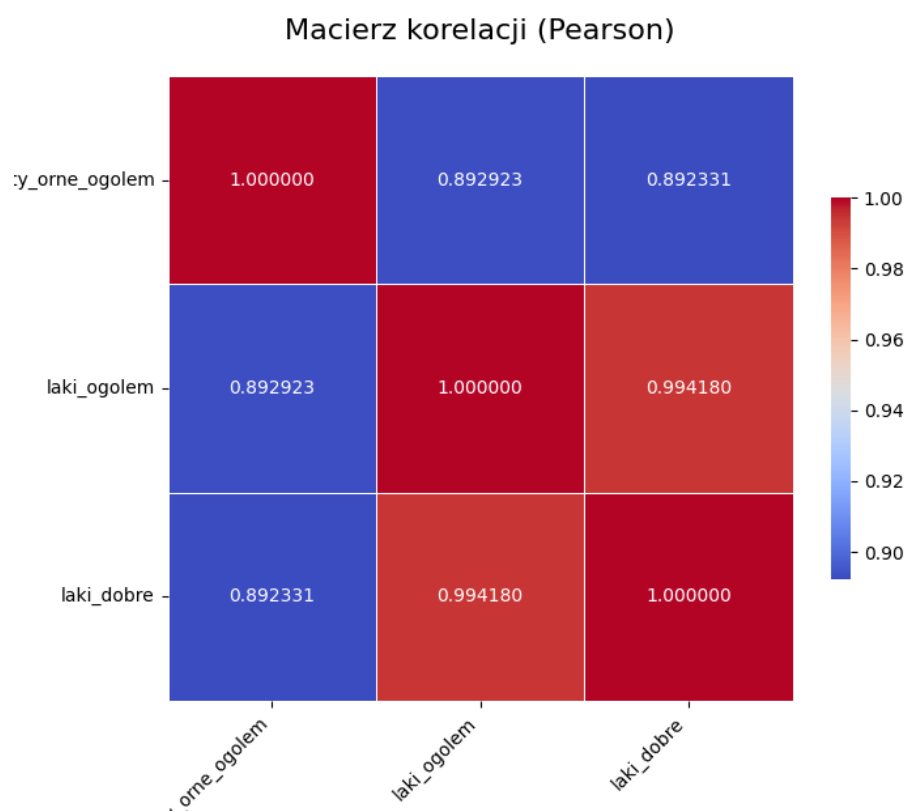


Rys. 3. Wykres prognozy dla wybranego województwa

#### 5. Sekcja: Korelacje między zmiennymi

Ostatnia sekcja aplikacji pozwala na badanie zależności pomiędzy zmiennymi. Użytkownik wybiera zestaw zmiennych, dla których obliczana jest korelacja (współczynnik korelacji Pearsona).

Wyniki mogą zostać wyświetlone w formie mapy ciepła, co umożliwia szybkie zidentyfikowanie silnych dodatnich lub ujemnych zależności pomiędzy zmiennymi.



Rys. 4. Mapa ciepła korelacji dla wybranych zmiennych

## **Wnioski**

Zrealizowana aplikacja pozwala na skuteczne pozyskiwanie, analizowanie oraz wizualizowanie danych statystycznych dla wybranych zmiennych z podziałem na województwa w Polsce. Dzięki wykorzystaniu publicznego API Banku Danych Lokalnych udostępnianego przez Główny Urząd Statystyczny, aplikacja umożliwia dynamiczne pobieranie danych bez potrzeby ręcznego ich przygotowywania, co znacząco usprawnia proces analizy oraz poprawia aktualność uzyskiwanych wyników.

Projekt stanowi praktyczne narzędzie do eksploracyjnej analizy danych statystycznych dla województw. W przyszłości aplikacja może zostać rozszerzona o możliwość analizy danych kwartalnych oraz miesięcznych. Ciekawym pomysłem na rozwój aplikacji byłoby również wzmocnienie interaktywności wykresów oferowanych przez stworzony program.