

## CELE:

- weryfikacja, które z województw charakteryzują się największym stopniem zachorowań na różnego typu choroby (zakaźne, psychiczne i weneryczne),
- sprawdzenie dostępu do specjalisty przez osoby chore wyrażonego w ułamku liczby specjalistów na osoby chore.

## PROCES PRZYGOTOWANIA DANYCH I UTWORZENIA MODELU DANYCH:

### 1. CENTRALNA TABELA W MODELU DANYCH – INFORMACJE O WOJEWÓDZTWACH

Tabela [**ww\_info**] w zeszycie **ww\_info** została utworzona z następujących atrybutów:

- id\_ww
- nazwa\_ww

Rys. 1 Tabela ww\_info

	A	B
1	ww_id	ww_nazwa
2		1 Dolnośląskie
3		2 Kujawsko-pomorskie
4		3 Lubelskie
5		4 Lubuskie
6		5 Łódzkie
7		6 Małopolskie
8		7 Mazowieckie
9		8 Opolskie
0		9 Podkarpackie
1		10 Podlaskie
2		11 Pomorskie
3		12 Śląskie
4		13 Świętokrzyskie
5		14 Warmińsko-mazurskie
6		15 Wielkopolskie
7		16 Zachodniopomorskie

Poprzez Power Query załadowano tą tabelę jako połączenie.

Aby zweryfikować postawiony cel należy uzyskać dane odnośnie liczby mieszkańców we wszystkich województwach w roku 2021. Dane takie przedstawia Główny Urząd Statystyczny [dalej GUS] w zeszycie dostępnym na stronie <https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/powierzchnia-i-ludnosc-w-przekroju-terytorialnym-w-2021-roku,7,18.html>. Z pobranego zeszytu pobrałem dane do pliku **ww\_info** przez Power Query.

Rys. 2 Dane nieobrobione z GUS.

Table.TransformColumnTypes(#"Nagłówki o podwyższonym poziomie",{{"TABL. 1. POWIERZCHNIA, LUDNOŚĆ ORAZ LOKATY WEDŁUG WOJEWÓDZTW#								
1	TABL. 1. POWIERZCHNIA, LUDNOŚĆ ORAZ LOKATY WEDŁUG...	Column2	Column3	Column4	Column5	Column6	Column7	Column8
2	Identyfikator terytorialny Territorial identifier	Województwa Voivodships	Powierzchnia Area	Ludność Population		Lokata według Ranking position by		
3			w ha in ha	w km <sup>2</sup> in km <sup>2</sup>	ogółem total	na 1 km <sup>2</sup> per 1 km <sup>2</sup>	powierzchni w ha area in ha	liczby ludność populat
4		P O L S K A	31270525	312705	38265013	122	x	x
5		P O L A N D	null	null	null	null	null	null
6	02	Dolnośląskie	1994670	19947	2891321	145	7	
7	04	Kujawsko-pomorskie	1797134	17971	2061942	115	10	
8	06	Lubelskie	2512246	25123	2095258	83	3	
9	08	Lubuskie	1398793	13988	1007145	72	13	
10	10	Łódzkie	1821895	18219	2437970	134	9	
11	12	Małopolskie	1518279	15183	3410441	225	12	
12	14	Mazowieckie	3555847	35559	5425028	153	1	
13	16	Opolskie	941187	9412	976774	104	16	
14	18	Podkarpackie	1784576	17846	2121229	119	11	
15	20	Podlaskie	2018702	20187	1173286	58	6	
16	22	Pomorskie	1832368	18323	2346671	128	8	
17	24	Śląskie	1233309	12333	4492330	364	14	
18	26	Świętokrzyskie	1171050	11710	1224626	105	15	
19	28	Warmińsko-mazurskie	2417347	24173	1416495	59	4	
20	30	Wielkopolskie	2982650	29826	3496450	117	2	
21	32	Zachodniopomorskie	2290472	22905	1688047	74	5	
22		null	null	null	null	null	null	null

W Power Query wykonano następujące kroki:

1. Usunięty pierwszego wiersza,
2. Usunięcie pierwszej kolumny,
3. Usunięcie kolumn dotyczących powierzchni województw,
4. Usunięcie niepotrzebnych kolumn zawierających wartości NULLowe,
5. Zmiana nazwy nagłówków,
6. Zmiana typu danych na liczby całkowite w kolumnach dot. Liczby ludności

Rys. 3 Dane z GUS obrobione.

1	Województwa...	1 <sup>2</sup> ww_ludnosc	1 <sup>2</sup> ww_ludnoscPerKm
1	Dolnośląskie	2891321	145
2	Kujawsko-pomorskie	2061942	115
3	Lubelskie	2095258	83
4	Lubuskie	1007145	72
5	Łódzkie	2437970	134
6	Małopolskie	3410441	225
7	Mazowieckie	5425028	153
8	Opolskie	976774	104
9	Podkarpackie	2121229	119
10	Podlaskie	1173286	58
11	Pomorskie	2346671	128
12	Śląskie	4492330	364
13	Świętokrzyskie	1224626	105
14	Warmińsko-mazurskie	1416495	59
15	Wielkopolskie	3496450	117
16	Zachodniopomorskie	1688047	74

W celu otrzymania pojedynczej tabeli zawierającej dane nt. numeru id województwa, nazwy województwa oraz liczby ludności w województwie wykonano scalenie połączenia tabeli ww\_info

z zapytaniem zawierającym dane z GUS obrobione. Kluczem łączącym obie tabelę jest atrybut zawierający nazwy województw (każda nazwa jest unikalna). Scalenie zostało przeprowadzone jako tzw. INNER JOIN. Tak scalone zapytanie załadowano do modelu danych – będzie ono stanowiło centralną tabelę w modelu.

Rys. 4 Scalone zapytania – pierwsza tabela w modelu danych.

	$1^2_3$ ww_info	$A^B_C$ ww_nazwa	$1^2_3$ ww_ludnosc	$1^2_3$ ww_ludnoscPerKm
1		1 Dolnośląskie	2891321	145
2		2 Kujawsko-pomorskie	2061942	115
3		3 Lubelskie	2095258	83
4		4 Lubuskie	1007145	72
5		5 Łódzkie	2437970	134
6		6 Małopolskie	3410441	225
7		7 Mazowieckie	5425028	153
8		8 Opolskie	976774	104
9		9 Podkarpackie	2121229	119
10		10 Podlaskie	1173286	58
11		11 Pomorskie	2346671	128
12		12 Śląskie	4492330	364
13		13 Świętokrzyskie	1224626	105
14		14 Warmińsko-mazurskie	1416495	59
15		15 Wielkopolskie	3496450	117
16		16 Zachodniopomorskie	1688047	74

2. Z pobranego arkusza kalkulacyjnego ze strony: <https://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie-i-ochrona-zdrowia-w-2021-roku,1,12.html> na potrzeby tego projektu utworzono bazę danych w MS SQL Server o nazwie Choroby. Na potrzeby szkoleniowe tabelę w tej bazie mają następującą formę:

Rys. 5 Tabele w bazie danych projekt23

SQLQuery1.sql - DE...BQP(Mateusz (80))\*

```

select top(5)* from psychiczne
select top(5)* from zakazne
select top(5)* from weneryczne
  
```

200 %

Results Messages

ww_id	Zaburzenia psychiczne i zaburzenia zachowania (bez uzależnień)	organiczne zaburzenia psychiczne	schizofrenia	inne zaburzenia psychiczne (niepsychotyczne)	zaburzenia nastroju	zaburzenia nerwicowe, odżywiania i zespoły _ behawio
1	3718,0912764587	474,107742467885	286,18642864435	51,722091849291	607,80317617546	1185,3551023432
2	4861,19592070356	734,520553185792	439,849113312646	57,7277437967199	1445,44973443766	1377,66745582503
3	3267,17972757279	347,410546868947	277,095504486216	77,1431560025414	1078,11555686414	1127,51042134788
4	3749,94605368938	462,683706006156	337,126804441362	58,613293778051	625,275377097257	1729,89508822605
5	4499,95240173431	718,878049122234	414,784886747036	56,4997595360344	979,370375910909	1736,45066944591

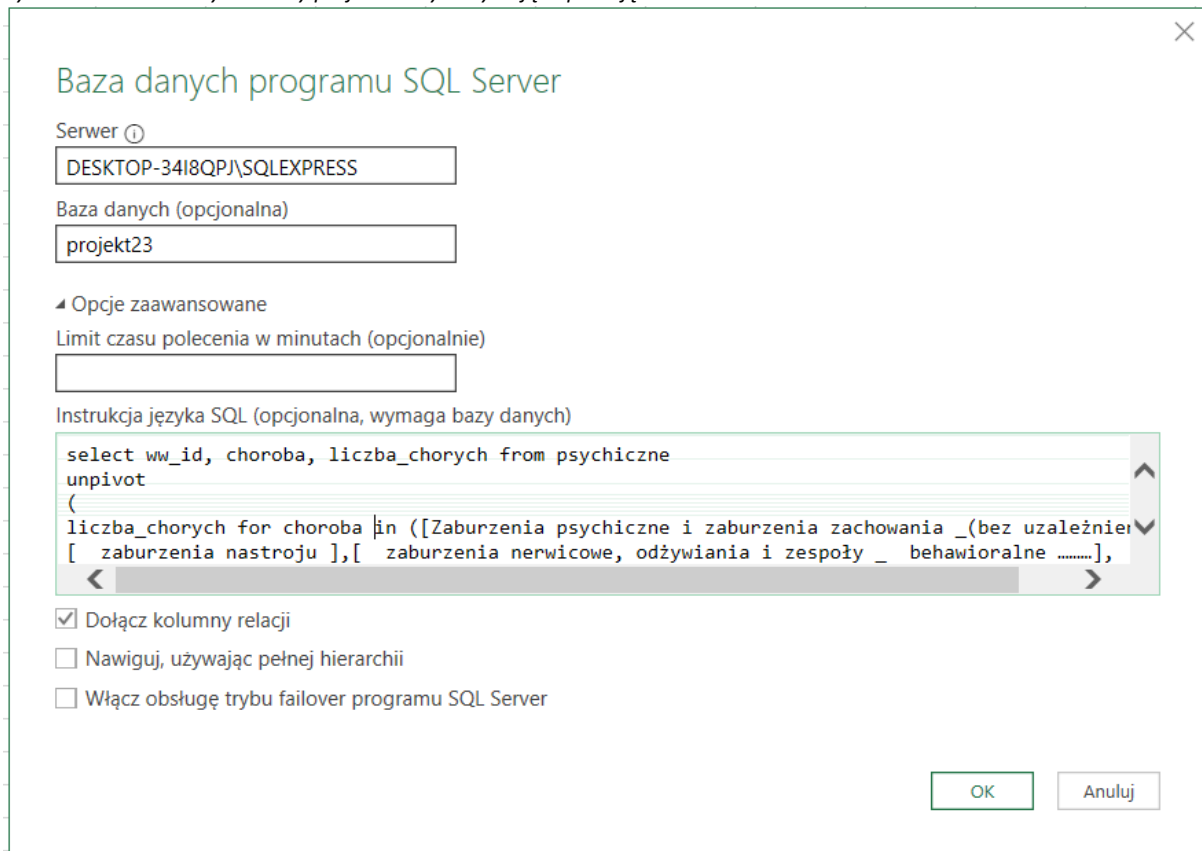
ww_id	Łęczę	Krzusiec	Odra	WZW typu B	WZW typu C	WZW typu A	Różyczka	AIDS	Salmonellozy	Czerwonka bakteryjna	Inne bakteryjne zatrucia	pokarmowe	Zapalenie jelit
1	-	0,137886470721562	0,0	2,51609959066851	3,55011312108023	0,206802706082343	0,0689342353607811	0,172335588401953	17,8884340761227	0,1	0,0344671176803906	NULL	7,3759631836
2	-	0,296445177652184	0,0	8,00401979660896	7,50994450052199	0,197630118434789	0,296445177652184	0,0988150592173946	22,3322033831312	-	0,296445177652184	NULL	22,628648560
3	-	0,195281987189502	-	1,51343540071864	3,51507576941103	0,0488204967973754	NULL	0,0978409935947508	20,3581471645055	-	0,390563974379003	NULL	20,162865177
4	0,2	0,10106808754922	-	3,94165941441959	3,73951923932115	0,10106808754922	0,30320426264766	0,404272350196681	12,2292385934556	-	4,14379158951803	NULL	9,3993321420
5	0,1	1,41380959311392	-	7,44329168139385	2,99394972659417	0,207913175457929	0,166330540366343	21,4150570721667	-	-	0,207913175457929	NULL	11,518389920

ww_id	kila	rzeżączka
1	4,13605412104867	0,792743706648983
2	2,91504424991314	0,494075296089973
3	1,95281987189502	0,485204967973754
4	4,75020011481335	0,404272350196681
5	1,66330540366343	0,322661080732686

W bazie znajdują się 3 tabele dotyczące ilości zachorowań w poszczególnych województwa na 100 tys. osób . W nagłówkach tabel znajduje się numer identyfikacyjny województwa oraz nazwy chorób. Nie jest to idealny format danych. Mając to na uwadze na tabeli **psychiczne** i **weneryczne** dokonano transformacji UNPIVOT. Tabela **zakazne** zostanie bezpośrednio zaimportowana do Power Query w celu szybkiego przetransformowania danych (scalenie kolumn oraz unpivot).

Rys. 6 Pobieranie danych z bazy projekt23 wykorzystując operację UNPIVOT.



Baza danych programu SQL Server

Server ⓘ  
DESKTOP-34I8QPJ\SQLEXPRESS

Baza danych (opcjonalna)  
projekt23

▲ Opcje zaawansowane

Limit czasu polecenia w minutach (opcjonalnie)

Instrukcja języka SQL (opcjonalna, wymaga bazy danych)

```
select ww_id, choroba, liczba_chorych from psychiczne
unpivot
(
  liczba_chorych for choroba in ([Zaburzenia psychiczne i zaburzenia zachowania _(bez uzależnień),
  [ zaburzenia nastroju ],[ zaburzenia nerwicowe, odżywiania i zespoły _ behawioralne .....],
```

☒ Dołącz kolumny relacji

☐ Nawiguj, używając pełnej hierarchii

☐ Włącz obsługę trybu failover programu SQL Server

OK Anuluj

Po pobraniu danych w Power Query zamieniono nazwy chorób na numery identyfikacyjne chorób. W związku z tym zostanie dodana kolejna tabela zawierające atrybuty:

-choroba\_id,

-nazwa\_choroby

Rys.7 Przekształcona tabela psychiczne

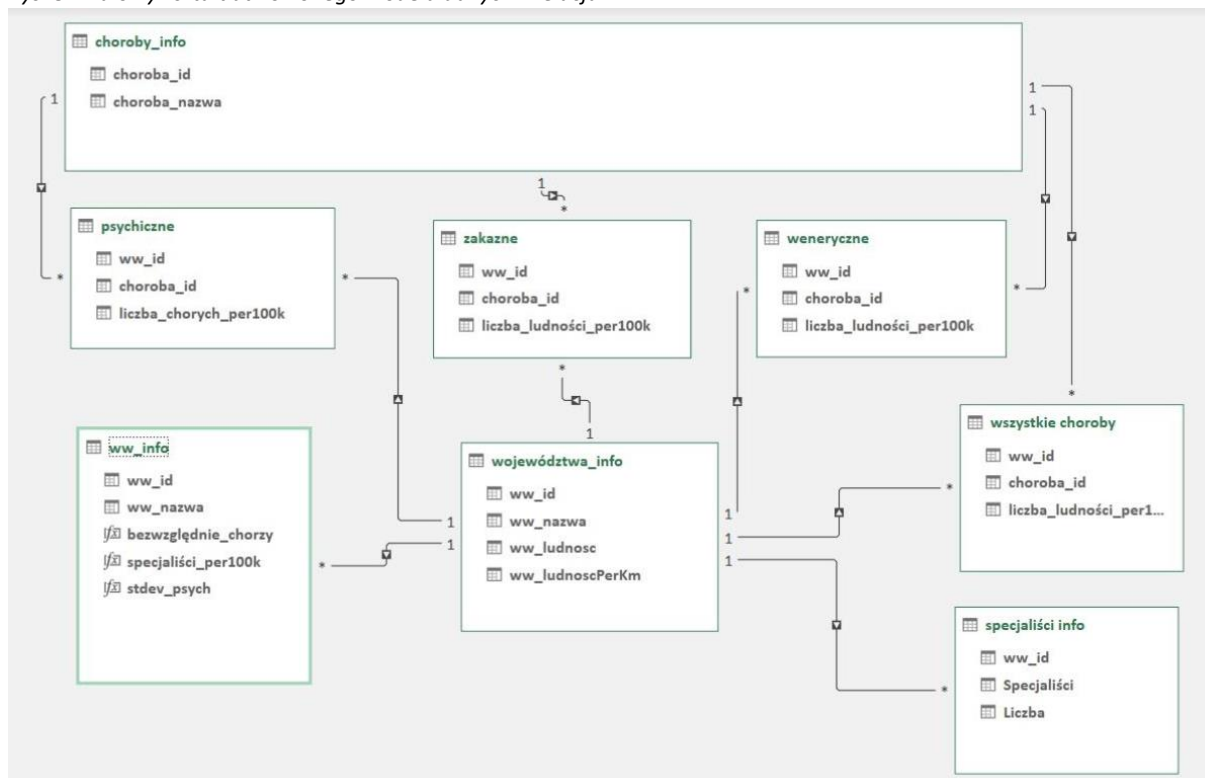
	1.2 ww_id	A <sup>B</sup> <sub>C</sub> choroba_id	1.2 liczba_chorych_per100k
1	8	1	6,35521925
2	7	1	121,9066195
3	10	1	16,46720655
4	15	1	31,13114872
5	12	1	22,49401232
6	3	1	4,116238707
7	1	1	30,38844389
8	4	1	51,98918866
9	11	1	44,48115636
10	13	1	31,95015444
11	6	1	39,95713848
12	5	1	19,41020185
13	16	1	29,58748563
14	14	1	4,101778805
15	9	1	2,237202488
16	2	1	11,77116271
17	9	2	146,1321625
18	7	2	109,0457429
19	13	2	164,730017
20	2	2	239,6412542
21	10	2	58,54048821

Tak przekształconą tabelę załadowano do modelu danych. Te same kroki przeszły tabele **zakazne** oraz **weneryczne**.

3. Z pobranego arkusza ze strony <https://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/zdrowie-i-ochrona-zdrowia-w-2021-roku,1,12.html> pobrano również dane z tabeli 8 zawierające informacje o liczbie lekarzy specjalistów w poszczególnych województwach. Dane przetransformowano w Power Query oraz dodano do modelu danych. Zastosowano również miarę, która daje informację o ilości lekarzy specjalistów na 100 tys. Osób w województwach. Wybrano takie specjalizacje lekarzy, które odpowiadają rodzajom chorób znajdujących się w bazie danych (przykład – choroby psychiczne i lekarz psychiatra).

Tabele dodawane do modelu były formatu bazodanowego z możliwością utworzenia między nimi odpowiednich relacji.

Rys. 8 Finałowy kształt utworzonego modelu danych z relacjami



## WIZUALIZACJA:

Elementy dashboardu:

1. Userforma – przygotowane w VBA narzędzie z przyciskami (W tym projekcie nie było potrzeby wykorzystywać większej ilości tego języka programowania. Pierwotnie chciałem przedstawić również wydatki służby zdrowia w różnych walutach podłączając się przez API w VBA, jednak zrezygnowałem z tego pomysłu – dashboard mógłby być zbyt szczegółowy i nudny),
2. Konceptualizacja dashboardu na osobnym arkuszu,
3. Przygotowanie tła – pobranie obrazu z internetu na licencji Creative Commons i ustawienie go jako tła. Stworzenie w Power Poincie planszy, na której zostały rozmieszczone poszczególne elementy dashboardu.
4. Fragmentator – przygotowano oryginalny styl fragmentatora. Fragmentator został połączony z tabelami przestawnymi z ukrytego arkusza na podstawie, których uzyskano dane do wizualizacji.
5. Nawigacja (górną część dashboardu obok fragmentatora). W celu obliczenia bezwzględnej liczby chorych osób w każdym województwie posłużono się następującą miarą w języku DAX:

```
=CALCULATE(sum('wszystkie choroby'[liczba_ludności_per100k. 1]) * sum('województwa_info'[ww_ludnosc]) / 100000).
```

Zastosowano również miarę w celu obliczenia ilości specjalistów na 100 tys. Mieszkańców województw

```
=DIVIDE(sum('specjaliści info'[Liczba]);SUM('województwa_info'[ww_ludnosc]))*100000
```

6. Wykresy słupkowe – dane zaczerpnięte z modelu danych,

7. Wykres mapka – niedynamiczny, utworzony „z ręki” (jeśli byłyby dostępne dane za inne lata, istniałaby możliwość utworzenia dynamicznie zmieniających wielkość kół i fragmentatora reprezentującego lata),
8. Dobór odpowiedniej kolorystyki do biznesowej wizualizacji (uzyskanie metalicznego gradientu tła, koloru fragmentatora oraz kolorów wykresów).
9. Ustawienie wyświetlania jako tryb pełnoekranowy, zablokowanie arkusza oraz skoroszytu, schowanie możliwości przełączania się po arkuszach.