

Additional data resources

Support for clustering Case study

SE labs 2018

Jerzy Stefanowski IIPP

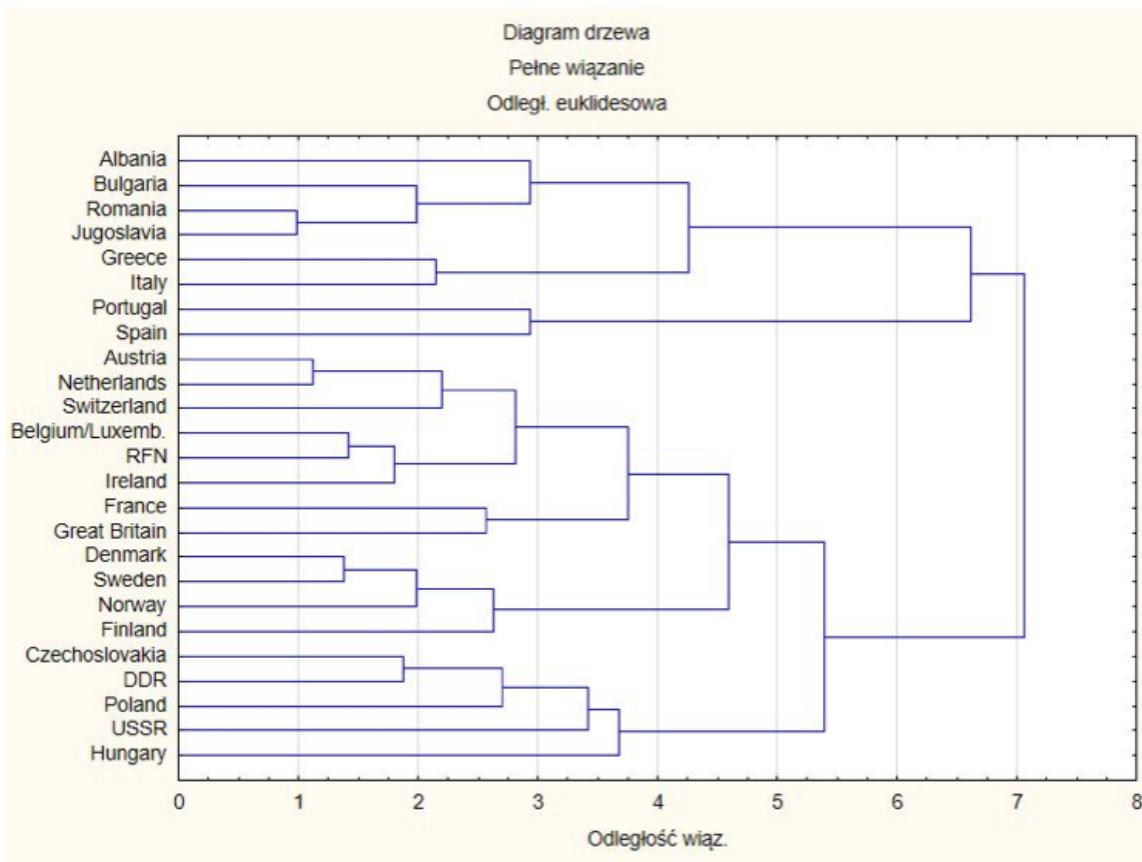
A bit more than step 1

What is described in lab instructions

The goal of this exercise is to find regions in Europe where consumptions of various products is similar. Our analysis is based on the “*proteins*” file which contains estimated protein consumption per person per day in European countries a few decades ago. We are considering 9 products:

- beef,
- pork and poultry,
- eggs,
- milk,
- fish,
- flour foods,
- starchy, foods,
- nuts, leguminous plants, seeds of oil plants,
- fruit and vegetables.

Perhaps - agglomerative, k-means



Country	Cluster no.	Distance to the center of the cluster
Denmark	1	0,44
Finland	1	0,52
Norway	1	0,36
Sweden	1	0,27
Czechoslovakia	2	0,36
DDR	2	0,65
Hungary	2	0,75
Poland	2	0,53
USSR	2	0,7
Greece	3	0,36
Italy	3	0,36
Austria	4	0,33

Analyse data structure, estimate poss. k

The additional aims od Case study 2

- Gain experiences to become a data analysts
- Try to discover a meaningful structure in your data and provide its description + interpretation
- Show that you can use appropriately at least two clustering algorithms
- Find an interesting data problem:
 - Acquire the good data set
 - Preprocess data and decide on attribute space

Past examples

- Students were able to find very interesting data problems (in Poland and outside / EU / USA and other regions)
- Other demo admin data repositories (OECD, WHO, etc)
- Specialized services – football, sport, music fans, etc.

General recommendations

Data size

- More than 4 attributes / criteria

- At least 15 examples – not too large (software limits)

Be careful with too many outliers or anomalies

Attribute characteristics

- Numerical attributes (or numbered coded)

- No redundancy and inter-correlations

You need to understand and interpret problem

GUS data services

Urzędy Statystyczne ▾ BIP Kontakt Kontrast ▾ 🔍

  Główny Urząd Statystyczny

Badania gospodarstw domowych i rolnicze • Sprawozdawczość podmiotów • Dla mediów

 ABC 123
Podstawowe dane

 Opracowania sygnalne

 Publikacje

 Bank Danych Lokalnych

 Bank Danych Makro - ekonomicznych

 Wskaźniki Zrównoważonego Rozwoju

 Dziedzinowe Bazy Wiedzy

 STRATEG

 Portal Geostatystyczny

 Portal Edukacyjny

 TERYT

 REGON

« na główna » Bazy danych

Bazy danych

- » Handel zagraniczny
- » Platforma Analityczna SWAiD - Dziedzinowe Bazy Wiedzy
- » Atlas Regionów
- » Demografia
- » SDDS
- » Baza organizacji oraz instytucji polskich i polonijnych za granicą
- » Obszary tematyczne

Bazy danych

🖨️ 🖌️ A A QR

Handel zagraniczny

Platforma Analityczna SWAiD - Dziedzinowe Bazy Wiedzy

Atlas Regionów

Demografia

<https://stat.gov.pl/banki-i-bazy-danych/>

Local GUS data

BANK DANYCH LOKALNYCH

DANE ▾

METADANE ▾

Szukaj cechy



aktualizacja na 2018.05.17

[Załącznik do danych](#)

Uprzejmie informujemy, że od 2 lutego 2018 r. Bank Danych Lokalnych dostępny jest w układzie jednostek statystycznych [NUTS 2016](#) (makroregiony, regiony i podregiony). Dane w układzie jednostek NTS 2013 (jednostki terytorialne wg 72 podregionów) przeniesione są do zasobów archiwalnych. Układ jednostek administracyjnych (kraj, województwa, powiaty i gminy) pozostał bez zmian.

maj 2018						
W	Ś	C	P	S	N	
1	2	3	4	5	6	
8	9	10	11	12	13	
15	16	17	18	19	20	
22	23	24	25	26	27	
29	30	31	1	2	3	
5	6	7	8	9	10	

aktualizowane

aktualizowane podgrupy

aktualizowane podgrupy

aktualizowane podgrupy

Dane według dziedzin



Dane roczne i krótkookresowe z wybranej dziedziny tematycznej dla wielu jednostek terytorialnych (administracyjnych i statystycznych). Informacje można prezentować w tablicach i na wykresach

Dane dla jednostki terytorialnej



Dane dla wybranej jednostki administracyjnej (miejscowości, gminy, powiatu, województwa, kraju) lub statystycznej (makroregionu, regionu, podregionu) z wielu dziedzin tematycznych.

BANK DANYCH LOKALNYCH jest największą w Polsce bazą danych o gospodarce, społeczeństwie i środowisku. BDL oferuje ponad 40 tys. cech statystycznych pogrupowanych tematycznie. Pierwsze dane pochodzą z 1995 roku.

Zapraszamy! [Więcej informacji...](#)



Atlas Regionów



Przekroje Terytorialne
Bank Danych Lokalnych



Portal
Geostatystyczny



STRATEG

<http://ec.europa.eu/eurostat/>

The screenshot shows the Eurostat website interface. At the top, there is a navigation bar with links for "Sign In | Register", "Legal notice | RSS | Cookies | Links | Contact", and a language selector set to "English". Below the navigation bar is the Eurostat logo and the tagline "Your key to European statistics". A search bar is also present. The main menu includes "News", "Data", "Publications", "About Eurostat", and "Help". Under the "Data" menu, the "EDUCATION AND TRAINING" section is expanded, showing various categories like "Overview", "Data", "Main tables", "DATABASE", "EU benchmarks", "Indicators", "Quality", "Quality reports", "Methodology", "Legislation", "Publications", "Statistics illustrated", and "Links". The "DATABASE" section is currently selected and expanded, displaying a hierarchical list of datasets: "Education and training (educ)", "Participation in education and training (educ_part)", "Learning mobility (educ_uee_mob)", "Education personnel (educ_uee_per)", "Education finance (educ_uee_fin)", "Education and training outcomes (educ_outc)", "Languages (educ_lang)", "Education-administrative data until 2012 (ISCED1997) (educ_uee_h)", and "Past series (trng_h)". At the bottom of the page, there is a footer with links for "News", "Data", "Publications", "About us", and "Opportunities", along with social media icons and the European Commission logo.

Many databases – other api and formats

Other countries

Instituto Nacional de Estadística

Instituto Nacional de Estadística

01 INE 03 Methods and Projects 05 Press
 02 INEbase 04 Training and Employment 06 Products and Services

ē Electoral Census
 :) Electronic Office

A | S | E

Participants in INE surveys

The statistical information

culture and environment

nce and technology

ography and population

onomy

stry, energy and

ction

our market

dard and life conditions

ices

ety

cal compilation

ean statistical information

mic indicators. IMF

Calendar 2018

M	T	W	T	F	S	S
1	2	3	4	5	6	
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

2018 European Statistics Competition European phase is already running!!

Quality and Good Practice

másINÉ Digital magazine

Explica

Territorial Statistics

CPI in one click

Names and surnames

How many people live in ..?

News flash

- 17 May 18. Competitiveness Guarantee Index. CGI
- 11 May 18. Consumer Price Index. CPI
- 11 May 18. Harmonised Index of Consumer Prices. HICP
- 11 May 18. Statistics on Transfer of Property Rights. STPR
- 10 May 18. Passenger Transport Statistics. PT
- 10 May 18. Mercantile Companies. MC

More

Indicator	Period	Value	Variation (%)
CPI	2018M04	103.174	1.1
EAPS. Employed (thousands)	2018Q1	18,874.2	2.36
EAPS. Unemployment rate	¹ 2018Q1	16.74	-2.01
GDP	^{2,4} 2018Q1	---	2.9
Total population (thousands)	³ 2017	46,549.0	0.21

1. Value as a %. The variation is the difference as compared with the rate for the same period the previous year
 2. Chained-linked volume measures, referring to the year 2010. Seasonal and calendar effects adjusted data
 3. Population Figures at 1st July 2017. Provisional data
 4. Preview data

EAPS. Unemployment rate (%)

Period	Value (%)
Q1/15	~23.5
Q2/15	~23.0
Q3/15	~21.5
Q4/15	~20.5
Q1/16	~20.5
Q2/16	~19.5
Q3/16	~18.5
Q4/16	~18.0
Q1/17	~18.0
Q2/17	~17.0
Q3/17	~16.5
Q4/17	~16.5
Q1/18	~16.5

Access to all our videos and Infographics

OECD – whole world development



› Français

Find, compare and share the latest OECD data: charts, maps, tables and related publications ...

Search

Browse by

topic ▾ or country ▾

Search tips

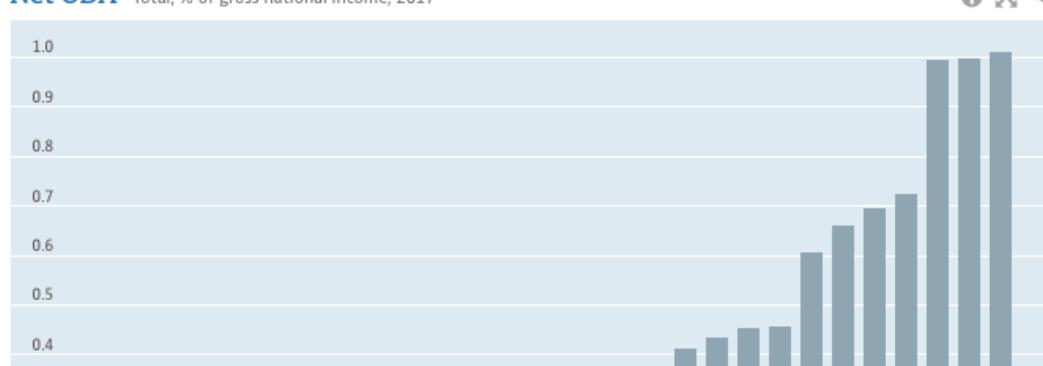
› Catalogue of OECD databases

OECD Data

Featured Charts

DAC members; Denmark, Luxembourg, Norway, Sweden and the United Kingdom; met the United Nations target in 2017 for an ODA/GNI ratio of at least 0.7%. Germany achieved the target in 2016, but joined the 24 other DAC donors under the threshold in 2017.

Net ODA Total, % of gross national income, 2017



Latest news

Statistical news releases

See recent statistical news releases.

Compare your income

Try OECD's [Compare your income](#) to check your perception of income inequality against reality!

Statistical resources

Database access

A screenshot of the OECD Database interface. The title is "Households' financial and non-financial assets and liabilities - Annual and Quarterly". The interface includes a toolbar with "Customize", "Export", and "Draw chart" options. A dropdown menu "My Databases" is open, showing "United States" selected. Below the toolbar, there are sections for "Variables" and "Periods". The main area displays a table with data for the United States, including columns for "Capital stock" and years from 2000 to 2016.

<https://data.oecd.org>

Poland admin info

Voidodships and counties



Local GUS repository

For instance, look at

Ocena poziomu życia w powiatach województwa wielkopolskiego =
Quality of life (selected criteria) in 34 powiaty (county) in Greater
Poland voivodeship

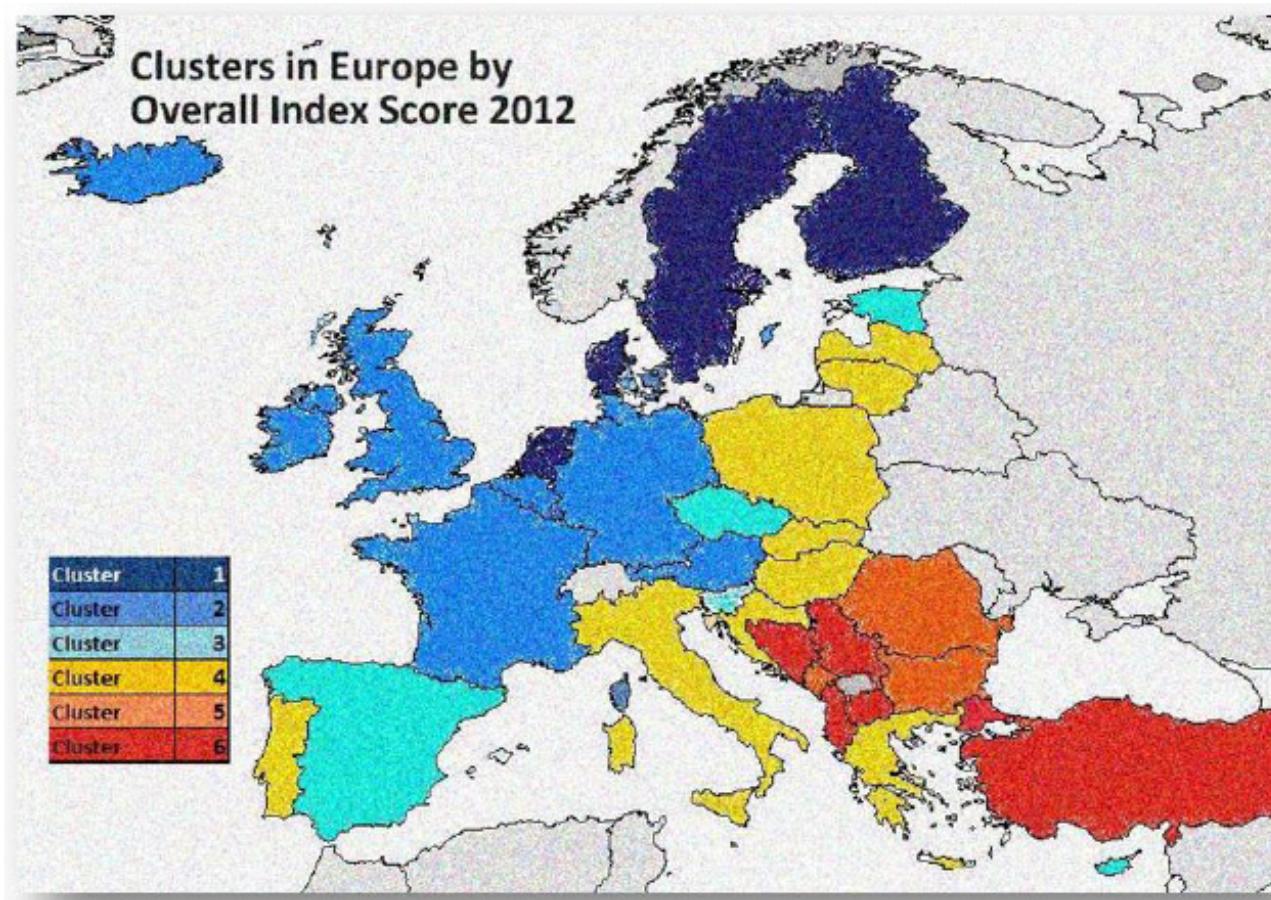
Different perspectives (sorry for Polish)

1. Ludność.
2. Dane demograficzne o przyroście naturalnym i licznie mażeństw.
3. Płace.
4. Zatrudnienie w wybranych przez siebie działach gospodarki (np. przemysł, usługi, handel, transport, ...).
5. Stopień bezrobocia
6. Bezpieczeństwie publicznym.
7. Infrastrukturze mieszkaniowej i komunalnej (np. wodociągi, drogi)
8. Liczbie placówek kulturalnych (muzea, kina itp.).
9. Ochronie zdrowia i opiece społecznej
10. Dochodach i wydatkach budżetów gmin (nakłady inwestycyjne).
11. Edukacji (różne możliwości zbudowania cech).

The Greater Poland counties (choose year)

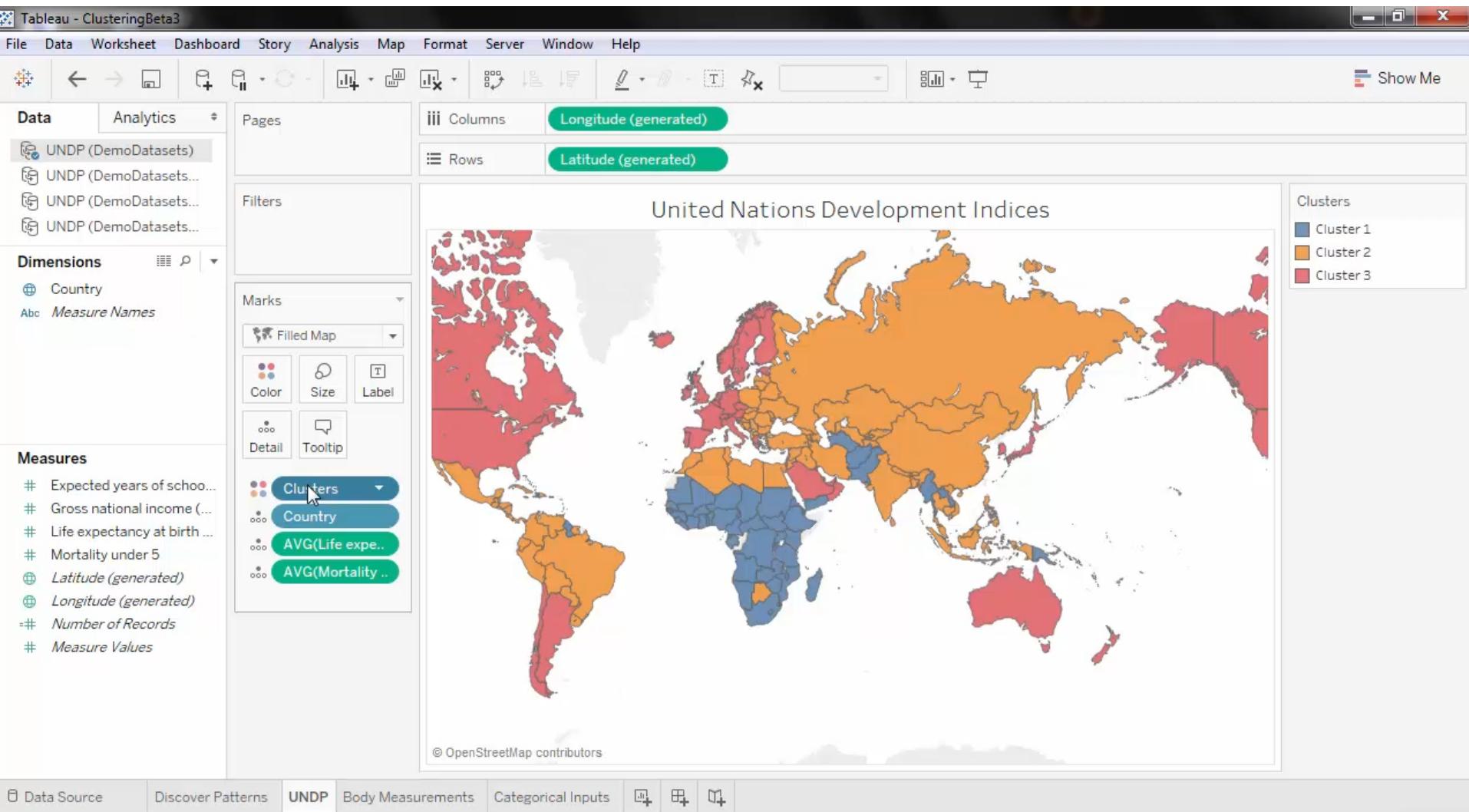
Lp	Powiaty	Ilość muzeów	Ilość kin	Ilość bibliotek	Ilość szkół podstawowych	Ilość gimnazjów	Ilość przedszkoli	Ilość łóżek przypadająca na 10 tys os / i szpitali	Nakład budżetów gmin na oświatę i wychowanie w tys. zł	Nakład budżetów gmin na ochronę zdrowia w tys. zł	Nakład budżetów gmin i kulturę i ochr. dziedzictwa narodowego w tys. zł
1	chodzieski	0	2	27	14	7	19	74.2	9202.2	455.3	74.6
2	czarnkowsko-trzcianecki	2	2	33	39	18	33	32.3	22730.2	2664.1	104.7
3	pilski	2	3	43	52	21	53	56.8	31929.5	1665.4	1541.9
4	wągrowiecki	1	1	22	27	11	27	29.6	12049.4	1391.6	156.9
5	złotowski	1	2	32	37	16	39	41.1	13364.5	780.6	29.1
6	gnieźnieński	4	1	32	56	24	59	25.4	33719.2	2295.4	181.7
7	grodziski	1	0	15	22	11	22	20.3	6617.9	529.4	61.2
8	kościński	1	2	24	26	16	19	36.9	10682.9	903.2	89.1
9	leszczyński	0	0	27	29	19	26	16.2	3349.9	7.4	27.0
10	miedzychodzki	2	0	26	10	6	12	42.7	8224.9	898.5	191.8
11	nowotomyski	2	2	34	34	11	36	22.4	11393.9	426.7	86.2
12	obornicki	2	0	31	24	12	26	74.4	12895.1	1172.7	67.4
13	poznański	9	1	72	94	49	119	21.8	18897.9	3600.6	238.8
14	szamotulski	1	1	45	34	16	42	28.2	14201.9	940.6	1015.3
15	średzki	2	1	41	25	13	31	24.3	10337.7	1167.0	117.0
16	śremski	1	1	17	25	10	22	76.9	11032.8	871.9	100.5
17	wolsztyński	1	0	17	27	14	36	39.1	11046.2	421.7	133.0
18	wrześniński	2	1	23	30	15	33	23.00	17715.6	523.4	127.9
19	Mnppleszno	1	1	18	12	10	19	79.3	61898.1	1659.8	3101.4
20	gostyński	1	2	24	34	16	39	39.00	14087.9	991.6	87.7
21	jarociński	2	1	16	33	17	34	22.9	13391.9	1221.8	191.7
22	kaliski	1	0	33	61	21	51	16.7	4320.6	16.6	546.7
23	kępiński	1	1	19	36	11	40	44.2	10299.9	699.6	138.0
24	krotoszyński	3	3	28	38	16	39	31.3	17835.0	1631.7	48.5
25	ostrowski	2	1	57	79	26	85	32.5	34038.0	1476.9	288.2
26	ostrzeszowski	1	1	14	33	10	35	33.00	12342.5	482.8	86.1
27	pleszewski	3	1	59	38	14	44	39.9	13304.6	899.5	187.4
28	rawicki	1	1	22	24	10	18	31.6	9061.0	722.2	53.0
29	Mnppkalisz	1	2	10	25	15	41	92.5	76840.1	3986.2	4982.2
30	kolski	2	2	38	52	21	54	29.9	16811.6	1424.6	126.1
31	koniński	2	1	37	87	24	91	0.00	6084.4	1792.9	99.4
32	skarpecki	1	2	16	33	11	37	39.3	11846.2	764.6	76.3
33	turecki	1	1	35	49	15	52	29.1	15601.3	1297.3	140.3
34	Mnppkonin	1	2	20	13	12	19	109.6	73556.3	1753.6	3251.1

You can use extra visualization

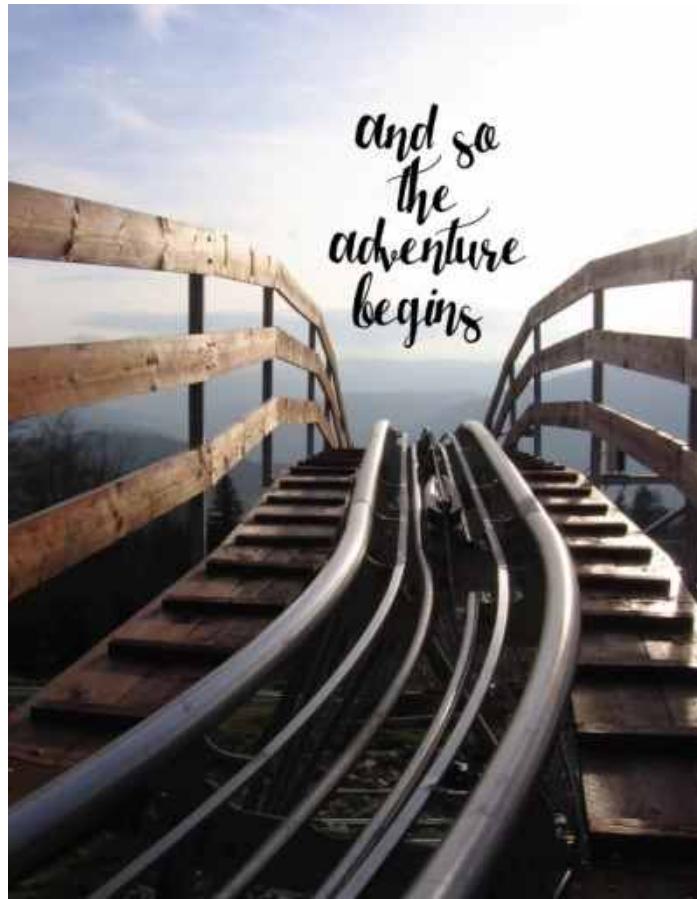


Explore and find something better!

Sophisticated tools



Have fun!



Brilliant ideas and simple implementations,...