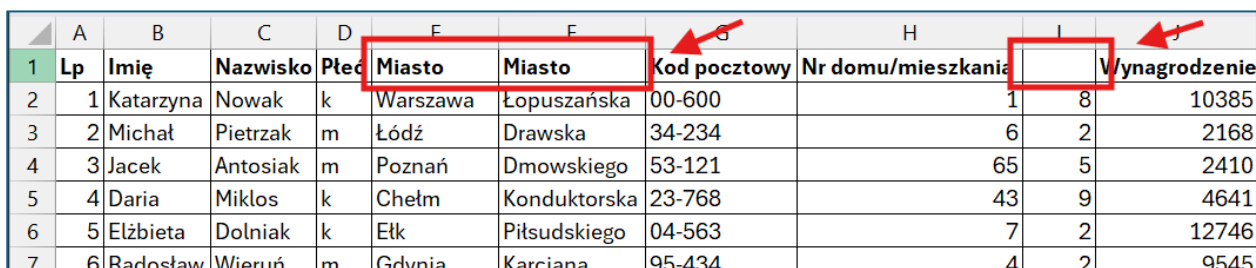


## Wskazówki dotyczące przygotowania danych do publikacji w portalu

W portalu należy udostępniać dane o możliwie najwyższym stopniu otwartości (w formatach takich jak np.: XLSX, CSV, JSON). **Nie powinno się udostępniać danych w formatach utrudniających maszynowy odczyt, takich jak PDF, JPG, PNG, TIF (graficznych).**

### Wymagania techniczne dot. ustrukturyzowania danych są następujące:

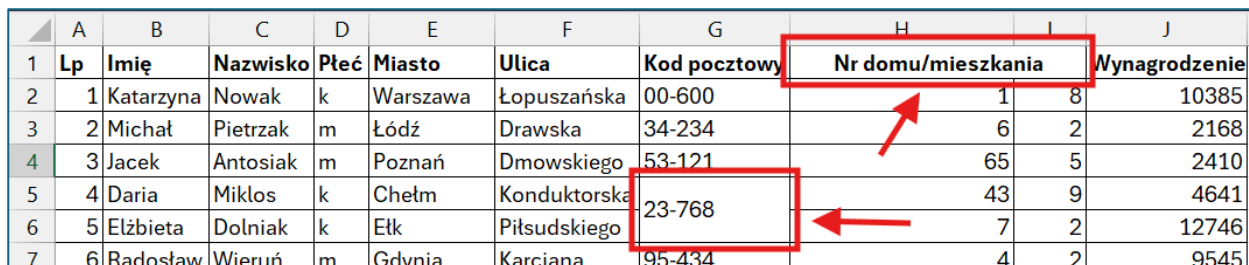
- 1) **Wiersz nagłówek** – arkusz powinien posiadać jeden wiersz nagłówek zawierający opisy wszystkich kolumn; opisy nie powtarzają się oraz kolumny nagłówek nie są puste



	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Miasto	Kod pocztowy	Nr domu/mieszkania		Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Drawska	34-234	6	2	2168
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Etł	Piłsudskiego	04-563	7	2	12746
7	6	Radosław	Wieruń	m	Gdynia	Karciana	95-434	4	2	9545

Rysunek 1. Błąd polegający na braku nagłówków i powtarzających się nagłówkach.

- 2) **Scalone komórki** – arkusz nie może zawierać scalonych komórek



	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania		Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Drawska	34-234	6	2	2168
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Etł	Piłsudskiego		7	2	12746
7	6	Radosław	Wieruń	m	Gdynia	Karciana	95-434	4	2	9545

Rysunek 2. Błąd polegający na scalaniu komórek.

- 3) **Umieszczenie treści w pliku** - treść dotycząca jednego rekordu powinna znajdować się w jednym wierszu, a nie w wielu wierszach



	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Drawska	34-234	6	2	2168
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121			
5		65		5	2410					
6	4	Elżbieta	Dolniak	k	Etł	Piłsudskiego	04-563	7	2	12746
7	5	Radosław	Wieruń	m	Gdynia	Karciana	95-434	4	2	9545
8	6	Piotr	Szewczyk	m	Tarnów	Mazowiecka	56-320	87	4	11334
9	7	Alicja	Nowak	k	Siedlce	Bracka	82-510	45	2	4680

Rysunek 3. Błąd obrazujący dane rekordu ujęte w dwóch wierszach.

- 4) **Formuły agregujące i filtrowanie danych** – dane nie powinny zawierać formuł agregujących, np. funkcji SUMA, lub funkcji filtrowania danych

	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Etł	Piłsudskiego	04-563	7	2	12746
7	6	Radostaw	Wieruń	m	Gdynia	Karciana	95-434	4	2	9545
8	7	Piotr	Szewczyk	m	Tarnów	Mazowiecka	56-320	87	4	11334
9										53229

Rysunek 4. Błąd obrazujący użycie filtrowania oraz formuł agregujących.

- 5) **Dodatkowe informacje poza obszarem danych tabeli** – nie należy umieszczać dodatkowych tytułów, opisów, przypisów i komentarzy poza obszarem danych tabeli.

	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Dawska	34-234	6	2	2168
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Etł	Piłsudskiego	04-563	7	2	12746
7										
8		H3 - do weryfikacji								
9		Tabela zawiera fikcyjne dane								
10										

Rysunek 5. Błąd polegający na umieszczeniu informacji poza obszarem danych.

- 6) **Formuły pobierające dane spoza tabeli z danymi** – nie należy umieszczać formuł pobierających dane spoza źródła danych (do innych arkuszy, plików), które będą niedostępne po przeniesieniu arkusza na inną platformę.

	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Dawska	34-234	6	2	2168
4	3	Jacek	Antosiak	m	Poznań	link	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Etł	Piłsudskiego	04-563	7	2	12746
7	6	Radostaw	Wieruń	m	Gdynia	Karciana	95-434	4	2	9545
8	7	Piotr	Szewczyk	m	Tarnów	Mazowiecka	56-320	87	4	11334

Rysunek 6. Błąd ukazujący formuły pobierające dane z zewnętrznych źródeł.

- 7) **Jednorodność danych w kolumnie** – każda kolumna tabeli powinna zawierać dane jednego rodzaju, tzn. liczbę, tekst lub ciąg cyfrowo-tekstowy o określonej strukturze i określonym znaczeniu, np. numer telefonu, kod pocztowy.

	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Drawska	34-234	6	2	2168
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Ełk	Piłsudskiego	04-563	siedem	2	dwa tysiące
7	6	Radostaw	Wieruń	m	Gdynia	Karciana	95 - 434	4	2	9545
8	7	Piotr	Szewczyk	m	Tarnów	Mazowiecka	56-320	87	IV	11334

Rysunek 7. Błąd ukazujący niejednorodne dane w kolumnach.

- 8) **Wartości umieszczane w kolumnie** – Każda kolumna tabeli powinna zawierać dane dotyczące pojedynczej wielkości (elementarnej), np. numer kodu, nazwę miejscowości, co ułatwia sortowanie i analizę danych.

	A	B	C	D	E	F	G	H	I	J		
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie		
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600		1	8	10385	
3	2	Michał	Pietrzak	m	Łódź	Drawska	34-234		6	2	2168	
4	3	Jacek	Antosiak	m	Poznań, Dmowskiego 65, 53-121					5	2410	
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768		43	9	4641	
6	5	Elżbieta	Dolniak	k	Elk	Piłsudskiego	04-563		7	2	12746	
7	6	Radosław	Wieruń	m	Gdynia, Karciana 4, 95-434		Karciana	95-434		4	2	9545
8	7	Piotr	Shewczyk	m	Tarnów	Mazowiecka	56-320		87	4	11334	

Rysunek 8. Błąd ukazujący łączenie w kolumnie wielu różnych danych.

- 9) **Format danych w kolumnie** – zapis w każdym polu w kolumnie powinien być zgodny z założonym formatem właściwym dla danego typu zawartości.

	A	B	C	D	E	F	G	H	I	J
1	Lp	Imię	Nazwisko	Płeć	Miasto	Ulica	Kod pocztowy	Nr domu/mieszkania	Status	Wynagrodzenie
2	1	Katarzyna	Nowak	k	Warszawa	Łopuszańska	00-600	1	8	10385
3	2	Michał	Pietrzak	m	Łódź	Drawska	34-234	6	2,00 zł	2168
4	3	Jacek	Antosiak	m	Poznań	Dmowskiego	53-121	65	5	2410
5	4	Daria	Miklos	k	Chełm	Konduktorska	23-768	43	9	4641
6	5	Elżbieta	Dolniak	k	Elk	Piłsudskiego	04-563	7	2	12746
7	6	Radostaw	Wieruń	m	Gdynia	Karciana	95-434	4	2	954500%
8	7	Piotr	Szewczyk	m	Tarnów	Mazowiecka	56-320	87	4	11334

Rysunek 9. Błąd ukazujący niejednolite i nieprawidłowe formaty w kolumnach.

- 10) **Liczba arkuszy w pliku** – zaleca się opracowywanie danych w postaci pojedynczych arkuszy, a nie tzw. skrószków z wieloma arkuszami.

19	18 Alicja	Miklos	k	Gdynia	Karciana	95-435		34	6	2992
20	19 Marzena	Dolniak	k	Tarnów	Mazowiecka	56-321		30	8	6497
21	20 Tomasz	Wieruń	m	Siedlce	Bracka	82-511		21	6	6316
22	21 Katarzyna	Szewczyk	k	Bełchatów	Terespolska	94-934		26	2	9461
23	22 Michał	Nowak	m	Szczecin	Bojańska	31-823		1	3	7331

Rysunek 10. Błąd ukazujący stosowanie wielu arkuszy w pliku (skoroszyście).

- 11) **Znaki specjalne i separatory w plikach CSV**– należy używać cudzysłowów do otaczania znaków specjalnych (takich jak przecinki czy znaki nowej linii wewnątrz pól).

```
Lp,Imię,Nazwisko,Płeć,Miasto,Ulica,Kod pocztowy,Nr domu/mieszkania,Status,Wynagrodzenie
1,Katarzyna,Nowak,k,Warszawa,Łopuszańska,00-600,1,"8,16,24",10385
2,Michał,Pietrzak,m,Łódź,Drawska,34-234,6,2,2168
3,Jacek,Antosiak,m,Poznań,Dmowskiego,53-121,65,5,2410
4,Daria,Miklos,k,Chełm,Konduktorska,23-768,43,9,4641
```

Rysunek 11. Właściwe oznaczanie znaków specjalnych (przecinków) poprzez ujęcie w cudzysłow.

## Rekomendacje w zakresie formatowania danych w wybranych kolumnach

- a) **Data** – zaleca się zapis dat w postaci yyyy-mm-dd, a łącznie dat i godzin w postaci yyyy-mm-ddThh:m lub yyyy-mm-dd hh:mm:ss (spacja między datą a czasem),
- b) **Numer telefonu** – w polskiej strefie numeracyjnej zaleca się zapisywanie numeru telefonu jako ciągu 9 cyfr, bez wyróżniania tzw. numeru kierunkowego miejscowości i bez spacji lub łączników. Dopuszczalne jest stosowanie prefiksu międzynarodowego, poprzedzonego znakiem plus „+”,
- c) **Nr REGON** – wymaga się zapisywania identyfikatora podmiotu jako ciągu 9 lub 14 cyfr, bez spacji lub łączników,
- d) **Nr NIP** – wymaga się zapisywania numeru identyfikacji podatkowej jako ciągu 10 cyfr, bez spacji lub łączników,
- e) **Nr KRS** – wymaga się zapisywania numeru jako ciągu 10 cyfr, bez spacji lub łączników.

Data	Nr telefonu	Nr REGON	Nr NIP	Nr KRS
2025-02-12	225970927	525189465	5252955037	0000372110

Rysunek 12. Prawidłowe formatowanie danych.

Data	Nr telefonu	Nr REGON	Nr NIP	Nr KRS
1.1.2022	(22) 5970927	525-189-465	52-52-955-037	000-037-2110
3 IV 2019	46-88-11-115	525 189 465	52 52 955 037	00 00 37 21 10
11/05/2024	+48 22 816-87-25	5 2 5 1 8 9 4 6 5	5 2 5 2 9 5 5 0 3 7	0 0 0 0 3 7 2 1 1 0

Rysunek 13. Błędne formatowanie danych.

Wymagania odnośnie udostępnianych danych zawarte są w Standardzie technicznym – załączniku do uchwały nr 28 Rady Ministrów z 18 lutego 2021 r. w sprawie Programu otwierania danych na lata 2021-2027: <https://dane.gov.pl/pl/knowledgebase/useful-materials/program-otwierania-danych-na-lata-2021-2027>.