

13. Importowanie i formatowanie danych

Informacje w naszym komputerze przechowujemy dzięki wielu różnym programom i narzędziom. Bardzo często jednak potrzebujemy przenieść dane z jednej aplikacji do innej. Jeśli jest ich bardzo dużo, zadania tego nie da się wykonać ręcznie. Niestety, zwykle każdy program ma własną strukturę i określoną specyfikę przechowywania danych. Dlatego często potrzebujemy narzędzia pośredniego, za pomocą którego ujednolicimy i dostosujemy przetwarzane informacje do naszych potrzeb. Takim narzędziem może być arkusz kalkulacyjny. W bardzo prosty sposób pozwoli on analizować duże ilości danych, a także przedstawiać je na wykresach.

CSV (ang. *Comma Separated Values*) – to jeden z najbardziej znanych formatów wymiany danych pomiędzy aplikacjami. Jest to zwykły plik tekstowy, w którym wartości są rozdzielane przecinkiem (średnikiem, tabulatorem lub innym znakiem), a wszystkie wpisy dotyczące jednego obiektu są umieszczane w jednym wierszu. Pierwszy wiersz może zawierać nagłówki.

13.1. Wartości rozdzielane przecinkiem

Załóżmy, że pewna spółdzielnia mieszkaniowa używa specjalnego programu do wystawiania faktur za zużytą wodę. Z tego programu dane dotyczące wszystkich mieszkańców wyeksportowano do pliku **V.13.1.Import.csv**, który znajdziesz na płycie dołączonej do podręcznika. Kliknij go prawym przyciskiem myszy i wybierz **Otwórz za pomocą → Notatnik**. Teraz możesz zobaczyć strukturę pliku .csv (zob. ryc. 1). Ponownie otwórz plik - tym razem jednak skorzystaj z Microsoft Excel. Arkusz powinien automatycznie i bez problemu rozpoznać zawartość i w pierwszym wierszu wyświetlić nagłówki, a w kolejnych dane dotyczące poszczególnych osób. Zauważ, że każda wartość rozdzielona średnikiem znalazła się w osobnej kolumnie (zob. ryc. 2). Teraz możesz dowolnie zarządzać dostępnymi danymi - zastosować je do korespondencji seryjnej lub podsumować zużycie wody przez poszczególne rodziny lub po prostu sformatować dane i wydrukować zestawienie.

5.1. Import.csv - Notatnik						
Plik	Edycja	Format	Widok	Pomo		
Imię;Nazwisko;Płeć;Adres;Zużycie wody;Wartość	Dorota;Nowak;K;ul. Wymyślona 13 m.1;18;72					
Łukasz;Kowalski;M;ul. Wymyślona 13 m.2;13;52						
Artur;Wiśniewski;M;ul. Wymyślona 13 m.3;20;80						
Krzysztof;Dąbrowski;M;ul. Wymyślona 13 m.4;23;92						
Sebastian;Lewandowski;M;ul. Wymyślona 13 m.5;17;68						
Paulina;Wójcik;K;ul. Wymyślona 13 m.6;13;52						
Mateusz;Kamiński;M;ul. Wymyślona 13 m.7;13;52						
Marcin;Kowalczyk;M;ul. Wymyślona 13 m.8;17;68						
Adrian;Zieliński;M;ul. Wymyślona 13 m.9;11;44						
Jarosław;Szymański;M;ul. Wymyślona 13 m.10;7;28						
Sebastian;Woźniak;M;ul. Wymyślona 13 m.11;14;56						
Adam;Kozłowski;M;ul. Wymyślona 13 m.12;16;64						
Michał;Jankowski;M;ul. Wymyślona 13 m.13;0;0						
Krzysztof;Wojciechowski;M;ul. Wymyślona 13 m.14;0;0						
Adam;Kwiatkowski;M;ul. Wymyślona 13 m.15;19;76						
Magdalena;Kaczmarek;K;ul. Wymyślona 13 m.16;5;20						
Karolina;Mazur;K;ul. Wymyślona 13 m.17;0;0						
Anna;Krawczyk;K;ul. Wymyślona 13 m.18;11;44						

Ryc. 1. Struktura pliku import.csv

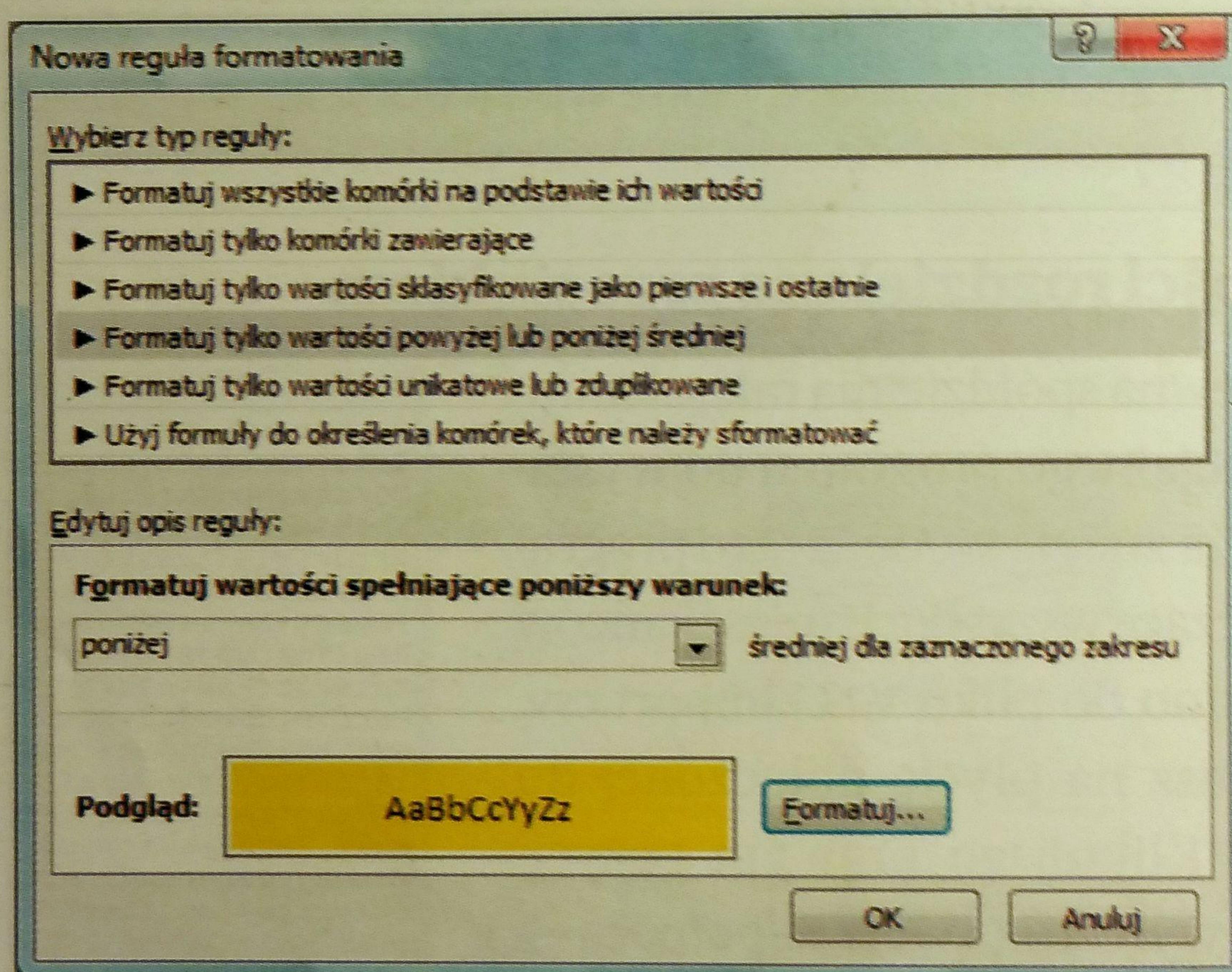
A	B	C	D	E	F
1 Imię	Nazwisko	Płeć	Adres	Zużycie w	Wartość
2 Dorota	Nowak	K	ul. Wymy:	18	72
3 Łukasz	Kowalski	M	ul. Wymy:	13	52
4 Artur	Wiśniews	M	ul. Wymy:	20	80
5 Krzysztof	Dąbrowsk	M	ul. Wymy:	23	92
6 Sebastian	Lewandow	M	ul. Wymy:	17	68
7 Paulina	Wójcik	K	ul. Wymy:	13	52
8 Mateusz	Kamiński	M	ul. Wymy:	13	52
9 Marcin	Kowalczyk	M	ul. Wymy:	17	68
10 Adrian	Zieliński	M	ul. Wymy:	11	44
11 Jarosław	Szymańsk	M	ul. Wymy:	7	28
12 Sebastian	Woźniak	M	ul. Wymy:	14	56
13 Adam	Kozłowski	M	ul. Wymy:	16	64
14 Michał	Jankowski	M	ul. Wymy:	0	0
15 Krzysztof	Wojciech	M	ul. Wymy:	0	0
16 Adam	Kwiatkow	M	ul. Wymy:	19	76

Ryc. 2. Import.csv otwarty w Excelu

13.2. Formatowanie warunkowe

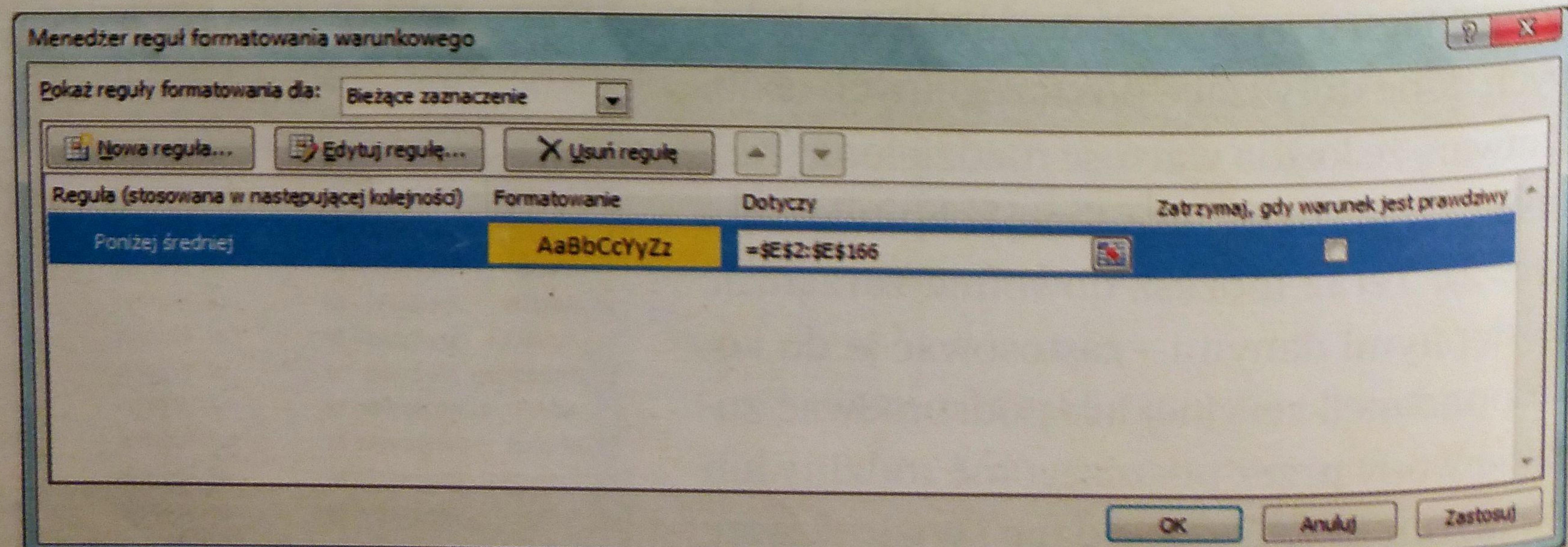
Formatowanie warunkowe stosuje się wtedy, gdy poszczególne elementy komórek (kolory, obramowania i czcionki) mają być im nadane nie na stałe, lecz w zależności od ich zawartości, na przykład w arkuszu ocen ucznia tło każdej komórki z wpisaną oceną niedostateczną automatycznie stanie się czerwone, a tło każdej piątki przyjmie na przykład barwę niebieską.

Powiedzmy, że chcielibyśmy, aby na wydruku żółte tło miały te komórki, w których jest wpisane zużycie wody niższe od średniego zużycia wody w całej spółdzielni. W tym celu zaznacz komórki E2:E166 i wybierz Narzędzia główne → Style → Formatow. warunk. → Zarządzaj regułami → Nowa reguła... → Formatuj tylko wartości powyżej lub poniżej średniej. Następnie z listy Formatuj wartości spełniające poniższy warunek wybierz poniżej i kliknij przycisk Formatuj.... W nowym oknie przejdź do zakładki Wypełnienie i wskaż kolor żółty. Naciśnij OK. Sprawdź jeszcze raz ustawienia (zob. ryc. 3) i ponownie zaakceptuj OK.



Ryc. 3. Reguła formatowania: „poniżej średniej”

W widocznym menedżerze reguł formatowania warunkowego (zob. ryc. 4) naciśnij OK. Czy widzisz różnice w formatowaniu komórek?



Ryc. 4. Menedżer reguł formatowania warunkowego

13.3. Import pliku tekstu

Import tekstu z pliku .txt jest możliwy pod warunkiem, że został on uprzednio odpowiednio przygotowany, to znaczy, że zawiera logiczną strukturę, analogiczną jak .csv. W przypadku plików tekstowych (tak zwanych rozdzielanych) separatorem jest zazwyczaj tabulator, średnik lub przecinek.

Istnieje również możliwość importu danych z pliku .txt, w którym dane zawarte są w polach o stałej szerokości.

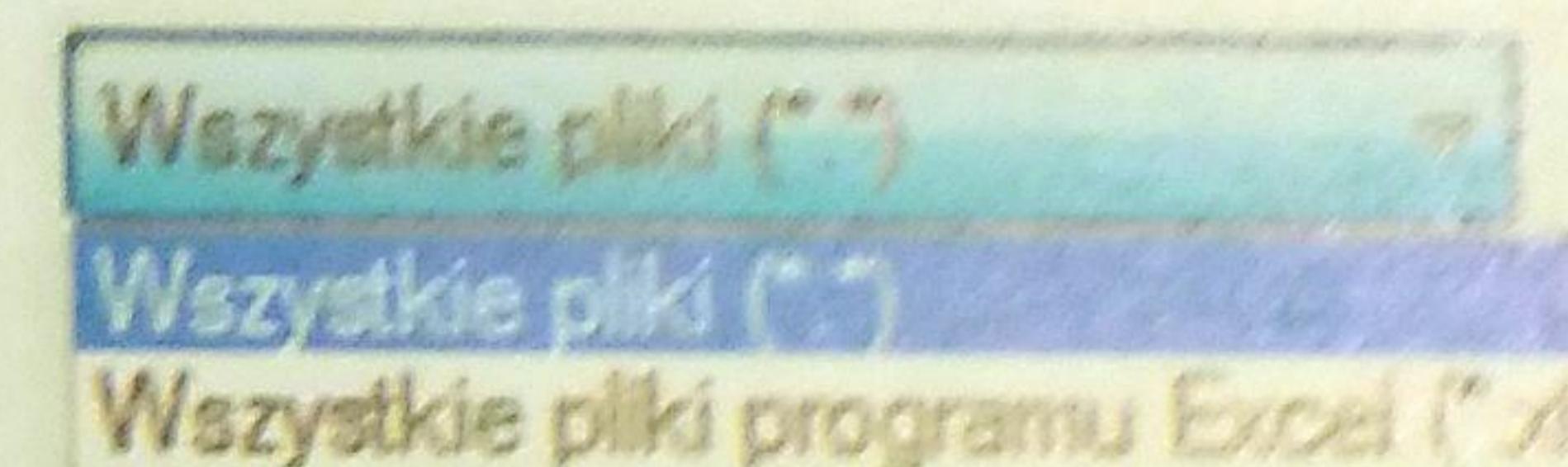


Załóżmy, że Twoja szkoła dysponuje specjalnym systemem, który po zakończeniu sprawdzianu zapisuje wyniki osiągnięte przez wszystkich uczniów do jednego pliku według zasady: każdy uczeń w osobnym wierszu, poszczególne dane dotyczące jednego ucznia rozdzielane kropką. Przeanalizowanie rezultatów będzie znacznie łatwiejsze po zimportowaniu pliku do arkusza kalkulacyjnego.

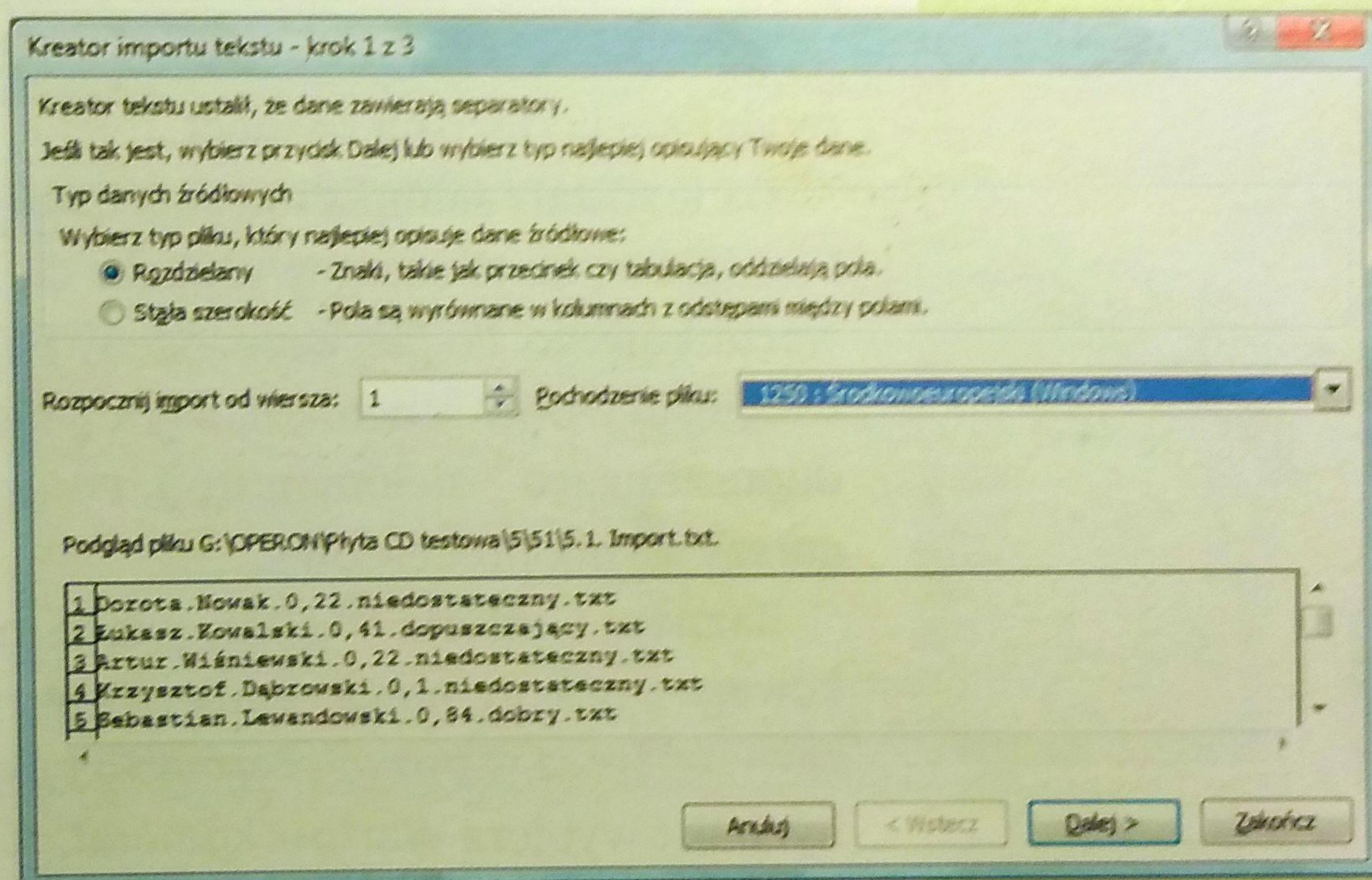
Z górnego menu Excela wybierz Plik → Otwórz i wskaż folder, w którym jest zapisany plik V.13.3.Import.txt. Okazuje się, że ten plik początkowo nie jest widoczny w oknie dialogowym. Musisz zmienić filtrowanie z Wszystkie pliki programu Excel na Wszystkie pliki (*.*) (zob. ryc. 5).

Teraz zaznacz omawiany plik i naciśnij Otwórz. Będziesz musiał najpierw odpowiedzieć na kilka pytań. W kroku pierwszym wybierz Typ pliku: Rozdzielany, a Pochodzenie pliku: 1250 Środkowoeuropejski (Windows) (zob. ryc. 6).

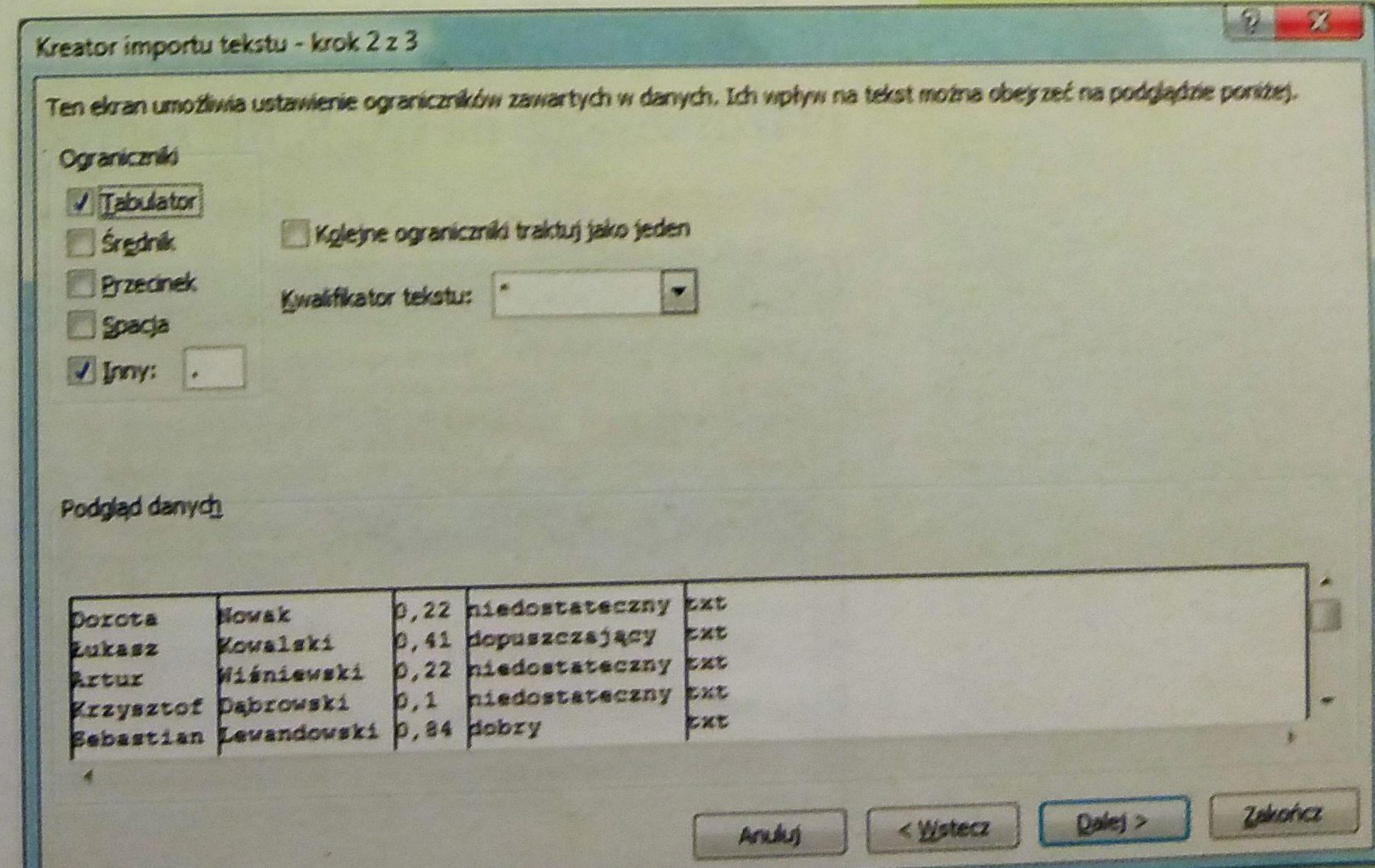
Kliknij Dalej. W kroku drugim w pozycji Ograniczniki zaznacz Inny i w polu znajdującym się tuż obok wpisz kropkę (zob. ryc. 7). Kliknij Dalej i przejdź do ostatniego kroku, w którym możesz szczegółowo określić format danych (ogólny, tekstowy, daty i czasu).



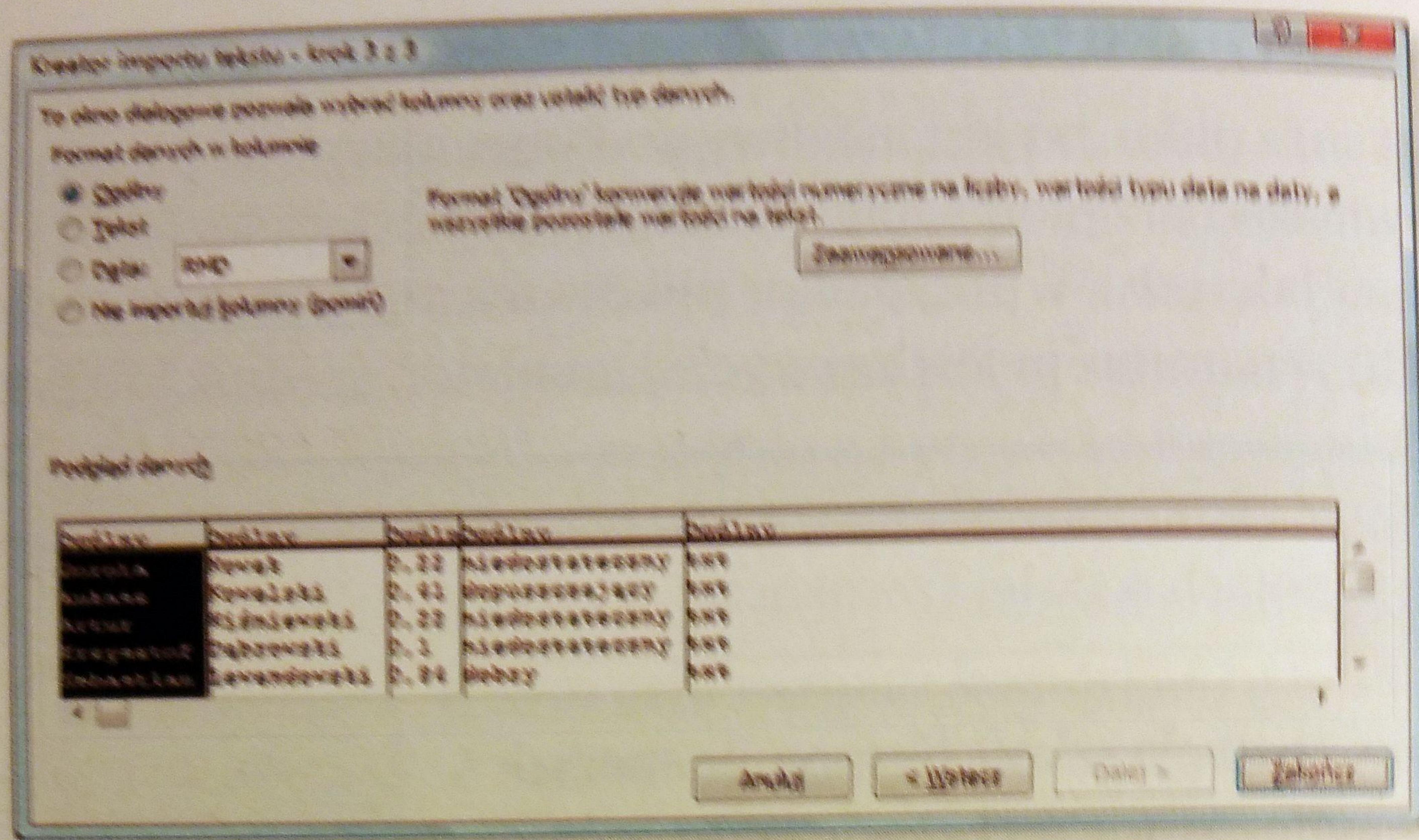
Ryc. 5. Filtr „wszystkie pliki” w oknie otwierania plików



Ryc. 6. Kreator importu - krok 1



Ryc. 7. Kreator importu - krok 2

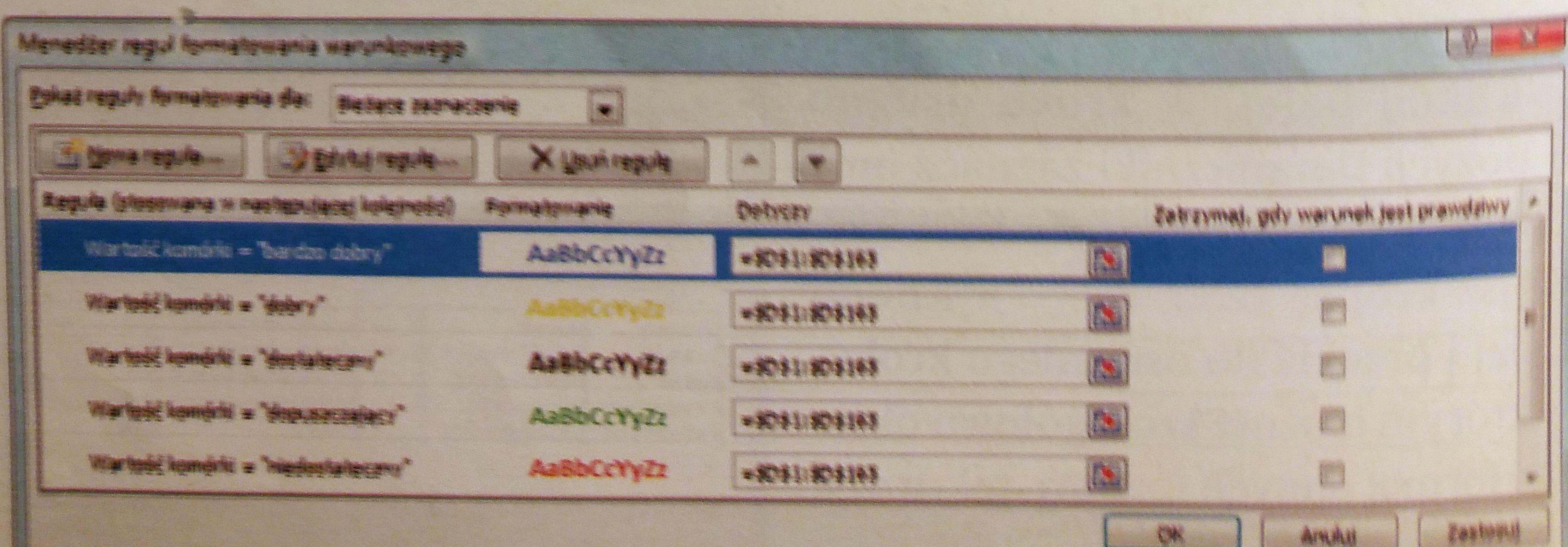


Ryc. 8. Kreator importu - krok 3

Krok ten (zob. ryc. 8) jest opcjonalny i przydatny bardzo rzadko - zazwyczaj program bardzo inteligentnie i bez problemu rozpoznaje typy pól. Gdyby jednak zaszła taka potrzeba, to w tym oknie można dokonać końcowych modyfikacji. Naciśnij **Zakończ**.

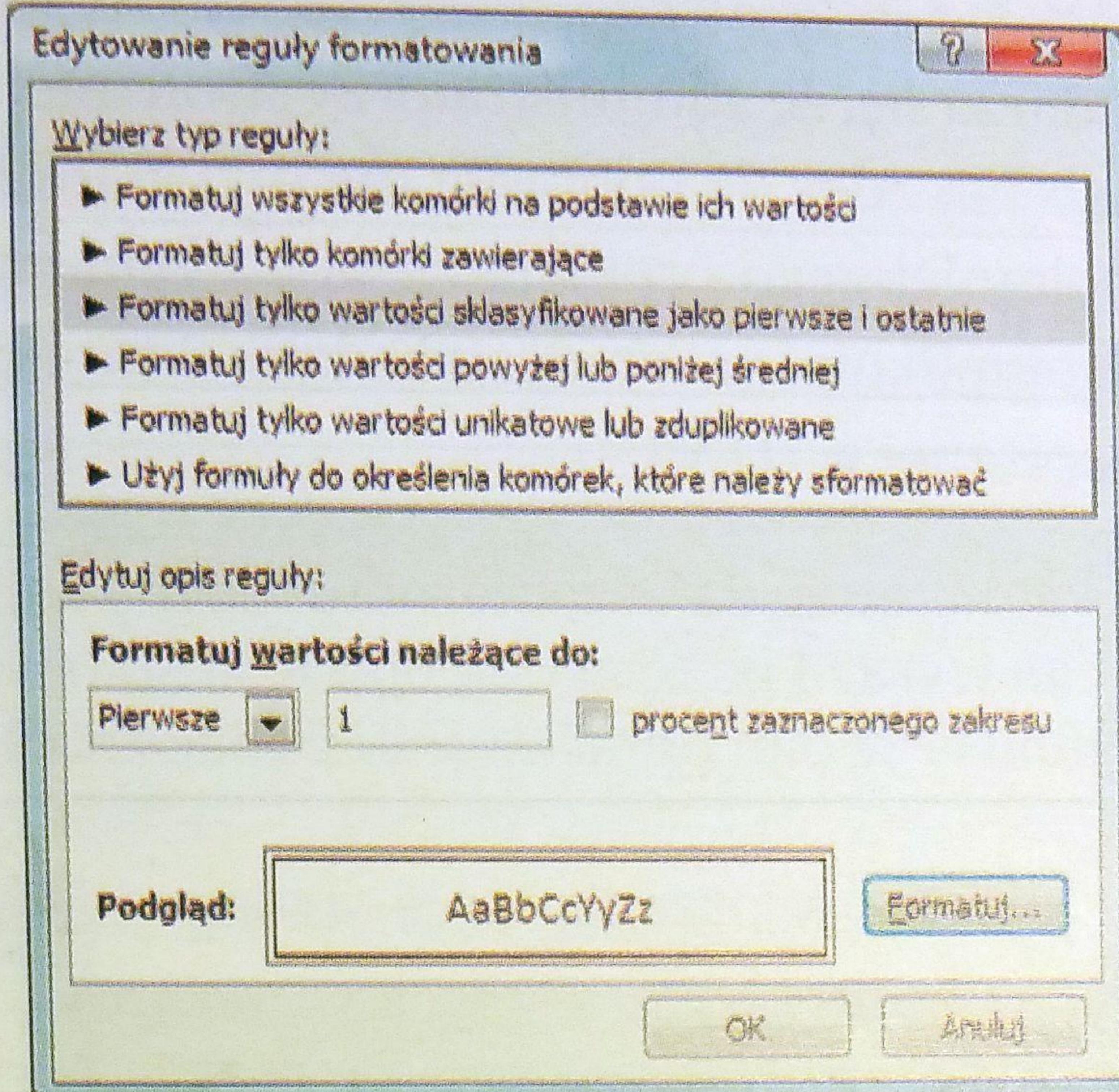
Dane zostały zaimportowane. Usuń zbędną kolumnę **E**, a liczbom z kolumny **C** nadaj format procentowy. Czeka cię teraz zadanie dotyczące formatowania warunkowego. Po pierwsze, oceny powinny być zapisane kolorową czcionką według pewnego schematu (na przykład niedostateczne kolorem czerwonym, dopuszczające - zielonym itp.). Po drugie, należy wyróżnić najlepszy i najgorszy wynik procentowy.

Zaznacz komórki **D1:D165**, włącz menedżera reguł formatowania warunkowego i utwórz nową regułę. Tym razem wybierz **Formatuj tylko komórki zawierające**, a w pozycji **Formatuj tylko komórki spełniające warunek z kolejnych list** wybierz **Wartość komórki, równa** i na końcu wpisz **niedostateczny**. Skorzystaj z przycisku **Formatuj** w celu ustawienia czerwonej czcionki. Naciśnij **OK**. W oknie menedżera reguł ponownie wybierz **Nowa reguła...** i postępuj analogicznie, tworząc regułę dla oceny dopuszczającej, potem nową regułę dla dostatecznej, a następnie dla dobrej i bardzo dobrej. Ostateczny efekt przedstawiono na ryc. 9. Nie zamkaj jeszcze okna menedżera.

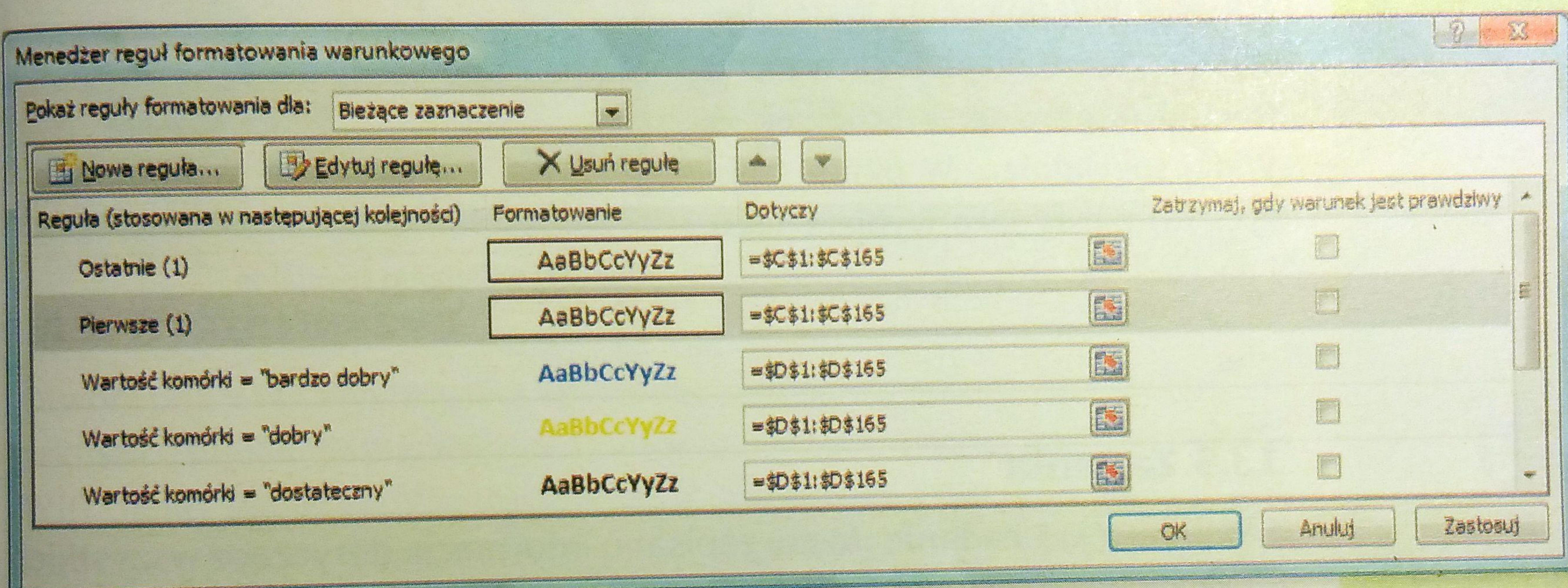


Ryc. 9. Menedżer reguł - formatowanie dla ocen

Dodaj kolejną nową regułę (typ: Formatuj tylko wartości sklasyfikowane jako pierwsze i ostatnie). Z listy Formatuj tylko wartości należące do wybierz Pierwsze, a w polu obok wpisz 1 (zob. ryc. 10). Założymy, że wyróżnienie będzie polegało na obramowaniu komórki konturem. Ustaw takie właśnie formatowanie. Żeby stworzyć regułę dla najgorszego wyniku - postępuj analogicznie, ale z jedną tylko różnicą: wartość Pierwsze zamień na Ostatnie.



Ryc. 10. Reguła formatowania „pierwsze i ostatnie”



Ryc. 11. Menedżer reguł formatowania warunkowego po zmianie adresowania

Dwie ostatnio dodane reguły nie będą jeszcze niestety działać poprawnie, zostały bowiem zdefiniowane dla adresów **\$D\$1:\$D\$165**, a wyniki uczniów są przechowywane w **\$C\$1:\$C\$165**. Trzeba więc odpowiednio zmienić wpisy w kolumnie **Dotyczy** menedżera reguł. Po tej operacji uzyskasz taki rezultat, jak to przedstawiono na ryc. 11.

Jeżeli wszystko się zgadza, naciśnij **OK** i sprawdź działanie formatowania warunkowego (zob. ryc. 12).

	A	B	C	D
139	Justyna	Piątek	53%	dostateczny
140	Agnieszka	Jóźwiak	17%	niedostateczny
141	Marta	Urban	33%	dopuszczający
142	Marcin	Dobrowolski	83%	dobry
143	Agnieszka	Pawlik	23%	niedostateczny
144	Sandra	Kruk	99%	bardzo dobry
145	Katarzyna	Domagała	48%	dopuszczający
146	Rafał	Piąsek	17%	niedostateczny
147	Piotr	Wierzbicki	95%	bardzo dobry
148	Jarosław	Karpinski	32%	dopuszczający
149	Mariusz	Jastrzębski	36%	dopuszczający
150	Kamila	Polak	99%	bardzo dobry
151	Małgorzata	Ignaczak	46%	dopuszczający
152	Dominika	Franc	15%	niedostateczny
153	Mariusz	Ignaczewski	85%	bardzo dobry
154	Joanna	Erazmus	14%	niedostateczny

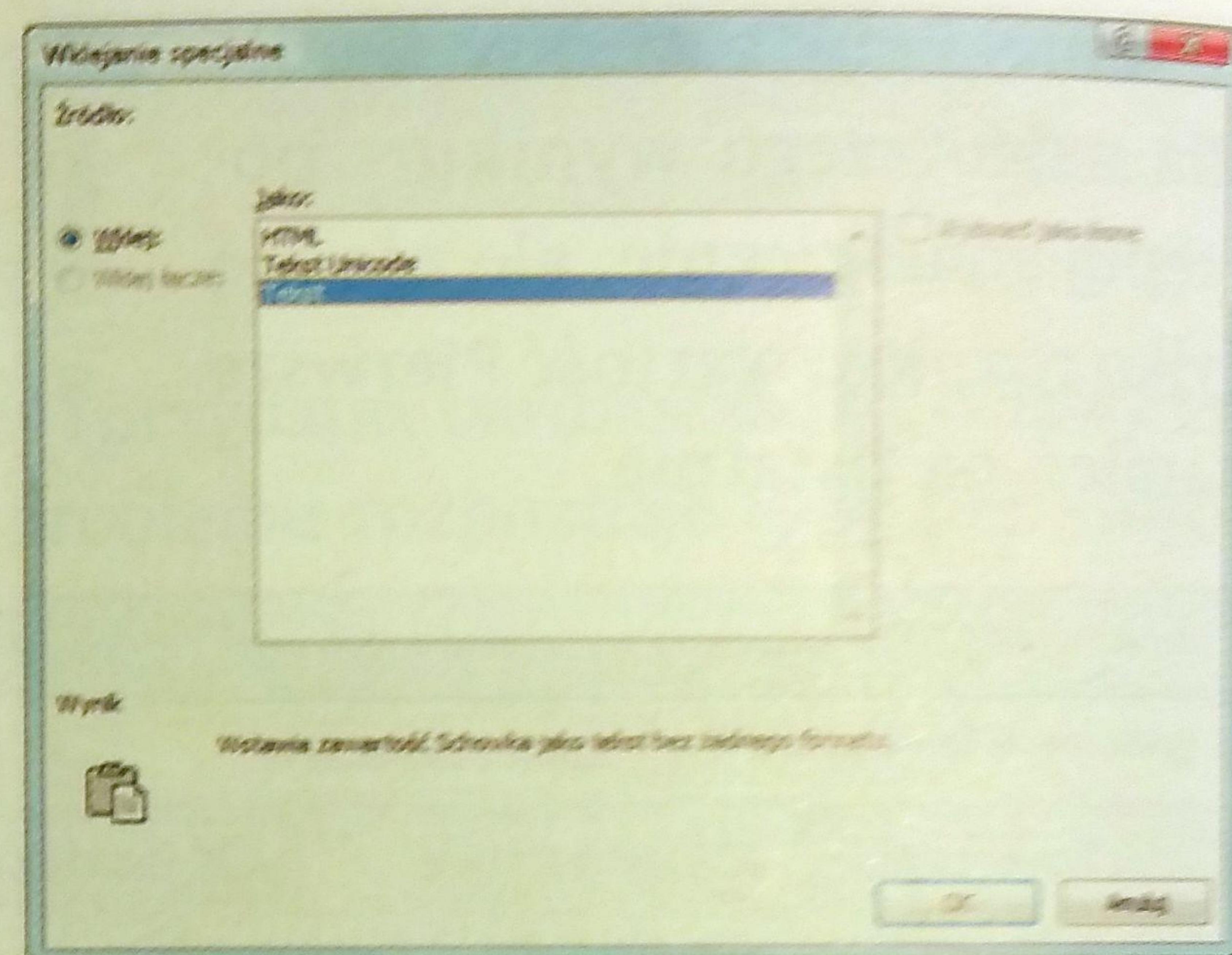
Ryc. 12. Efekt zastosowania formatowania warunkowego

13.4. Pobieranie danych z internetu

Zdarza się, że administratorzy stron internetowych udostępniają gotowe pliki .xls lub .xlsx do pobrania (na przykład GUS, NBP). Wtedy oczywiście nie ma najmniejszego problemu z otworzeniem ich w Excelu. Staje się to bardziej skomplikowane wtedy, gdy dane, które cię interesują, są fragmentem strony zapisanej w HTML-u.

Pamiętaj, że dane z internetu można kopiować tylko wtedy, gdy ich właściciel wyraził na to zgodę (lub jeśli stanowią własność publiczną). Nie zapomnij w swoim opracowaniu podać źródła tych danych.

Otwórz w przeglądarce stronę <http://www.nbp.pl/home.aspx?f=/kursy/kursya.html>, zaznacz tabelę z kursami walut tam się znajdująącą i naciśnij [Ctrl]+[C]. W nowym arkuszu użyj polecenia **Wklej specjalnie...**, które znajduje się w menu **Narzędzia główne → Schowek → Wklej specjalnie...**. W oknie (zob. ryc. 13) z listy **Jako** wybierz **Tekst** i naciśnij **OK**.



Ryc. 13. Wklejanie specjalne

13.5. Zadania

1. W pliku **V.13.5.Zadanie_M.csv** zapisano informacje dotyczące wszystkich pracowników pewnej firmy (każdej osoby w innym wierszu). Znajdują się tam następujące dane (rozdzielone średnikiem): imię, nazwisko, wydajność produkcji (w sztukach), premia (w złotych). Pierwszy wiersz zawiera nagłówki. Zaimportuj ten plik do Excela, a następnie wyróżnij:

- komórki z wydajnością wyższą niż średnia wydajność tego zakładu - żółte wypełnienie;
 - komórki z premią od 1 zł do 499 zł - czcionka pogrubiona, niebieska;
 - komórki z premią co najmniej 500 zł - czcionka pochylona, zielona.
- Zapisz plik pod nazwą **V.13.5.Zadanie_M.xlsx**.

2. W pliku **V.13.5.Zadanie_K.txt** zapisano informacje dotyczące wyników użytkanych przez dziewczęta w szkolnych mistrzostwach w biegach (każda zawodniczka w osobnym wierszu). Znajdują się tam następujące dane (rozdzielone znakiem #): imię, nazwisko, płeć, wynik. Pierwszy wiersz zawiera nagłówki. Zaimportuj ten plik do Excela, a następnie wyróżnij:

- komórki z siedmioma najlepszymi (najniższymi) wynikami - biała czcionka na niebieskim tle;
- komórki z siedmioma najsłabszymi (najwyższymi) wynikami - niebieska czcionka na białym tle.

Zapisz plik pod nazwą **V.13.5.Zadanie_K.xlsx**.