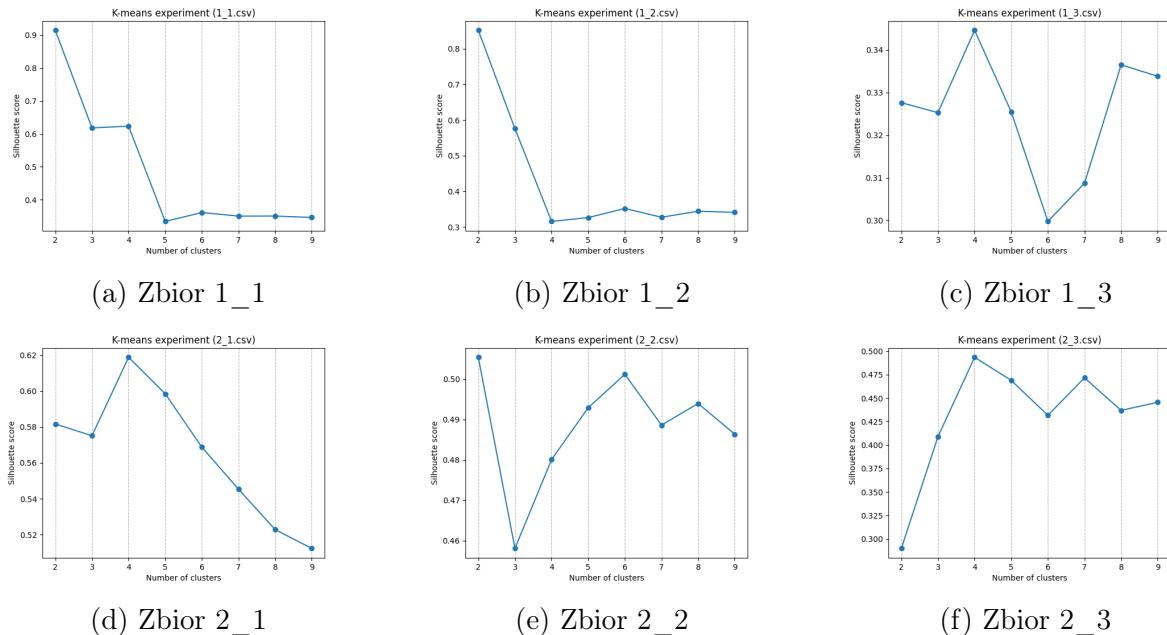
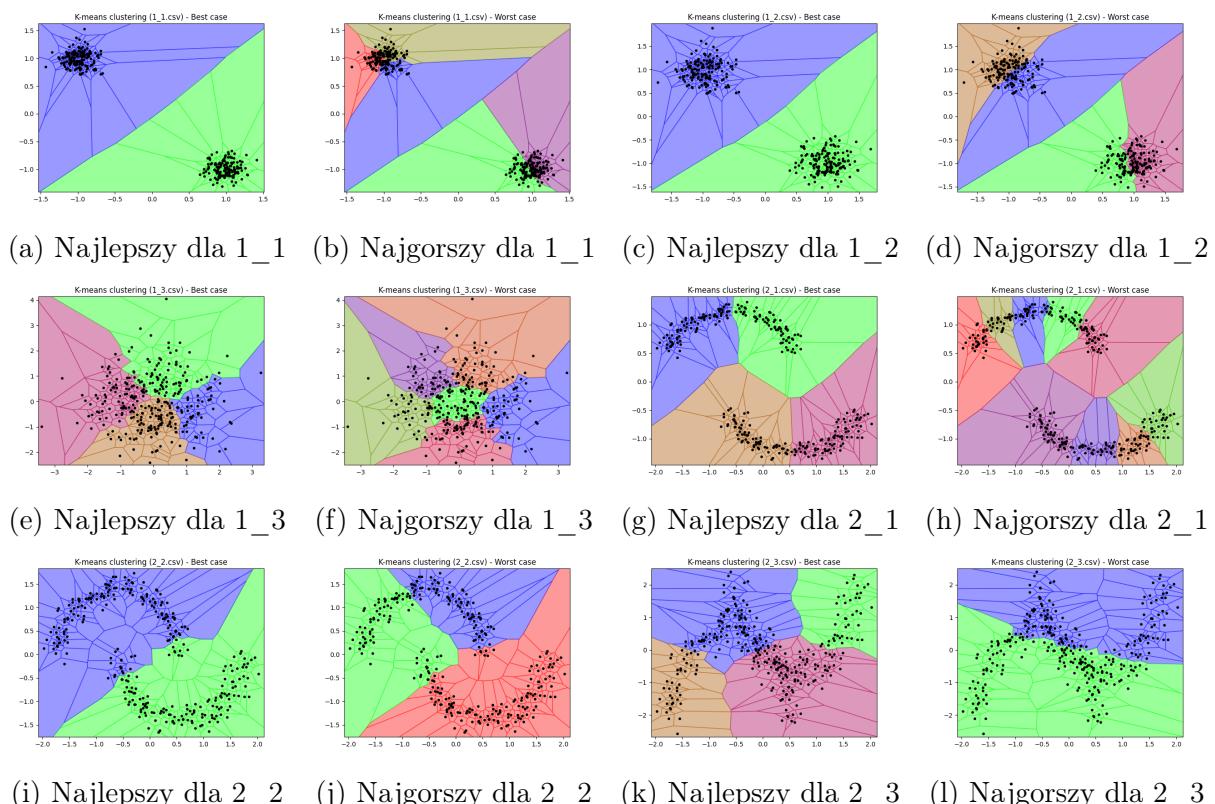


1 Eksperyment 1: Metoda K-Means

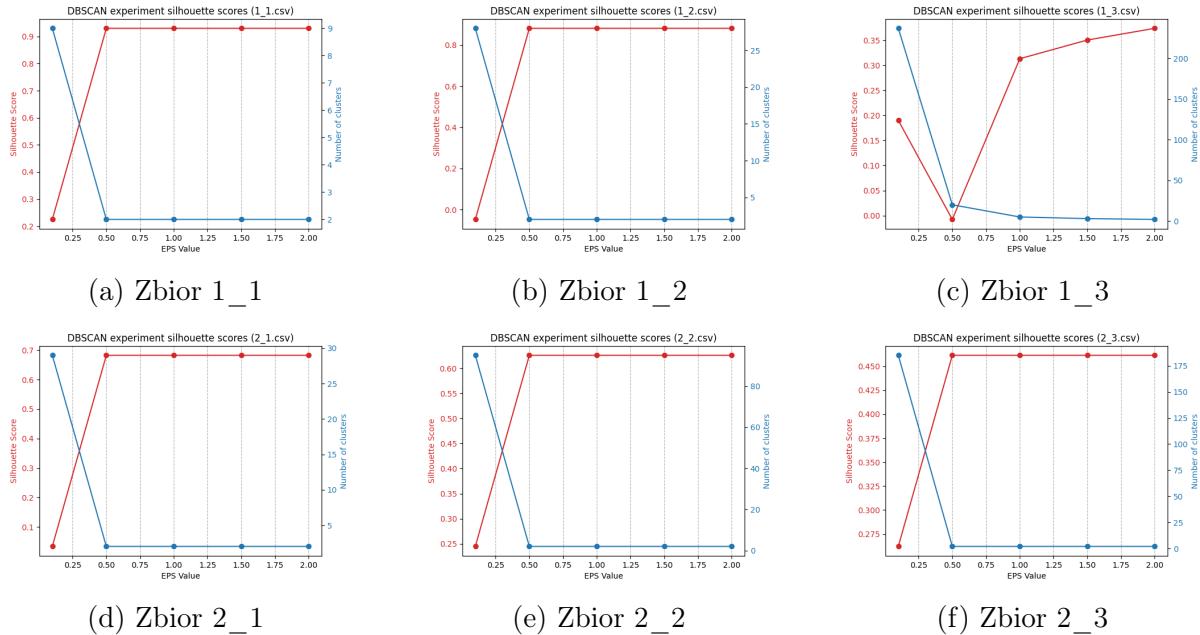


Rysunek 1: Zmiana wartości silhouette score dla wszystkich zbiorów w zależności od parametru n_clusters w metodzie K-means

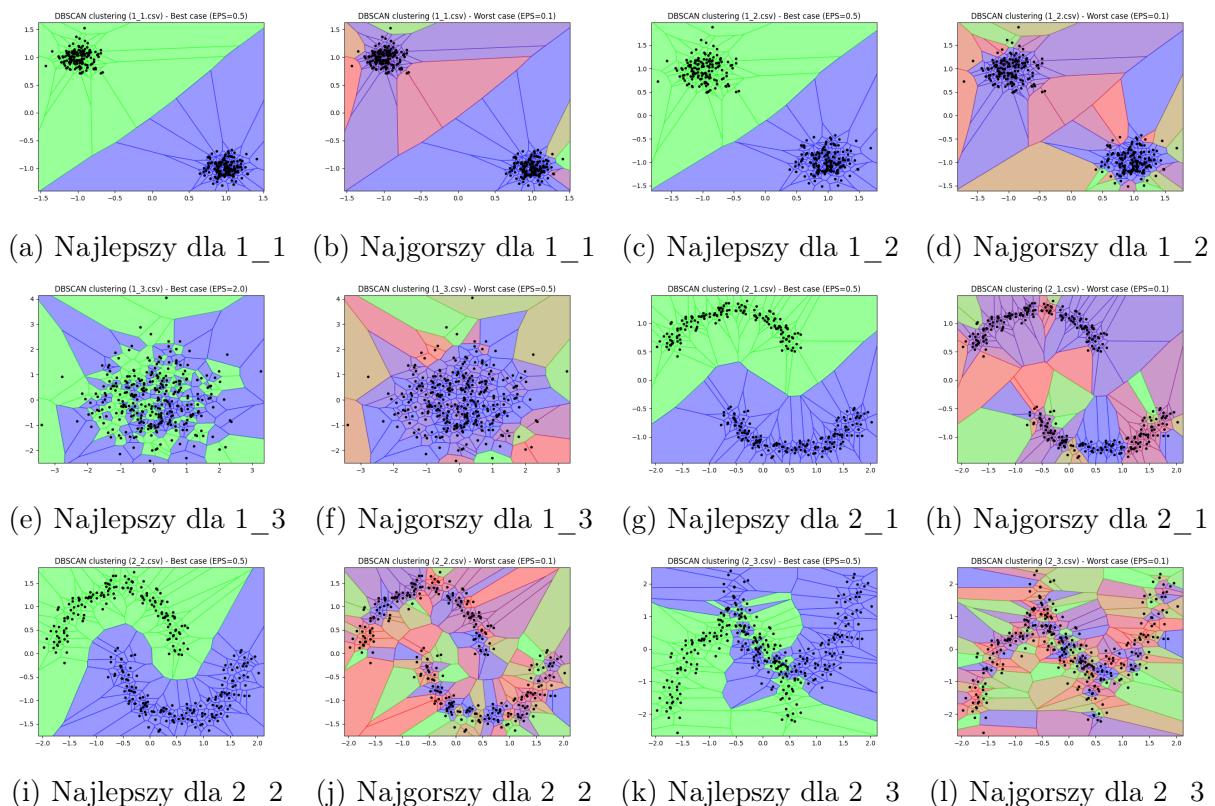


Rysunek 2: Wizualizacja klastrów dla wszystkich zbiorów na diagramie Voronoia dla najlepszego i najgorszego przypadku w metodzie K-means

2 Eksperyment 1: Metoda DBSCAN

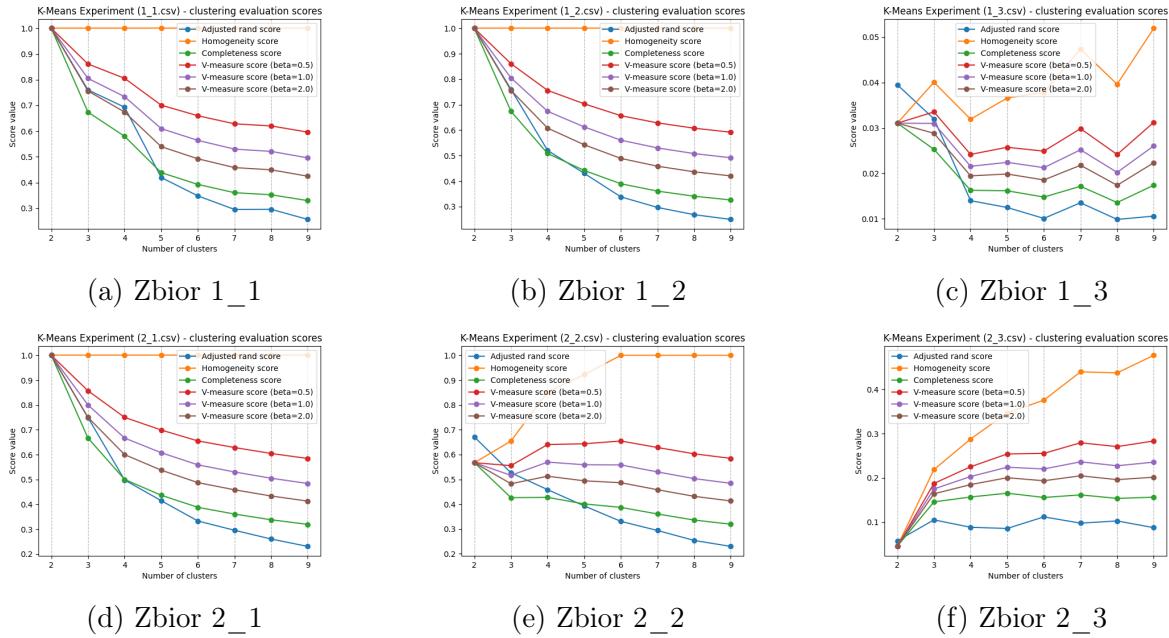


Rysunek 3: Zmiana wartości silhouette score oraz n_clusters dla wszystkich zbiorów w zależności od zmieniającego się parametru eps w metodzie DBSCAN

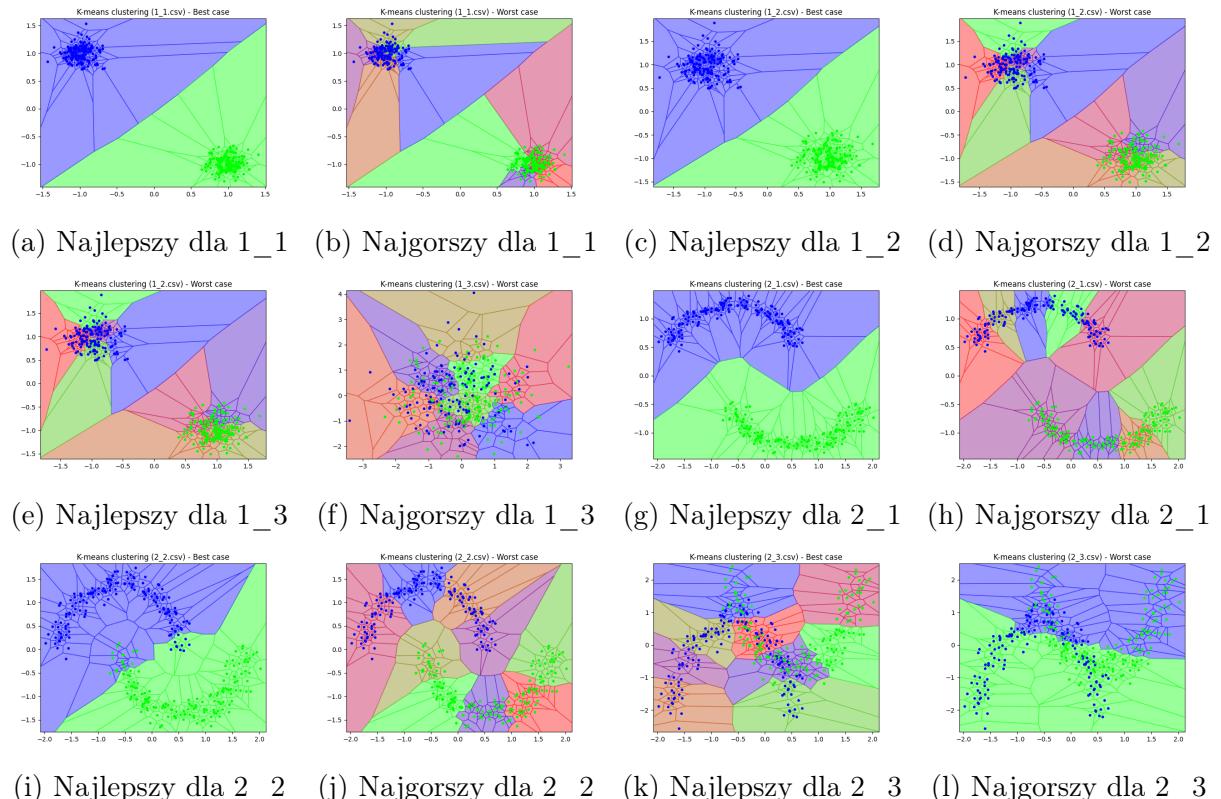


Rysunek 4: Wizualizacja klastrów dla wszystkich zbiorów na diagramie Voronoia dla najlepszego i najgorszego przypadku w metodzie DBSCAN

3 Eksperyment 2: Metoda K-Means z etykietami

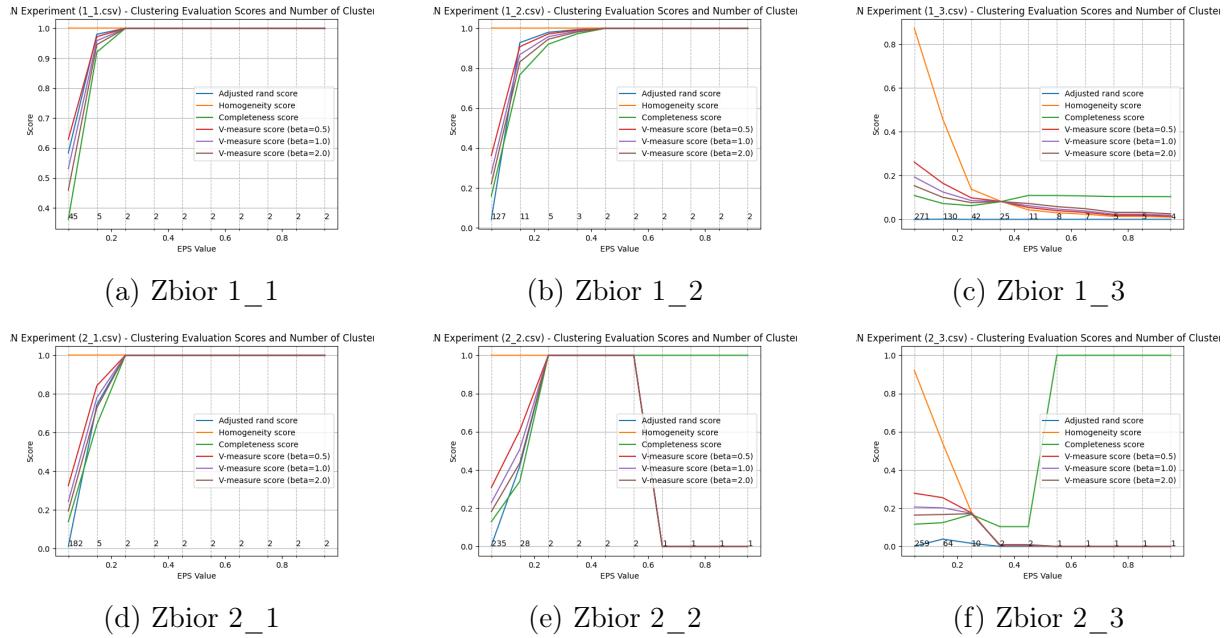


Rysunek 5: Zmiana wartości miar jakości dla wszystkich zbiorów w zależności od liczby klastrów w metodzie K-means

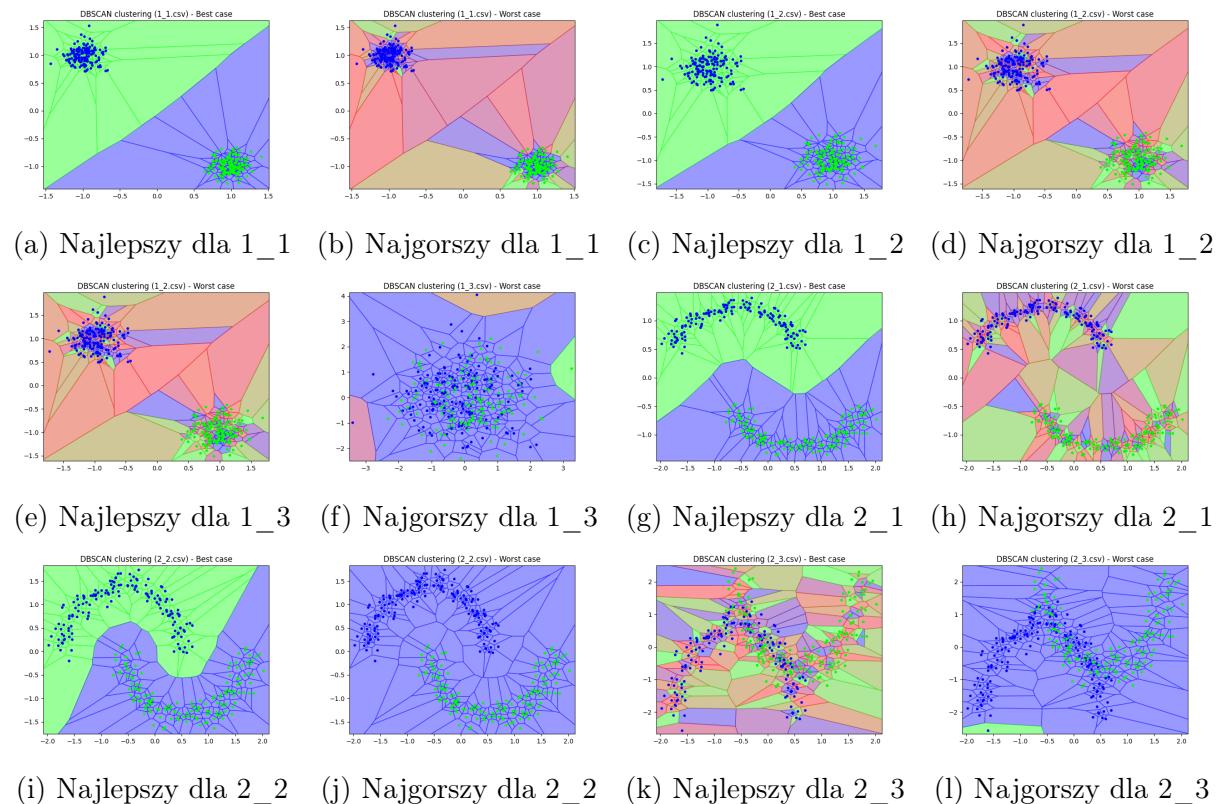


Rysunek 6: Wizualizacja klastrów wraz z prawdziwymi etykietami dla wszystkich zbiorów dla najlepszego i najgorszego przypadku w metodzie K-means

4 Eksperyment 2: Metoda DBSCAN z etykietami



Rysunek 7: Zmiana wartości miar jakości oraz liczba klastrów dla wszystkich zbiorów w zależności od wartości eps w metodzie DBSCAN



Rysunek 8: Wizualizacja klastrów wraz z prawdziwymi etykietami dla wszystkich zbiorów dla najlepszego i najgorszego przypadku w metodzie DBSCAN

5 Wnioski

Opis wniosków z eksperymentów przeprowadzonych na sześciu sztucznie wygenerowanych zbiorach. W przypadku pierwszego eksperymentu należy stwierdzić, jak, obserwując silhouette score, można dobrąć właściwe parametry metod klasteryzacji, aby klastry dobrze odkrywały strukturę danych. Dla eksperymentu drugiego należy wskazać jakie informację o separowalności klas można wyciągnąć obserwując wartości miar: adjusted rand score, homogeneity score, completeness score oraz V-measure score. Wnioski powinny mieć charakter ogólny, pozwalający przenieść je na przypadek, w którym nie ma możliwości zwizualizowania danych. Każdy wniosek powinien być poparty odniesieniami do wyników przedstawionych na pierwszych czterech stronach raportu.

Eksperiment 1

Eksperiment 2

Analiza

Analiza pozostałych, rzeczywistych zbiorów danych powinna uwzględniać wnioski z poprzednich eksperymentów. Można je również uzasadnić poprzez odwołanie do wartości miar uzyskanych na tych zbiorach.

6 Analiza pozostałych zbiorów danych

Opis analizy pozostałych, rzeczywistych zbiorów danych, w której to zastosowane zostaną wnioski z wcześniejszych eksperymentów. Wyniki tej analizy należy również uzasadnić poprzez odwołanie do wartości miar uzyskiwanych na tych zbiorach (warto wykorzystać tabele i/lub wykresy i się do nich odwołać). W przypadku zbioru Iris da się swoje wnioski podeprzeć wizualizacjami rzutów cech obiektów na dwuwykładowe przestrzenie wybranych kombinacji dwóch z nich.

Wnioski

Na podstawie przeprowadzonych eksperymentów możemy wnioskować o strukturze rzeczywistych zbiorów danych i skuteczności wybranych metod klasteryzacji.

Zbior Iris

Zbior Wine

Zbior Breast Cancer Wisconsin