

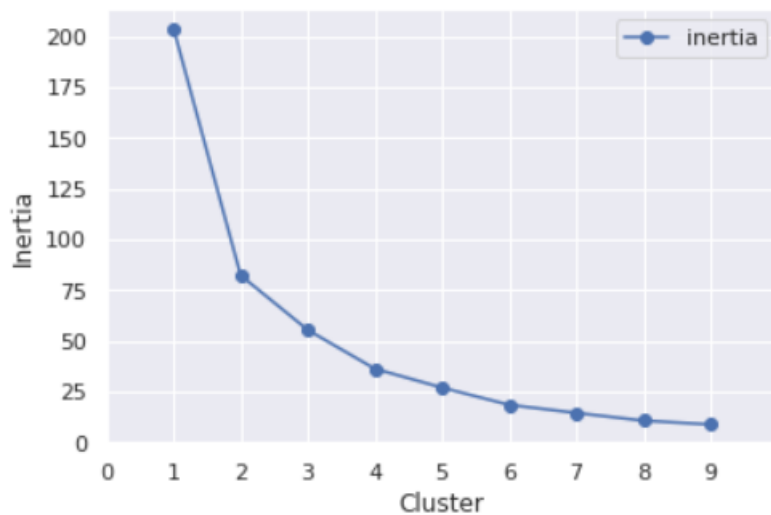
Zad 1

Dla każdego z algorytmów zamieszczam najlepsze uzyskane rozwiązania.

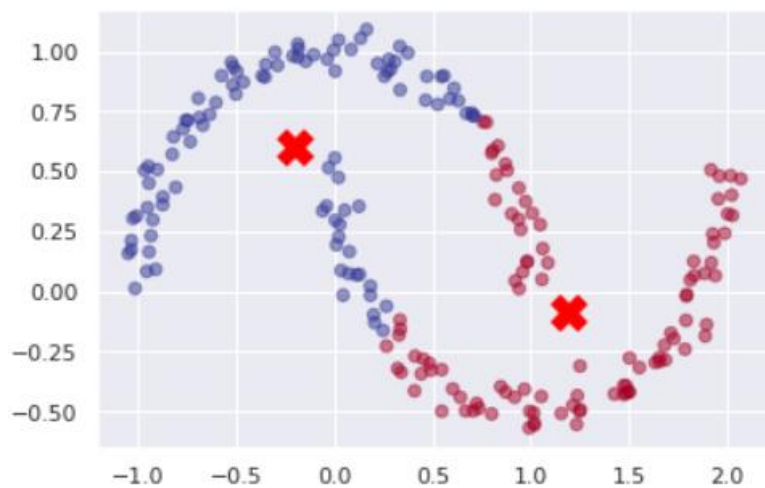
Księżyce

K-means

Wykres bezwładności od liczby klastrów

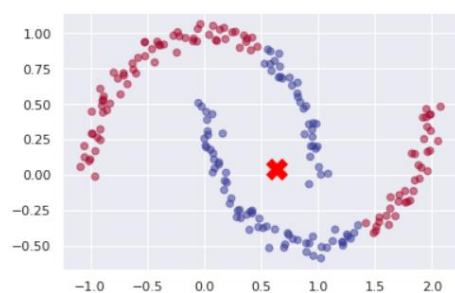


Najlepszy wynik dla księżyców osiągamy, gdy mamy 2 klastry, chociaż księżyce nie są rozdzielane na osobne klastry.

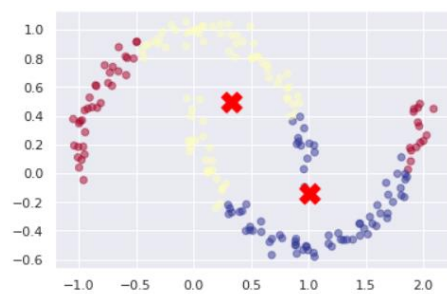


MeanShift

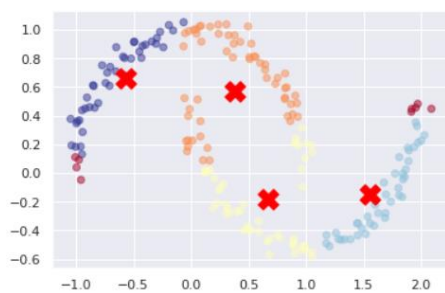
Liczba klastrów - 1



Liczba klastrow – 2



Liczba klastrow – 4



Nie udało się rozdzielić półksiężyców na osobne klastry.

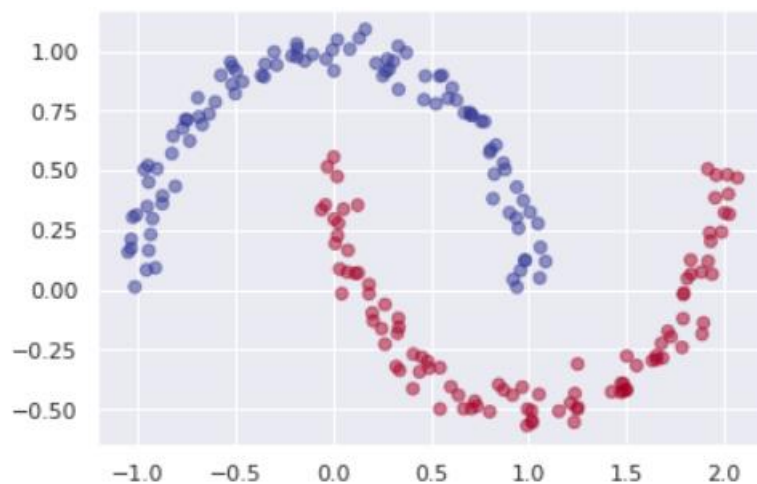
DBSCAN

Eps = 0,19 do 0,36

Liczba klastrow – 2

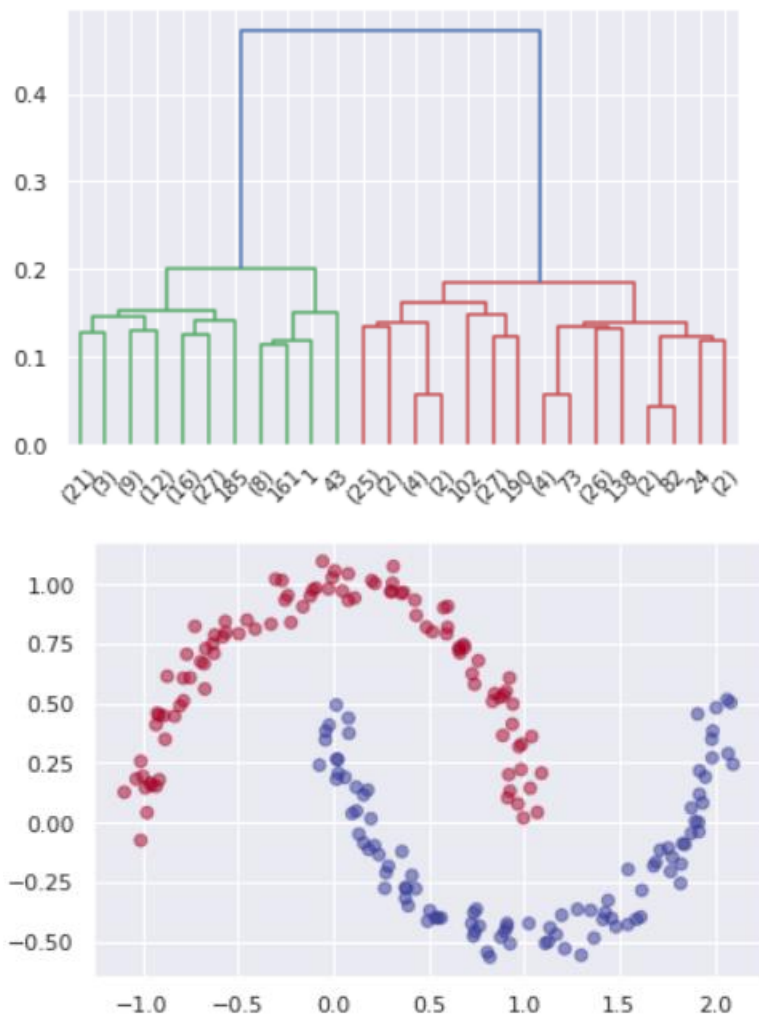
Liczba odstających elementów – 0

Jest to najlepszy wynik, ponieważ każdy półksiężyc ma swój własny klaster i nie ma żadnych odstających punktów.



Algorytm aglomeratywny

Najlepszy wynik udało mi się uzyskać dla affinity – manhattan, linkage – single, distance_threshold = 0,3.

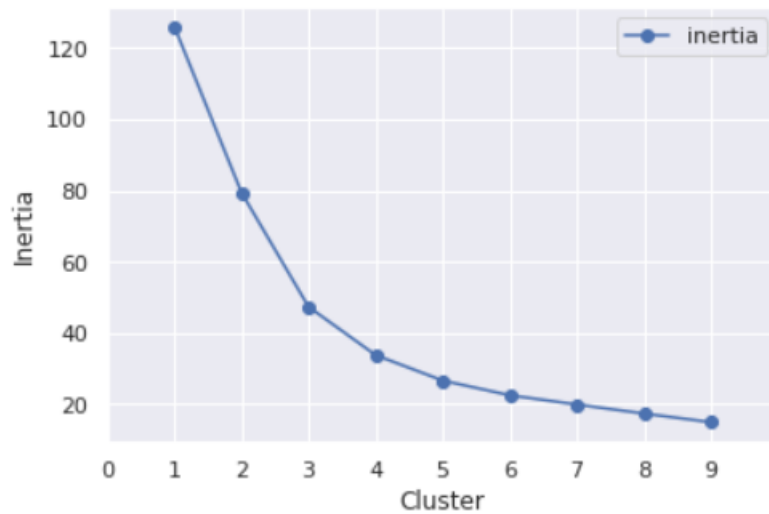


Jest to najlepszy wynik, ponieważ każdy księżyc ma swój klaster.

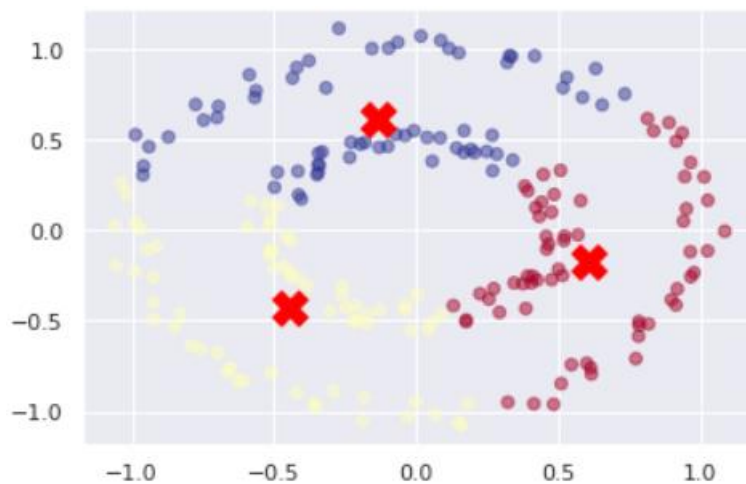
Kółka

k-means

Wykres bezwładności od liczby klastrów:

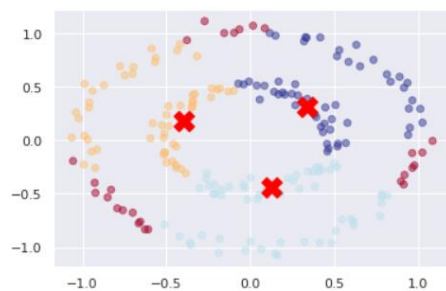


Dla 3 klastrów powinien być najlepszy wynik, ponieważ na wykresie bezwładności od ilości klastrów widać kolanko.

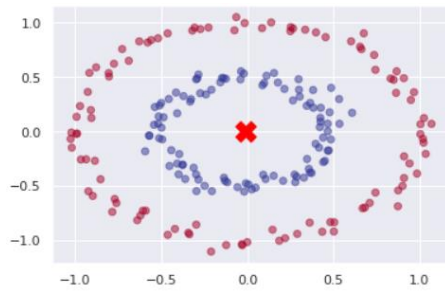


Meanshift

Liczba klastrów - 3



Liczba klastrow – 1



Nie udało się rozdzielić kółek na osobne klaster.

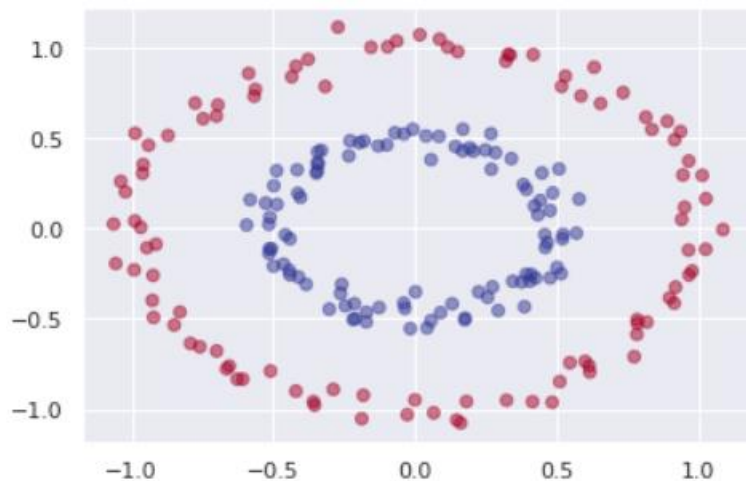
DBSCAN

Eps = 0,22 do 0,31

Liczba klastrow – 2

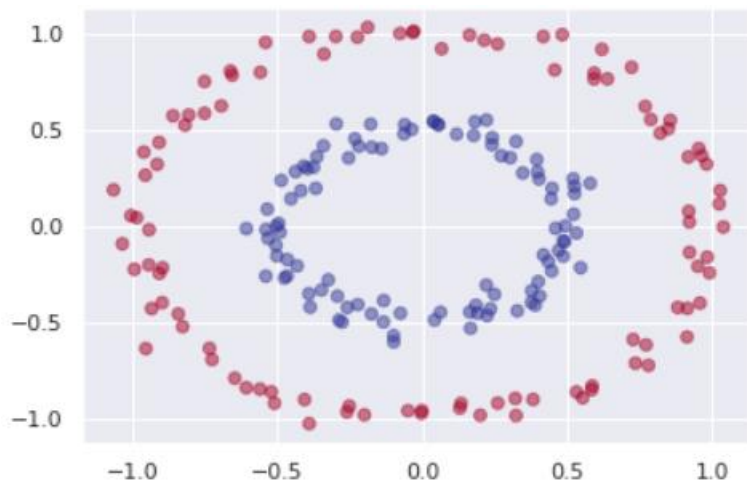
Liczba odstających elementów – 0

Jest to najlepszy wynik, ponieważ nie ma żadnych odstających elementów i każde kółko ma swój klaster.



Algorytm aglomeratywny

Najlepszy wynik udało mi się uzyskać dla affinity – manhattan, linkage – single, distance_threshold = 0,3.
Każde kółko jest rozdzielone na osobny klaster.



Dla półksiężyców i kółek dobrze sprawdzają się algorytmy DBSCAN i aglomeratywny – za ich pomocą można rozdzielić zarówno kółka jak i półksiężyce na osobne klastry. Nie za dobrze natomiast sprawdzają się algorytmy k-means i meanshift. Nie udało mi się za ich pomocą tak dobrze rozdzielić kształtów na klastry. Oprócz tego nawet jeśli w k-means i meanshift była taka sama liczba klastrów, to w meanshift zostawało trochę punktów, które nie należały do żadnego z klastrów.

Zadanie 2

cat.jpg

Zredukowano do 8 kolorów

Oryginał:

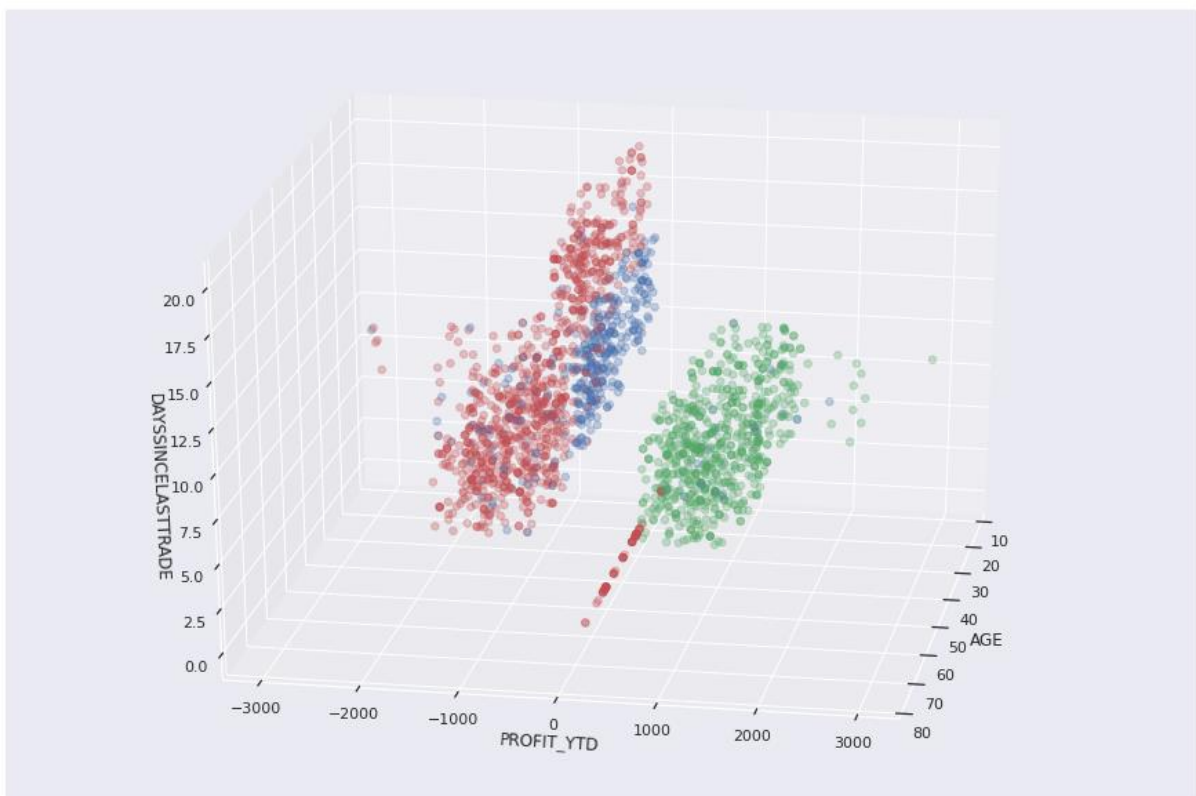


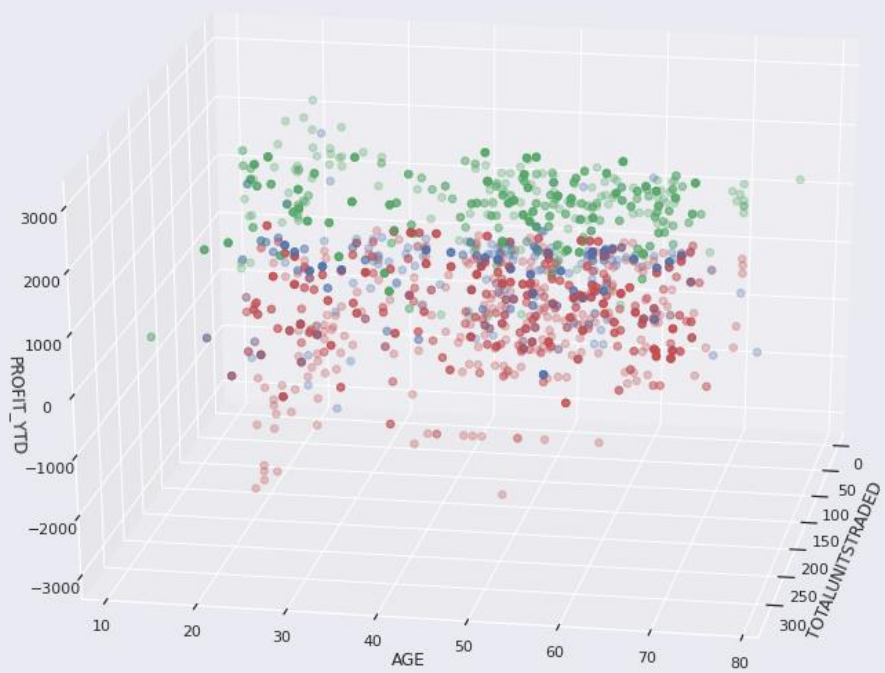
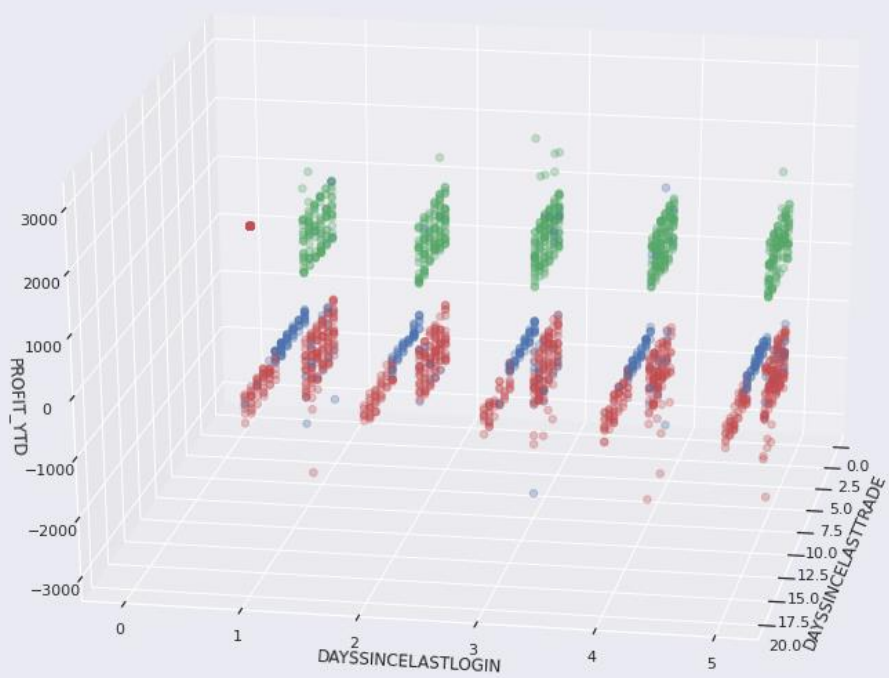
Oryginał obok kompresji:

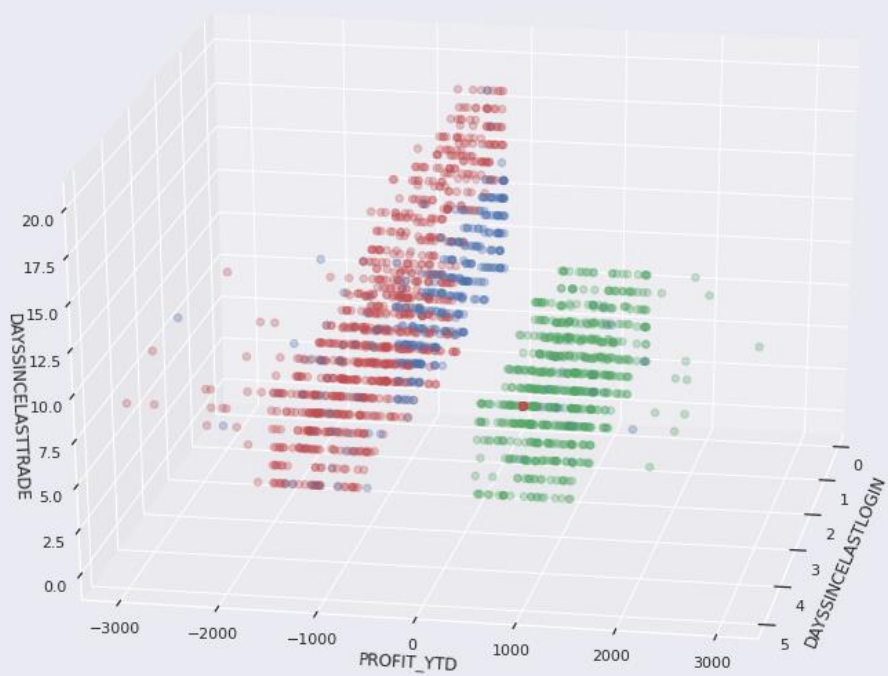
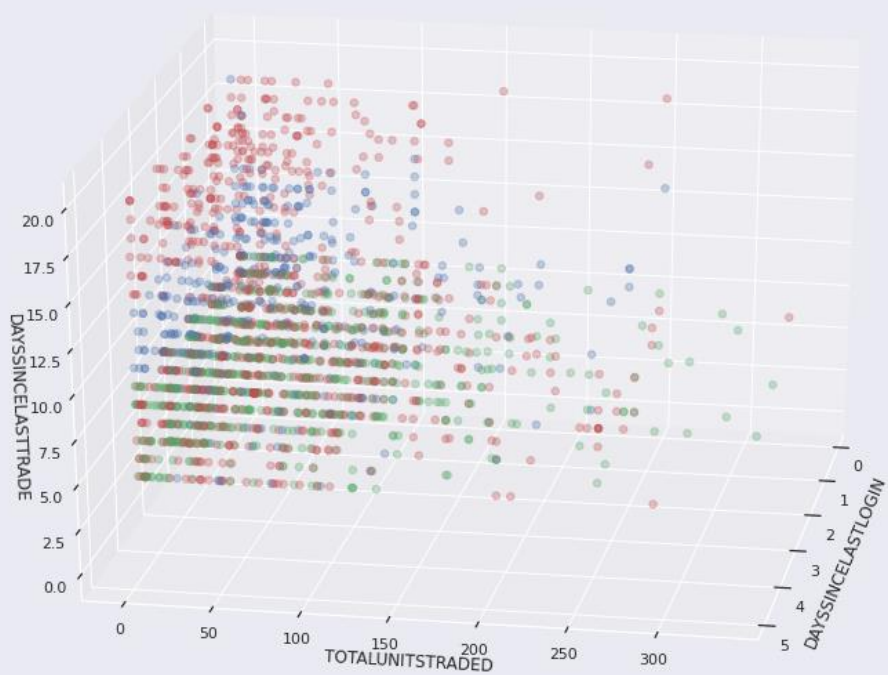


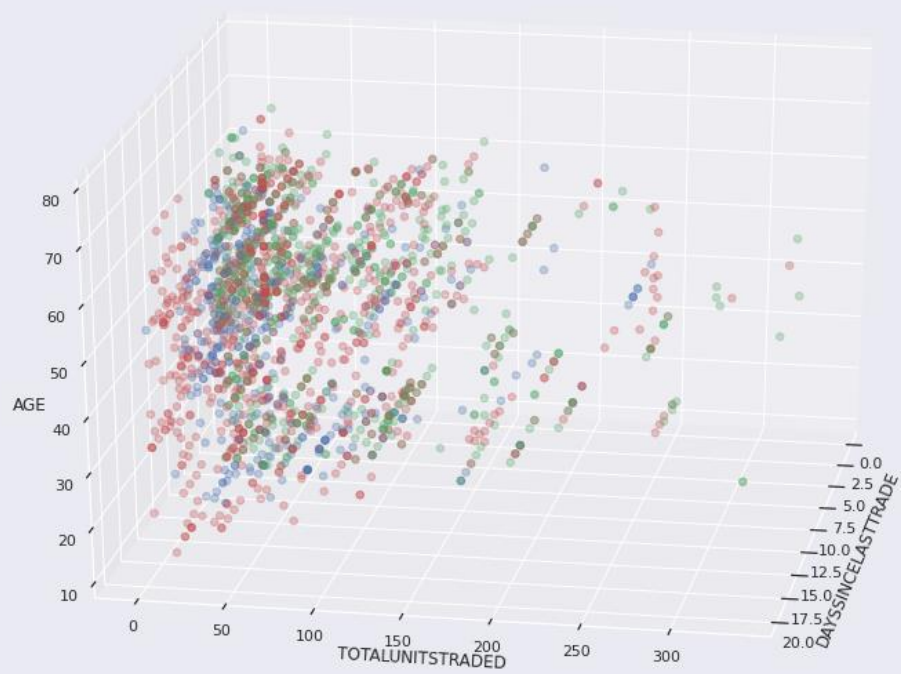
Zadanie 3

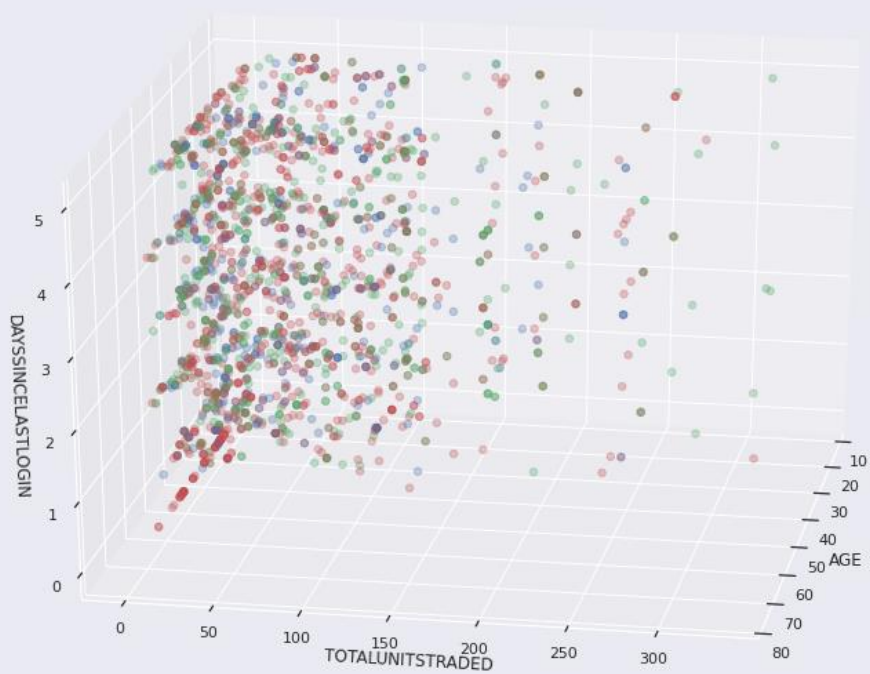
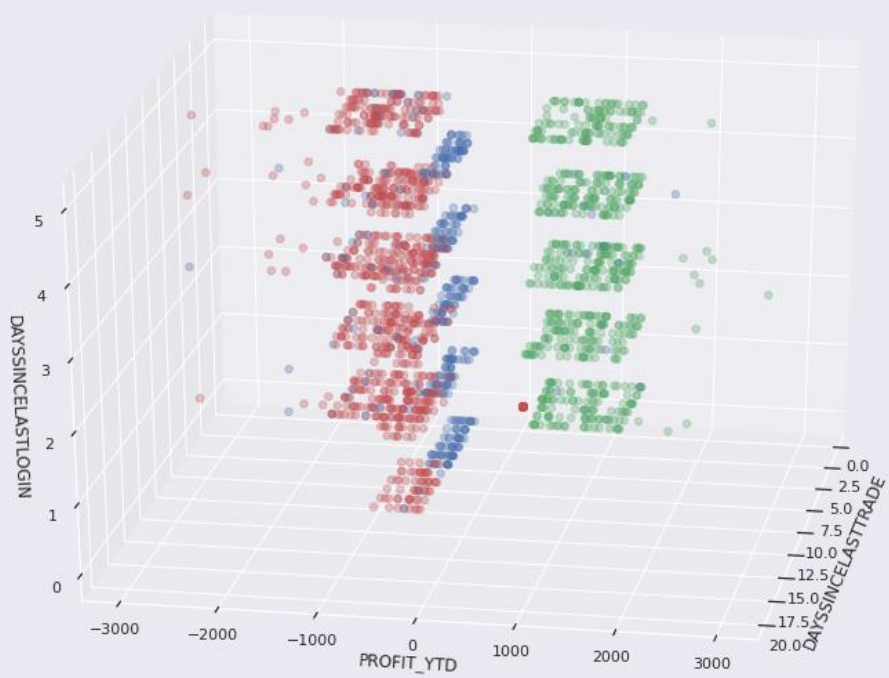
Zrobiłam trochę wykresów i na ich podstawie stwierdziłam, że parametr PROFIT_YTD ma największe znaczenie dla podziału na klastry. DAYSSINCELASTTRADE jest mniej ważny, ale też ma znaczenie.



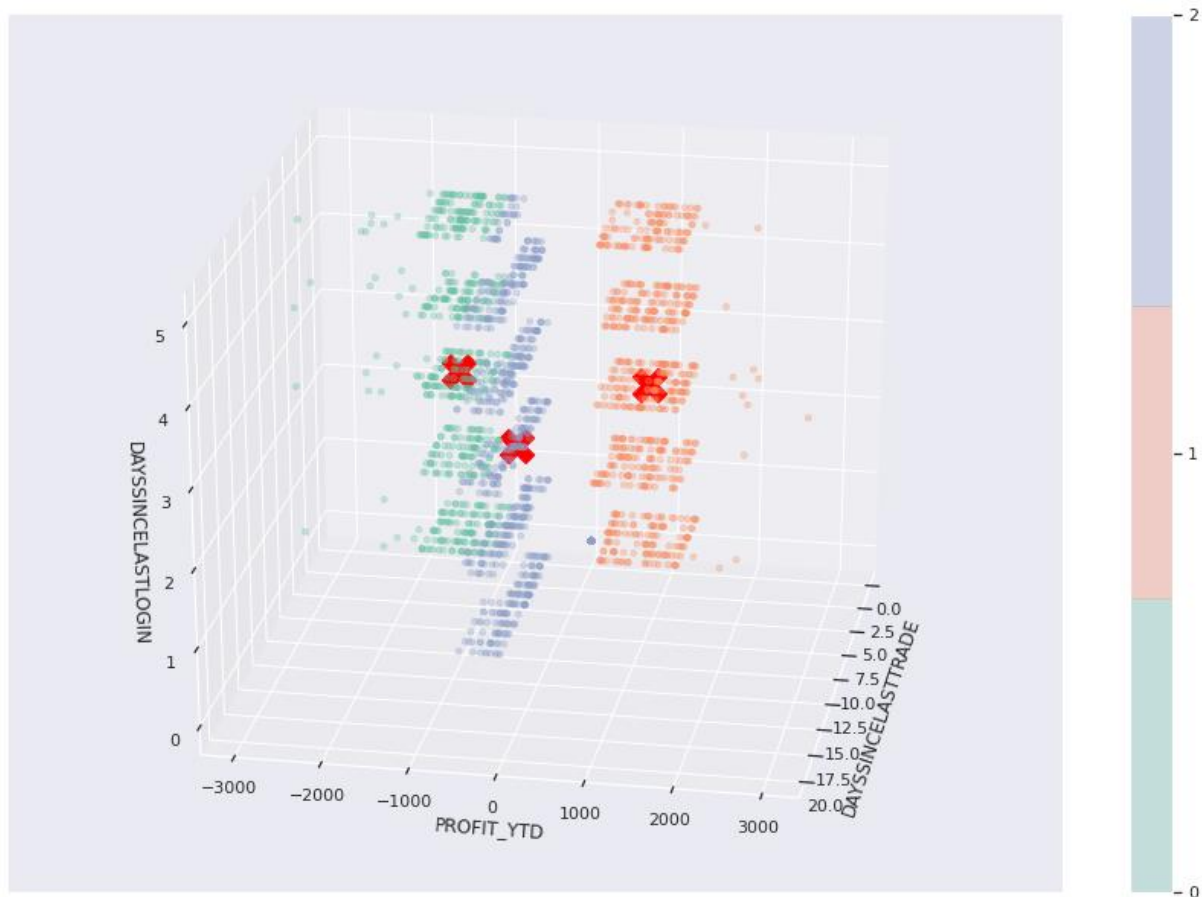
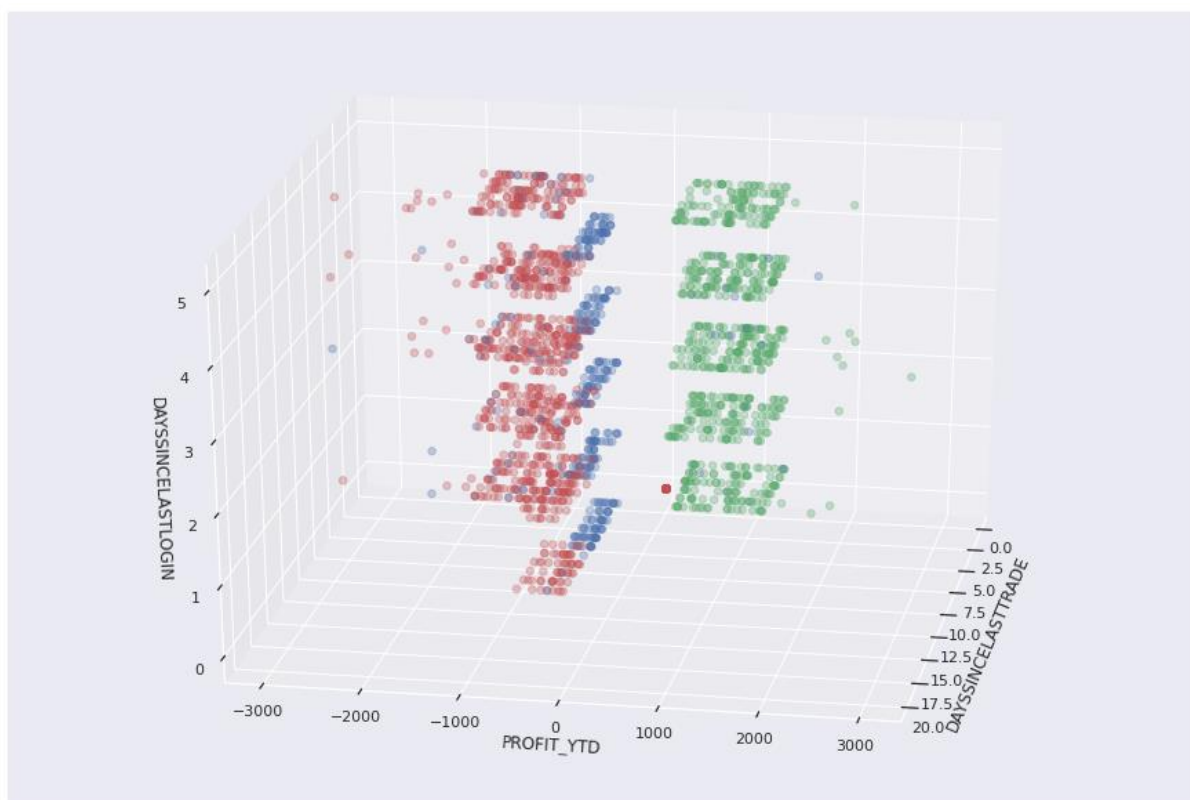


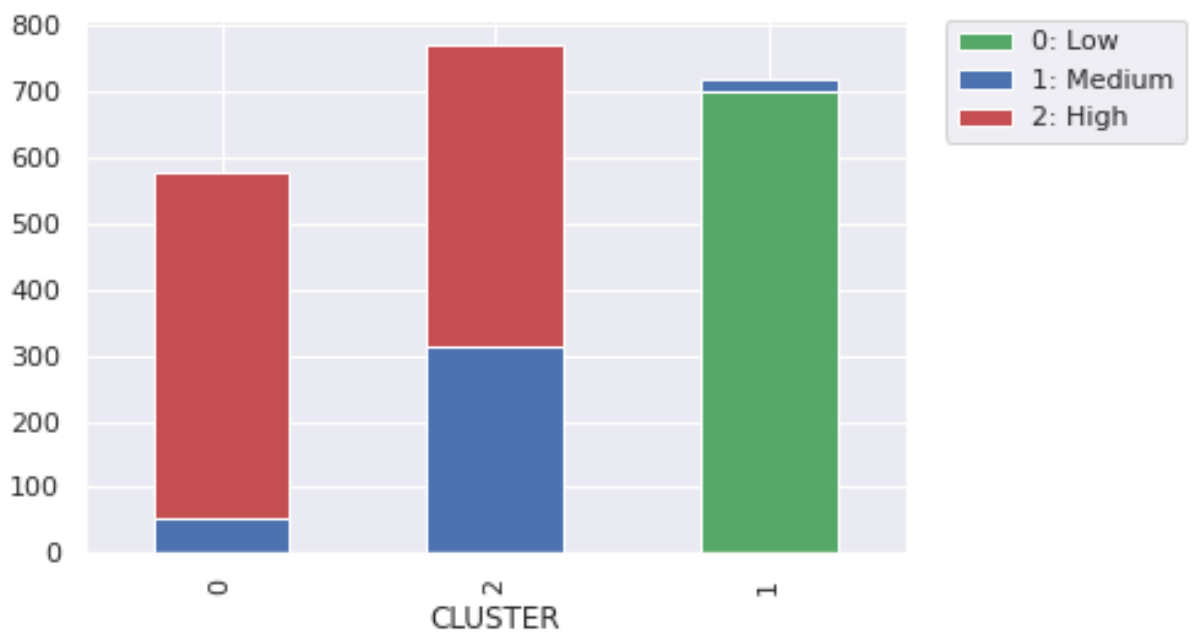






Następnie zrobiłam klasteryzację wybranego wykresu:





Najlepsze dopasowanie udało mi się uzyskać dla

distance_threshold=20000,
affinity='euclidean',
linkage='ward'

