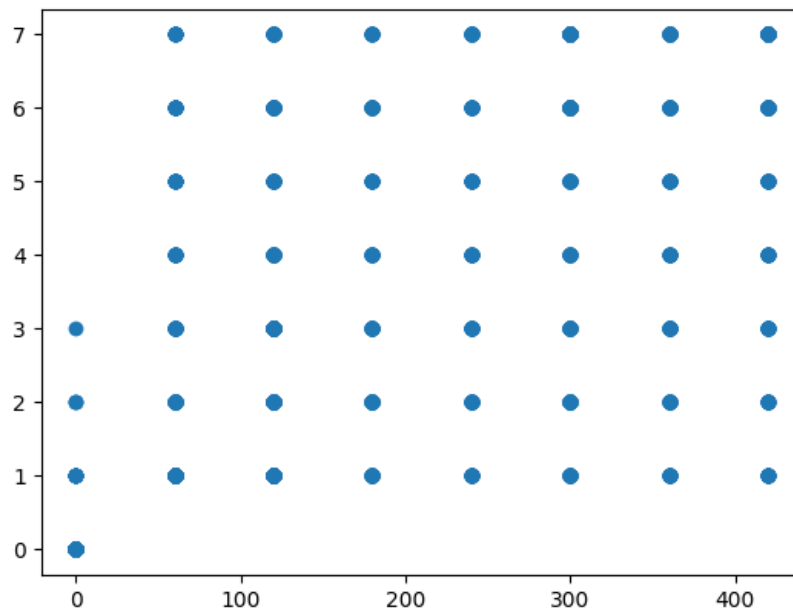


Uczenie nadzorowane vs. uczenie nienadzorowane

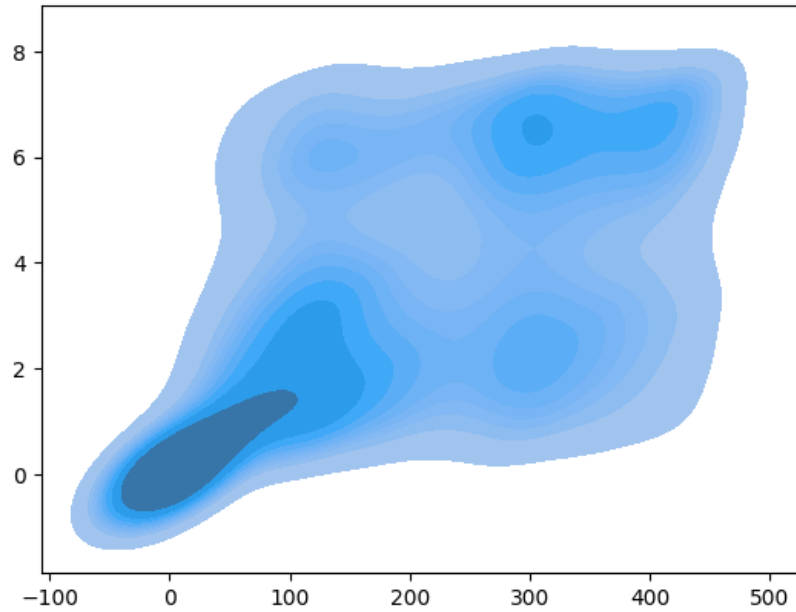
Kamil Pyla



Wyniki zadania 1:

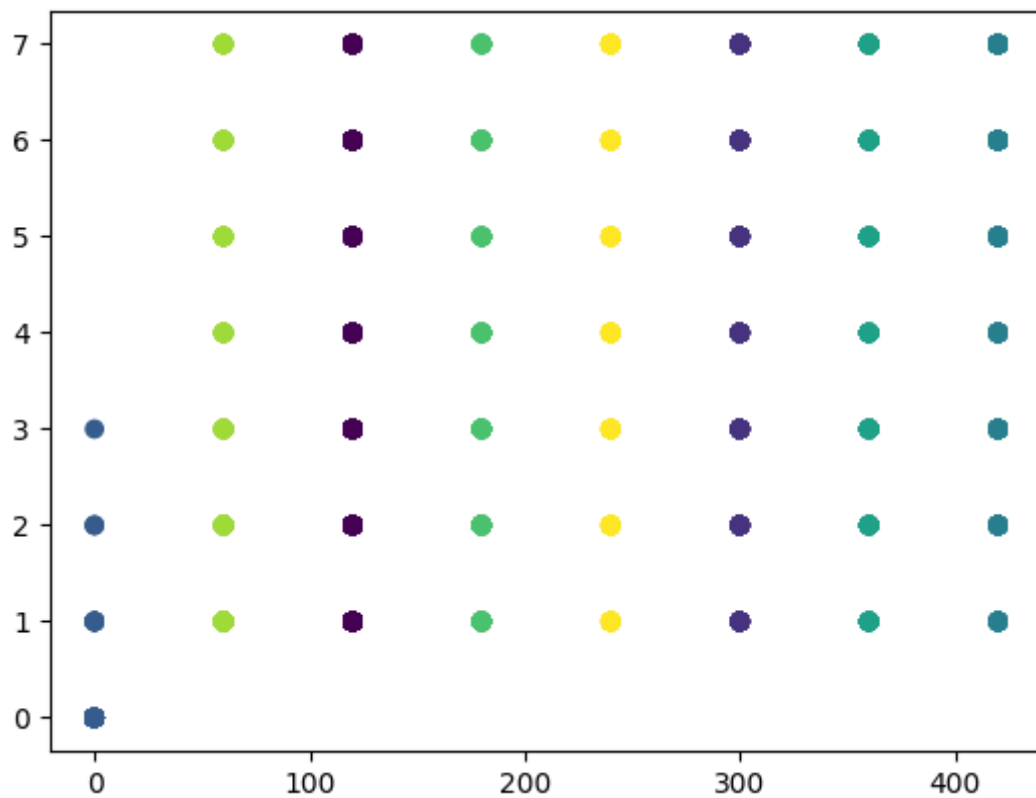


Wykres danych nieznormalizowanych



Wykres danych nieznormalizowanych z uwzględnieniem gęstości

Do klasyfikacji użyłem 8 klastrów:

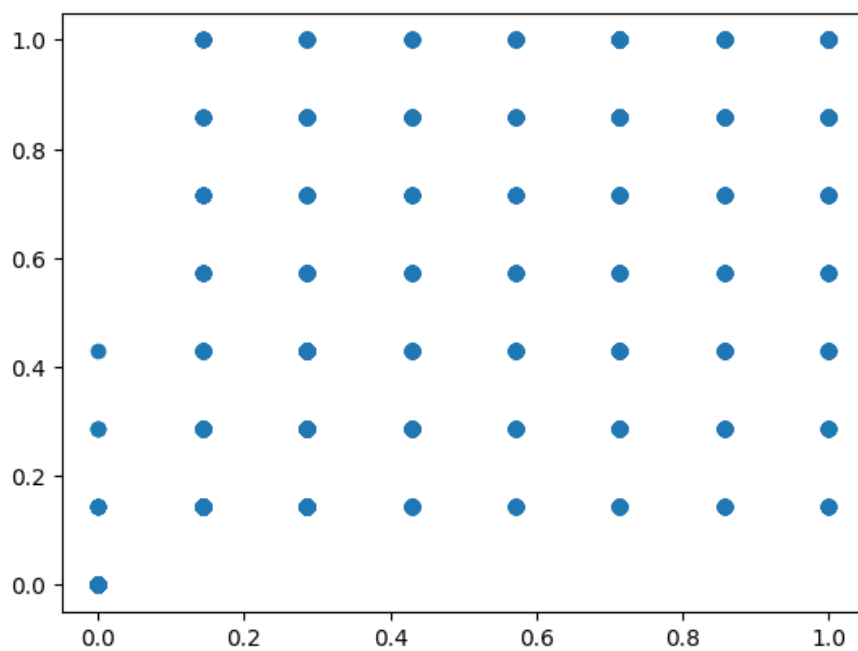


Wykres sklasyfikowanych danych

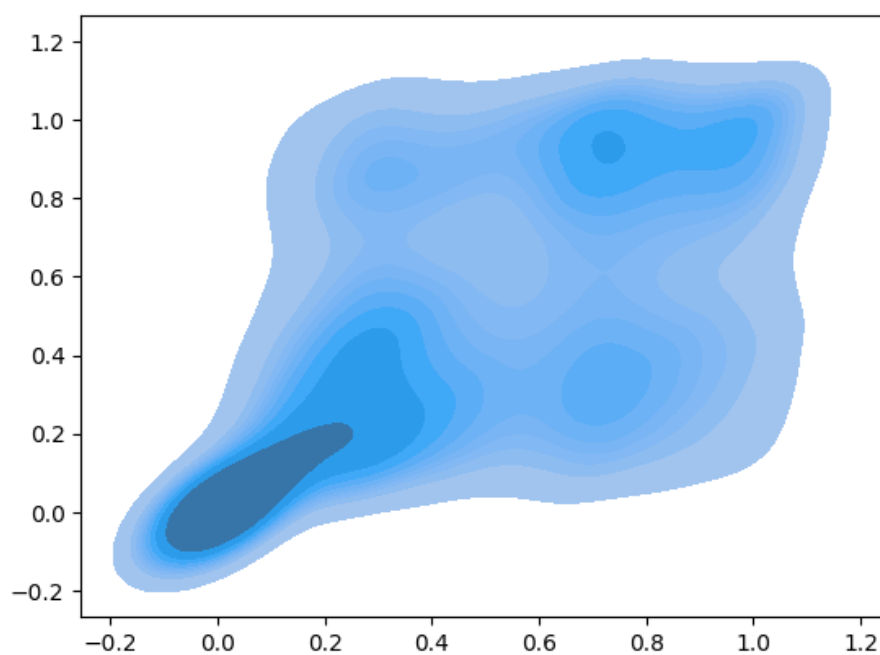
Sihouette Score: 0.967

DB Score : 0.055

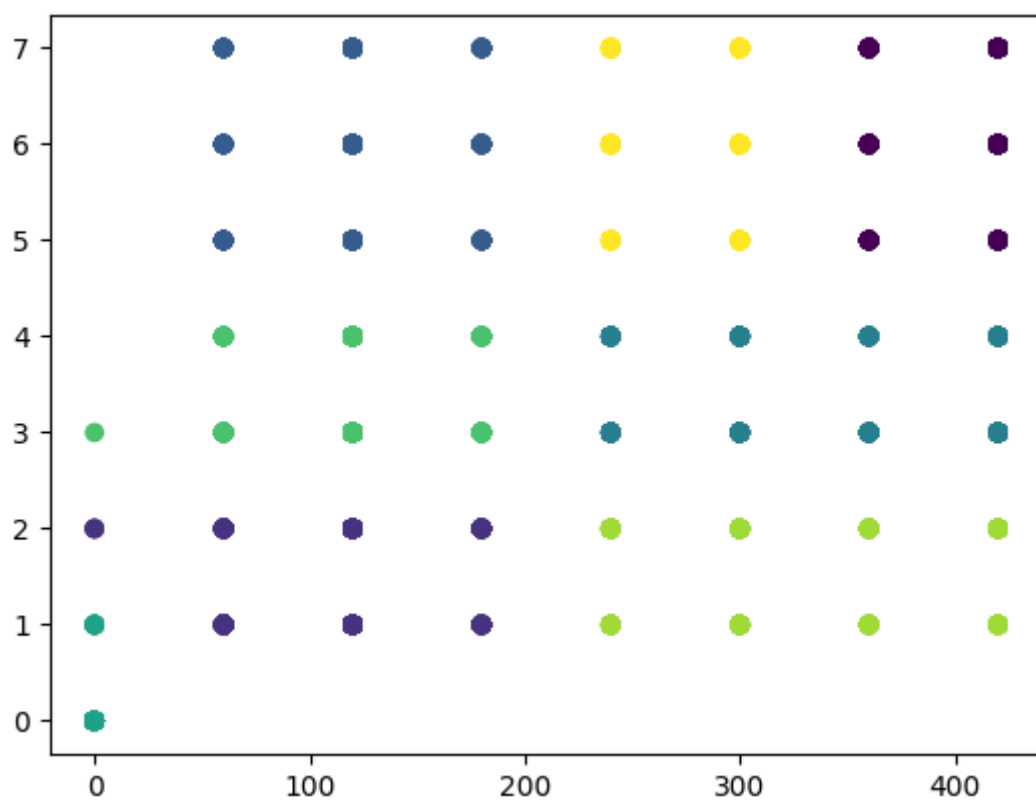
Wyniki dla danych znormalizowanych



Wykres danych znormalizowanych



Wykres danych znormalizowanych z uwzględnieniem gęstości



Sihouette Score: 0.487

DB Score : 0.857

Wyniki klasyfikacji dodanych czytelników (dane czytelników zostały wcześniej znormalizowane)

ilość czasu	ilość sesji	klasyfikacja
123	2	4
321	3	4
445	1	7
132	8	2

Wnioski: jendym z ważniejszych aspektów przed dokonaniem klasyfikacji danych jest ich normalizacja, o czym świadczy różnica klasyfikacji.

Wyniki zadania 2:

pl_orbper	pl_orbsmax	pl_rade	pl_masse	pl_orbeccen	pl_eqt	st_teff	st_mass	sy_dist
4.756285	0.0528	4.8535	25.6215	0.047	855.5	4801.5	0.78	115.6985
3.35524	0.048	15.8	352.78	0	1770	6279	1.31	367.658
5.6334729	0.0694	11.982	2294.7326	0.151	1198	5880	1.13	313.996
14.893291	0.1206	4.88	20.1	0.05	806	5547	0.95	168.067
3.2888	0.04501	12.644	190.698	0	1431	5670	1.04	321.296
3.6977104	0.03623	2.12	8.17	0.04	581	3514	0.45	36.0118
3.474119	0.057	15.513	5720.7	0.112	2260	7000	1.68	1057.41
2.73417876	0.047195	17.962	988.448815	0	2436	7670	1.825	282.5745
3.901645	0.04972	13.6765	304.629	0.04265	1509	5961	1.13	1084.015

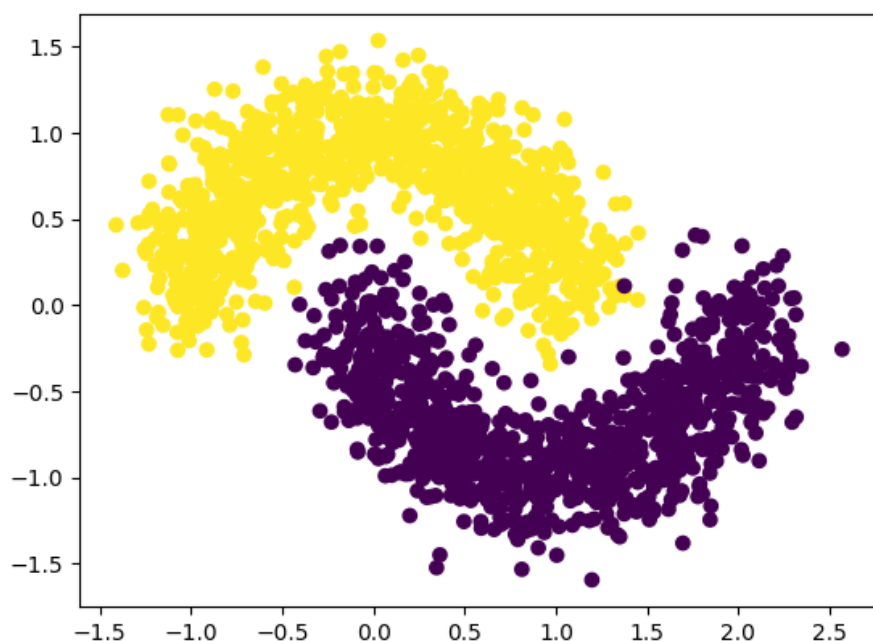
W tabeli przedstawiono medianę poszczególnych wartości

Niestety nie jestem w stanie powiedzieć jednoznacznie czym charakteryzuje się każdy klastery, dostrzegam różnicę, w długości obrotu wokół własnej osi w każdym klastrze, również każdy klastery zawiera planety różniące się masą

Wyniki zadania 3:

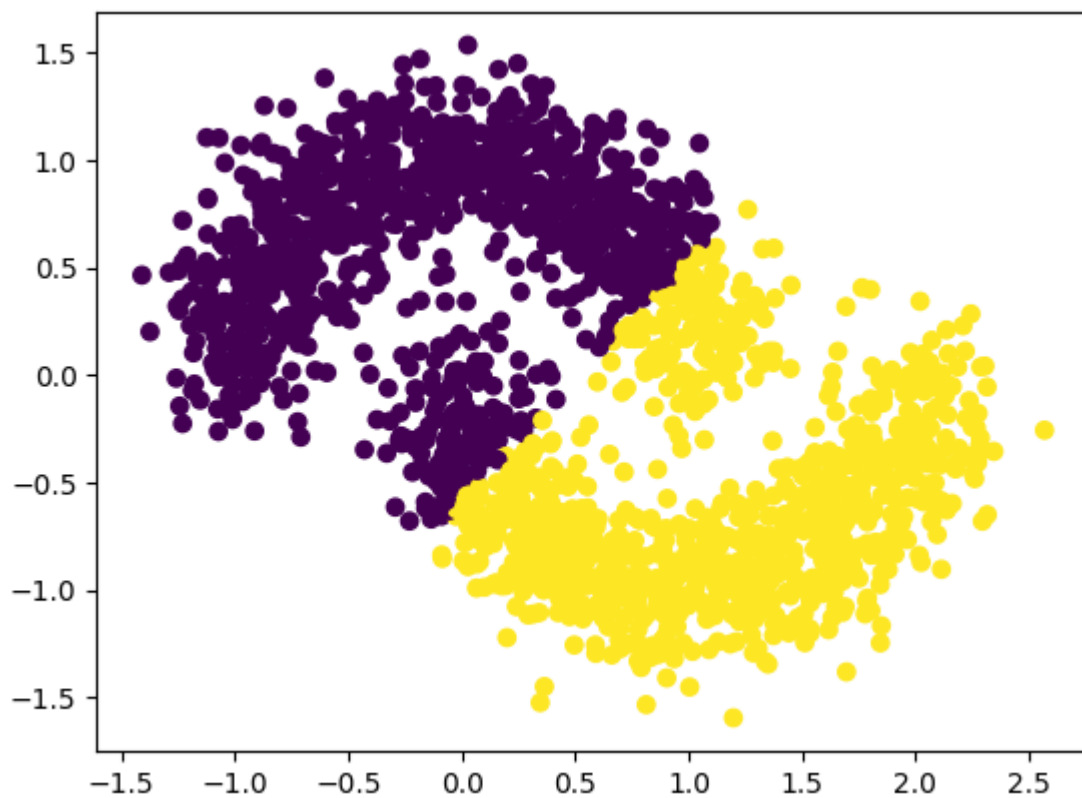
Zacząłem pisać kod, jednak nie doszedłem do wyników

Wyniki zadania 4:



Wygenerowane punkty

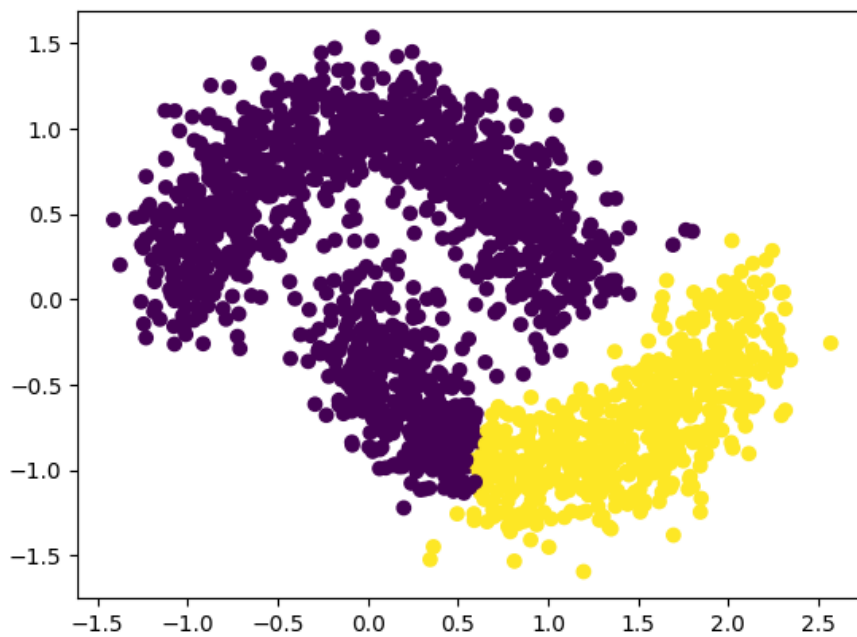
Klasyfikacja Kmeans



Sihouette Score: 0.476

DB Score : 0.816

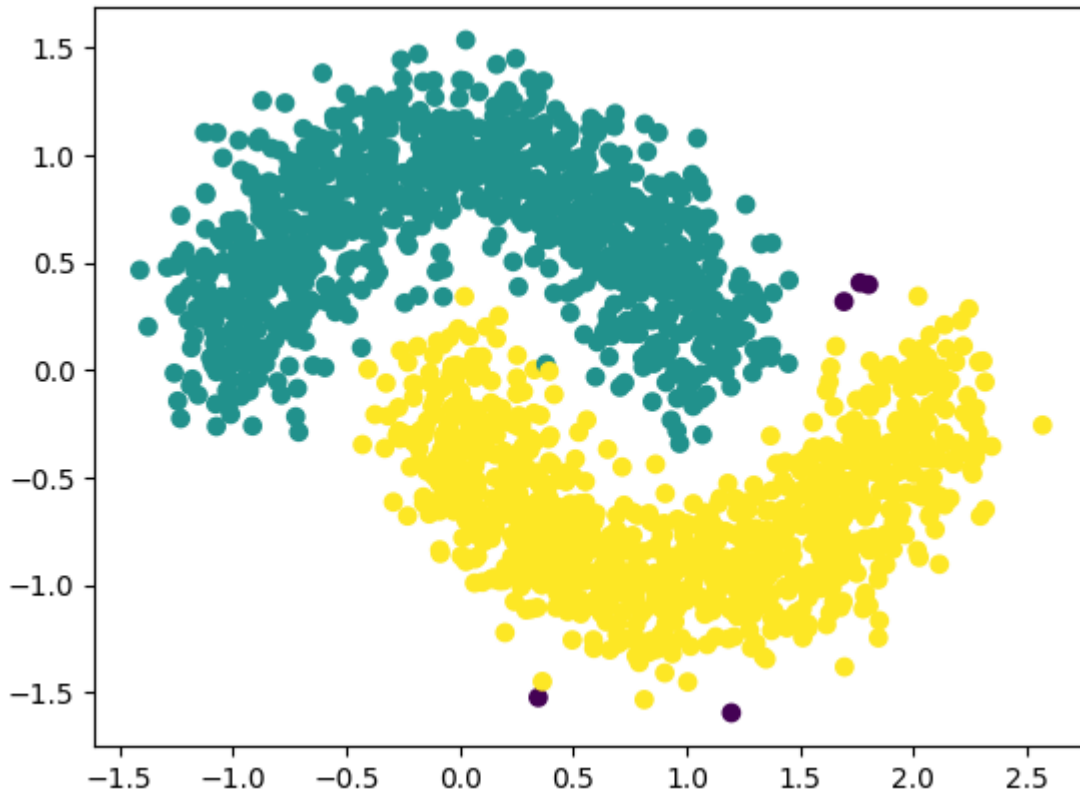
Wyniki dla algorytmu AgglomerativeClustering



Sihouette Score: 0.433

DB Score : 0.786

Wyniki dla klasyfikacji DBSCAN



Sihouette Score: 0.307

DB Score : 3.085

Wnioski: przy klasyfikacji klastrów które posiadają pewien punkt / obszar skupienia najlepszą klasyfikację wykonuje algorytm DBSCAN, który dokonuje klasyfikacji na podstawie właśnie tych punktów skupienia. Pozostałe algorytmy dokonują podziału przed dzielenie hiperpłaszczyzny co widać na wykresach. Zastawane funkcje do porównania poprawności klasyfikacji nie sprawdzają się w przypadku algorytmu dbscan.

Link do repozytorium z kodem:

https://github.com/KamilPyla/MIO_2023/tree/master/lab_06

