



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

Laboratorinis darbas Nr. 4

Parengė:

IFF-7/11 Martynas Jodeika

Darbą priėmė:

lekt. Budnikas Germanas

KAUNAS, 2020

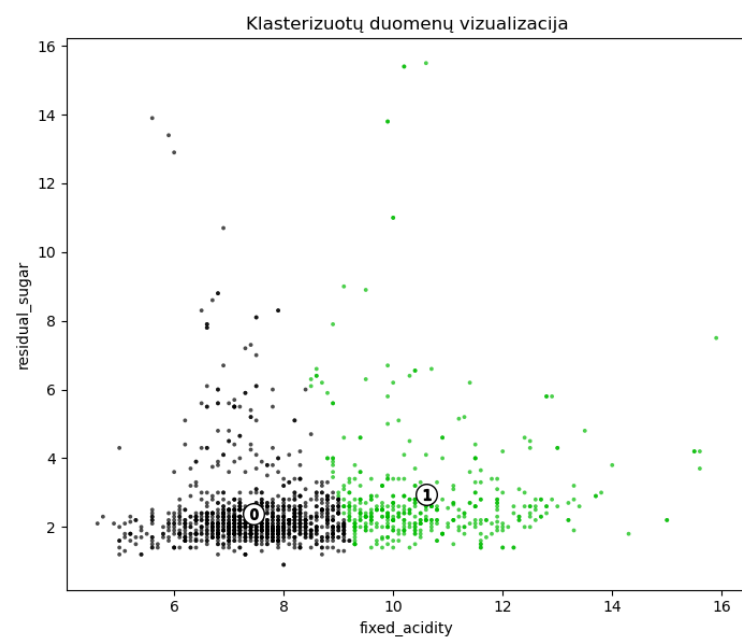
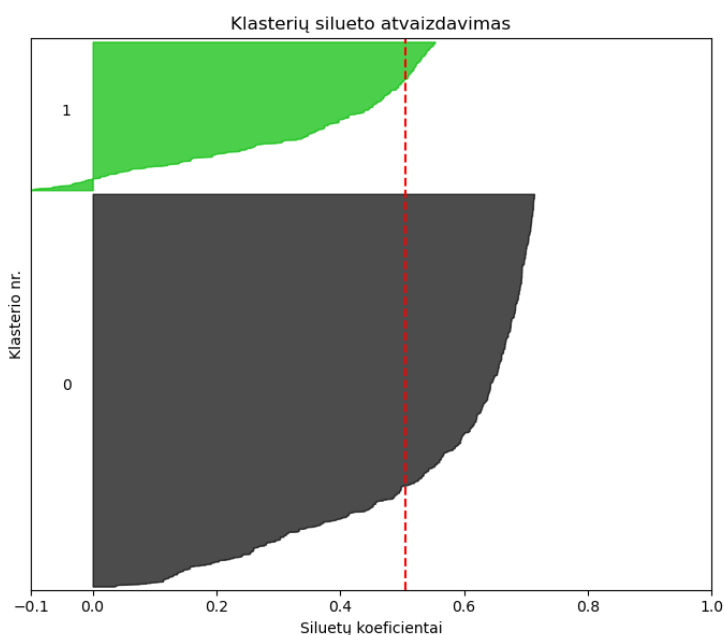
1. Duomenų rinkinio aprašas

Duomenų rinkinys yra skirtas nuspėti vyno kokybę pagal jo atributus. Šiame rinkinyje yra 12 tolydinio tipo atributų:

1. Pastovus rūgštumas
2. Nepastovus rūgštumas
3. Citrinos rūgštis
4. Likęs cukrus
5. Chloridai
6. Laisvas sieros dioksidas
7. Išviso sieros dioksido
8. Tankis
9. Rūgštingumas
10. Sulfatai
11. Alkoholis
12. Kokybė [0-10]

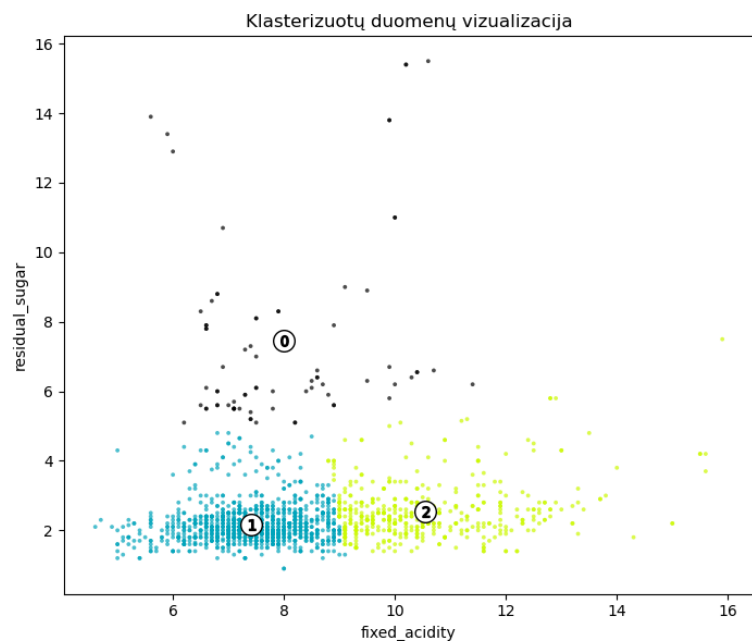
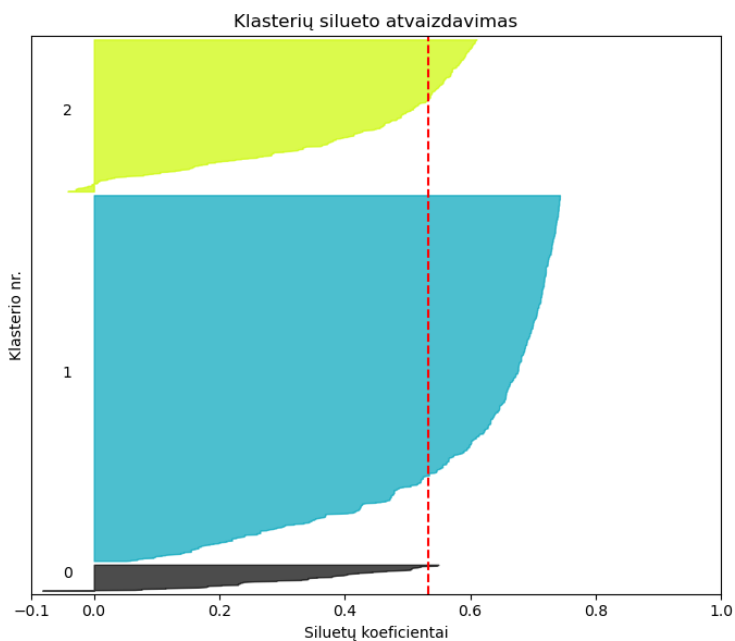
2. Klasterių grafikai

2.1 Pastovus rūgštingumas ir likęs cukrus



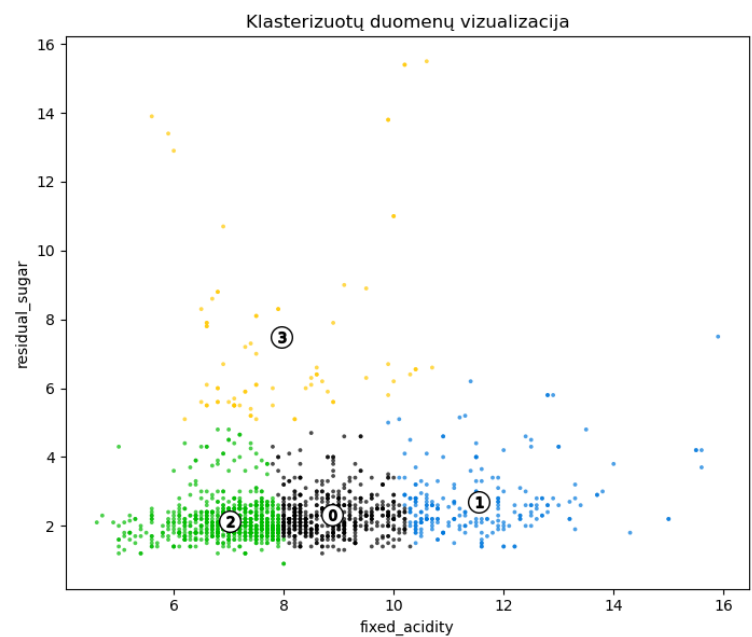
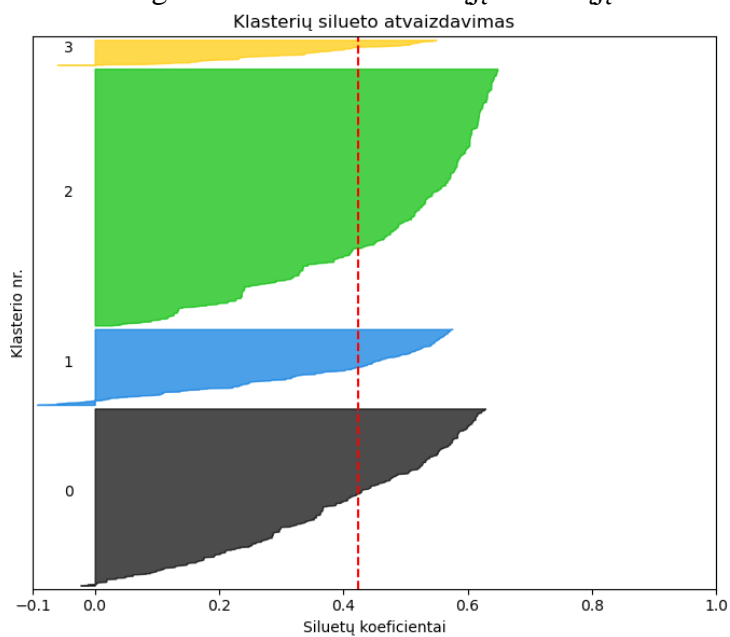
pav. 1 Likęs cukrus ir pastovus rūgštingumas - du klasteriai

Dviejų klasterių atveju pirmasis klasteris gaunasi gerokai didesnis už pirmąjį.



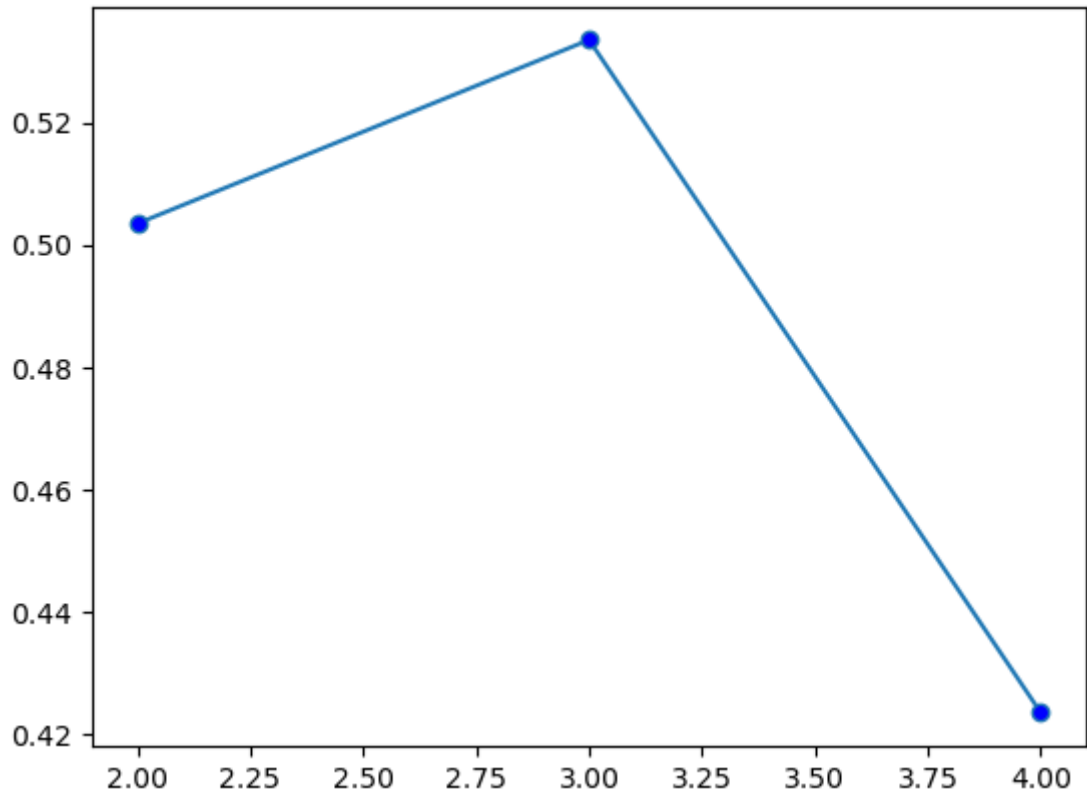
pav. 2 Likęs cukrus ir pastovus rūgštingumas - trys klasteriai

Skirstant į tris klasterius elementų skaičius visdar stipriai skiriasi, pirmasis klasteris gerokai didesnis už antrąjį ir trečiąjį.



pav. 3 Likęs cukrus ir pastovus rūgštingumas - keturi klasteriai

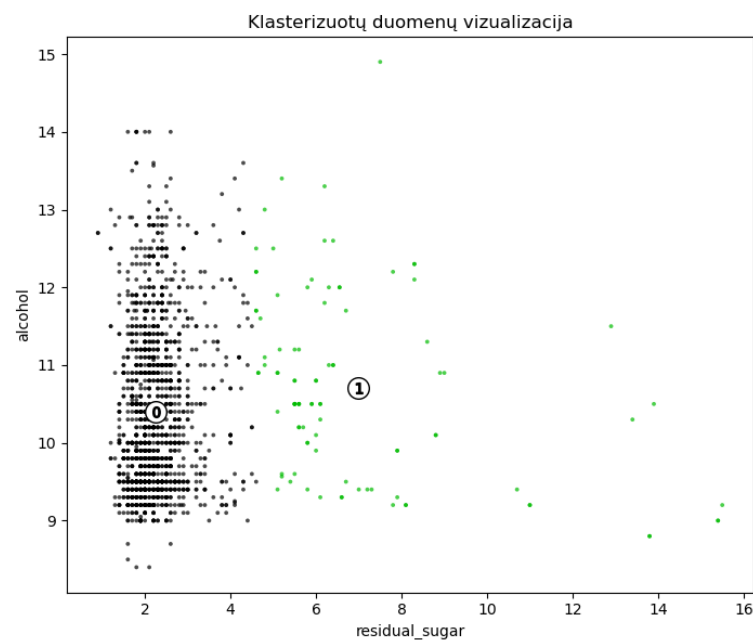
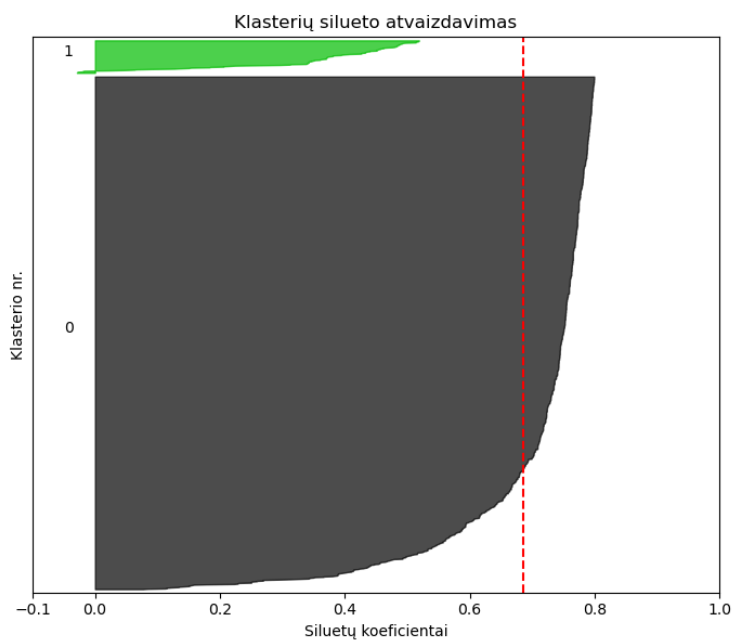
Naudojant keturis klasterius matome, kad elementai išsiskirstę tolygiausiai iš visų atvejų, nors trečiasis klasteris visdar matomai dominuoja duomenų rinkinį.



pav. 4 Klasterių siluetų įvertis

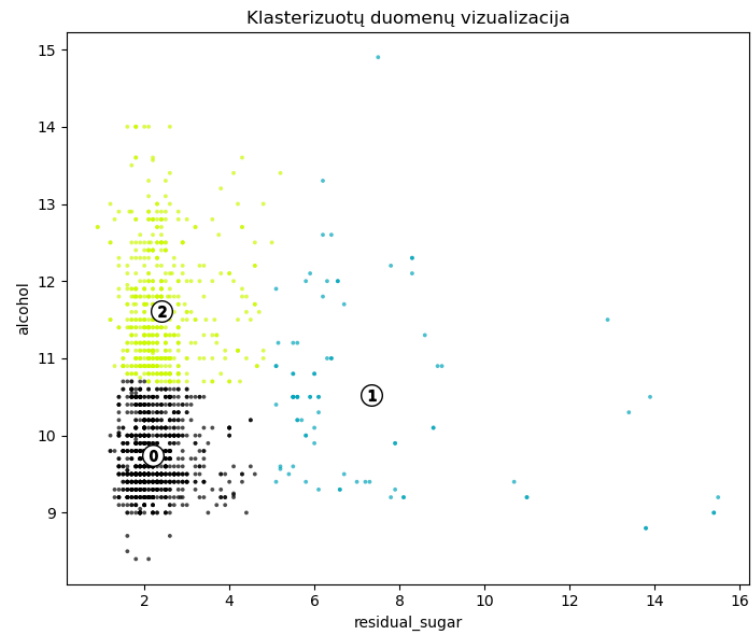
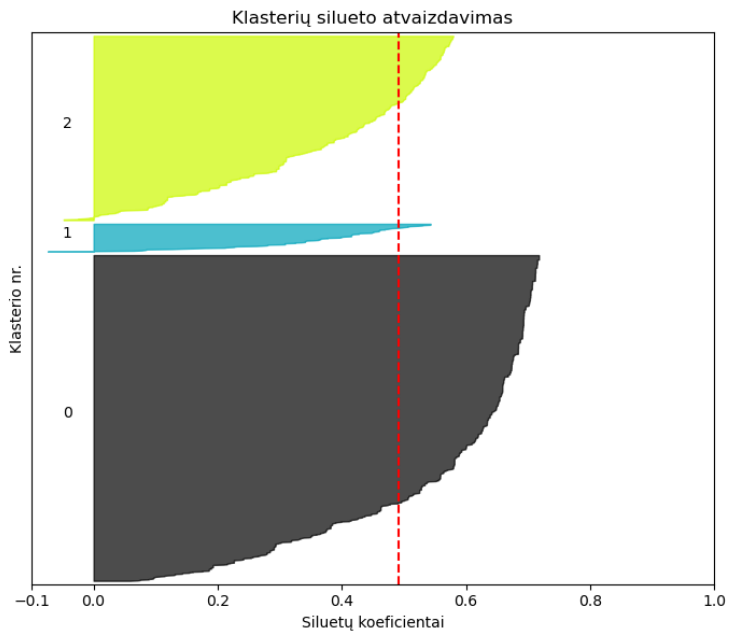
Pagal siluetų įverčius galime spręsti, kad šie atributai geriausiai susiklasterizavo kai buvo trys klasteriai.

2.2 Likęs cukrus ir alhokolio kiekis

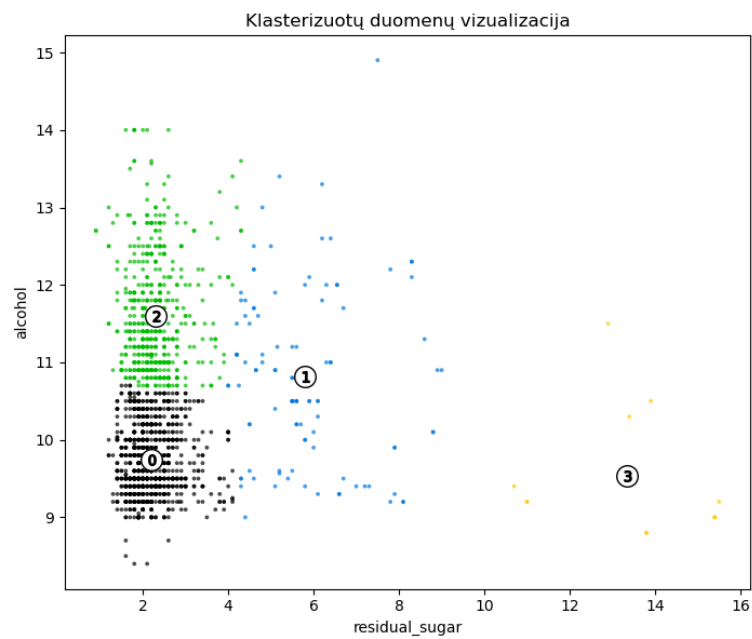
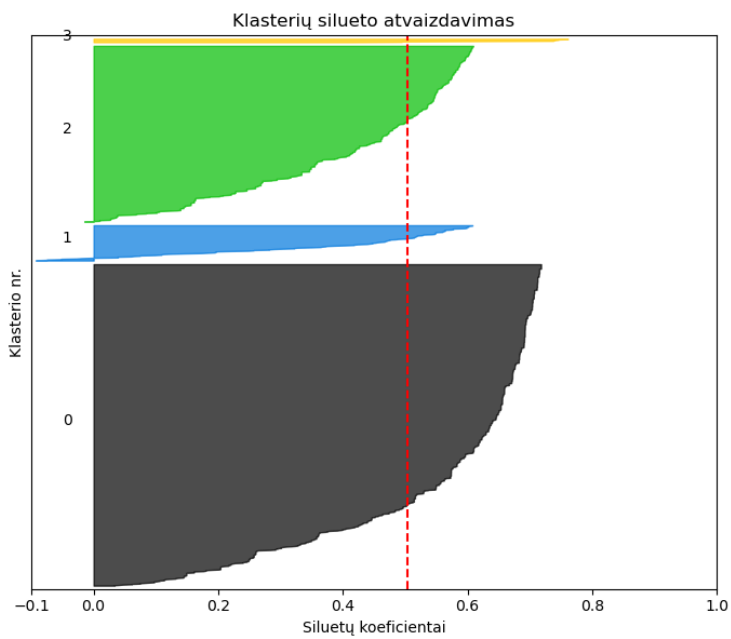


pav. 5 Likęs cukrus ir alkoholio kiekis - du klasteriai

Klasteriai nepastovūs – pirmas klasteris gerokai didesnis.

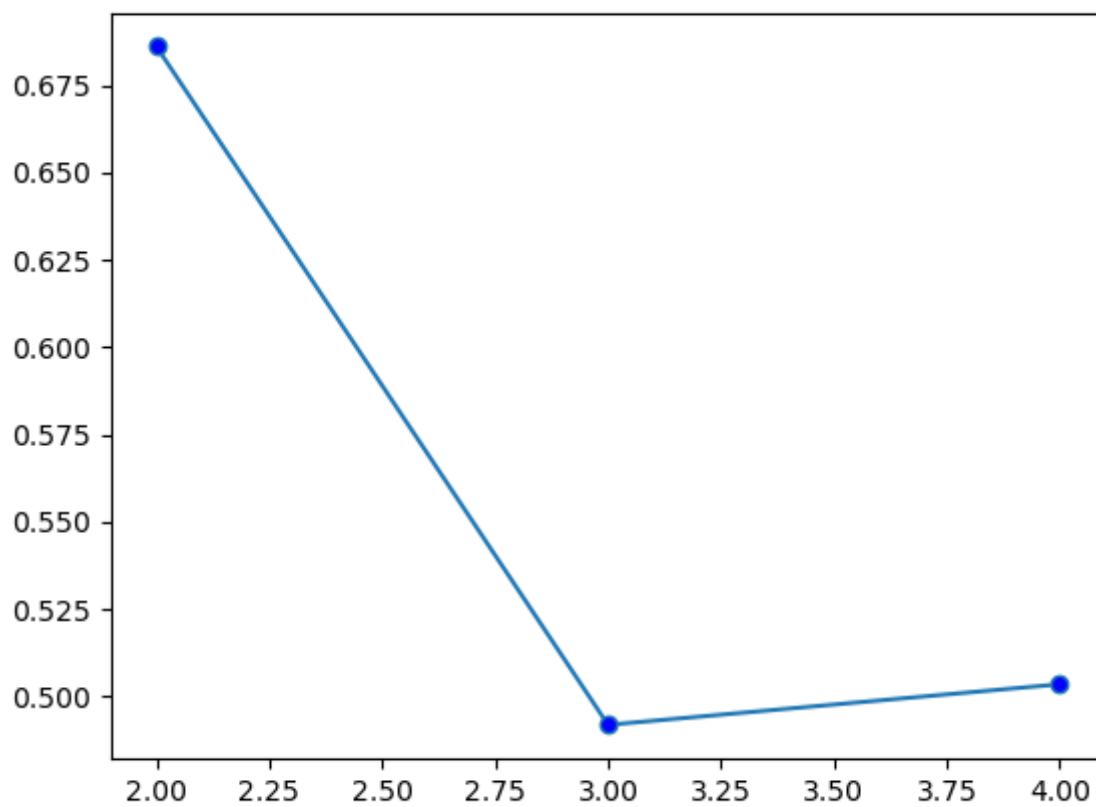


pav. 6 Likęs cukrus ir alkoholio kiekis - trys klasteriai



pav. 7 Likęs cukrus ir alkoholio kiekis - keturi klasteriai

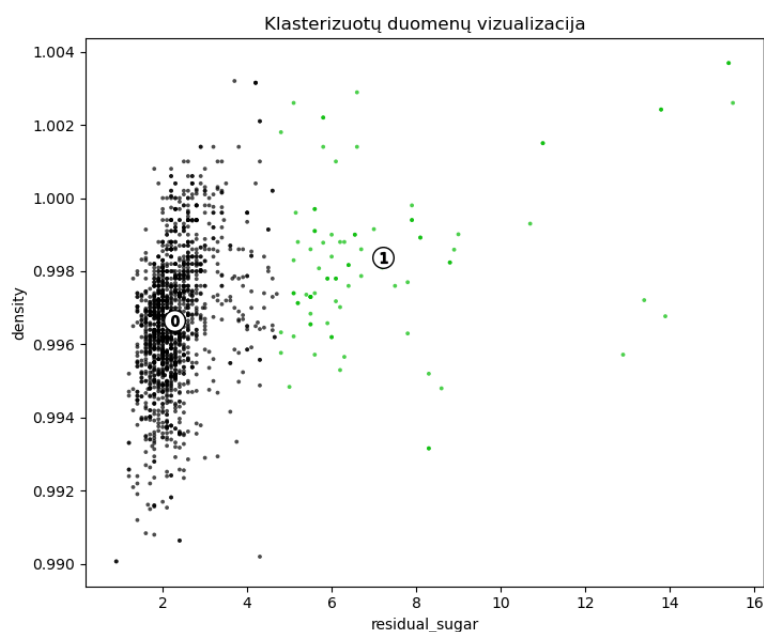
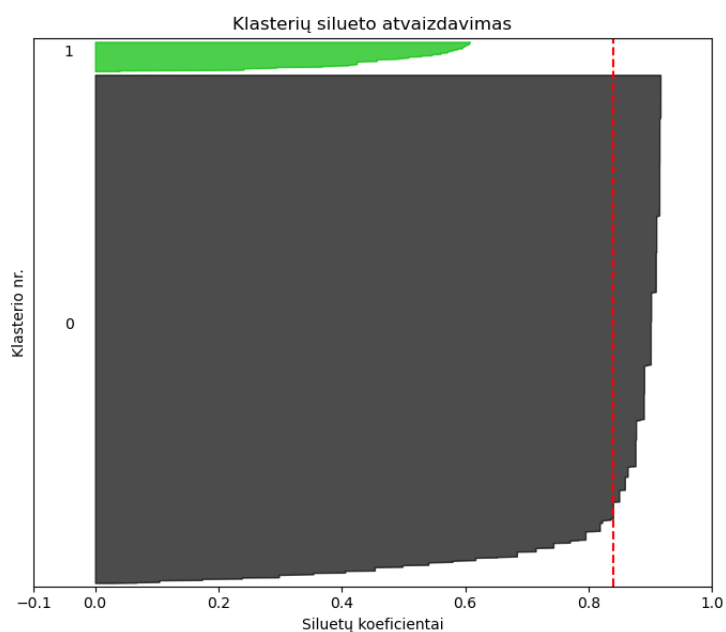
Klasteriai vos ne identiškai naudojant tris klasterius, tik iš antro klasterio atsiskyrė maža dalis sudaryti ketvirtajam.



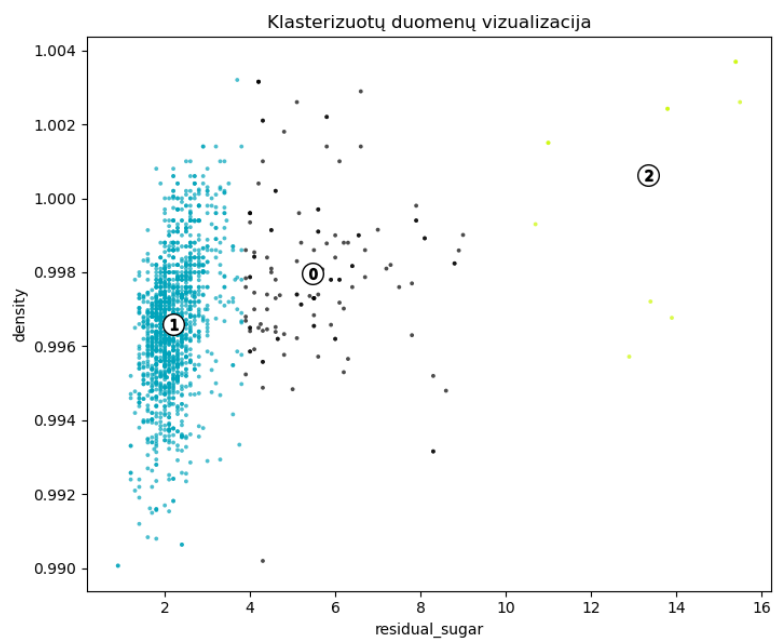
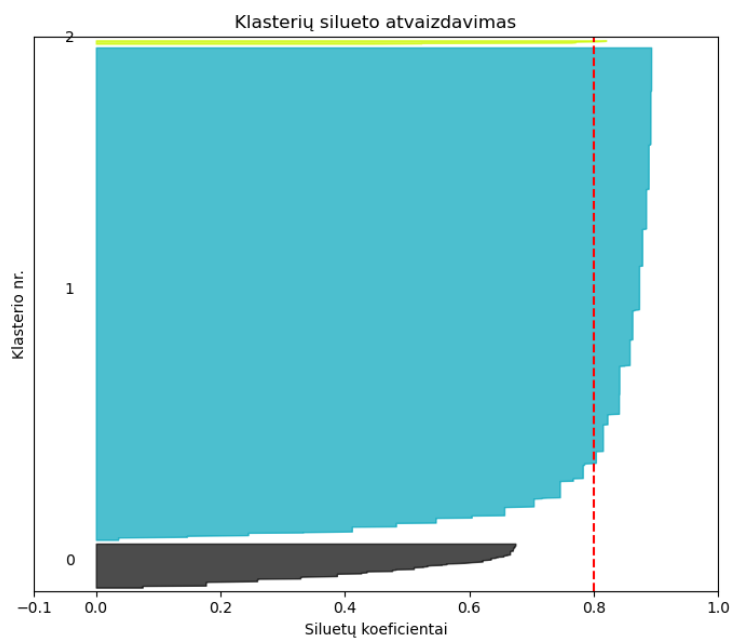
pav. 8 Klasterių siluetų įvertis

Naudojant silueto vidurkio metodą išsiaiškinau, kad šiuo atveju optimaliausia yra naudoti du klasterius.

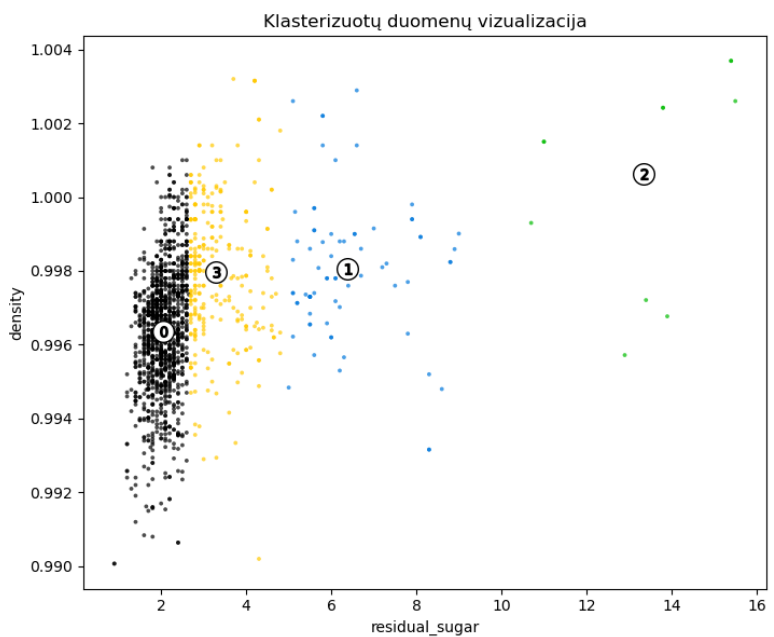
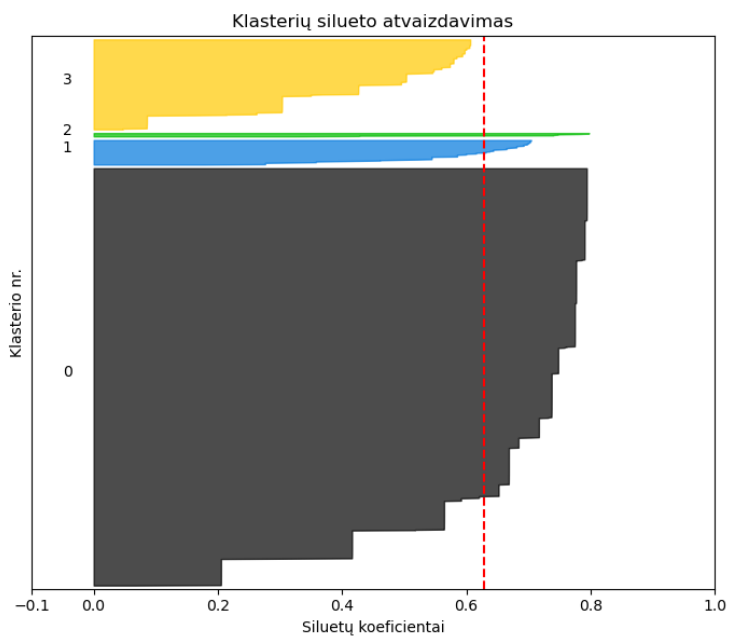
2.3 Likęs cukrus ir tankis



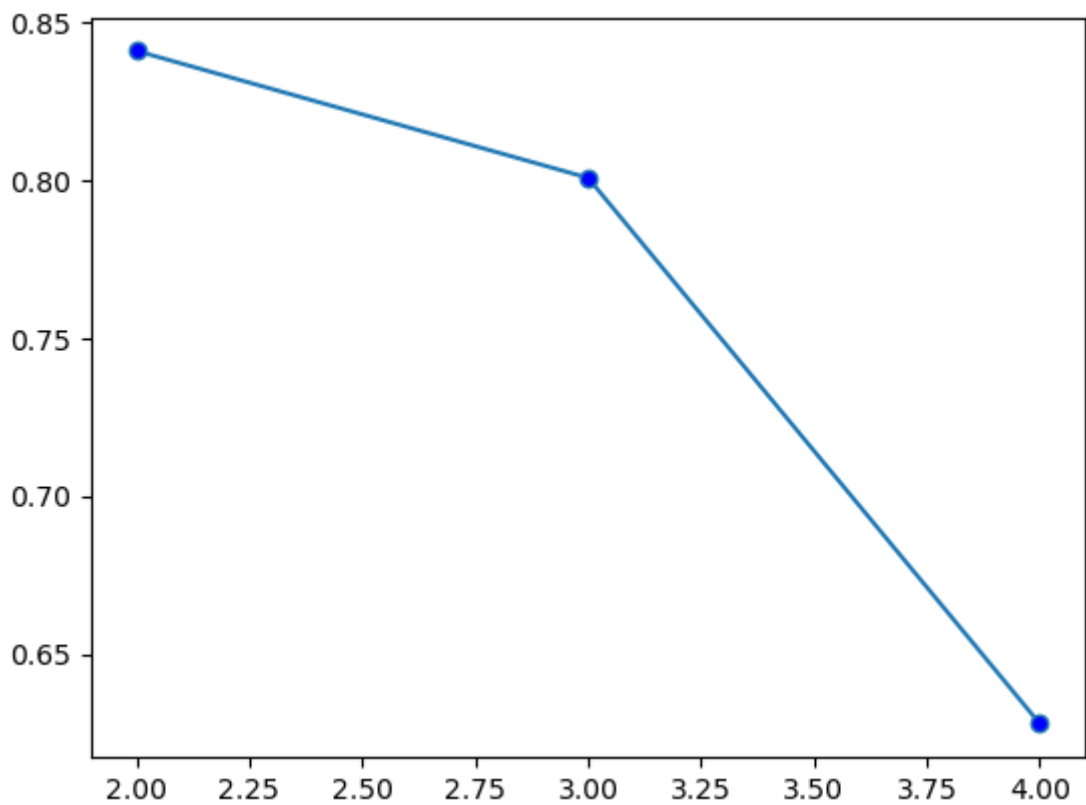
pav. 9 Likęs cukrus ir tankis - du klasteriai



pav. 10 Likęs cukrus ir tankis - trys klasteriai



pav. 11 Likęs cukrus ir tankis - keturi klasteriai



pav. 12 Klasterių siluetų įvertis

Optimaliausias klasterių kiekis likusiam cukrui ir tankiui: 2.

3. Išvados

Iš mano pasirinktų atributų geriausiai klasterizavosi likes cukrus ir tankis, kai buvo skirstoma į du klasterius. Blogiausiai – pastovus rūgštingumas ir likes cukrus. Šio laboratorinio darbo metu išmokau taikyti siluetų analizės metodą, kad rasti optimalų klasterių kiekį.