

2 Labaratorinis

Matas Gaulia, Vainius Gataveckas, Dovydas Martinkus

Naudotos bibliotekos

matplotlib

numpy

pandas

seaborn

sklearn

Duomenys

decade - dainos sukūrimo metų dešimtmetis

tempo - greitis

energy - energiškumas

danceability - šokamumas

loudness - garsumas

liveness - gyvumas

valence - pozityvumas

duration - trukmė

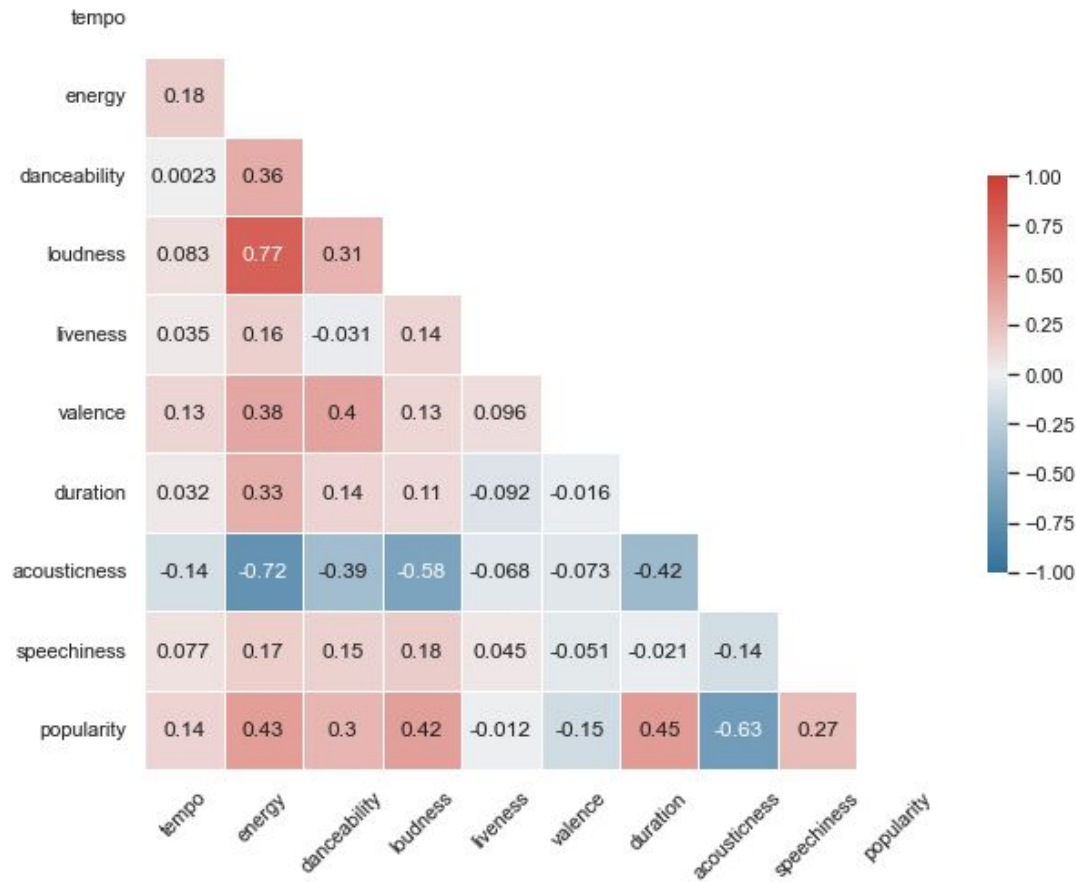
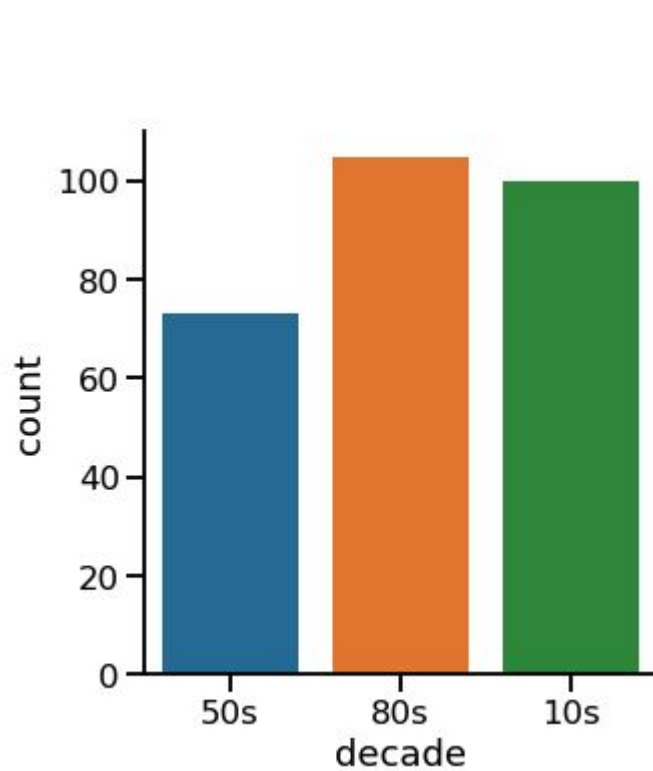
acousticness - akustika

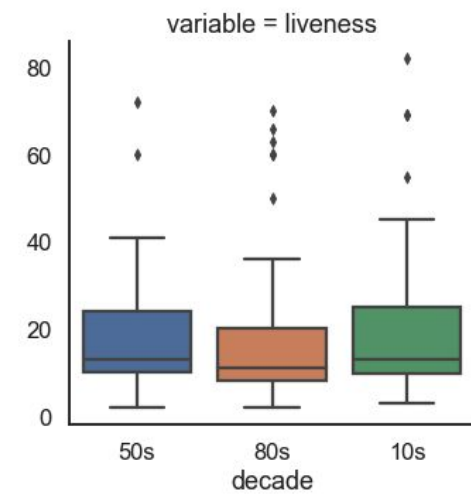
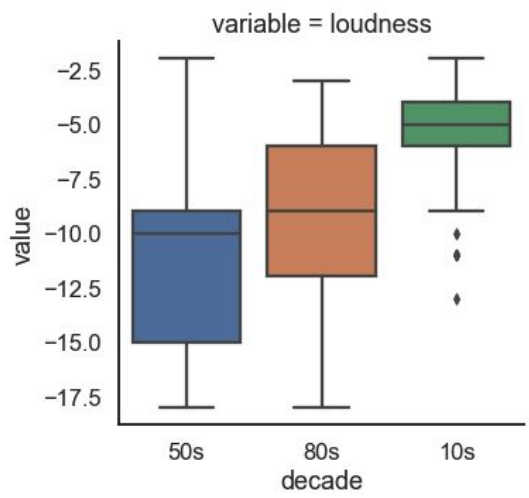
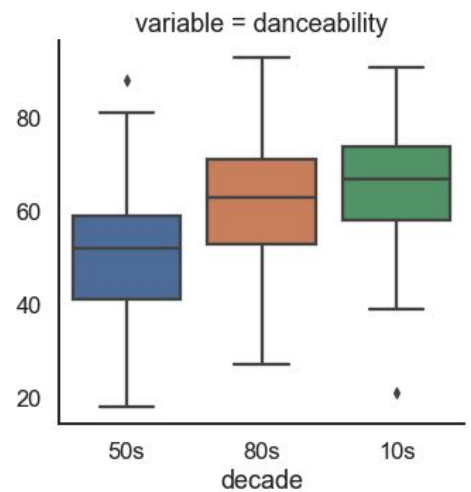
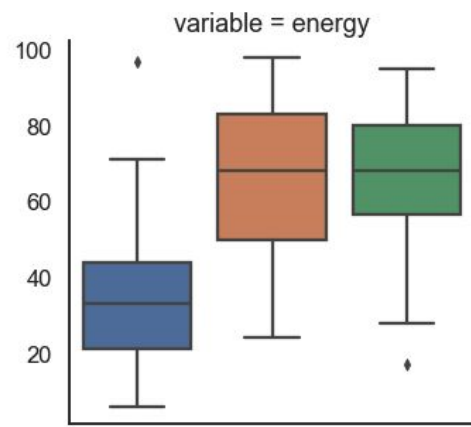
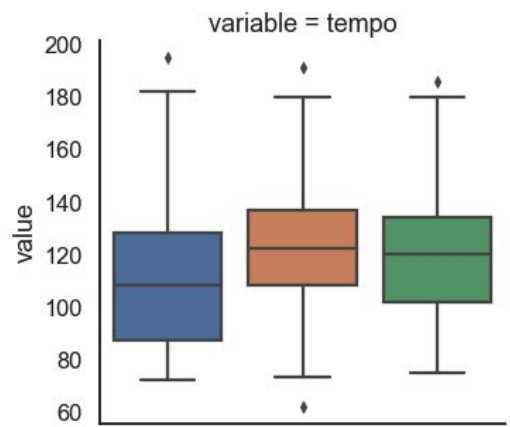
speechiness - žodžių kiekis dainoje

popularity - populiarumas

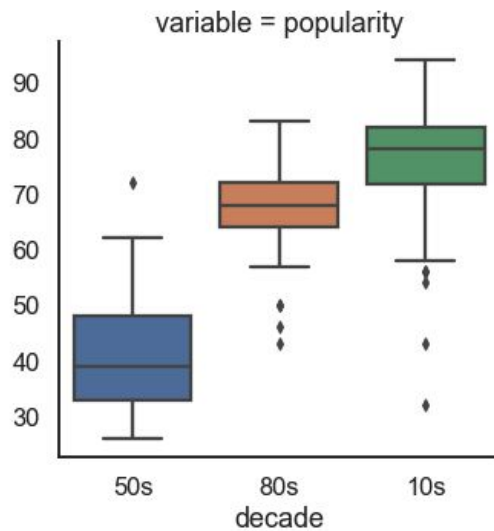
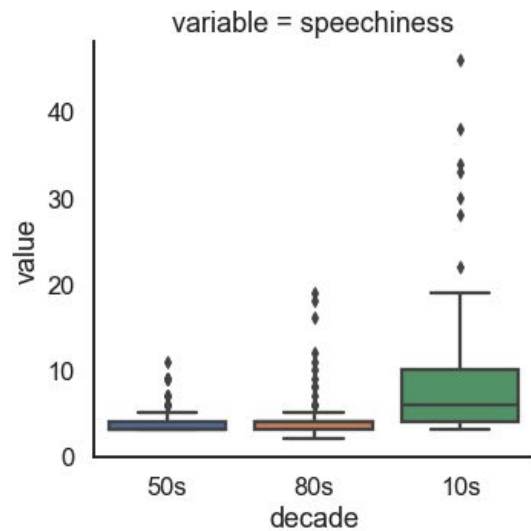
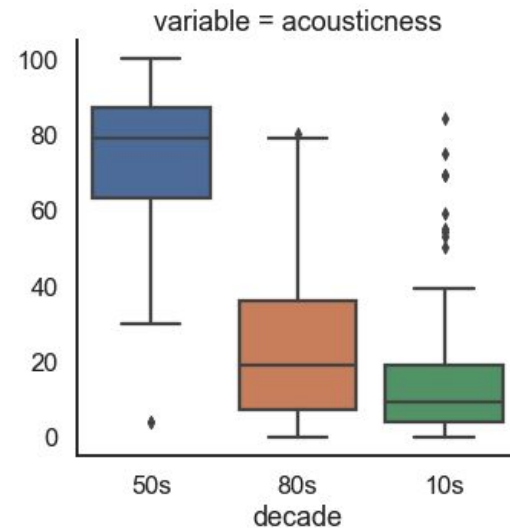
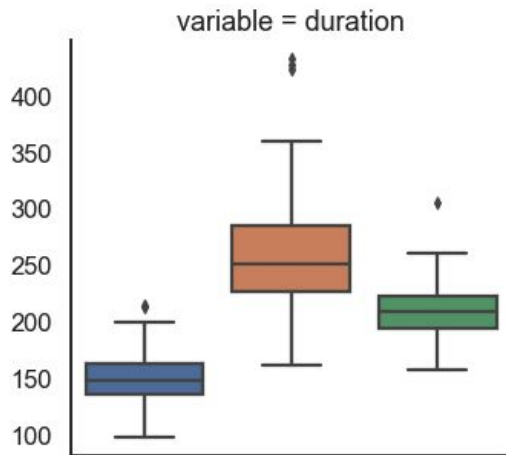
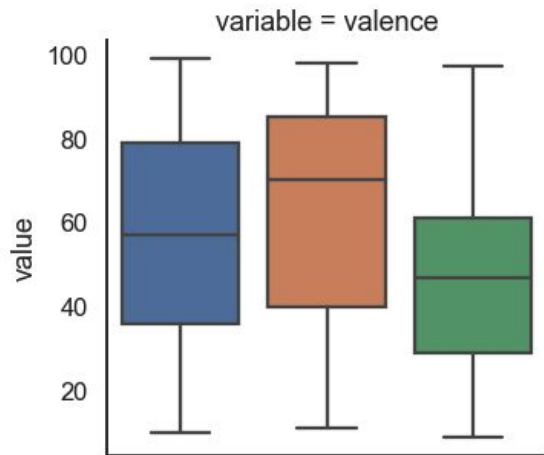
Iš duomenų aibės išskirtos trys grupės pagal dainų sukūrimo datą:

- 1940 - 1950-ųjų metų muzika
- 1970 - 1980-ųjų metų muzika
- 2000 - 2010-ųjų metų muzika





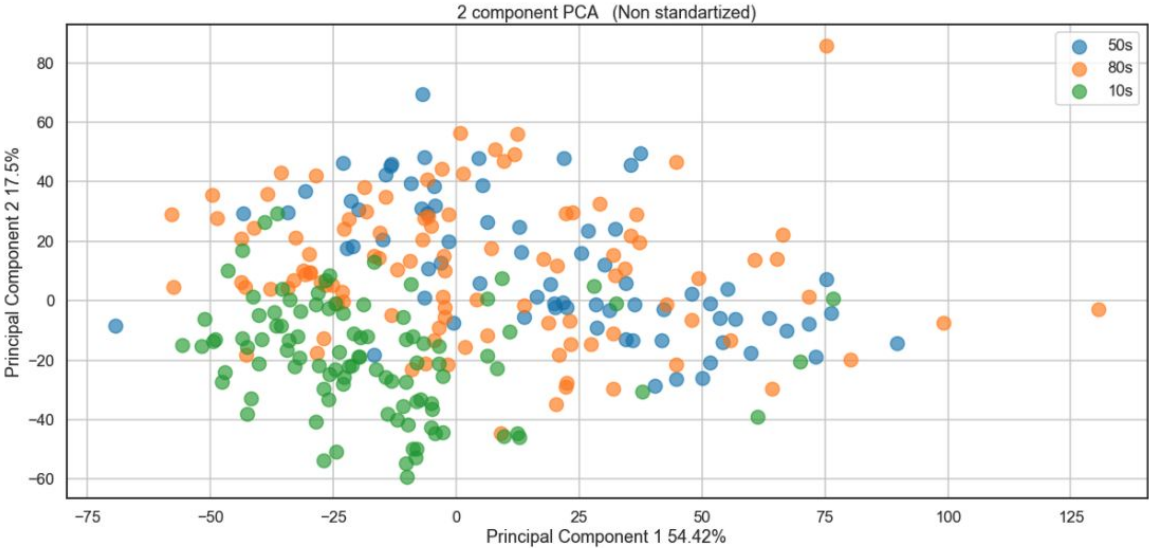
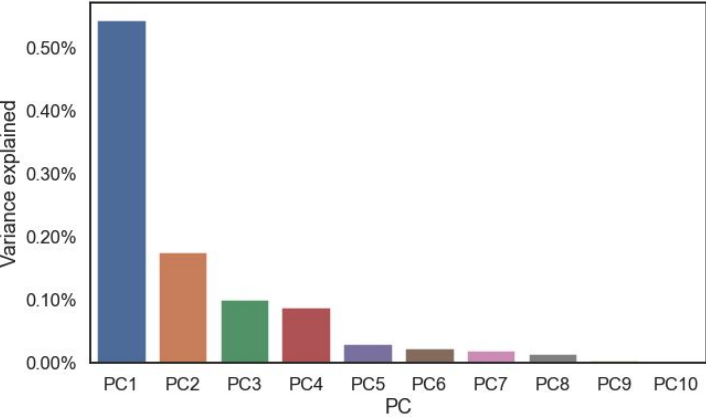
	mean
tempo	118.2
energy	58.1
danceability	60.4
loudness	-8.5
liveness	17.4



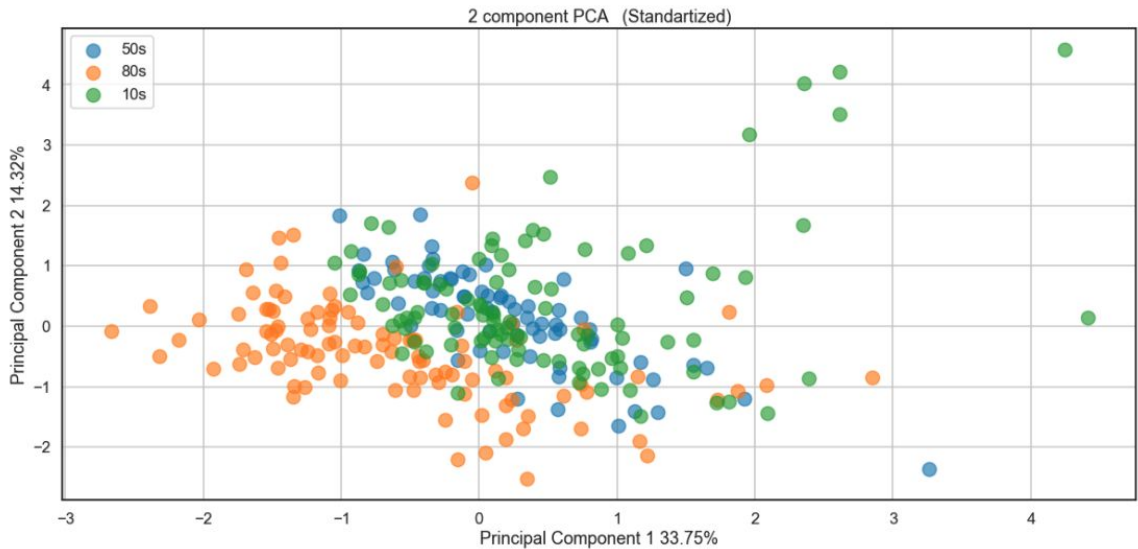
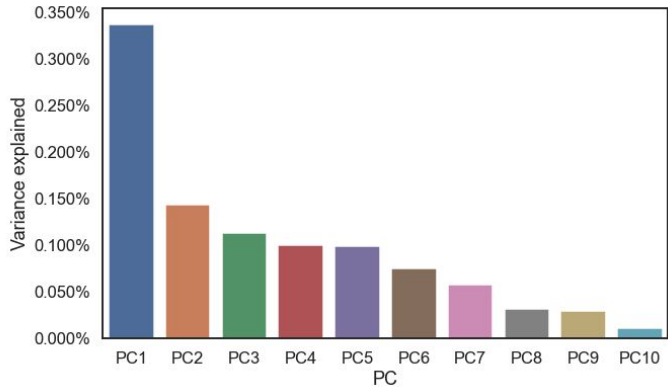
	mean
valence	55.6
duration	212.3
acousticness	33.7
speechiness	5.8
popularity	63.6

PCA

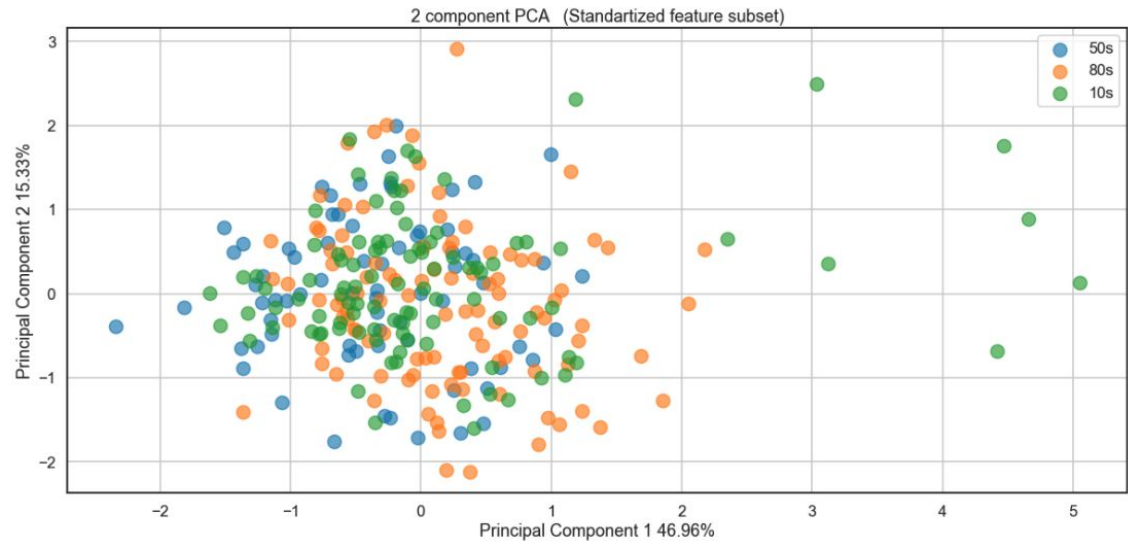
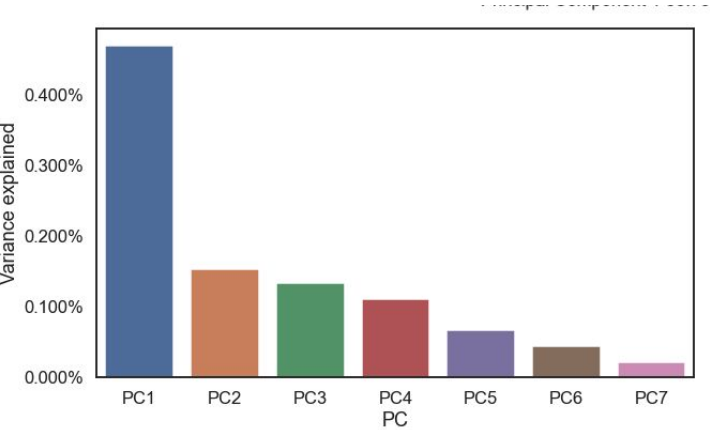
Nestandardizuoti duomenys



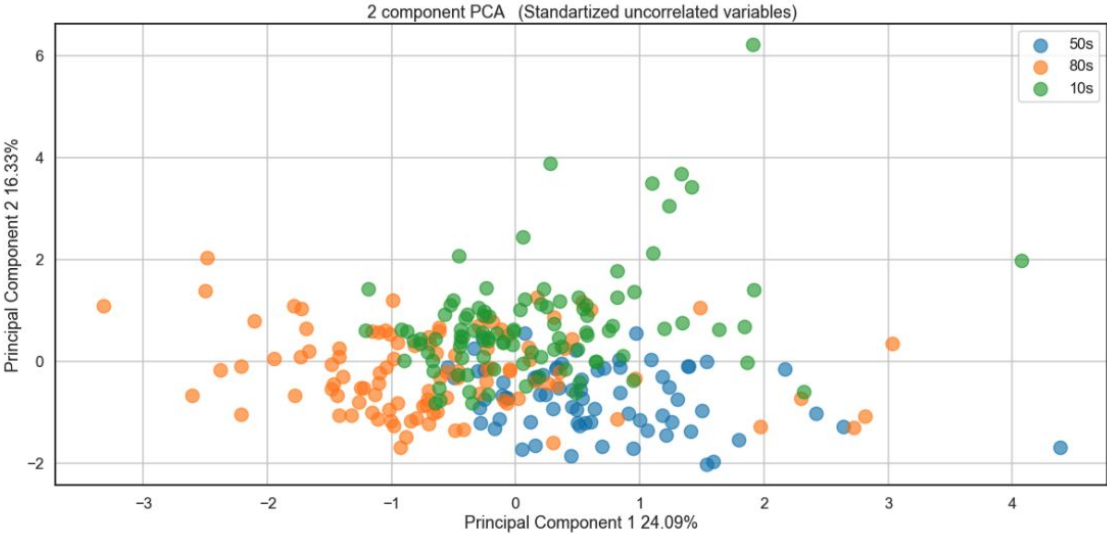
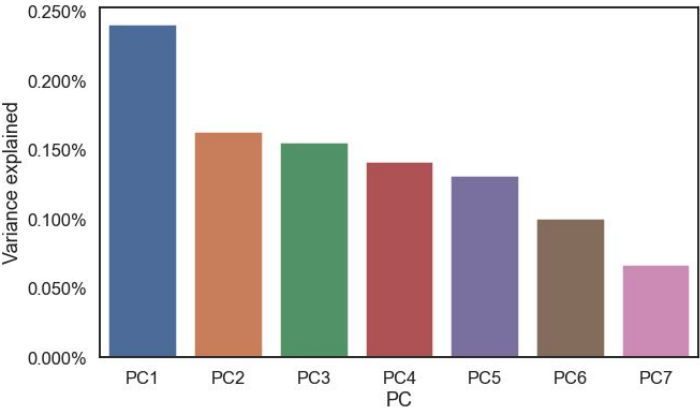
Standartizuoti duomenys



Standartizuoti duomenys be greičio, gyvumo, pozityvumo

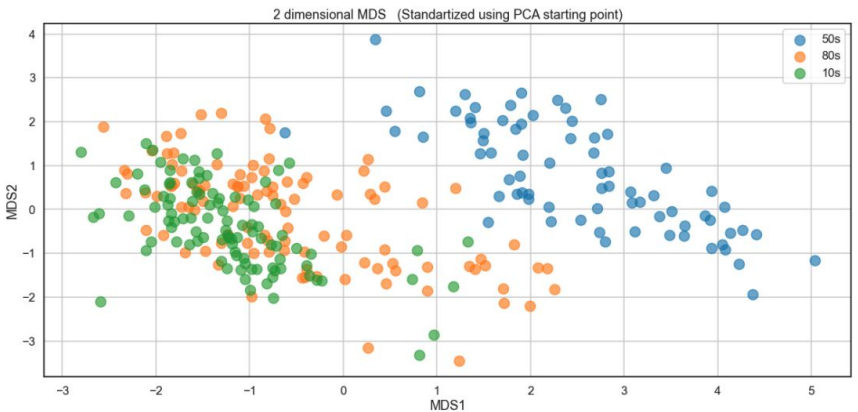
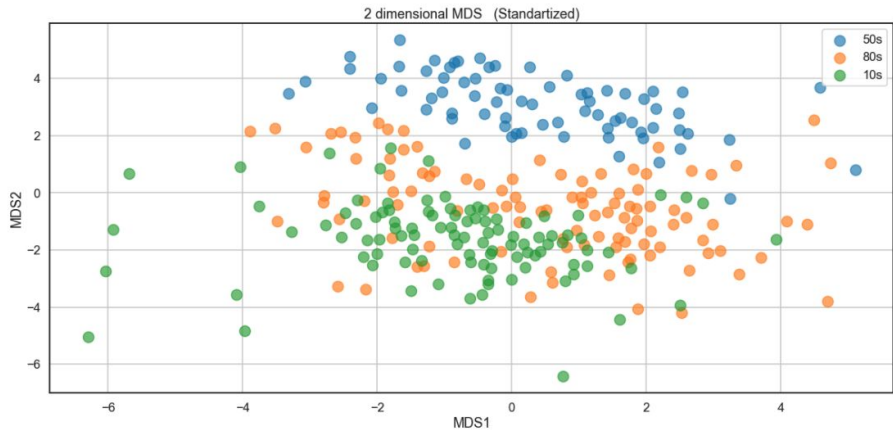
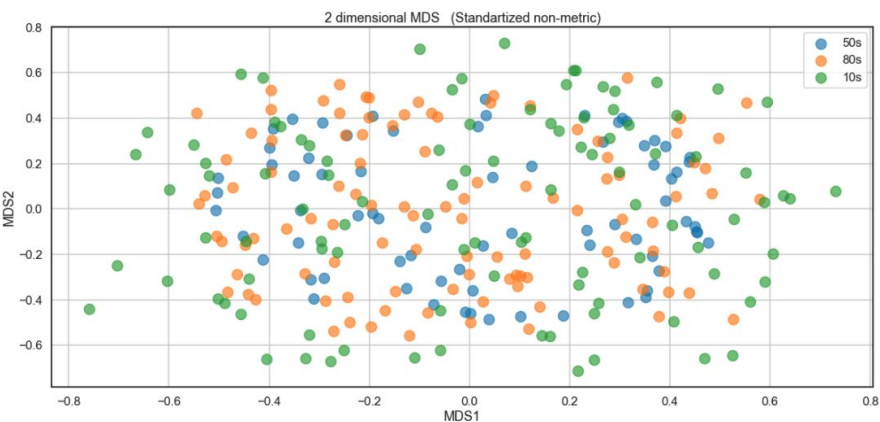
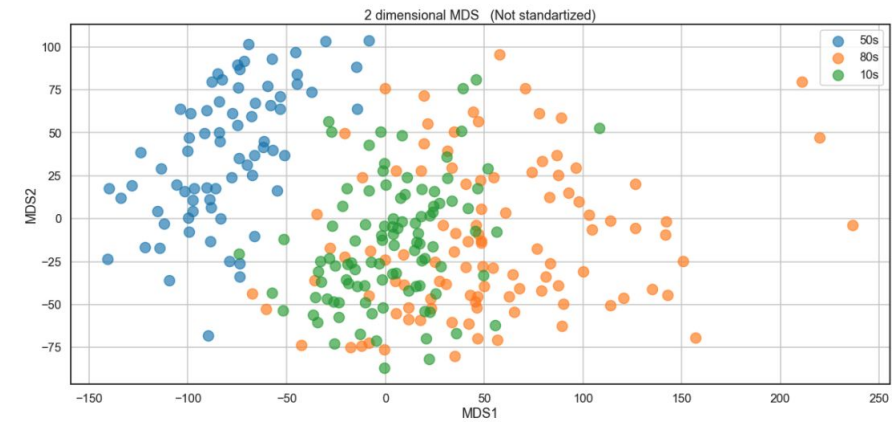


Standartizuoti duomenys be koreliuojančių kintamųjų
(akustika, energingumas, populiarumas)



MDS

Nestandardizueti duomenys

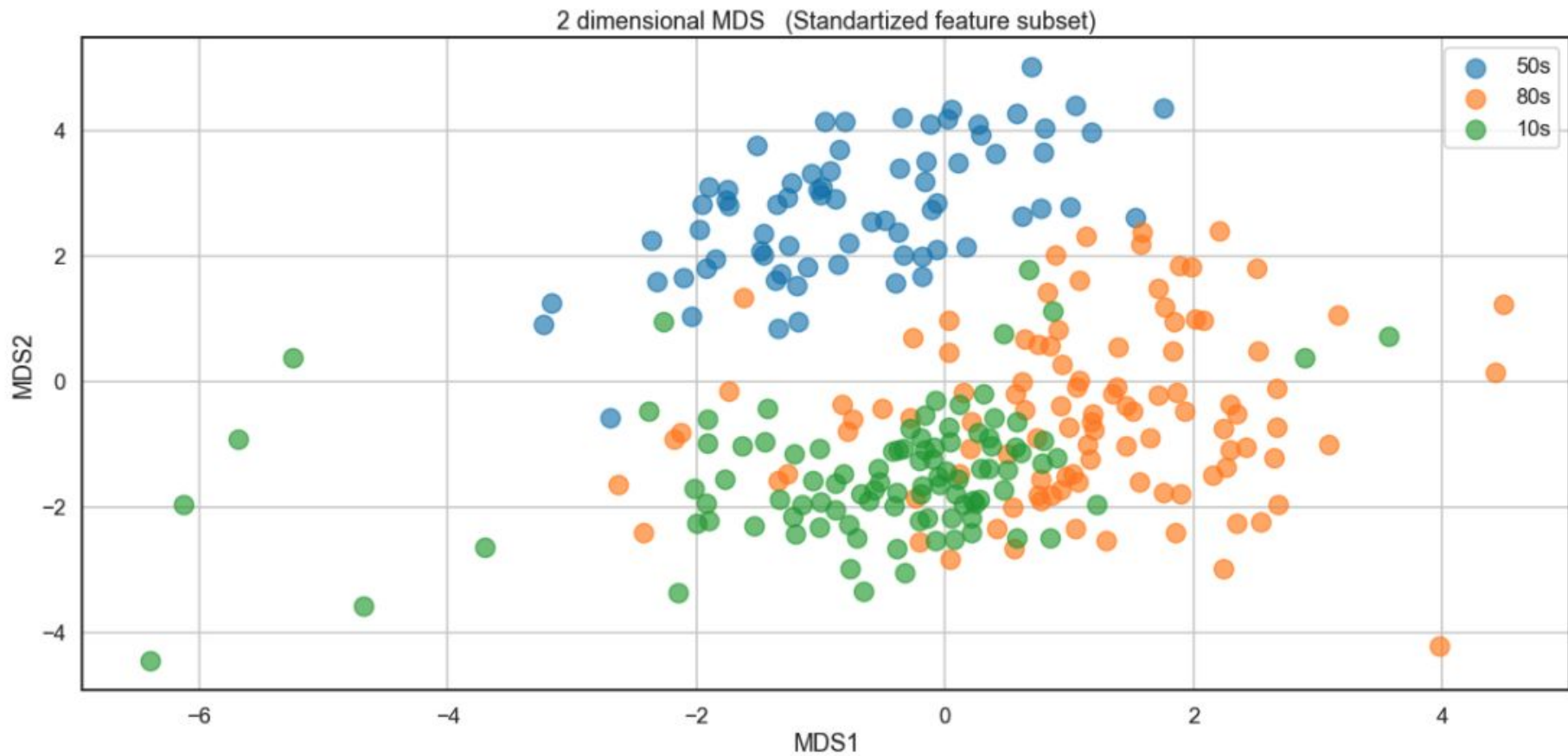


Standartizuoti duomenys, ne metrikinis MDS

Standartizuoti duomenys, metrikinis MDS

Standartizuoti duomenys, pradedama nuo PCA pradinės konfigūracija

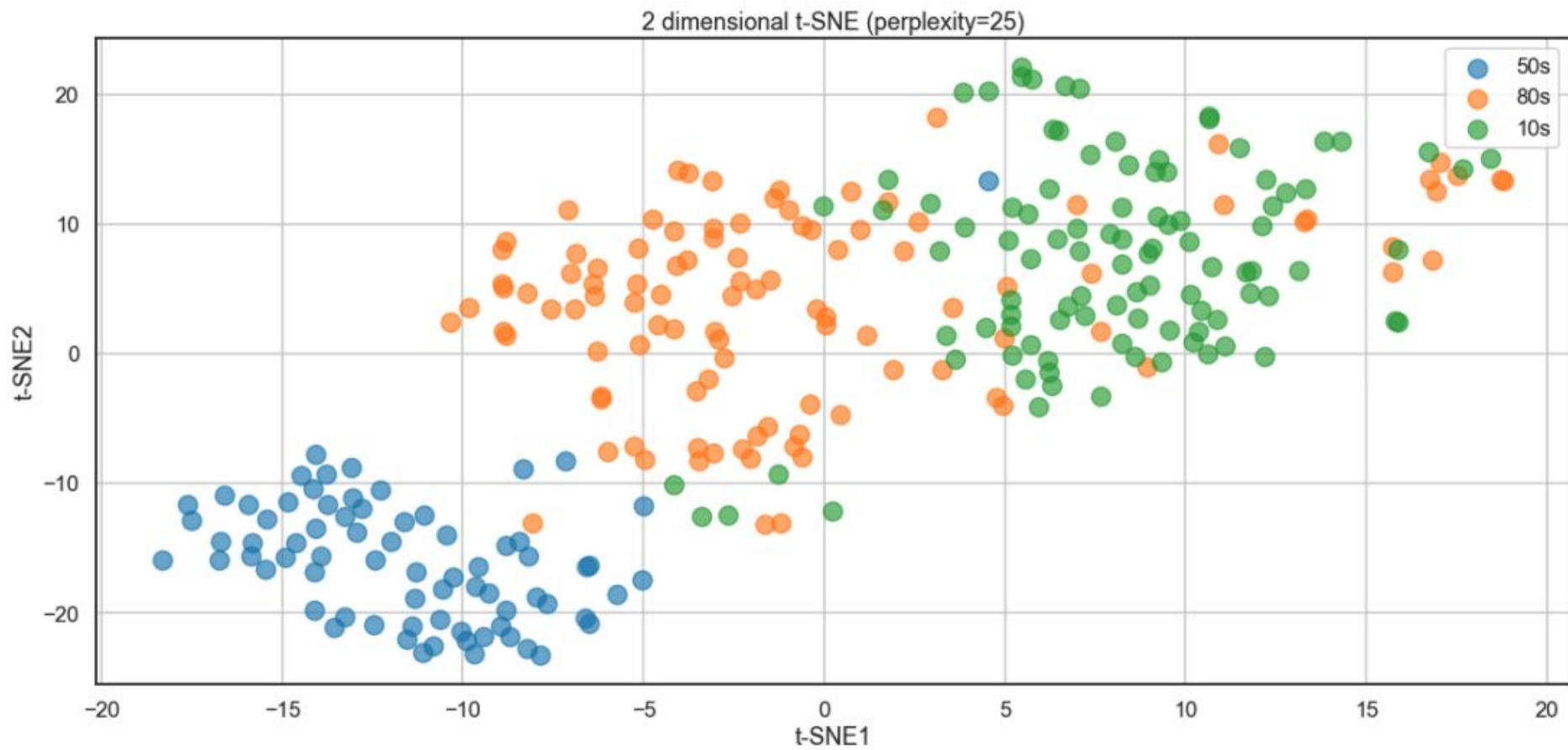
Standartizuoti duomenys be greičio, gyvumo, pozityvumo



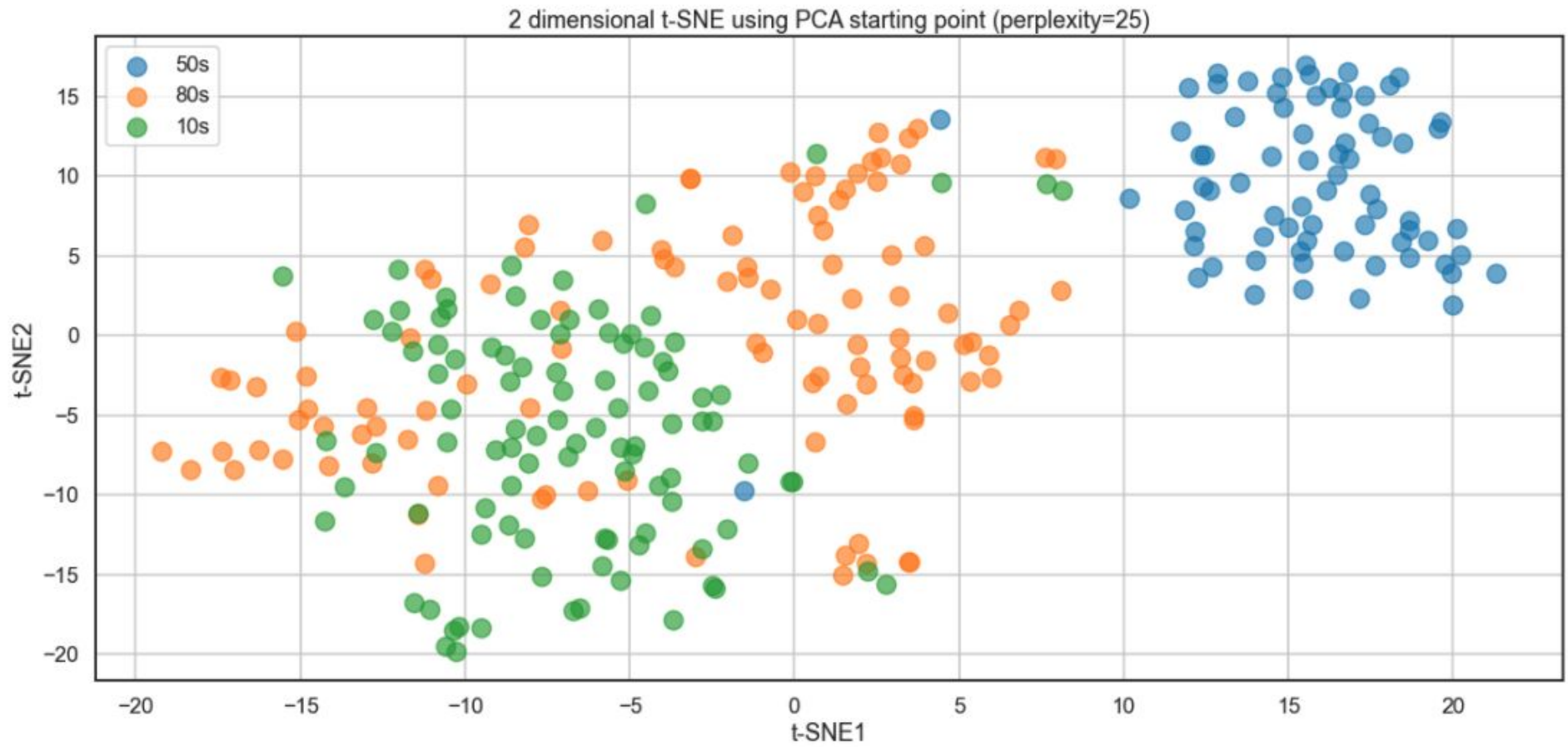
t-SNE

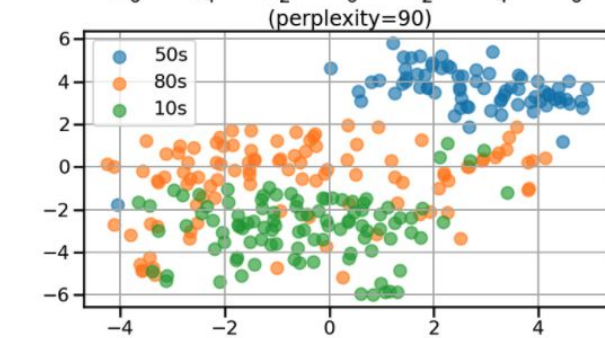
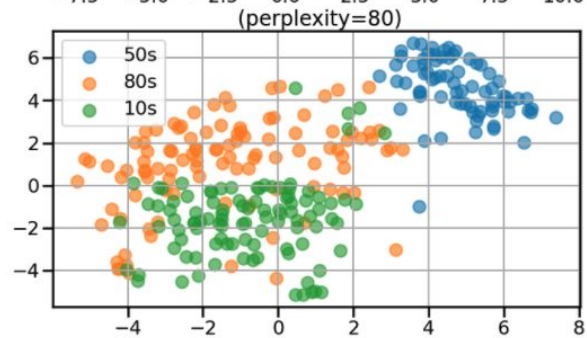
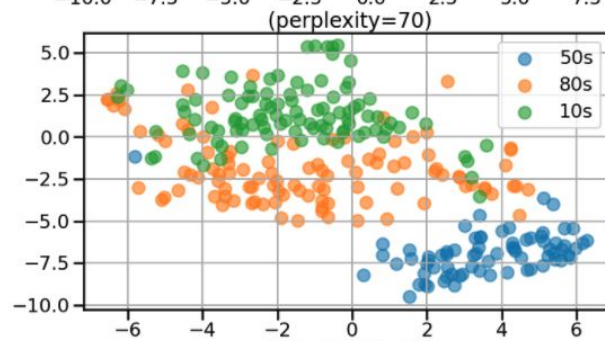
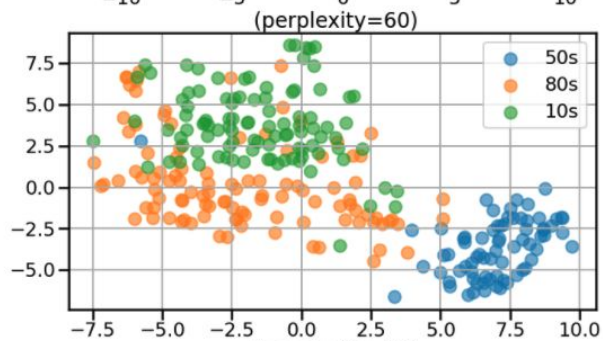
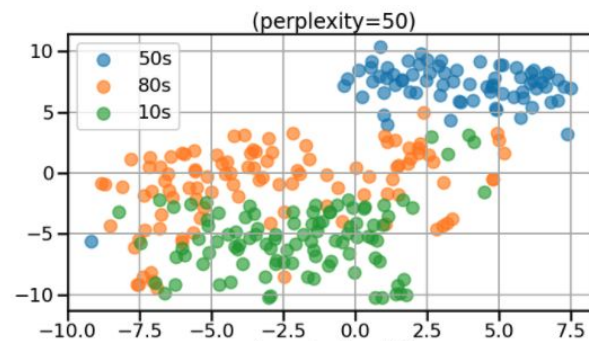
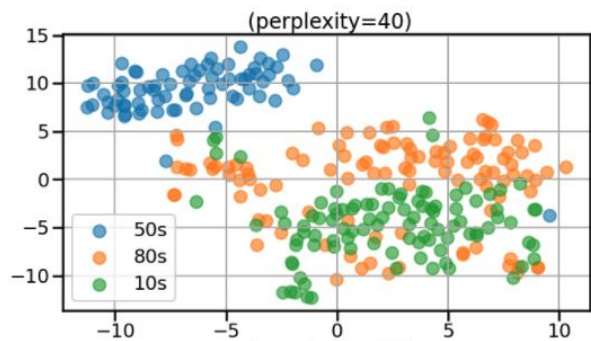
(visur standartizuoti duomenys)

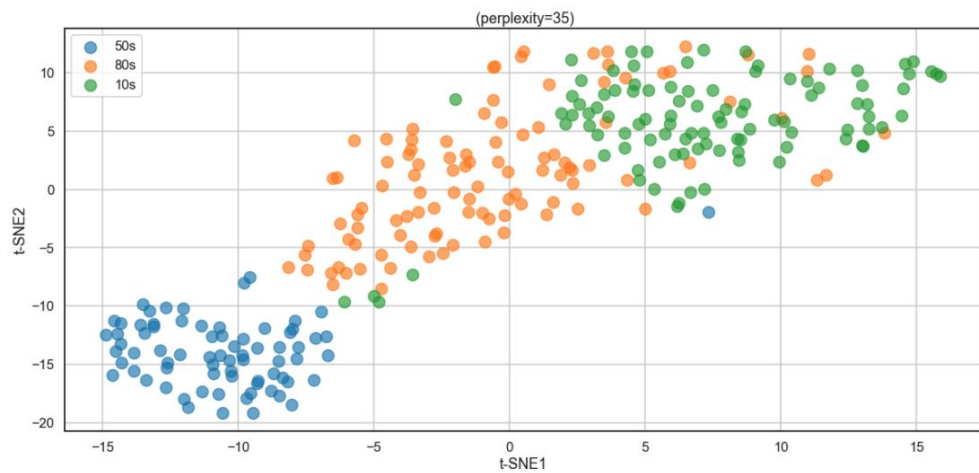
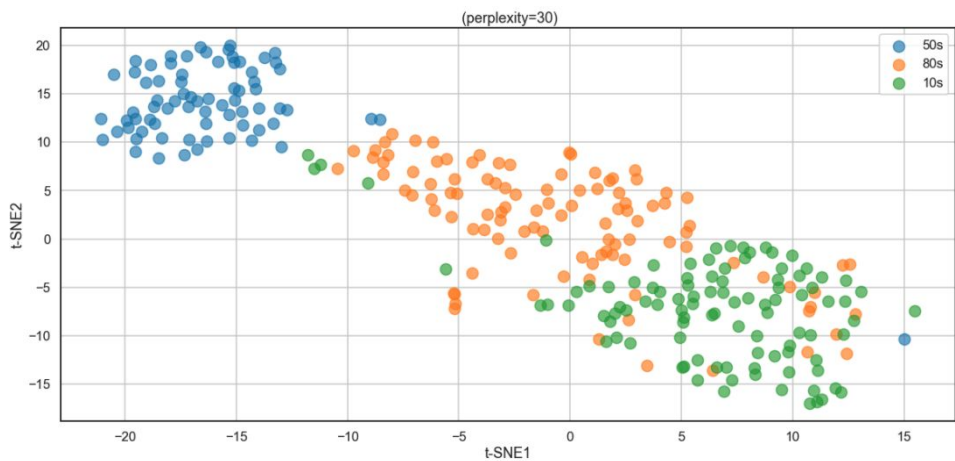
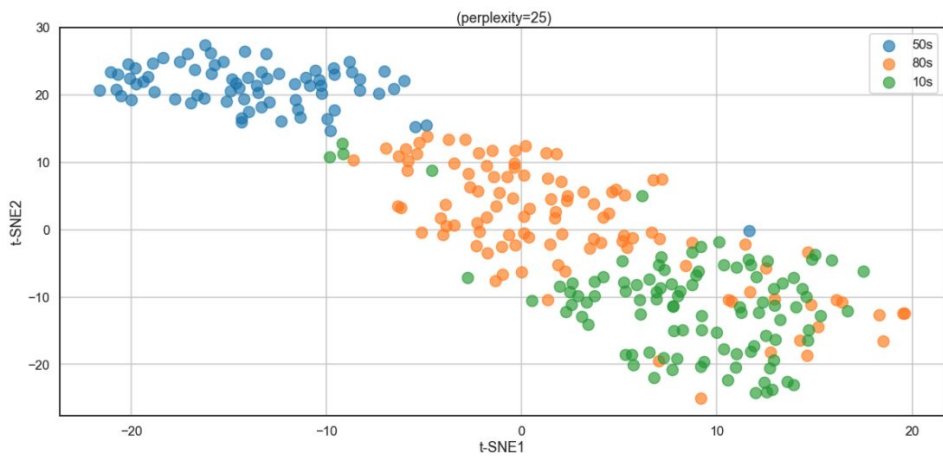
perplexity=25



perplexity=25, startuojama nuo PCA gautų koordinatų







Išvados

- Standartizuotų duomenų PCA metodo pirmosios dvi komponentės paaiškina 47% ir 15% variacijos.
 - t-SNE metodas geriausiai atskiria klases, kai "Perplexity" reikšmė yra (25;35) intervale.
 - Standartizuoti duomenys geriau atskiriami.
 - Panaikinus požymius, kurių empirinės charakteristikos nedaug skiriasi tarp grupių, buvo gauti geresni rezultatai.
-
- Metrikinis MDS metodas geriau atskiria grupes nei PCA
 - t-SNE vizualiai atskirdamas grupes pralenkia PCA ir MDS metodus