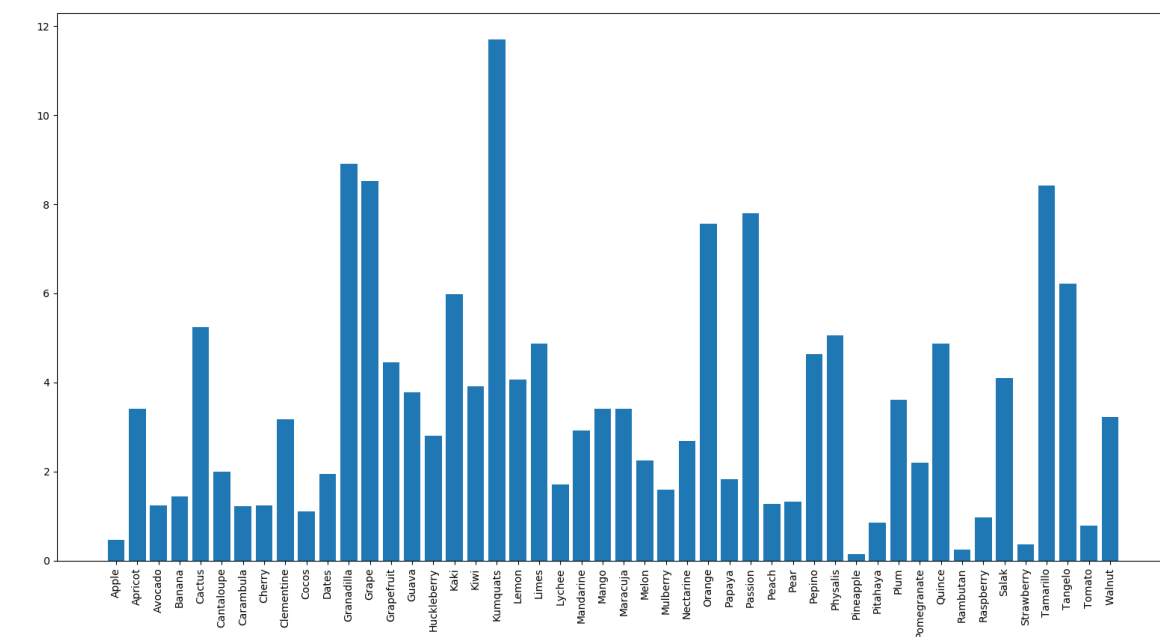


SUNS - Zadanie 2 dokumentácia

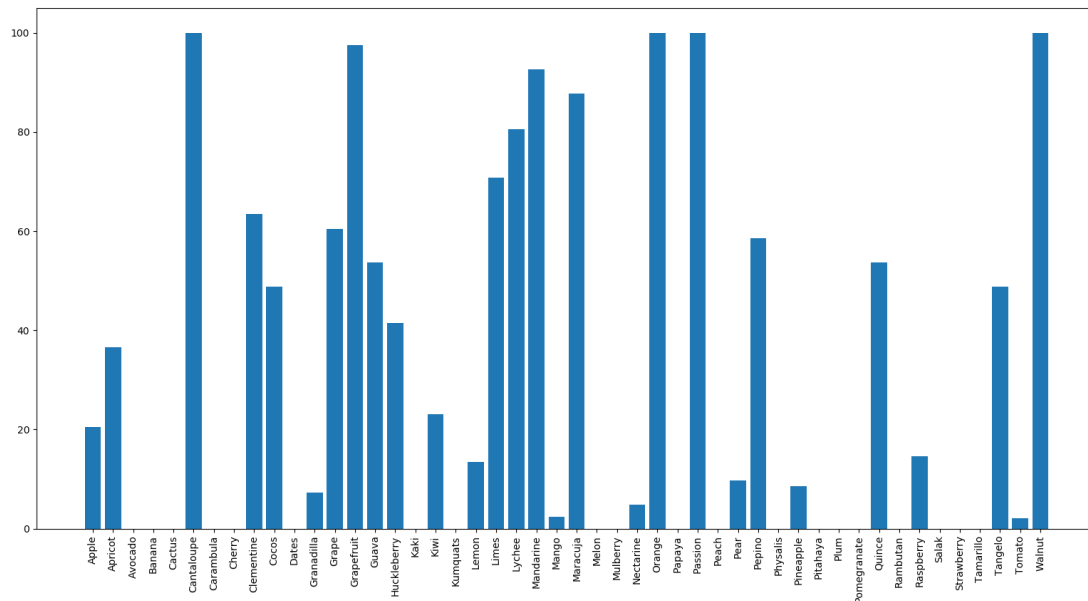
Funkcionalita celého zadanie bola implementovaná v knižnici MyLibs/clustering.py, ktorej funkcie sa starajú o analýzu predpripravených dát z prvého zadania.

V prvej časti behu program sa porovnávajú každý obrázok s každým v danej triede. Dáta sa porovnávajú pomocou euklidovskej vzdialenosti. Kvôli veľkému počtu dát v niektorých triedach bol počet obrázkov z každej triedy vynásobený koeficientom 0.25. Po porovnaní program vykreslí nasledujúci graf.



Tento graf nám znázorňuje koľko percent obrázkov z jednej triedy sa na seba podobajú podľa euklidovskej vzdialenosti. Z vybraných dát boli k sebe najpodobnejšie obrázky z triedy Kumquats a to necelých 12% z vybranej vzorky.

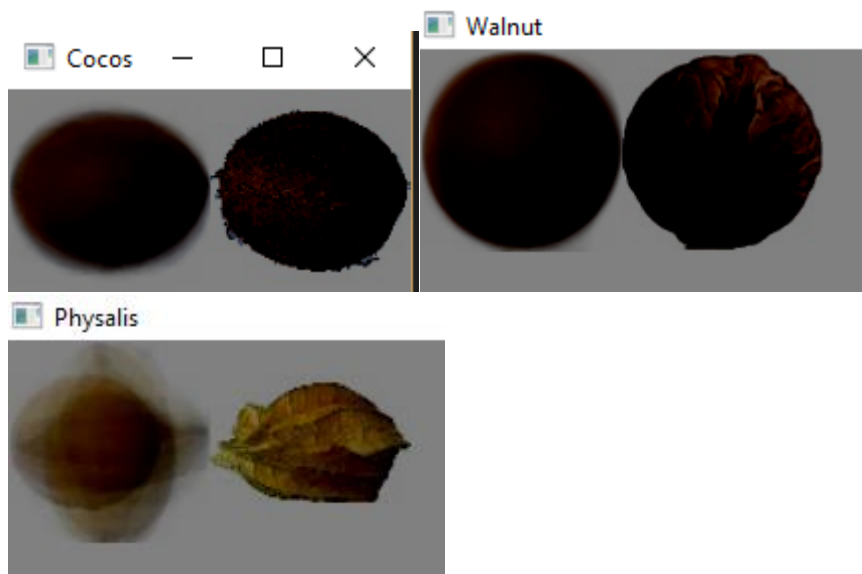
Následne program vytvorí pre každú triedu priemerný obrázok a podobným spôsobom ich porovná s každým obrázkom v danej triede a vykreslí nasledujúci graf.



Ako môžeme vidieť veľa tried sa vôbec nepodobá na svoj priemerný obrázok a veľa tried dosiahlo 100 percentnú úspešnosť. Toto mohla spôsobiť normalizácia zo zadanie č.1

V ďalšej časti program prešiel nachystané testovacie dáta pomocou zhukovacích algoritmov K-Means a DBScan

Pre rýchlosť testovanie bol kMeans nastavený len na 5 zhukov. Z Každého zhuku nájde najbližší obrázok k priemernému obrázku, ktorý je výstupom z algoritmu a následne vykreslí priemerný a najbližší:



Vstupné dáta pre DB scan boli upravené rovnako ako pre k-means. Po tom ako nám algoritmus vráti všetky zhľady vstupov tak nájde k nim najbližší a vykreslí rovnako ako v predchádzajúcom prípade. Pri aktuálnej konfigurácii algoritmus našiel 36 zhľadov zo 48 tried

