

Projekt podatkovnih kart in analize slik

Poročilo

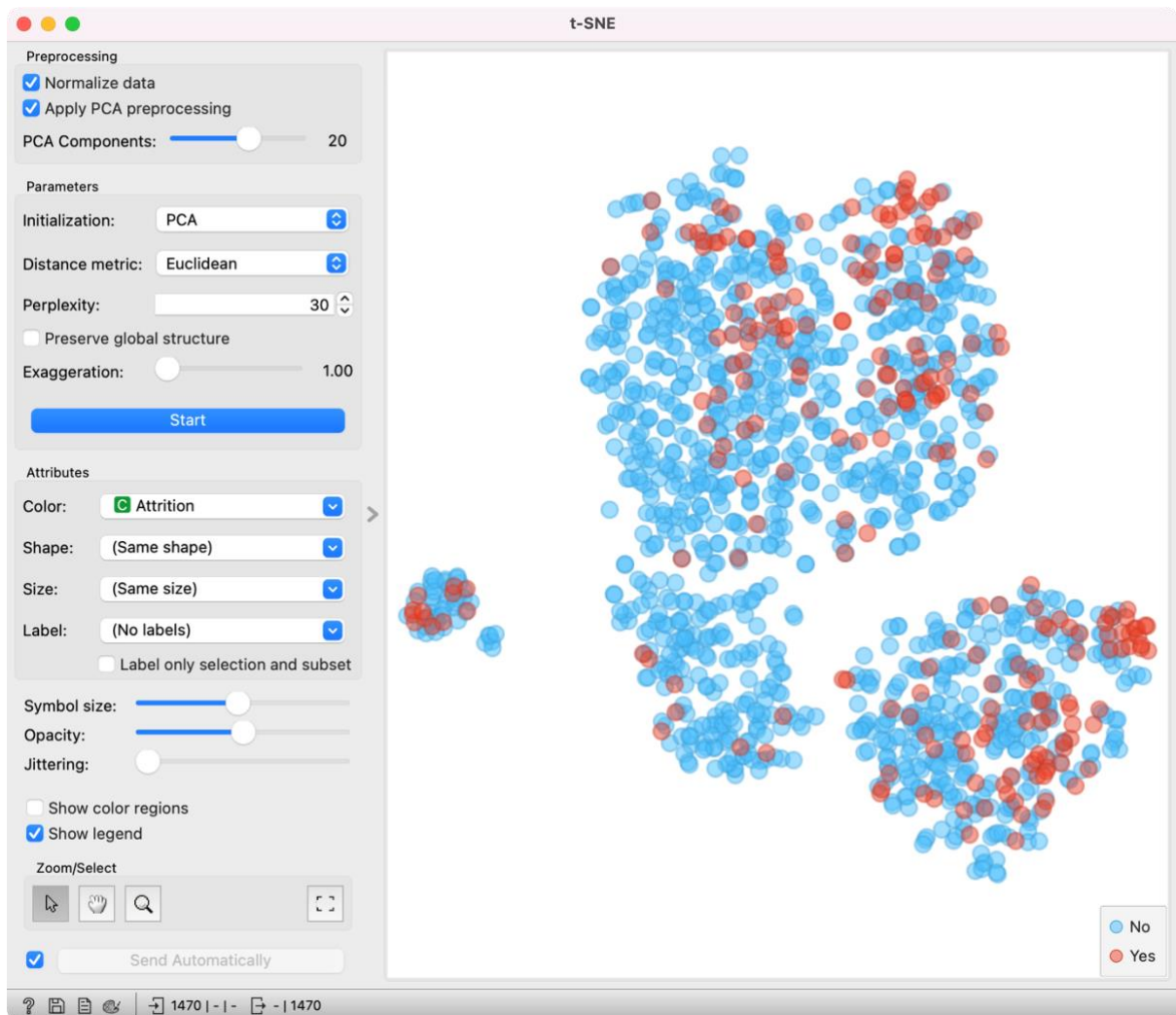
Ajda Pretnar Žagar
16. 5. 2025

Podatkovne karte

1. Podatki o odhodu zaposlenih iz podjetja (Employee Attrition) zajemajo 1470 primerov ter 33 spremenljivk. Zaposleni so opisani s starostjo, pogostostjo službenih potovanj, dnevno postavko, oddelkom itd. Posebej označena je spremenljivka o odhodu zaposlenega iz podjetja, ki je binarna.

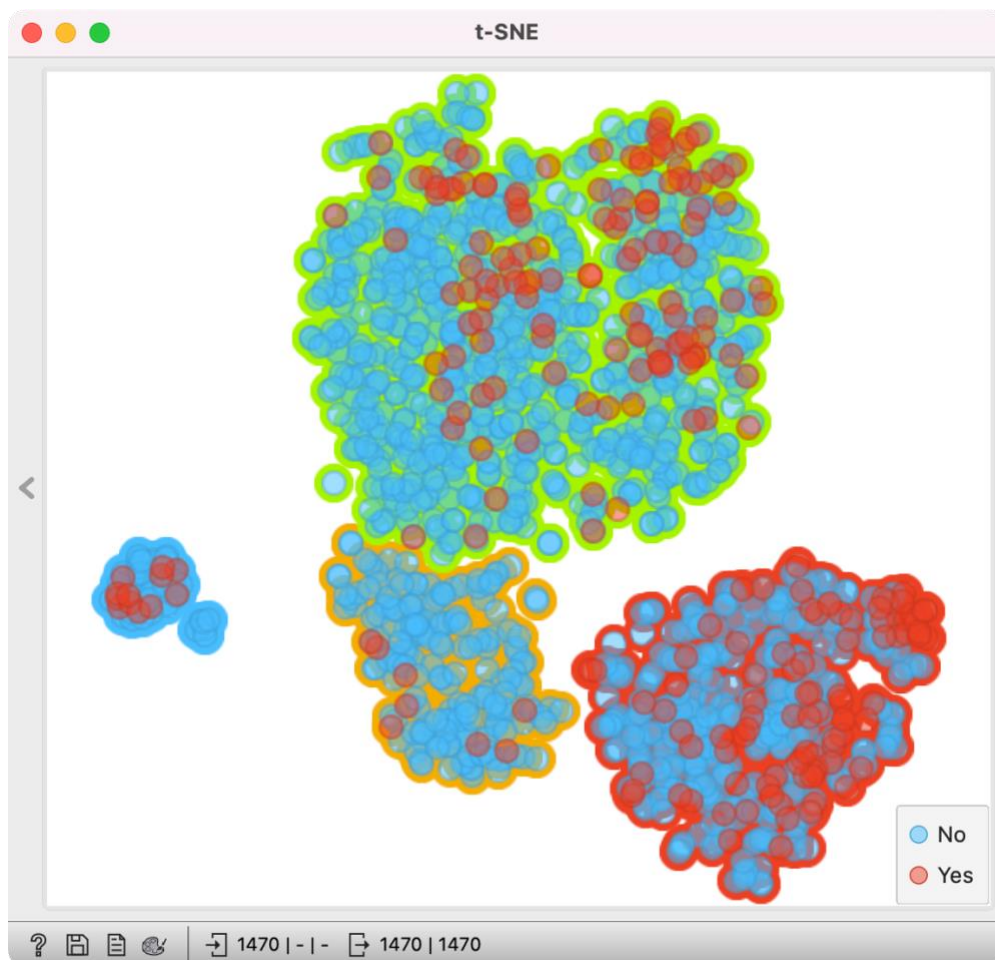
Data Table								
	Attrition	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
1	Yes	41	Travel_Rarely	1102	Sales	1	2	Life Sciences
2	No	49	Travel_Frequ...	279	Research & De...	8	1	Life Sciences
3	Yes	37	Travel_Rarely	1373	Research & De...	2	2	Other
4	No	33	Travel_Frequ...	1392	Research & De...	3	4	Life Sciences
5	No	27	Travel_Rarely	591	Research & De...	2	1	Medical
6	No	32	Travel_Frequ...	1005	Research & De...	2	2	Life Sciences
7	No	59	Travel_Rarely	1324	Research & De...	3	3	Medical
8	No	30	Travel_Rarely	1358	Research & De...	24	1	Life Sciences
9	No	38	Travel_Frequ...	216	Research & De...	23	3	Life Sciences
10	No	36	Travel_Rarely	1299	Research & De...	27	3	Medical

2. Gradnik t-SNE s privzetimi nastavitvami in točkami, obarvanimi glede na odhod zaposlenega.

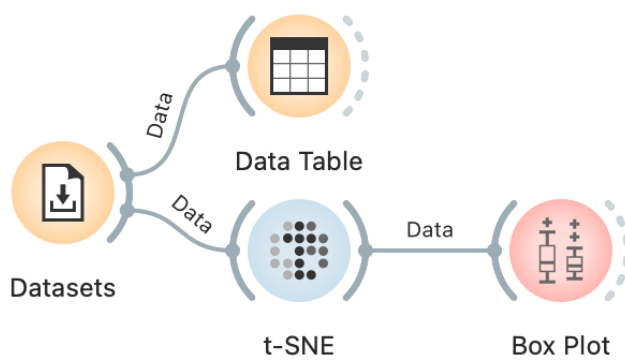


3. Gruča na skrajni levi predstavlja ljudi iz kadrovske službe (Human Resources). Gruča na skrajno desni (spodaj) predstavlja zaposlene iz oddelka prodaje. Večjo gručo na sredini sem razdelila na dve podgruči, večjo gručo zgoraj in manjšo spodaj. Večja gruča zajema tehnične poklice (Healthcare Representative, Laboratory Technician, Manufacturing Director, Research Scientist), medtem ko manjša gruča predstavlja vodje

(Manager, Research Director).



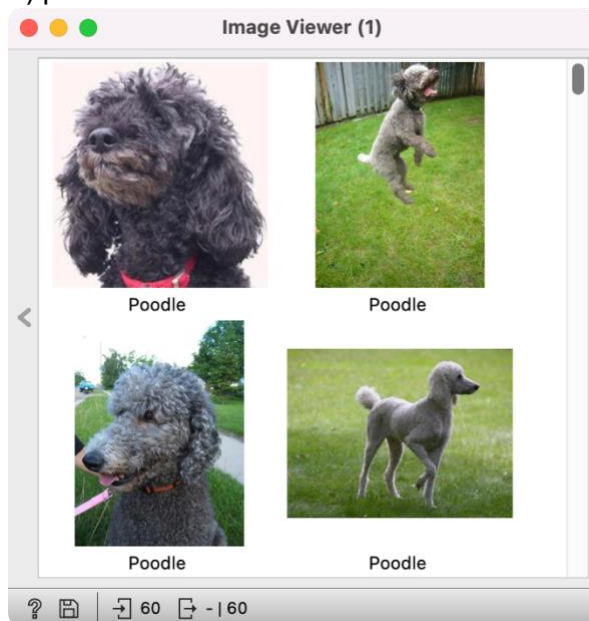
4. Prikaz končnega delotoka.



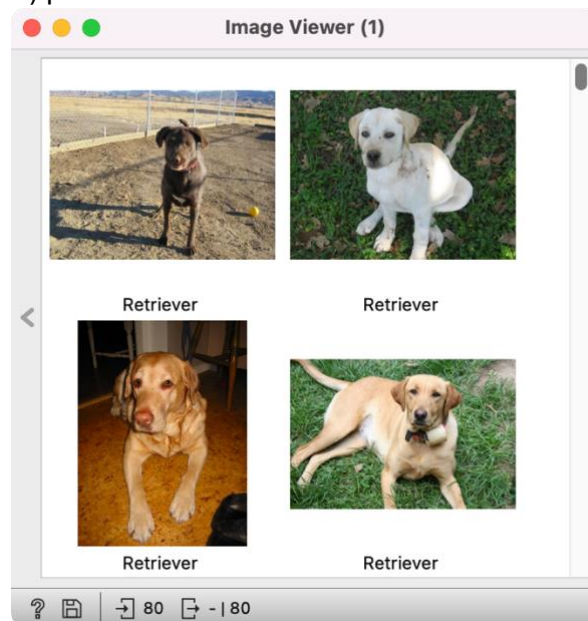
Analiza slik

1. Podatki zajemajo 260 slik psov, ki so razvrščene v štiri kategorije. Kategorije predstavljajo pasme, in sicer pudle (Poodle), prinašalce (Retriever), setre (Setter) ter terierje (Terrier).

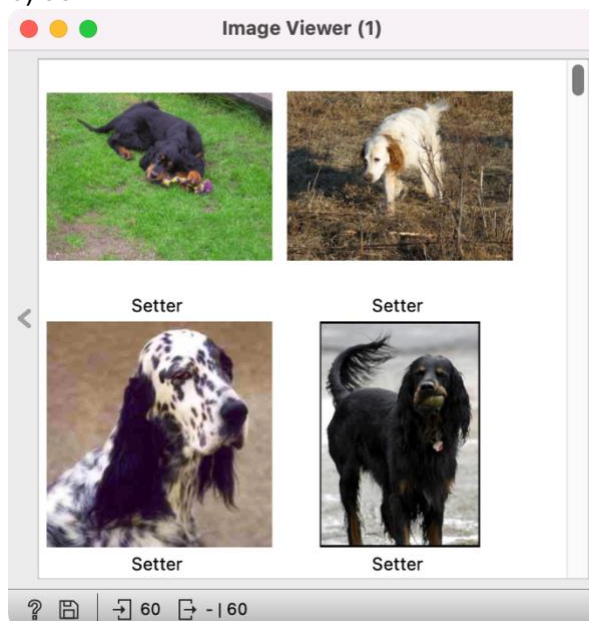
a) pudli



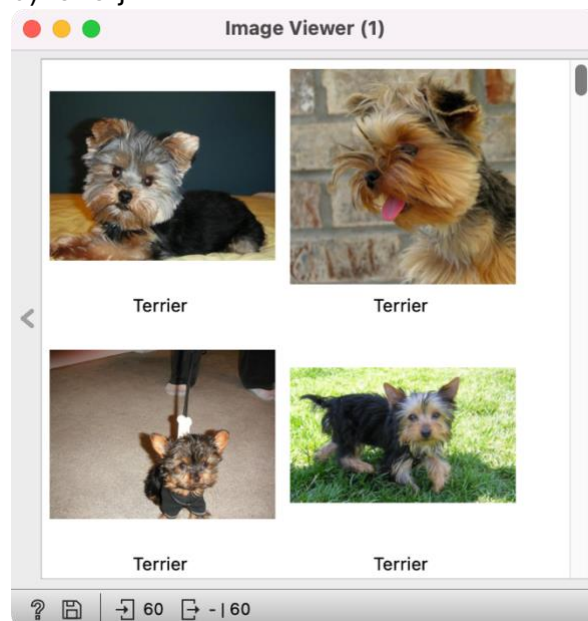
b) prinašalci



c) setri



d) terierji



2. Slike predstavimo s številskim opisom tako, da podatke pošljemo v gradnik **Image Embedding**. Uporabila sem model Inception v3. Ta vrne številске opise slik z 2048 dimenzijami.

Data Table

Info
260 instances (no missing data)
2048 features
Target with 4 values
5 meta attributes

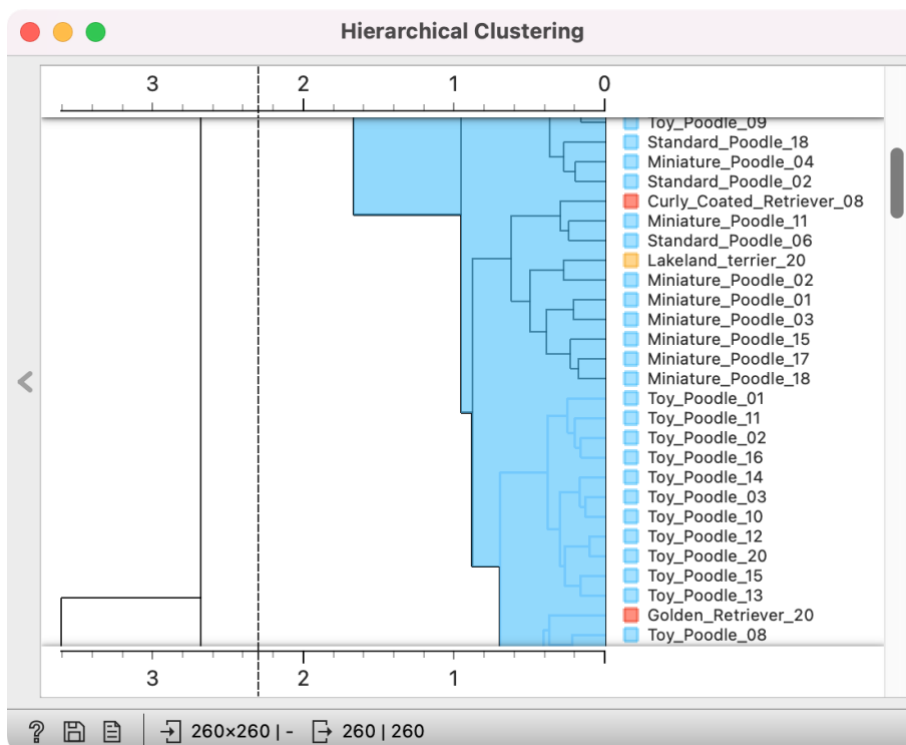
Variables
☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

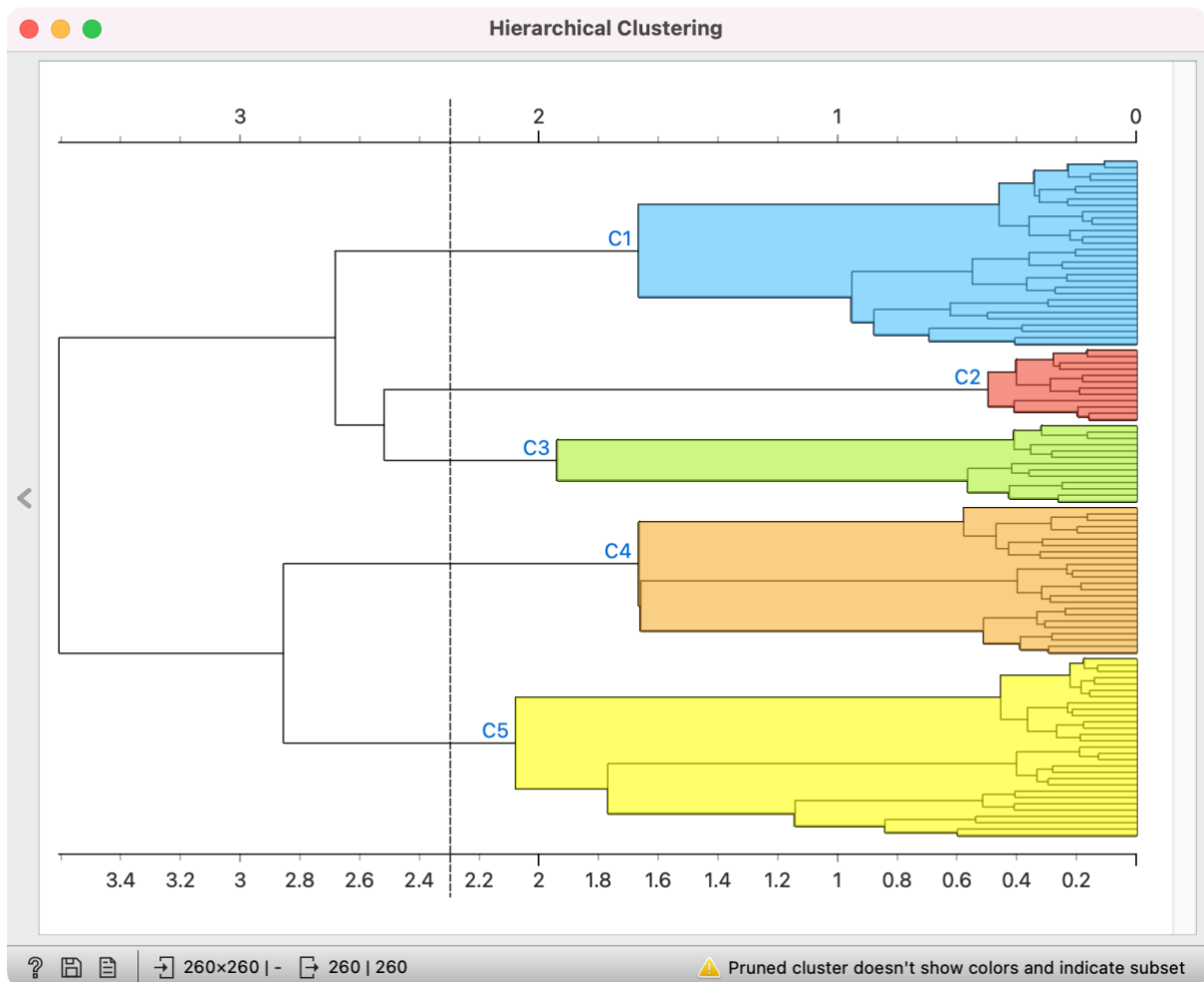
Selection
☒ Select full rows
Restore Original Order
☒ Send Automatically

hidden origin type	category	image name	image ownloads/Images, image	size	width	height	n0 True	n1 True	n2 True
1	Setter	Gordon_Sett...	Setter/Gordo...	51487	500	374	0.594767	0.0151315	0.30198
2	Setter	English_Sett...	Setter/Englis...	70075	500	375	0.103279	0.624603	0.0512919
3	Setter	English_Sett...	Setter/Englis...	11423	354	354	0.0763681	0.150287	0.127533
4	Setter	Gordon_Sett...	Setter/Gordo...	33900	357	500	0.0770098	0.059684	0.0184555
5	Setter	Gordon_Sett...	Setter/Gordo...	28650	500	375	0.229875	0.0811693	0.0264853
6	Setter	Gordon_Sett...	Setter/Gordo...	39807	500	375	0.257274	0.257542	0.0722097
7	Setter	English_Sett...	Setter/Englis...	20113	500	375	0.148363	0.458478	0.126872
8	Setter	English_Sett...	Setter/Englis...	67134	500	375	0.147101	0.27746	0.281376
9	Setter	English_Sett...	Setter/Englis...	478861	1992	1356	0.0633737	0.099071	0.45349
10	Setter	Gordon_Sett...	Setter/Gordo...	39093	500	375	0.104463	0.0966881	0.330212
11	Setter	Irish_Setter_...	Setter/Irish_...	51161	500	333	0.386734	0.140084	0.052087

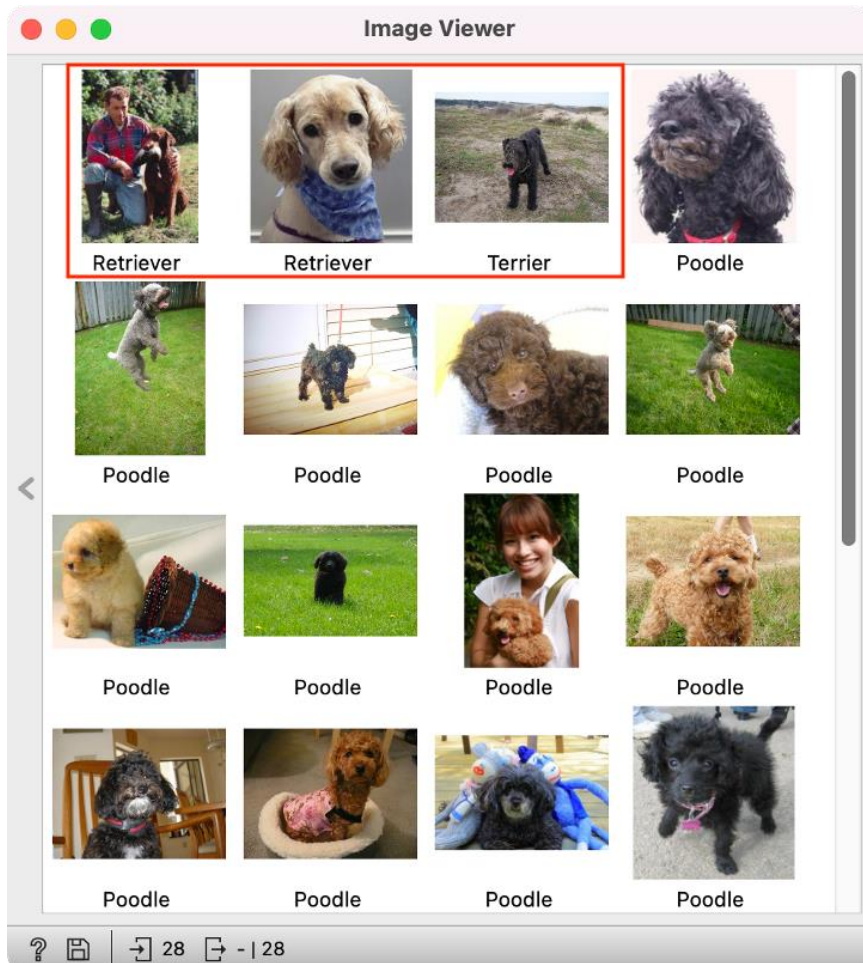
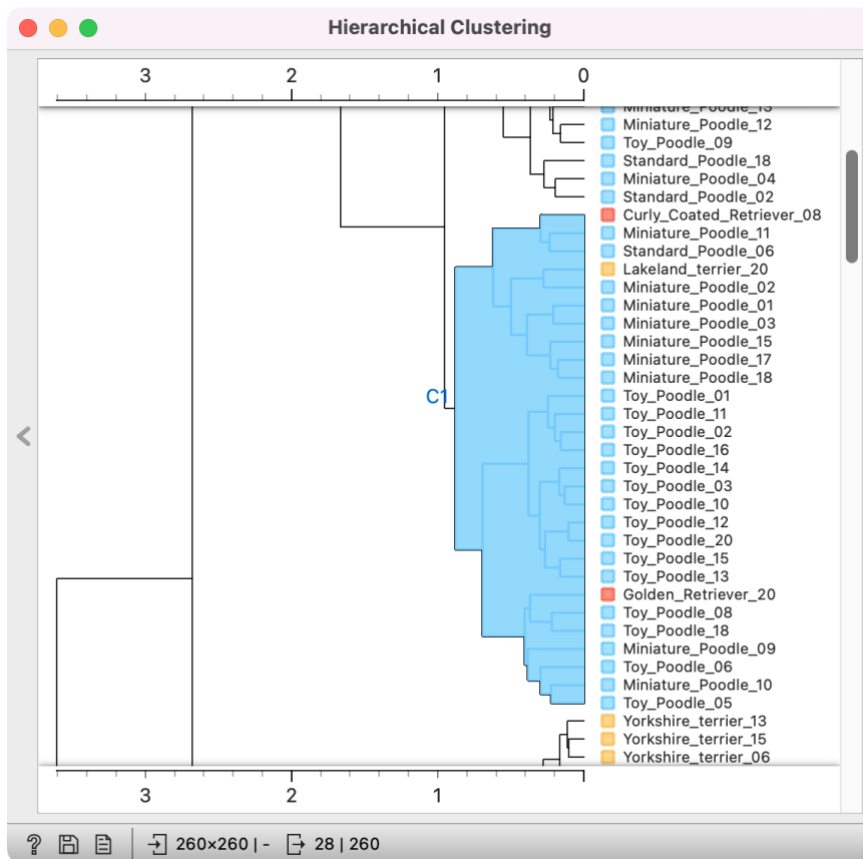
260 260 | 260

3. Za gručenje sem uporabila kosinusno razdaljo, saj ta bolje predstavi razdalje med primeri za visokodimenzionalne podatke kot evklidska razdalja. Razdalje sem izračunala z gradnikom **Distances**. Nato sem gruče določila z gradnikom **Hierarchical Clustering**, kjer sem izbrala Wardovo povezovanje med gručami. Izbrala sem pet skupin, ki dokaj dobro ustrezajo izhodiščnim kategorijam. Skupina C1 predstavlja pudlje, a vsebuje nekaj primerkov drugih pasem. Skupina C2 zajema izključno terierje, ravno tako skupina C3. Pomembno je opozoriti, da se skupini združita v naslednjem koraku. Skupina C4 predstavlja izključno setre, skupina C5 pa prinašalce. “Napačno” uvrščena v skupino C1 sta zgolj dva prinašalca ter en terier, kar pomeni, da gruče dokaj dobro ustrezajo kategorijam.





4. Ker gruče skoraj popolnoma ustrezajo kategorijam oziroma pasmam psov, kar sem prikazala že v točki 1, sem dodatno raziskala zgolj napačno uvrščene primere. Izbrala sem gručo, ki take primere vsebuje, ter slike preverila v gradniku **Image Viewer**.



5. Uporabila sem gradnika **t-SNE** ter **PCA**, oba s privzetimi nastavitvami. V PCA sem izbrala prvi dve glavni komponenti, ki predstavljata zgolj 15 % variance v podatkih. t-SNE tako lepše prikaže karto podatkov, saj so meje med kategorijami slik (pasmami) zelo jasne. Še več, setri so celo razdeljeni na angleške, irske ter škotske setre, terierji pa na jorkširske, jezerske ter bedlingtonske.



6. Slika končnega delotoka.

