

3 - Extraterrestrial Life - oplossingen

June 14, 2021

1 Extraterrestrial life...

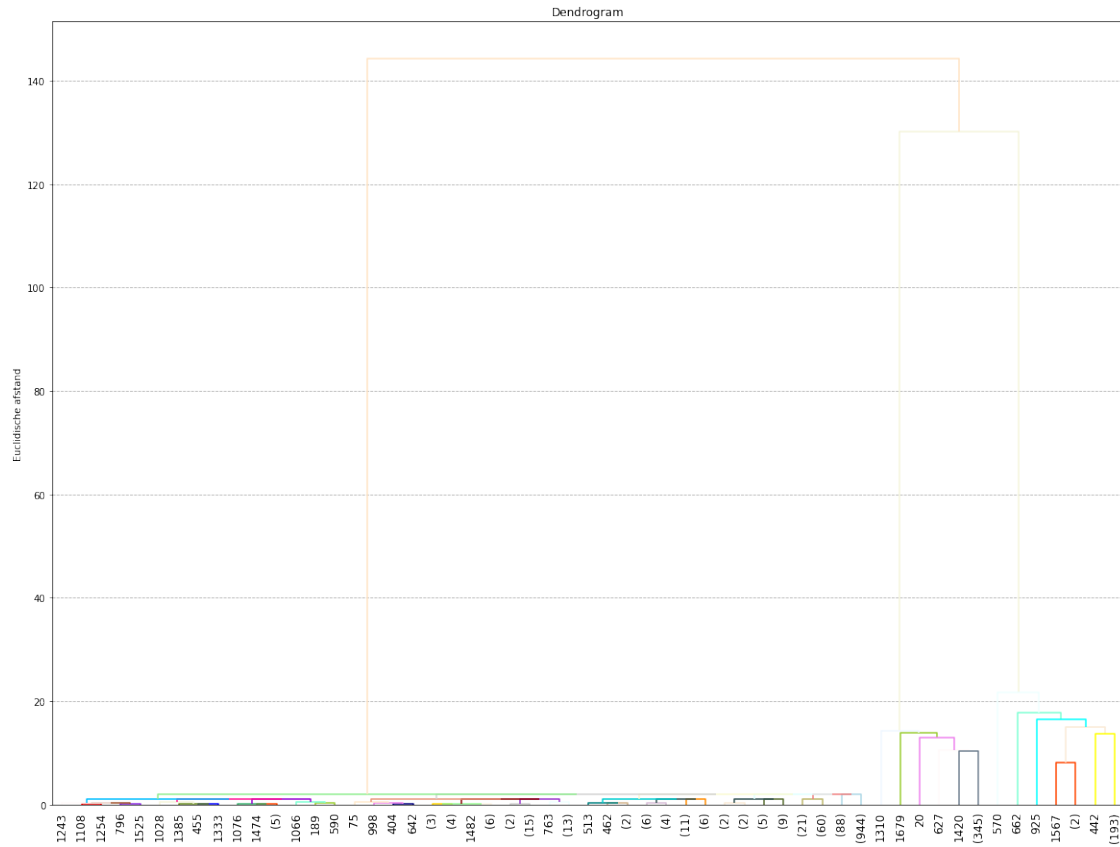
Na een lange reis, stort het ruimteschip van spaceman Spiff neer op een verre planeet. Alles lijkt rustig, maar Spiff ontdekt dat er levende wezens op deze planeet rondlopen. Als goed onderzoeker, begint hij onmiddellijk data te verzamelen over deze wezens. Met zijn “Mertilizer” kan hij de wezens vangen. Per wezen schrijft hij het volgende op: * het aantal poten dat ze hebben, * het aantal ogen die ze hebben, * hun lengte, * hun breedte, * hun hoogte, * de aanwezigheid van een staart, * hun kleur, * het al dan niet aanwezig zijn van vleugels

Je vindt deze data in het bestand: “spiffs metingen.csv”.

1. Lees het bestand in

	poten	ogen	lengte	breedte	hoogte	staart	kleur	vleugels
0	6	2	1.6	0.8	0.7	False	1	False
1	6	2	2.4	0.7	0.8	False	2	False
2	6	2	2.1	0.6	0.9	False	1	False
3	6	2	2.2	0.8	0.7	False	2	False
4	6	2	1.7	0.7	0.7	False	1	False

2. Maak een dendrogram van deze data. Je merkt dat het dendrogram teveel takken heeft. Gebruik de `truncate_mode` optie ‘level’ en parameter `p` in `dendrogram` om aantal niveaus te beperken. Zoek op in de `scipy` documentatie (google) hoe je dit kan doen.



3. Bepaal de hoogte waarbij er 3 clusters zijn

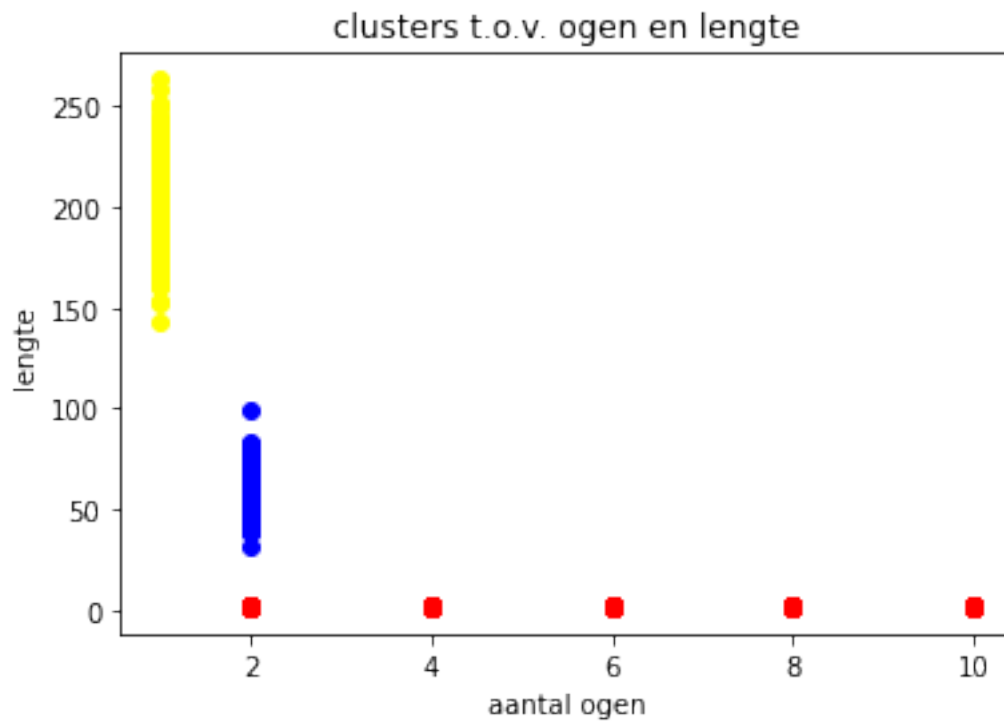
```
array([ 21.7324642 , 130.23878071, 144.25175909])
```

4. Maak een scatterplot waarbij je het aantal ogen in de X-as zet en de lengte in de Y-as. Gebruik de cluster als kleur. Wat zie je?

```
1    1230
2    350
0    200
dtype: int64

1    1230
2    350
3    200
dtype: int64
```

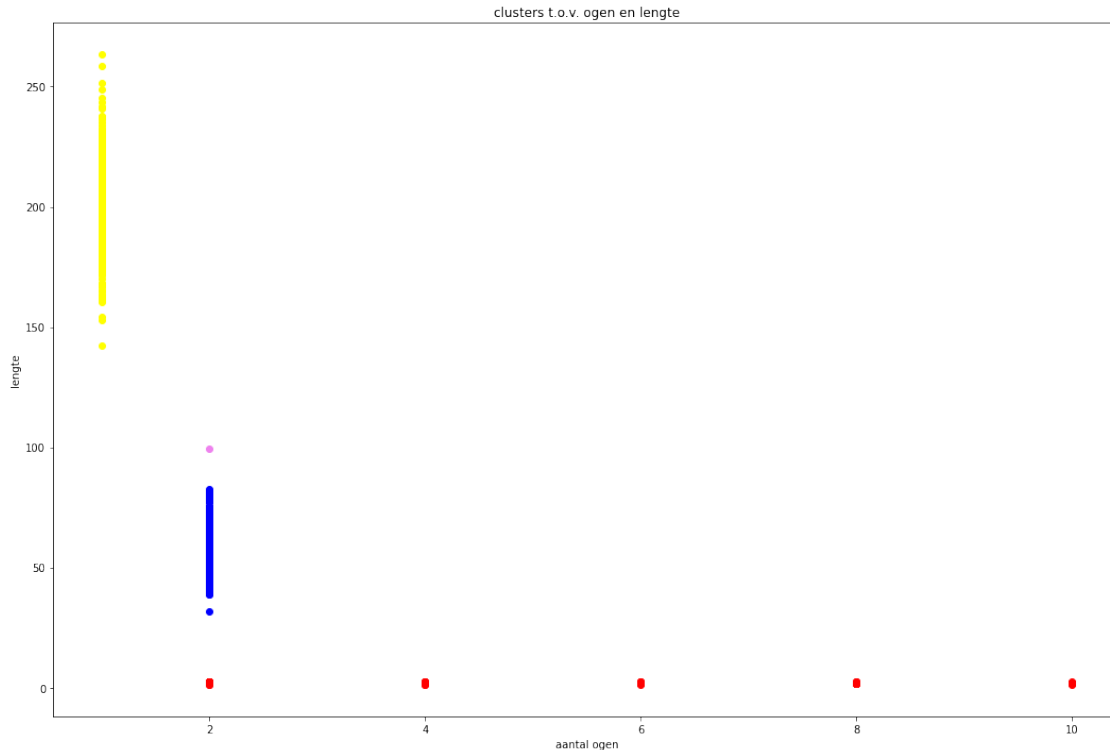
```
Text(0.5, 1.0, 'clusters t.o.v. ogen en lengte')
```



5. Doe dit ook voor 4 clusters. Welke cluster werd er opgesplitst?

```
1    1230
2     350
3     199
4         1
dtype: int64
```

```
Text(0.5, 1.0, 'clusters t.o.v. ogen en lengte')
```



6. Zoek met k-means 3 clusters. Bekijk de centroids. Welke soorten wezens zijn er op deze planeet?

	poten	ogen	lengte	...	staart	kleur	vleugels
0	6.621138	3.614634	2.00187	...	2.498002e-15	1.483740	3.707317e-01
1	9.942857	1.000000	201.40600	...	5.342857e-01	1.522857	2.775558e-16
2	4.000000	2.000000	60.22700	...	-5.551115e-17	1.495000	2.775558e-16

[3 rows x 8 columns]

7. Voeg een kolom toe aan de data met de clusternummer. Zoek nu een beslissingsboom die bepaalt tot welke categorie een ruimtewezen behoort.

```
[Text(133.92000000000002, 181.2, 'hoogte <= 8.7\nentropy = 1.184\nsamples =
1780\nvalue = [1230, 350, 200]\nclasse = cluster 0'),
Text(66.96000000000001, 108.72, 'entropy = 0.0\nsamples = 1230\nvalue = [1230,
0, 0]\nclasse = cluster 0'),
Text(200.88000000000002, 108.72, 'ogen <= 1.5\nentropy = 0.946\nsamples =
550\nvalue = [0, 350, 200]\nclasse = cluster 1'),
Text(133.92000000000002, 36.239999999999998, 'entropy = 0.0\nsamples =
350\nvalue = [0, 350, 0]\nclasse = cluster 1'),
Text(267.84000000000003, 36.239999999999998, 'entropy = 0.0\nsamples =
200\nvalue = [0, 0, 200]\nclasse = cluster 2')]
```

