

# clusteringmodel

April 18, 2024

## 1 Week 9-1: Unsupervised Machine Learning

Van Pjotr en Sennen.

Deze week hebben we het over ongesuperviseerde machine learning. Dit is een vorm van machine learning waarbij we geen gelabelde data hebben. Dit betekent dat we geen data hebben waarbij we weten wat de juiste output is. In plaats daarvan gaan we op zoek naar patronen in de data. Dit kan bijvoorbeeld zijn dat we clusters van data vinden, of dat we de data kunnen reduceren naar een kleinere dimensie.

Maar deze keer gaan we clustering gebruiken, een techniek waarbij we data in groepen verdelen. Dit kan bijvoorbeeld handig zijn als we een dataset hebben met verschillende soorten bloemen, en we willen weten welke bloemen bij elkaar horen. Of als we een dataset hebben met verschillende soorten klanten, en we willen weten welke klanten bij elkaar horen.

PR9-1: Great Outdoors wil graag weten in welke segmenten verkoopafdelingen ('sales\_branches') opgedeeld kan worden. Er bestaan al retailersegmenten (table 'retailer\_segment'), Great Outdoors wil dus óók segmenten creëren voor verkoopafdelingen: - Pas waar nodig Dummy Encoding toe. - Train het initiële clustermodel. - Experimenteer met meerdere k's door het berekenen van de inter- en intraclusterafstand.

Maar eerst een voorbeeld met de titanic.

### 1.1 Bibliotheken importeren

```
[ ]: import pandas as pd
import sqlite3
from sklearn.cluster import KMeans
import math
import matplotlib.pyplot as plt
import warnings
import numpy as np
warnings.simplefilter('ignore')
```

## 2 De Sales Branches clusteren

Om de segmenten van de verkoopafdelingen te bepalen, gaan we de volgende stappen doorlopen: - We kunnen de segmenten op basis van producten categorieën en het verkoop van die producten, bepalen welke segmenten we nodig zullen hebben voor onze verkoopafdeling. - We pakken daarbij

ook de branches van de retailers erbij, zodat we de segmenten van de verkoopafdelingen kunnen bepalen. - Daarna zullen we gebaseerd op de omzet en kosten van de verkoopafdelingen, de segmenten bepalen. - Waar we dan de gemixte data in een clustering model zullen stoppen en de segmenten bepalen.

Eerst gaan we de data inladen.

```
[ ]: # Verbinding maken met de databases
conn_sales = sqlite3.connect('data/go_sales.sqlite')
conn_crm = sqlite3.connect('data/go_crm.sqlite')
conn_staff = sqlite3.connect('data/go_staff.sqlite')

# Gegevens ophalen uit de databases (Helaas verouderd)
sales_branch = pd.read_sql_query("SELECT * FROM sales_branch", conn_staff)
# country1 = pd.read_sql_query("SELECT * FROM country", conn_sales)
# country2 = pd.read_sql_query("SELECT * FROM country", conn_crm)
# territory = pd.read_sql_query("SELECT * FROM sales_territory", conn_crm)

orders_header = pd.read_sql_query("SELECT * FROM order_header;", conn_sales)
order_details = pd.read_sql_query("SELECT * FROM order_details;", conn_sales)
```

Hier gaan we de data van de great outdoors inlezen en bepaalde data dropen.

```
[ ]: # (Helaas Verouderd)
# country = country1[['CURRENCY_NAME', 'LANGUAGE']]
# country = pd.concat([country2, country], axis = 1)
# sales_branch.drop('TRIAL633', axis=1, inplace=True)
# country.drop('TRIAL219', axis=1, inplace=True)
# territory.drop('TRIAL222', axis=1, inplace=True)
orders_header.drop('TRIAL885', axis=1, inplace=True)
order_details.drop('TRIAL879', axis=1, inplace=True)
sales_branch.drop('TRIAL633', axis=1, inplace=True)
sales_branch.drop(["ADDRESS1", "ADDRESS2", "POSTAL_ZONE"], axis=1)
```

```
[ ]:
```

	SALES_BRANCH_CODE	CITY	REGION	COUNTRY_CODE
0	13	Hamburg	None	2
1	14	München	None	2
2	15	Kista	None	10
3	17	Calgary	Alberta	4
4	18	Toronto	Ontario	4
5	19	Boston	Massachusetts	3
6	20	Seattle	Washington	3
7	21	Los Angeles	California	3
8	22	Miami	Florida	3
9	23	Lyon	None	1
10	24	Distrito Federal	México	20
11	25	Tokyo	Tokyo	11
12	26	Osaka City	Osaka	11

13	28	Melbourne	VIC	16
14	29	Bilbao	Vizcaya	19
15	30	São Paulo	SP	21
16	31	Kuopio	None	22
17	32	Seoul	None	13
18	33	San Chung	Taipei	12
19	34	Shanghai	None	14
20	35	London	None	9
21	36	Birmingham	West Midlands	9
22	37	Zürich	None	8
23	38	Heverlee	None	17
24	39	Wien	None	5
25	6	Paris	None	1
26	7	Milano	None	6
27	9	Amsterdam	Noord-Holland	7

Nu gaan we data mergen. Orders worden samengevoegd met details om een compleet beeld van de bestellingen te krijgen.

Productinformatie wordt geladen en samengevoegd om een lookup-tabel te maken voor productnummers naar productcategorieën.

Verkoopgegevens worden voorbereid, zoals de omzet- en afzetgegevens per verkoopfiliaal en de verhouding van de verkoop per productlijn.

Bij de productlijn wordt ook dummy encoding toegepast.

We zullen later de branches dummies geven.

```
[ ]: order_full = pd.merge(order_details, orders_header, on='ORDER_NUMBER')
order_full['UNIT_SALE_PRICE'] = order_full['UNIT_SALE_PRICE'].astype(float)
order_full['UNIT_PRICE'] = order_full['UNIT_PRICE'].astype(float)
order_full['UNIT_COST'] = order_full['UNIT_COST'].astype(float)

# Lees productstabellen
product = pd.read_sql_query("SELECT * FROM product;", conn_sales)
product_type = pd.read_sql_query("SELECT * FROM product_type;", conn_sales)
product_line = pd.read_sql_query("SELECT * FROM product_line;", conn_sales)

# Maak lookuptabel voor productnumber -> naam productscategorie
product_line_lookup = pd.merge(product, product_type, on='PRODUCT_TYPE_CODE')
product_line_lookup = pd.merge(product_line_lookup, product_line,
    ↪on='PRODUCT_LINE_CODE')
product_line_lookup = product_line_lookup.loc[:,['PRODUCT_NUMBER',
    ↪'PRODUCT_LINE_EN']]

# Maak dummies van product_line
product_line_dummies = pd.get_dummies(product_line_lookup['PRODUCT_LINE_EN'])
```

```

product_line_lookup = product_line_lookup.drop(['PRODUCT_LINE_EN'], axis=1)
product_line_lookup = pd.concat([product_line_dummies, product_line_lookup],
    ↪axis = 1)

# Omzet/Afzetsdata
sales_profit_data = order_full.groupby('SALES_BRANCH_CODE').aggregate('sum') #.
    ↪reset_index()
sales_profit_data = sales_profit_data[['QUANTITY', 'UNIT_COST', 'UNIT_PRICE',
    ↪'UNIT_SALE_PRICE']]

# Productsverkoopdata (Verhouding van verkoop per productlijn)
sales_product_data = pd.merge(order_full, product_line_lookup,
    ↪on='PRODUCT_NUMBER')
sales_product_data = sales_product_data[['SALES_BRANCH_CODE', 'Camping',
    ↪Equipment', 'Golf Equipment', 'Mountaineering Equipment', 'Outdoor',
    ↪Protection', 'Personal Accessories']]
sales_product_data = sales_product_data.groupby('SALES_BRANCH_CODE').
    ↪aggregate(np.mean)

```

## 2.1 Clusteringmodel bouwen met 2 dimensies

Eerst wordt een 2D-dataset gemaakt met hoeveelheid en eenheidsprijs van de verkochte producten.

Waar we ook dummy encoding toepassen.

```

[ ]: table = sales_profit_data[['QUANTITY', 'UNIT_SALE_PRICE']]
df = pd.get_dummies(table) # Dummy encoding

table = pd.concat([table, df], axis=1)
table

```

```

[ ]:

```

	QUANTITY	UNIT_SALE_PRICE	QUANTITY	UNIT_SALE_PRICE
SALES_BRANCH_CODE				
13	102648	188463.46	102648	188463.46
14	116778	206523.20	116778	206523.20
15	158062	517067.65	158062	517067.65
17	73848	88668.11	73848	88668.11
18	171852	324804.56	171852	324804.56
19	88444	321688.92	88444	321688.92
20	76848	342898.94	76848	342898.94
21	83110	403109.89	83110	403109.89
22	33550	168475.23	33550	168475.23
23	65862	195489.13	65862	195489.13
24	62844	225383.89	62844	225383.89
25	19214	114444.77	19214	114444.77
26	44992	175456.95	44992	175456.95
28	66630	189991.35	66630	189991.35
29	75426	211154.37	75426	211154.37

30	97282	81545.55	97282	81545.55
31	56488	165810.09	56488	165810.09
32	36358	133803.80	36358	133803.80
33	41796	130358.39	41796	130358.39
34	41108	123299.60	41108	123299.60
35	87954	266665.35	87954	266665.35
36	86514	221279.16	86514	221279.16
37	67242	216831.04	67242	216831.04
38	49692	179281.69	49692	179281.69
39	74940	261826.18	74940	261826.18
6	89324	171166.01	89324	171166.01
7	115570	339516.33	115570	339516.33
9	130978	402590.49	130978	402590.49

Een k-means clusteringmodel wordt geïnstantieerd met het gewenste aantal clusters (3 of 4 in het eerste geval, met een willenkeurigheid van 42) en toegepast op de 2D-dataset.

```
[ ]: # Train het clustermodel
kmeans = KMeans(n_clusters=4, random_state=42)
prediction_results = kmeans.fit_predict(df)
prediction_results

kmeans_centra = pd.DataFrame(kmeans.cluster_centers_)
kmeans_centra

for i in range(len(kmeans_centra.columns)):
    kmeans_centra = kmeans_centra.rename(columns = {i : f'{df.columns[i]}'})

kmeans_centra
```

```
[ ]:          QUANTITY  UNIT_SALE_PRICE
0    84293.250000    224765.831250
1   117837.714286    378810.968571
2    51601.000000    112020.036667
3    58076.857143    177952.921429
```

De centra van de clusters worden geëxtraheerd en weergegeven. Voor elk punt in de dataset wordt de afstand tot elk clustercentrum berekend, en het punt wordt toegewezen aan het dichtstbijzijnde cluster.

```
[ ]: for src_index, _ in df.iterrows():
    euclidian_distances = dict()
    print(f"Afstand van bronindex {src_index} tot...")

    for centrumindex, _ in kmeans_centra.iterrows():
        print(f"\tCentrumindex {centrumindex}:")
        euclidian_sum = 0
```

```

    for column_name in kmeans_centra.columns:
        current_difference = df.at[src_index, column_name] - kmeans_centra.
↪at[centrumindex, column_name]
        print(f'\t\t{df.at[src_index, column_name]} - {kmeans_centra.
↪at[centrumindex, column_name]} = {current_difference}')
        euclidian_sum += current_difference ** 2

    print(f'\tTotale euclidische som: {euclidian_sum}')
    euclidian_distance = math.sqrt(euclidian_sum)
    print(f'\tEuclidische afstand: {euclidian_distance}')
    euclidian_distances[centrumindex] = euclidian_distance
    print('-----')

print(euclidian_distances)
centrum_number = min(euclidian_distances, key = euclidian_distances.get)
print(centrum_number)
df.at[src_index, 'Centrum'] = centrum_number
print("=====")

```

df

Afstand van bronindex 13 tot...

Centrumindex 0:

102648 - 84293.25 = 18354.75

188463.46 - 224765.83125 = -36302.37125

Totale euclidische som: 1654759005.9353263

Euclidische afstand: 40678.729158312286

Centrumindex 1:

102648 - 117837.71428571429 = -15189.71428571429

188463.46 - 378810.96857142856 = -190347.50857142857

Totale euclidische som: 36462901439.43171

Euclidische afstand: 190952.61569151576

Centrumindex 2:

102648 - 51601.0 = 51047.0

188463.46 - 112020.03666666668 = 76443.42333333331

Totale euclidische som: 8449393179.919208

Euclidische afstand: 91920.58082888297

Centrumindex 3:

102648 - 58076.857142857145 = 44571.142857142855

188463.46 - 177952.92142857146 = 10510.538571428537

Totale euclidische som: 2097058196.6533234

Euclidische afstand: 45793.6479945999

{0: 40678.729158312286, 1: 190952.61569151576, 2: 91920.58082888297, 3:

45793.6479945999}

0

=====

Afstand van bronindex 14 tot...

Centrumindex 0:

$116778 - 84293.25 = 32484.75$

$206523.2 - 224765.83125 = -18242.631249999977$

Totale euclidische som: 1388052577.4859757

Euclidische afstand: 37256.57764054524

-----

Centrumindex 1:

$116778 - 117837.71428571429 = -1059.7142857142899$

$206523.2 - 378810.96857142856 = -172287.76857142855$

Totale euclidische som: 29684198193.68947

Euclidische afstand: 172291.02760645858

-----

Centrumindex 2:

$116778 - 51601.0 = 65177.0$

$206523.2 - 112020.03666666668 = 94503.163333333333$

Totale euclidische som: 13178889209.006678

Euclidische afstand: 114799.34324292399

-----

Centrumindex 3:

$116778 - 58076.857142857145 = 58701.142857142855$

$206523.2 - 177952.92142857146 = 28570.278571428556$

Totale euclidische som: 4262084990.3837233

Euclidische afstand: 65284.64590073016

-----

{0: 37256.57764054524, 1: 172291.02760645858, 2: 114799.34324292399, 3: 65284.64590073016}

0

=====

Afstand van bronindex 15 tot...

Centrumindex 0:

$158062 - 84293.25 = 73768.75$

$517067.65 - 224765.83125 = 292301.81875000003$

Totale euclidische som: 90882181721.12038

Euclidische afstand: 301466.717435143

-----

Centrumindex 1:

$158062 - 117837.71428571429 = 40224.28571428571$

$517067.65 - 378810.96857142856 = 138256.68142857146$

Totale euclidische som: 20732903120.86599

Euclidische afstand: 143989.2465459348

-----

Centrumindex 2:

$158062 - 51601.0 = 106461.0$

$517067.65 - 112020.03666666668 = 405047.613333333334$

Totale euclidische som: 175397513588.0295  
Euclidische afstand: 418804.8633767635

-----  
Centrumindex 3:

158062 - 58076.857142857145 = 99985.14285714286

517067.65 - 177952.92142857146 = 339114.72857142857

Totale euclidische som: 124995827926.23694

Euclidische afstand: 353547.4903407418

-----  
{0: 301466.717435143, 1: 143989.2465459348, 2: 418804.8633767635, 3:  
353547.4903407418}

1

=====

Afstand van bronindex 17 tot...

Centrumindex 0:

73848 - 84293.25 = -10445.25

88668.11 - 224765.83125 = -136097.72125

Totale euclidische som: 18631692977.005203

Euclidische afstand: 136497.95960747986

-----  
Centrumindex 1:

73848 - 117837.71428571429 = -43989.71428571429

88668.11 - 378810.96857142856 = -290142.8585714286

Totale euclidische som: 86117973342.93878

Euclidische afstand: 293458.63991870946

-----  
Centrumindex 2:

73848 - 51601.0 = 22247.0

88668.11 - 112020.03666666668 = -23351.92666666668

Totale euclidische som: 1040241488.0453784

Euclidische afstand: 32252.77488907549

-----  
Centrumindex 3:

73848 - 58076.857142857145 = 15771.142857142855

88668.11 - 177952.92142857146 = -89284.81142857145

Totale euclidische som: 8220506498.855971

Euclidische afstand: 90667.00887784912

-----  
{0: 136497.95960747986, 1: 293458.63991870946, 2: 32252.77488907549, 3:  
90667.00887784912}

2

=====

Afstand van bronindex 18 tot...

Centrumindex 0:

171852 - 84293.25 = 87558.75

324804.56 - 224765.83125 = 100038.72875000001

Totale euclidische som: 17674281951.478577

Euclidische afstand: 132944.6574762543



```

-----
Centrumindex 1:
    171852 - 117837.71428571429 = 54014.28571428571
    324804.56 - 378810.96857142856 = -54006.40857142856
Totale euclidische som: 5834235228.008562
Euclidische afstand: 76382.16564099609
-----

Centrumindex 2:
    171852 - 51601.0 = 120251.0
    324804.56 - 112020.03666666668 = 212784.52333333332
Totale euclidische som: 59737556371.19387
Euclidische afstand: 244412.67637173377
-----

Centrumindex 3:
    171852 - 58076.857142857145 = 113775.14285714286
    324804.56 - 177952.92142857146 = 146851.63857142854
Totale euclidische som: 34510186883.27675
Euclidische afstand: 185769.17635408934
-----

{0: 132944.6574762543, 1: 76382.16564099609, 2: 244412.67637173377, 3:
185769.17635408934}
1
=====
Afstand van bronindex 19 tot...
Centrumindex 0:
    88444 - 84293.25 = 4150.75
    321688.92 - 224765.83125 = 96923.08875
Totale euclidische som: 9411313858.402876
Euclidische afstand: 97011.9263719821
-----

Centrumindex 1:
    88444 - 117837.71428571429 = -29393.71428571429
    321688.92 - 378810.96857142856 = -57122.048571428575
Totale euclidische som: 4126918872.50685
Euclidische afstand: 64241.099558669215
-----

Centrumindex 2:
    88444 - 51601.0 = 36843.0
    321688.92 - 112020.03666666668 = 209668.88333333333
Totale euclidische som: 45318447287.24693
Euclidische afstand: 212881.29858502586
-----

Centrumindex 3:
    88444 - 58076.857142857145 = 30367.142857142855
    321688.92 - 177952.92142857146 = 143735.99857142853
Totale euclidische som: 21582200650.631824
Euclidische afstand: 146908.81747067405
-----

```

{0: 97011.9263719821, 1: 64241.099558669215, 2: 212881.29858502586, 3: 146908.81747067405}

1

=====

Afstand van bronindex 20 tot...

Centrumindex 0:

$$76848 - 84293.25 = -7445.25$$

$$342898.94 - 224765.83125 = 118133.10875000001$$

Totale euclidische som: 14010863130.50183

Euclidische afstand: 118367.49186538435

-----

Centrumindex 1:

$$76848 - 117837.71428571429 = -40989.71428571429$$

$$342898.94 - 378810.96857142856 = -35912.02857142856$$

Totale euclidische som: 2969830473.339591

Euclidische afstand: 54496.15099563997

-----

Centrumindex 2:

$$76848 - 51601.0 = 25247.0$$

$$342898.94 - 112020.03666666668 = 230878.90333333332$$

Totale euclidische som: 53942479013.40267

Euclidische afstand: 232255.20233872626

-----

Centrumindex 3:

$$76848 - 58076.857142857145 = 18771.142857142855$$

$$342898.94 - 177952.92142857146 = 164946.01857142855$$

Totale euclidische som: 27559544846.729317

Euclidische afstand: 166010.6769058223

-----

{0: 118367.49186538435, 1: 54496.15099563997, 2: 232255.20233872626, 3: 166010.6769058223}

1

=====

Afstand van bronindex 21 tot...

Centrumindex 0:

$$83110 - 84293.25 = -1183.25$$

$$403109.89 - 224765.83125 = 178344.05875000003$$

Totale euclidische som: 31808003371.985962

Euclidische afstand: 178347.9839302535

-----

Centrumindex 1:

$$83110 - 117837.71428571429 = -34727.71428571429$$

$$403109.89 - 378810.96857142856 = 24298.921428571455$$

Totale euclidische som: 1796451722.1020932

Euclidische afstand: 42384.56938677203

-----

Centrumindex 2:

$$83110 - 51601.0 = 31509.0$$

```

403109.89 - 112020.03666666668 = 291089.85333333333
Totale euclidische som: 85726119794.6215
Euclidische afstand: 292790.23172677995
-----
Centrumindex 3:
83110 - 58076.857142857145 = 25033.142857142855
403109.89 - 177952.92142857146 = 225156.96857142856
Totale euclidische som: 51322318737.58139
Euclidische afstand: 226544.2975172436
-----
{0: 178347.9839302535, 1: 42384.56938677203, 2: 292790.23172677995, 3:
226544.2975172436}
1
=====
Afstand van bronindex 22 tot...
Centrumindex 0:
33550 - 84293.25 = -50743.25
168475.23 - 224765.83125 = -56290.60124999998
Totale euclidische som: 5743509209.648999
Euclidische afstand: 75785.9433513168
-----
Centrumindex 1:
33550 - 117837.71428571429 = -84287.71428571429
168475.23 - 378810.96857142856 = -210335.73857142855
Totale euclidische som: 51345541699.898544
Euclidische afstand: 226595.54651382394
-----
Centrumindex 2:
33550 - 51601.0 = -18051.0
168475.23 - 112020.03666666668 = 56455.19333333333
Totale euclidische som: 3513027455.304044
Euclidische afstand: 59270.797660433454
-----
Centrumindex 3:
33550 - 58076.857142857145 = -24526.857142857145
168475.23 - 177952.92142857146 = -9477.691428571445
Totale euclidische som: 691393356.1213392
Euclidische afstand: 26294.359777742055
-----
{0: 75785.9433513168, 1: 226595.54651382394, 2: 59270.797660433454, 3:
26294.359777742055}
3
=====
Afstand van bronindex 23 tot...
Centrumindex 0:
65862 - 84293.25 = -18431.25
195489.13 - 224765.83125 = -29276.701249999984
Totale euclidische som: 1196836212.6442506

```

Euclidische afstand: 34595.32067555164

-----  
Centrumindex 1:

65862 - 117837.71428571429 = -51975.71428571429

195489.13 - 378810.96857142856 = -183321.83857142855

Totale euclidische som: 36308371372.71912

Euclidische afstand: 190547.5567219877  
-----

Centrumindex 2:

65862 - 51601.0 = 14261.0

195489.13 - 112020.03666666668 = 83469.09333333332

Totale euclidische som: 7170465662.888709

Euclidische afstand: 84678.60215478706  
-----

Centrumindex 3:

65862 - 58076.857142857145 = 7785.142857142855

195489.13 - 177952.92142857146 = 17536.20857142855

Totale euclidische som: 368127060.3667666

Euclidische afstand: 19186.637547177634  
-----

{0: 34595.32067555164, 1: 190547.5567219877, 2: 84678.60215478706, 3:  
19186.637547177634}

3

=====

Afstand van bronindex 24 tot...

Centrumindex 0:

62844 - 84293.25 = -21449.25

225383.89 - 224765.83125 = 618.0587500000256

Totale euclidische som: 460452322.1809516

Euclidische afstand: 21458.15281381302  
-----

Centrumindex 1:

62844 - 117837.71428571429 = -54993.71428571429

225383.89 - 378810.96857142856 = -153427.07857142854

Totale euclidische som: 26564177049.902084

Euclidische afstand: 162985.2050030986  
-----

Centrumindex 2:

62844 - 51601.0 = 11243.0

225383.89 - 112020.03666666668 = 113363.85333333333

Totale euclidische som: 12977768291.58151

Euclidische afstand: 113920.00830223596  
-----

Centrumindex 3:

62844 - 58076.857142857145 = 4767.142857142855

225383.89 - 177952.92142857146 = 47430.96857142856

Totale euclidische som: 2272422430.644252

Euclidische afstand: 47669.93214432187

-----  
{0: 21458.15281381302, 1: 162985.2050030986, 2: 113920.00830223596, 3:  
47669.93214432187}

0

=====

Afstand van bronindex 25 tot...

Centrumindex 0:

$19214 - 84293.25 = -65079.25$

$114444.77 - 224765.83125 = -110321.06124999998$

Totale euclidische som: 16406045335.888748

Euclidische afstand: 128086.08564511896

-----

Centrumindex 1:

$19214 - 117837.71428571429 = -98623.71428571429$

$114444.77 - 378810.96857142856 = -264366.19857142854$

Totale euclidische som: 79616123966.6182

Euclidische afstand: 282163.2930886266

-----

Centrumindex 2:

$19214 - 51601.0 = -32387.0$

$114444.77 - 112020.03666666668 = 2424.7333333333227$

Totale euclidische som: 1054797100.7377777

Euclidische afstand: 32477.640011826254

-----

Centrumindex 3:

$19214 - 58076.857142857145 = -38862.857142857145$

$114444.77 - 177952.92142857146 = -63508.15142857145$

Totale euclidische som: 5543606963.180485

Euclidische afstand: 74455.4025117082

-----

{0: 128086.08564511896, 1: 282163.2930886266, 2: 32477.640011826254, 3:  
74455.4025117082}

2

=====

Afstand van bronindex 26 tot...

Centrumindex 0:

$44992 - 84293.25 = -39301.25$

$175456.95 - 224765.83125 = -49308.88124999998$

Totale euclidische som: 3975954021.6890993

Euclidische afstand: 63055.16649481706

-----

Centrumindex 1:

$44992 - 117837.71428571429 = -72845.71428571429$

$175456.95 - 378810.96857142856 = -203354.01857142855$

Totale euclidische som: 46659354958.944824

Euclidische afstand: 216007.76596906147

-----

Centrumindex 2:

```

44992 - 51601.0 = -6609.0
175456.95 - 112020.03666666668 = 63436.91333333333
Totale euclidische som: 4067920854.260844
Euclidische afstand: 63780.25442298615
-----

Centrumindex 3:
44992 - 58076.857142857145 = -13084.857142857145
175456.95 - 177952.92142857146 = -2495.9714285714435
Totale euclidische som: 177443359.82122463
Euclidische afstand: 13320.786756840776
-----

{0: 63055.16649481706, 1: 216007.76596906147, 2: 63780.25442298615, 3:
13320.786756840776}
3
=====
Afstand van bronindex 28 tot...
Centrumindex 0:
66630 - 84293.25 = -17663.25
189991.35 - 224765.83125 = -34774.48124999998
Totale euclidische som: 1521254946.7691004
Euclidische afstand: 39003.26841136651
-----

Centrumindex 1:
66630 - 117837.71428571429 = -51207.71428571429
189991.35 - 378810.96857142856 = -188819.61857142855
Totale euclidische som: 38275078359.82712
Euclidische afstand: 195640.1757304136
-----

Centrumindex 2:
66630 - 51601.0 = 15029.0
189991.35 - 112020.03666666668 = 77971.31333333332
Totale euclidische som: 6305396543.924843
Euclidische afstand: 79406.5270864105
-----

Centrumindex 3:
66630 - 58076.857142857145 = 8553.142857142855
189991.35 - 177952.92142857146 = 12038.42857142855
Totale euclidische som: 218080015.20408112
Euclidische afstand: 14767.532468360485
-----

{0: 39003.26841136651, 1: 195640.1757304136, 2: 79406.5270864105, 3:
14767.532468360485}
3
=====
Afstand van bronindex 29 tot...
Centrumindex 0:
75426 - 84293.25 = -8867.25
211154.37 - 224765.83125 = -13611.461249999993

```

Totale euclidische som: 263899999.92275137  
Euclidische afstand: 16244.99922815484

---

Centrumindex 1:  
75426 - 117837.71428571429 = -42411.71428571429  
211154.37 - 378810.96857142856 = -167656.59857142856  
Totale euclidische som: 29907488553.194202  
Euclidische afstand: 172937.81701292

---

Centrumindex 2:  
75426 - 51601.0 = 23825.0  
211154.37 - 112020.03666666668 = 99134.33333333331  
Totale euclidische som: 10395246670.44444  
Euclidische afstand: 101957.08249280401

---

Centrumindex 3:  
75426 - 58076.857142857145 = 17349.142857142855  
211154.37 - 177952.92142857146 = 33201.44857142854  
Totale euclidische som: 1403328945.1187654  
Euclidische afstand: 37461.03235521901

---

{0: 16244.99922815484, 1: 172937.81701292, 2: 101957.08249280401, 3:  
37461.03235521901}

0

=====

Afstand van bronindex 30 tot...

Centrumindex 0:  
97282 - 84293.25 = 12988.75  
81545.55 - 224765.83125 = -143220.28125  
Totale euclidische som: 20680756587.8916  
Euclidische afstand: 143808.05466972844

---

Centrumindex 1:  
97282 - 117837.71428571429 = -20555.71428571429  
81545.55 - 378810.96857142856 = -297265.41857142857  
Totale euclidische som: 88789266468.24254  
Euclidische afstand: 297975.27828369004

---

Centrumindex 2:  
97282 - 51601.0 = 45681.0  
81545.55 - 112020.03666666668 = -30474.48666666668  
Totale euclidische som: 3015448098.596845  
Euclidische afstand: 54913.09587518122

---

Centrumindex 3:  
97282 - 58076.857142857145 = 39205.142857142855  
81545.55 - 177952.92142857146 = -96407.37142857145  
Totale euclidische som: 10831424492.215515

```

Euclidische afstand: 104074.12979321765
-----
{0: 143808.05466972844, 1: 297975.27828369004, 2: 54913.09587518122, 3:
104074.12979321765}
2
=====
Afstand van bronindex 31 tot...
  Centrumindex 0:
    56488 - 84293.25 = -27805.25
    165810.09 - 224765.83125 = -58955.74124999999
  Totale euclidische som: 4248911353.899451
  Euclidische afstand: 65183.6739828268
-----
  Centrumindex 1:
    56488 - 117837.71428571429 = -61349.71428571429
    165810.09 - 378810.96857142856 = -213000.87857142856
  Totale euclidische som: 49133161715.13923
  Euclidische afstand: 221660.01379396155
-----
  Centrumindex 2:
    56488 - 51601.0 = 4887.0
    165810.09 - 112020.03666666668 = 53790.053333333315
  Totale euclidische som: 2917252606.6028423
  Euclidische afstand: 54011.59696401174
-----
  Centrumindex 3:
    56488 - 58076.857142857145 = -1588.857142857145
    165810.09 - 177952.92142857146 = -12142.831428571459
  Totale euclidische som: 149972822.12311095
  Euclidische afstand: 12246.339131475615
-----
{0: 65183.6739828268, 1: 221660.01379396155, 2: 54011.59696401174, 3:
12246.339131475615}
3
=====
Afstand van bronindex 32 tot...
  Centrumindex 0:
    36358 - 84293.25 = -47935.25
    133803.8 - 224765.83125 = -90962.03125
  Totale euclidische som: 10571879321.688477
  Euclidische afstand: 102819.64462926565
-----
  Centrumindex 1:
    36358 - 117837.71428571429 = -81479.71428571429
    133803.8 - 378810.96857142856 = -245007.16857142857
  Totale euclidische som: 66667456491.47005
  Euclidische afstand: 258200.419231786
-----

```



```

Centrumindex 2:
    36358 - 51601.0 = -15243.0
    133803.8 - 112020.03666666668 = 21783.763333333307
Totale euclidische som: 706881393.9626766
Euclidische afstand: 26587.241187507152
-----

Centrumindex 3:
    36358 - 58076.857142857145 = -21718.857142857145
    133803.8 - 177952.92142857146 = -44149.12142857147
Totale euclidische som: 2420853678.506585
Euclidische afstand: 49202.171481618425
-----

{0: 102819.64462926565, 1: 258200.419231786, 2: 26587.241187507152, 3:
49202.171481618425}
2
=====
Afstand van bronindex 33 tot...
Centrumindex 0:
    41796 - 84293.25 = -42497.25
    130358.39 - 224765.83125 = -94407.44124999999
Totale euclidische som: 10718781220.9347
Euclidische afstand: 103531.54698416662
-----

Centrumindex 1:
    41796 - 117837.71428571429 = -76041.71428571429
    130358.39 - 378810.96857142856 = -248452.57857142854
Totale euclidische som: 67511026110.30208
Euclidische afstand: 259828.84002801168
-----

Centrumindex 2:
    41796 - 51601.0 = -9805.0
    130358.39 - 112020.03666666668 = 18338.353333333318
Totale euclidische som: 432433227.9781772
Euclidische afstand: 20795.02892467758
-----

Centrumindex 3:
    41796 - 58076.857142857145 = -16280.857142857145
    130358.39 - 177952.92142857146 = -47594.531428571456
Totale euclidische som: 2530305731.211398
Euclidische afstand: 50302.14439973109
-----

{0: 103531.54698416662, 1: 259828.84002801168, 2: 20795.02892467758, 3:
50302.14439973109}
2
=====
Afstand van bronindex 34 tot...
Centrumindex 0:
    41108 - 84293.25 = -43185.25

```

123299.6 - 224765.83125 = -101466.23124999998  
Totale euclidische som: 12160361901.640972  
Euclidische afstand: 110274.03094854641

---

Centrumindex 1:  
41108 - 117837.71428571429 = -76729.71428571429  
123299.6 - 378810.96857142856 = -255511.36857142855  
Totale euclidische som: 71173508523.61176  
Euclidische afstand: 266783.6361616127

---

Centrumindex 2:  
41108 - 51601.0 = -10493.0  
123299.6 - 112020.03666666668 = 11279.563333333324  
Totale euclidische som: 237331597.9906776  
Euclidische afstand: 15405.57035590301

---

Centrumindex 3:  
41108 - 58076.857142857145 = -16968.857142857145  
123299.6 - 177952.92142857146 = -54653.32142857145  
Totale euclidische som: 3274927655.909441  
Euclidische afstand: 57226.983634553384

---

{0: 110274.03094854641, 1: 266783.6361616127, 2: 15405.57035590301, 3:  
57226.983634553384}  
2

=====

Afstand van bronindex 35 tot...

Centrumindex 0:  
87954 - 84293.25 = 3660.75  
266665.35 - 224765.83125 = 41899.51874999999  
Totale euclidische som: 1768970762.0441005  
Euclidische afstand: 42059.134109537976

---

Centrumindex 1:  
87954 - 117837.71428571429 = -29883.71428571429  
266665.35 - 378810.96857142856 = -112145.61857142858  
Totale euclidische som: 13469676144.278551  
Euclidische afstand: 116058.9339270293

---

Centrumindex 2:  
87954 - 51601.0 = 36353.0  
266665.35 - 112020.03666666668 = 154645.31333333333  
Totale euclidische som: 25236713544.964832  
Euclidische afstand: 158860.67337439067

---

Centrumindex 3:  
87954 - 58076.857142857145 = 29877.142857142855  
266665.35 - 177952.92142857146 = 88712.42857142852

```

Totale euclidische som: 8762538648.34693
Euclidische afstand: 93608.43257071945
-----
{0: 42059.134109537976, 1: 116058.9339270293, 2: 158860.67337439067, 3:
93608.43257071945}
0
=====
Afstand van bronindex 36 tot...
Centrumindex 0:
      86514 - 84293.25 = 2220.75
      221279.16 - 224765.83125 = -3486.671249999985
Totale euclidische som: 17088606.968076456
Euclidische afstand: 4133.8368337509955
-----
Centrumindex 1:
      86514 - 117837.71428571429 = -31323.71428571429
      221279.16 - 378810.96857142856 = -157531.80857142856
Totale euclidische som: 25797445788.43827
Euclidische afstand: 160615.8329319942
-----
Centrumindex 2:
      86514 - 51601.0 = 34913.0
      221279.16 - 112020.03666666668 = 109259.12333333332
Totale euclidische som: 13156473600.568542
Euclidische afstand: 114701.67217860663
-----
Centrumindex 3:
      86514 - 58076.857142857145 = 28437.142857142855
      221279.16 - 177952.92142857146 = 43326.23857142855
Totale euclidische som: 2685834042.625894
Euclidische afstand: 51825.03297274295
-----
{0: 4133.8368337509955, 1: 160615.8329319942, 2: 114701.67217860663, 3:
51825.03297274295}
0
=====
Afstand van bronindex 37 tot...
Centrumindex 0:
      67242 - 84293.25 = -17051.25
      216831.04 - 224765.83125 = -7934.79124999998
Totale euclidische som: 353706038.7435762
Euclidische afstand: 18807.074167546005
-----
Centrumindex 1:
      67242 - 117837.71428571429 = -50595.71428571429
      216831.04 - 378810.96857142856 = -161979.92857142855
Totale euclidische som: 28797423564.08673
Euclidische afstand: 169698.03641788766

```

```

-----
Centrumindex 2:
    67242 - 51601.0 = 15641.0
    216831.04 - 112020.03666666668 = 104811.00333333333
Totale euclidische som: 11229987300.74001
Euclidische afstand: 105971.63441572471
-----

Centrumindex 3:
    67242 - 58076.857142857145 = 9165.142857142855
    216831.04 - 177952.92142857146 = 38878.11857142855
Totale euclidische som: 1595507947.2458944
Euclidische afstand: 39943.809873945356
-----

{0: 18807.074167546005, 1: 169698.03641788766, 2: 105971.63441572471, 3:
39943.809873945356}
0
=====

Afstand van bronindex 38 tot...
Centrumindex 0:
    49692 - 84293.25 = -34601.25
    179281.69 - 224765.83125 = -45484.141249999986
Totale euclidische som: 3266053606.8124504
Euclidische afstand: 57149.39725677297
-----

Centrumindex 1:
    49692 - 117837.71428571429 = -68145.71428571429
    179281.69 - 378810.96857142856 = -199529.27857142856
Totale euclidische som: 44455771382.74495
Euclidische afstand: 210845.37315944344
-----

Centrumindex 2:
    49692 - 51601.0 = -1909.0
    179281.69 - 112020.03666666668 = 67261.65333333332
Totale euclidische som: 4527774290.13351
Euclidische afstand: 67288.73821178035
-----

Centrumindex 3:
    49692 - 58076.857142857145 = -8384.857142857145
    179281.69 - 177952.92142857146 = 1328.7685714285471
Totale euclidische som: 72071455.22253874
Euclidische afstand: 8489.490869453759
-----

{0: 57149.39725677297, 1: 210845.37315944344, 2: 67288.73821178035, 3:
8489.490869453759}
3
=====

Afstand van bronindex 39 tot...
Centrumindex 0:

```

$74940 - 84293.25 = -9353.25$   
 $261826.18 - 224765.83125 = 37060.348750000005$   
 Totale euclidische som: 1460952735.034127  
 Euclidische afstand: 38222.41142358926

---

Centrumindex 1:  
 $74940 - 117837.71428571429 = -42897.71428571429$   
 $261826.18 - 378810.96857142856 = -116984.78857142857$   
 Totale euclidische som: 15525654648.040619  
 Euclidische afstand: 124601.98492817287

---

Centrumindex 2:  
 $74940 - 51601.0 = 23339.0$   
 $261826.18 - 112020.03666666668 = 149806.14333333333$   
 Totale euclidische som: 22986589501.407204  
 Euclidische afstand: 151613.28932981833

---

Centrumindex 3:  
 $74940 - 58076.857142857145 = 16863.142857142855$   
 $261826.18 - 177952.92142857146 = 83873.25857142854$   
 Totale euclidische som: 7319089090.410118  
 Euclidische afstand: 85551.6749713886

---

{0: 38222.41142358926, 1: 124601.98492817287, 2: 151613.28932981833, 3:  
 85551.6749713886}  
 0

---

Afstand van bronindex 6 tot...

Centrumindex 0:  
 $89324 - 84293.25 = 5030.75$   
 $171166.01 - 224765.83125 = -53599.82124999998$   
 Totale euclidische som: 2898249283.5944495  
 Euclidische afstand: 53835.39062358932

---

Centrumindex 1:  
 $89324 - 117837.71428571429 = -28513.71428571429$   
 $171166.01 - 378810.96857142856 = -207644.95857142855$   
 Totale euclidische som: 43929460722.49763  
 Euclidische afstand: 209593.56078490967

---

Centrumindex 2:  
 $89324 - 51601.0 = 37723.0$   
 $171166.01 - 112020.03666666668 = 59145.973333333333$   
 Totale euclidische som: 4921270890.547377  
 Euclidische afstand: 70151.77040208876

---

Centrumindex 3:  
 $89324 - 58076.857142857145 = 31247.142857142855$

```

171166.01 - 177952.92142857146 = -6786.911428571446
Totale euclidische som: 1022446103.4739674
Euclidische afstand: 31975.711148838698
-----
{0: 53835.39062358932, 1: 209593.56078490967, 2: 70151.77040208876, 3:
31975.711148838698}
3
=====
Afstand van bronindex 7 tot...
Centrumindex 0:
115570 - 84293.25 = 31276.75
339516.33 - 224765.83125 = 114750.49875000003
Totale euclidische som: 14145912053.936258
Euclidische afstand: 118936.58837353735
-----
Centrumindex 1:
115570 - 117837.71428571429 = -2267.71428571429
339516.33 - 378810.96857142856 = -39294.63857142854
Totale euclidische som: 1549211148.5408323
Euclidische afstand: 39360.01967149956
-----
Centrumindex 2:
115570 - 51601.0 = 63969.0
339516.33 - 112020.03666666668 = 227496.29333333333
Totale euclidische som: 55846596441.406044
Euclidische afstand: 236318.8448715126
-----
Centrumindex 3:
115570 - 58076.857142857145 = 57493.142857142855
339516.33 - 177952.92142857146 = 161563.40857142856
Totale euclidische som: 29408196464.810192
Euclidische afstand: 171488.18170594203
-----
{0: 118936.58837353735, 1: 39360.01967149956, 2: 236318.8448715126, 3:
171488.18170594203}
1
=====
Afstand van bronindex 9 tot...
Centrumindex 0:
130978 - 84293.25 = 46684.75
402590.49 - 224765.83125 = 177824.65875
Totale euclidische som: 33801075142.11645
Euclidische afstand: 183850.68708633224
-----
Centrumindex 1:
130978 - 117837.71428571429 = 13140.28571428571
402590.49 - 378810.96857142856 = 23779.521428571432
Totale euclidische som: 738132748.0249491

```

Euclidische afstand: 27168.598565714594

-----  
Centrumindex 2:

130978 - 51601.0 = 79377.0

402590.49 - 112020.03666666668 = 290570.4533333333

Totale euclidische som: 90731896479.33884

Euclidische afstand: 301217.3575332916  
-----

Centrumindex 3:

130978 - 58076.857142857145 = 72901.14285714286

402590.49 - 177952.92142857146 = 224637.56857142854

Totale euclidische som: 55776613843.56081

Euclidische afstand: 236170.73028544584  
-----

{0: 183850.68708633224, 1: 27168.598565714594, 2: 301217.3575332916, 3:  
236170.73028544584}

1

=====

[ ]:

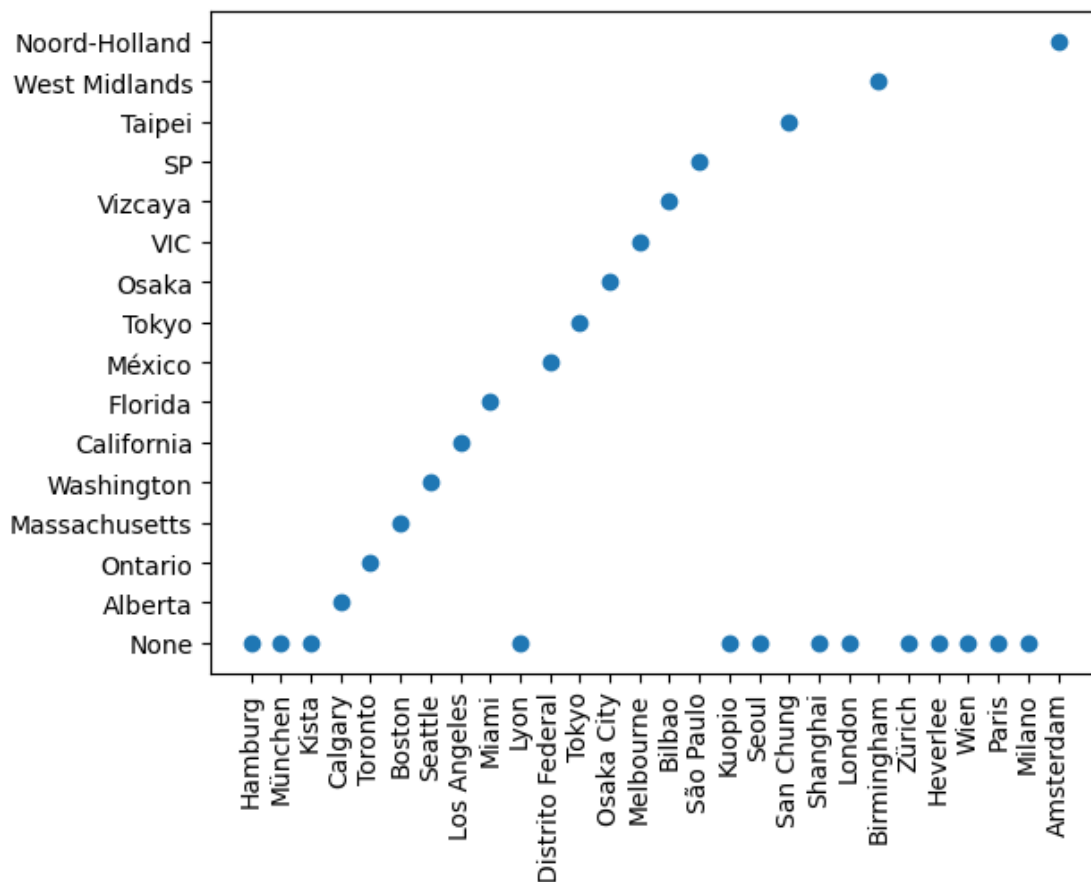
	QUANTITY	UNIT_SALE_PRICE	Centrum
SALES_BRANCH_CODE			
13	102648	188463.46	0.0
14	116778	206523.20	0.0
15	158062	517067.65	1.0
17	73848	88668.11	2.0
18	171852	324804.56	1.0
19	88444	321688.92	1.0
20	76848	342898.94	1.0
21	83110	403109.89	1.0
22	33550	168475.23	3.0
23	65862	195489.13	3.0
24	62844	225383.89	0.0
25	19214	114444.77	2.0
26	44992	175456.95	3.0
28	66630	189991.35	3.0
29	75426	211154.37	0.0
30	97282	81545.55	2.0
31	56488	165810.09	3.0
32	36358	133803.80	2.0
33	41796	130358.39	2.0
34	41108	123299.60	2.0
35	87954	266665.35	0.0
36	86514	221279.16	0.0
37	67242	216831.04	0.0
38	49692	179281.69	3.0
39	74940	261826.18	0.0
6	89324	171166.01	3.0

7	115570	339516.33	1.0
9	130978	402590.49	1.0

De resultaten worden gevisualiseerd door een scatterplot te maken van de kwantiteit versus de eenheidsprijs, waarbij de punten zijn gekleurd op basis van hun toegewezen cluster.

```
[ ]: # (Helaas verouderd)
# plt.scatter(table['TERRITORY_NAME_EN'], table['CURRENCY_NAME'], c =
# prediction_results, cmap = 'rainbow')
# plt.show()
# Nu gaan we een bar weergeven, waar alle quantiteiten worden weergegeven per
# sales_branch
# Create a DataFrame with the branch names mapping
branch_names = sales_branch[['SALES_BRANCH_CODE', 'CITY', 'REGION']]
branch_names = branch_names.drop_duplicates()
branch_names['CITY']

plt.scatter(branch_names['CITY'].astype(str), branch_names['REGION'].
# astype(str))
plt.xticks(rotation=90)
plt.show()
```





De centra van de clusters worden geëxtraheerd en weergegeven.

```
[ ]: df.groupby('Centrum', as_index = False)['Centrum'].count()
```

```
[ ]: Centrum
0      8
1      7
2      6
3      7
```

## 2.2 Clusteringmodel bouwen met meer dan 2 dimensies (alle kolommen uit de dataset)

Nu gaan we hetzelfde doen als hierboven, maar dan met alle kolommen uit de dataset. En misschien ook verder testen met verschillende k's.

```
[ ]: kmeans = KMeans(n_clusters = 5, random_state = 42)
prediction_results = kmeans.fit_predict(sales_product_data)
prediction_results
```

```
[ ]: array([4, 4, 2, 2, 4, 3, 2, 3, 4, 4, 0, 1, 1, 2, 0, 3, 2, 1, 1, 1, 3, 3,
          2, 2, 2, 4, 2, 4])
```

```
[ ]: sales_product_data['Centrum'] = prediction_results
sales_product_data
```

```
[ ]:
SALES_BRANCH_CODE    Camping Equipment    Golf Equipment  \
13                  0.409370          0.123656
14                  0.377478          0.155432
15                  0.411430          0.085079
17                  0.383126          0.042877
18                  0.401382          0.178628
19                  0.532534          0.084760
20                  0.391703          0.076419
21                  0.468878          0.087633
22                  0.405076          0.165482
23                  0.371622          0.207529
24                  0.370558          0.160986
25                  0.095070          0.330986
26                  0.089474          0.175439
28                  0.460123          0.095433
29                  0.316947          0.125485
30                  0.440063          0.123028
31                  0.331240          0.085557
32                  0.000000          0.315946
```

33	0.000000	0.332326
34	0.000000	0.264865
35	0.476429	0.068205
36	0.444041	0.095584
37	0.415531	0.106948
38	0.409861	0.087827
39	0.439621	0.081337
6	0.389522	0.123007
7	0.366025	0.113398
9	0.393523	0.105245

	Mountaineering Equipment	Outdoor Protection \
SALES_BRANCH_CODE		
13	0.132873	0.133641
14	0.161776	0.096749
15	0.213085	0.119214
17	0.237898	0.106501
18	0.106326	0.117491
19	0.048373	0.140839
20	0.253712	0.086463
21	0.142097	0.112203
22	0.081218	0.148223
23	0.143822	0.100386
24	0.099347	0.191443
25	0.068662	0.209507
26	0.114912	0.250877
28	0.194274	0.081118
29	0.206986	0.207633
30	0.077287	0.134069
31	0.285714	0.121664
32	0.000000	0.228018
33	0.000000	0.265861
34	0.000000	0.393243
35	0.127382	0.108325
36	0.099819	0.116757
37	0.212534	0.060627
38	0.284284	0.063945
39	0.173653	0.095309
6	0.135156	0.138193
7	0.232925	0.095009
9	0.165083	0.116156

	Personal Accessories	Centrum
SALES_BRANCH_CODE		
13	0.200461	4
14	0.208565	4
15	0.171192	2

17	0.229599	2
18	0.196172	4
19	0.193493	3
20	0.191703	2
21	0.189189	3
22	0.200000	4
23	0.176641	4
24	0.177665	0
25	0.295775	1
26	0.369298	1
28	0.169052	2
29	0.142950	0
30	0.225552	3
31	0.175824	2
32	0.456036	1
33	0.401813	1
34	0.341892	1
35	0.219659	3
36	0.243799	3
37	0.204360	2
38	0.154083	2
39	0.210080	2
6	0.214123	4
7	0.192644	2
9	0.219993	4

```
[ ]: sales_product_data.groupby('Centrum', as_index = False)['Centrum'].count()
```

```
[ ]: Centrum
0      2
1      5
2      9
3      5
4      7
```

Hier visualiseren we de gemiddelde verkoop van verschillende productcategorieën binnen elk cluster.

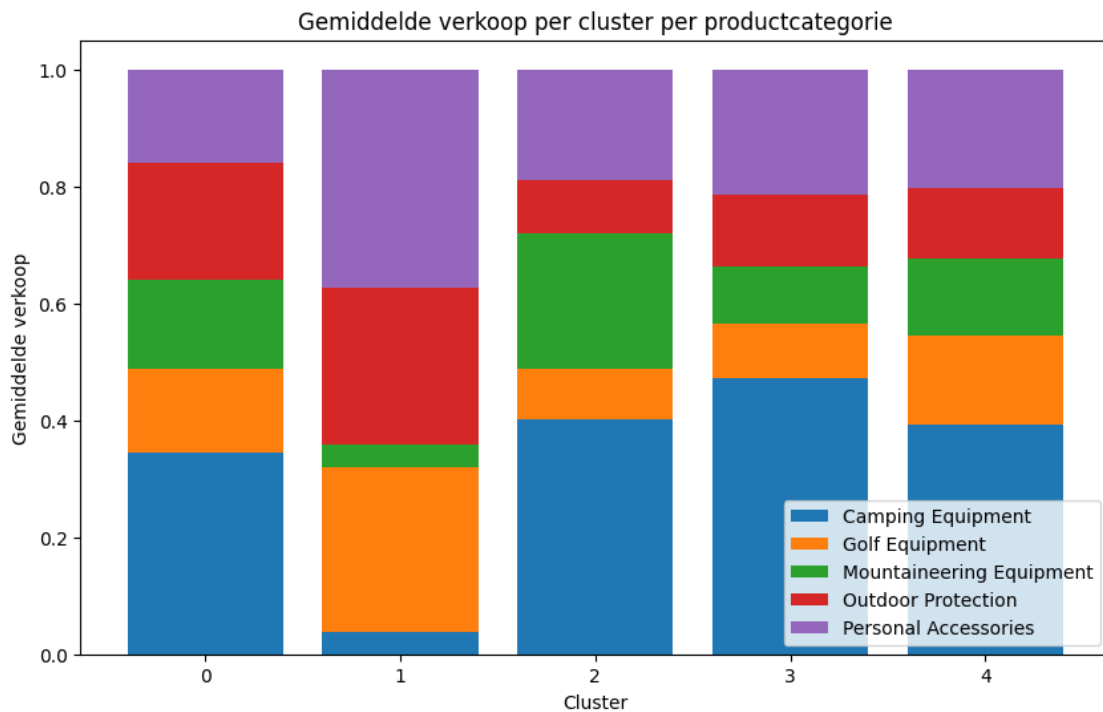
```
[ ]: cluster_sales_mean = sales_product_data.groupby('Centrum').mean()

# Visualisatie van gemiddelde verkoop per cluster
plt.figure(figsize=(10, 6))
plt.bar(cluster_sales_mean.index, cluster_sales_mean['Camping Equipment'],
        ↪label='Camping Equipment')
plt.bar(cluster_sales_mean.index, cluster_sales_mean['Golf Equipment'],
        ↪bottom=cluster_sales_mean['Camping Equipment'], label='Golf Equipment')
```

```

plt.bar(cluster_sales_mean.index, cluster_sales_mean['Mountaineering_
↳Equipment'], bottom=cluster_sales_mean['Camping_
↳Equipment']+cluster_sales_mean['Golf Equipment'], label='Mountaineering_
↳Equipment')
plt.bar(cluster_sales_mean.index, cluster_sales_mean['Outdoor Protection'],
↳bottom=cluster_sales_mean['Camping Equipment']+cluster_sales_mean['Golf_
↳Equipment']+cluster_sales_mean['Mountaineering Equipment'], label='Outdoor_
↳Protection')
plt.bar(cluster_sales_mean.index, cluster_sales_mean['Personal Accessories'],
↳bottom=cluster_sales_mean['Camping Equipment']+cluster_sales_mean['Golf_
↳Equipment']+cluster_sales_mean['Mountaineering_
↳Equipment']+cluster_sales_mean['Outdoor Protection'], label='Personal_
↳Accessories')
plt.xlabel('Cluster')
plt.ylabel('Gemiddelde verkoop')
plt.title('Gemiddelde verkoop per cluster per productcategorie')
plt.legend()
plt.show()

```



## 2.3 Evaluatie van de clustering

Hier gaan we de inter- en intraclusterafstand berekenen voor verschillende k's.

```

[ ]: # Lijst om interclusterafstanden op te slaan
intercluster_distances = []

# Lijst om intraclusterafstanden op te slaan
intracluster_distances = []

# Lijst van verschillende k's om te evalueren
k_values = [2, 3, 4, 5, 6]

for k in k_values:
    # K-means clusteringmodel toepassen
    kmeans = KMeans(n_clusters=k, random_state=42)
    prediction_results = kmeans.fit_predict(sales_product_data)

    # Interclusterafstand
    intercluster_distance = np.sum(np.min(kmeans.transform(sales_product_data),
↪axis=1)) / sales_product_data.shape[0]
    intercluster_distances.append(intercluster_distance)

    # Intraclusterafstand
    intracluster_distance = kmeans.inertia_ / sales_product_data.shape[0]
    intracluster_distances.append(intracluster_distance)

# Visualisatie van inter- en intraclusterafstanden voor verschillende k's
plt.plot(k_values, intercluster_distances, marker='o', label='Intercluster_
↪distance')
plt.plot(k_values, intracluster_distances, marker='x', label='Intracluster_
↪distance')
plt.xlabel('Number of clusters (k)')
plt.ylabel('Distance')
plt.title('Inter- and Intracluster distances for different k values')
plt.legend()
plt.show()

```

