

## We are again using the Adult Dataset

So we can use our results from Project 1 of how to preprepare the Data, select the features, reduce Dimensionality in a maximal fassion and prepare everything

In [88]:

```
# Import Libs
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import graphviz
import copy
import numpy
import time
import sys
import altair as alt

from sklearn import datasets, tree
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.cluster import DBSCAN
from sklearn import metrics
from sklearn.cluster import KMeans

from os import system
from IPython.display import Image

%matplotlib inline
numpy.set_printoptions(threshold=sys.maxsize)
```

In [3]:

```
# Import the data and name the collums, clear NA
adultDf = pd.read_csv("adult.data", header=None, index_col=False,
                      names=['age', 'workclass', 'fnlwgt', "education", "educat
                              "marital-stat", "occupation", "relationship", "r
                              "cap-gains", "cap_loss", "hpw", "native_country"]

adultDf = adultDf.dropna()
```

Now prepare Data:

In [67]:

```

# >50k -> 1, <=50k ->0
adultDf['50k'] = adultDf['50k'].replace(to_replace=' >50K', value=1)
adultDf['50k'] = adultDf['50k'].replace(to_replace=' <=50K', value=0)

# Copy so not to mess with the original dataframe
adultData = adultDf.copy()

# Remove fnlwgt(uncorrelated) and education (double data)
adultData.drop("fnlwgt", axis=1, inplace=True)
adultData.drop("education", axis=1, inplace=True)

# Add together Capital gains and losses
cap_gains = adultData["cap-gains"] - adultData["cap_loss"]
adultData["cap-gains"] = cap_gains
adultData.drop("cap_loss", axis=1, inplace=True)

# Repace Sex with 0,1:
adultData['sex'] = adultData['sex'].replace(to_replace=' Female', value=1)
adultData['sex'] = adultData['sex'].replace(to_replace=' Male', value=0)

# Repace all the complicated relationship status with -0,1 for single, married
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Wife', value=0)
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Husband', value=0)
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Unmarried', value=1)
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Other', value=1)
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Not in relationship', value=1)
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Own relationship', value=1)

# Remove marital-stat as this only doubles relationship status
adultData.drop("marital-stat", axis=1, inplace=True)

# Remove native Country as turning that into numerical data will either
# add unwanted ordering (like using the GDP or development Index)
# Or give us the curse of dimensionality if we add a "isfrom" variable for all
adultData.drop("native_country", axis=1, inplace=True)

# Add 3 binary collums for Workclass, isPrivate, isPublic, isSelfemployed, wh
adultData.insert(2, "isPrivate", ([0]*adultData.shape[0]), True)
adultData.insert(2, "isPublic", ([0]*adultData.shape[0]), True)
adultData.insert(2, "isSelfemployed", ([0]*adultData.shape[0]), True)

# Now fill them, if we mask multiple times we need to add the earlier 0/1s:
adultData['isPrivate'] = numpy.where(adultData['workclass']==' Private' , 1,
adultData['isPublic'] = numpy.where(adultData["workclass"]==" State-gov" , 1
adultData['isPublic'] = numpy.where(adultData['workclass']==' Local-gov' , 1
adultData['isPublic'] = numpy.where(adultData['workclass']==' Federal-gov', 1
adultData["isSelfemployed"] = numpy.where(adultData['workclass']==' Self-emp
adultData["isSelfemployed"] = numpy.where(adultData['workclass']==' Self-emp

# Now we can remove the Workclass
adultData.drop('workclass', axis=1, inplace=True)

# Add 3 binary collums for Race, isWhite, isBlack, isAsian,
# Where a 0 on all means AmerIndianEskimo or Other as we only have limited Da
adultData.insert(8, "isWhite", ([0]*adultData.shape[0]), True)
adultData.insert(8, "isBlack", ([0]*adultData.shape[0]), True)
adultData.insert(8, "isAsian", ([0]*adultData.shape[0]), True)
adultData['isWhite'] = numpy.where(adultData['race']==' White' , 1, 0)

```

```

# 2 High paying fields >40% over 50k
adultData["occField"] = numpy.where(adultData['occupation']== ' Exec-managerial', 1, 0)
adultData["occField"] = numpy.where(adultData['occupation']== ' Prof-specialty', 1, 0)
# 3 low paying fields, <10% over 50k
adultData["occField"] = numpy.where(adultData['occupation']== ' Handlers-cleaners', 1, 0)
adultData["occField"] = numpy.where(adultData['occupation']== ' Other-service', 1, 0)
adultData["occField"] = numpy.where(adultData['occupation']== ' Priv-house-serv', 1, 0)
adultData.drop('occupation', axis=1, inplace=True)

# STANDARDIZE NOW: <--- Mistake, not snormalized Now, but standardized numerical
adultData['education_num'] = adultData['education_num'].apply(lambda x: x/13*30)
column = adultData["age"]
max_age = column.max()
adultData['age'] = adultData['age'].apply(lambda x: x/max_age*30)
column = adultData["cap-gains"]
max_capgains = column.max()
adultData['cap-gains'] = adultData['cap-gains'].apply(lambda x: x/max_capgains*30)
column = adultData["hpw"]
max_hpw = column.max()
adultData['hpw'] = adultData['hpw'].apply(lambda x: x/max_hpw*30)
# Lets look at the Table Now
adultDataLabels = adultData["50k"]
adultData.drop('50k', axis=1, inplace=True)
print(adultData)
print(adultDataLabels)

```

	age	isSelfemployed	isPublic	isPrivate	education_num	\
0	13.000000	0	1	0	30.000000	
1	16.666667	1	0	0	30.000000	
2	12.666667	0	0	1	20.769231	
3	17.666667	0	0	1	16.153846	
4	9.333333	0	0	1	30.000000	
...	...	...	...	...	...	
32556	9.000000	0	0	1	27.692308	
32557	13.333333	0	0	1	20.769231	
32558	19.333333	0	0	1	20.769231	
32559	7.333333	0	0	1	20.769231	
32560	17.333333	1	0	0	20.769231	

  

	occField	relationship	isAsian	isBlack	isWhite	sex	cap-gains	\
0	0	0	0	0	1	0	0.652207	
1	1	1	0	0	1	0	0.000000	
2	-1	0	0	0	1	0	0.000000	
3	-1	1	0	1	0	0	0.000000	
4	1	1	0	1	0	1	0.000000	
...	...	...	...	...	...	...	...	
32556	0	1	0	0	1	1	0.000000	
32557	0	1	0	0	1	0	0.000000	
32558	0	0	0	0	1	1	0.000000	
32559	0	0	0	0	1	0	0.000000	
32560	1	1	0	0	1	1	4.507245	

```

      hpw
0      12.121212
1       3.939394
2      12.121212
3      12.121212
4      12.121212
...
32556  11.515152
32557  12.121212
32558  12.121212
32559   6.060606
32560  12.121212

```

```
[32561 rows x 13 columns]
```

```

0      0
1      0
2      0
3      0
4      0
...
32556  0
32557  1
32558  0
32559  0
32560  1

```

```
Name: 50k, Length: 32561, dtype: int64
```

## Now we can run the DBSCAN Algorithm

For a fast and simple way to get Metrics for our resulting clusterings we followed the Example from the scikit-learn website:

[https://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_dbscan.html#sphx-glr-auto-examples-cluster-plot-dbscan-py](https://scikit-learn.org/stable/auto_examples/cluster/plot_dbscan.html#sphx-glr-auto-examples-cluster-plot-dbscan-py)

We will first run the clustering with some Parameters, and then run the same Evaluation Code to learn how it went

In [70]:

```

# epsilon = 1 and min_samples = 10
clustering_1_10_e = DBSCAN(eps=1, min_samples=10, metric='euclidean').fit(adu

```

In [71]:

```

# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = clustering_1_10_e.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultDataLabels, labels))
# I have no idea why the Labels 0 and 1 are in places 1 and 2. Wierd, i think
print("Confusion Matrix")
print( metrics.confusion_matrix(adultDataLabels, labels)[1] )
print( metrics.confusion_matrix(adultDataLabels, labels)[2] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1

```

Estimated number of clusters: 79

Estimated number of noise points: 13966

Homogeneity: 0.069

Confusion Matrix

```

[9862 6261 1643 4113    4 103    5 10 391 215 246 83 30 88
 69 154 40 23 236 274 25 35 31 134 8 48 6 74
 21 16 3 16 11 1 7 10 18 10 7 31 11 1
 9 54 3 12 0 5 14 3 7 10 22 4 20 7
 15 5 3 2 6 22 12 17 5 10 4 4 0 4
 4 8 10 8 10 7 7 3 5 5]
[4104 1045 1110 767 13 239 5 0 102 42 0 1 22 80
 28 8 0 2 22 13 0 0 0 37 1 0 9 0
 16 0 3 0 14 13 3 8 13 0 0 2 0 13
 0 2 7 2 11 3 0 5 6 6 0 8 0 3
 0 7 2 12 4 2 2 1 1 0 4 0 5 0
 6 0 2 0 0 0 0 5 0 0]

```

Silhouette Coefficient: -0.540

Overview over the Cluster

Cluster 0 has Size: 7306

Silluete is: -0.3490971287620456

Cluster 1 has Size: 2753

Silluete is: -0.4231838541750693

Cluster 2 has Size: 4880

Silluete is: -0.2537262611786824

Cluster 3 has Size: 17

Silluete is: -0.8144877012559446

Cluster 4 has Size: 342  
Silluete is: -0.8686457989717775

Cluster 5 has Size: 10  
Silluete is: -0.0042507866560079245

Cluster 6 has Size: 10  
Silluete is: -0.5502472013953666

Cluster 7 has Size: 493  
Silluete is: -0.7314765860839526

Cluster 8 has Size: 257  
Silluete is: -0.6231630662090505

Cluster 9 has Size: 246  
Silluete is: -0.652548811900525

Cluster 10 has Size: 84  
Silluete is: -0.5662685047582137

Cluster 11 has Size: 52  
Silluete is: -0.8008871343741062

Cluster 12 has Size: 168  
Silluete is: -0.5343306484372548

Cluster 13 has Size: 97  
Silluete is: -0.6988809113595241

Cluster 14 has Size: 162  
Silluete is: -0.7842154613111251

Cluster 15 has Size: 40  
Silluete is: -0.7734166414829456

Cluster 16 has Size: 25  
Silluete is: -0.7474556177130561

Cluster 17 has Size: 258  
Silluete is: -0.33701951666341073

Cluster 18 has Size: 287  
Silluete is: -0.6745777405285928

Cluster 19 has Size: 25  
Silluete is: -0.7813142884108677

Cluster 20 has Size: 35  
Silluete is: -0.7534485107256563

Cluster 21 has Size: 31  
Silluete is: -0.4981763837228431

Cluster 22 has Size: 171  
Silluete is: -0.7819816520024891

Cluster 23 has Size: 9

Silluete is: -0.23533873644966333

Cluster 24 has Size: 48  
Silluete is: -0.7665152671005844

Cluster 25 has Size: 15  
Silluete is: -0.7352677051411972

Cluster 26 has Size: 74  
Silluete is: -0.6729658801440149

Cluster 27 has Size: 37  
Silluete is: -0.8092420744974352

Cluster 28 has Size: 16  
Silluete is: -0.5524886505746539

Cluster 29 has Size: 6  
Silluete is: -0.6034618350139074

Cluster 30 has Size: 16  
Silluete is: -0.7468637870113657

Cluster 31 has Size: 25  
Silluete is: -0.78174156462021

Cluster 32 has Size: 14  
Silluete is: -0.4083942380462062

Cluster 33 has Size: 10  
Silluete is: -0.5807290357258584

Cluster 34 has Size: 18  
Silluete is: -0.7784615639798542

Cluster 35 has Size: 31  
Silluete is: -0.6464106177572934

Cluster 36 has Size: 10  
Silluete is: -0.42616801237914786

Cluster 37 has Size: 7  
Silluete is: -0.5839198974440366

Cluster 38 has Size: 33  
Silluete is: -0.476785653112585

Cluster 39 has Size: 11  
Silluete is: -0.8101964424658494

Cluster 40 has Size: 14  
Silluete is: 0.5120717798658136

Cluster 41 has Size: 9  
Silluete is: 0.6480949450971449

Cluster 42 has Size: 56  
Silluete is: -0.5544309342265901

Cluster 43 has Size: 10  
Silluete is: -0.5309758706299244

Cluster 44 has Size: 14  
Silluete is: -0.40125426084791055

Cluster 45 has Size: 11  
Silluete is: -0.8283220679995854

Cluster 46 has Size: 8  
Silluete is: -0.61999231276368

Cluster 47 has Size: 14  
Silluete is: -0.15609890633586435

Cluster 48 has Size: 8  
Silluete is: 0.7568884176984447

Cluster 49 has Size: 13  
Silluete is: -0.35224081028901594

Cluster 50 has Size: 16  
Silluete is: -0.7776450032206813

Cluster 51 has Size: 22  
Silluete is: -0.4419815428051927

Cluster 52 has Size: 12  
Silluete is: -0.7269614766433077

Cluster 53 has Size: 20  
Silluete is: -0.7985967604169909

Cluster 54 has Size: 10  
Silluete is: -0.7626670688428169

Cluster 55 has Size: 15  
Silluete is: -0.5158562480175063

Cluster 56 has Size: 12  
Silluete is: 0.840732742491737

Cluster 57 has Size: 5  
Silluete is: -0.22383092815208686

Cluster 58 has Size: 14  
Silluete is: -0.3054678496533522

Cluster 59 has Size: 10  
Silluete is: -0.3345134491427981

Cluster 60 has Size: 24  
Silluete is: -0.3954336800149919

Cluster 61 has Size: 14  
Silluete is: -0.799749384004595

Cluster 62 has Size: 18  
Silluete is: -0.5765893327196382



```
Cluster 63 has Size: 6
Silluete is: -0.7953774679256995

Cluster 64 has Size: 10
Silluete is: -0.5335188563085873

Cluster 65 has Size: 8
Silluete is: -0.358373555978187

Cluster 66 has Size: 4
Silluete is: -0.5286843546371884

Cluster 67 has Size: 5
Silluete is: -0.5985978230877361

Cluster 68 has Size: 4
Silluete is: 0.36216905937859906

Cluster 69 has Size: 10
Silluete is: -0.5478353435883552

Cluster 70 has Size: 8
Silluete is: -0.6833812495719099

Cluster 71 has Size: 12
Silluete is: 0.30197600081240833

Cluster 72 has Size: 8
Silluete is: -0.7564583670574939

Cluster 73 has Size: 10
Silluete is: -0.4650489345817875

Cluster 74 has Size: 7
Silluete is: -0.44069295836395245

Cluster 75 has Size: 7
Silluete is: -0.4294926914865068

Cluster 76 has Size: 8
Silluete is: -0.14992173396108519

Cluster 77 has Size: 5
Silluete is: -0.3856176445708867

Cluster 78 has Size: 5
Silluete is: 0.22867395718103373

Cluster -1 has Size: 13966
```

Thats Bad, lots of clusters, bad silhouette, no homogeneity

```
In [72]: # epsilon = 1 and min_samples = 10
         clustering_05_5_e = DBSCAN(eps=0.5, min_samples=5, metric='euclidean').fit(ad
```

In [73]:

```

# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = clustering_05_5_e.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultData.labels, labels))
# I have no idea why the Labels 0 and 1 are in places 1 and 2. Wierd, i think
print("Confusion Matrix")
print( metrics.confusion_matrix(adultData.labels, labels)[1] )
print( metrics.confusion_matrix(adultData.labels, labels)[2] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1

```

Estimated number of clusters: 594

Estimated number of noise points: 19445

Homogeneity: 0.152

Confusion Matrix

14087	223	7	692	577	1050	12	14	295	1	34	16
295	10	382	59	20	14	14	115	5	47	92	147
9	39	6	10	6	12	5	18	30	23	18	17
10	128	12	16	14	62	10	53	11	17	19	22
10	5	1	24	32	8	67	10	3	43	13	9
147	6	12	17	12	13	9	50	10	0	62	17
18	15	4	9	46	64	15	74	7	5	24	7
5	50	2	6	66	9	19	56	36	5	101	17
46	24	11	10	13	8	14	123	17	21	57	6
11	50	16	1	15	4	5	49	36	8	11	36
9	4	17	5	11	77	18	12	37	33	84	9
3	76	20	12	14	11	5	17	4	24	15	5
5	14	30	18	30	15	31	7	18	13	4	6
8	14	40	41	6	8	13	22	5	9	3	66
14	12	6	2	4	10	18	18	8	32	20	2
20	1	13	13	1	31	32	5	5	5	3	40
2	3	9	2	7	58	15	12	1	33	37	4
5	5	5	9	8	5	7	7	18	2	14	21
4	9	20	8	5	20	37	4	15	14	6	11
46	1	27	17	9	13	11	2	5	4	4	4
59	17	8	3	5	0	25	2	26	6	6	1
5	5	1	4	4	17	5	8	7	17	3	1
7	3	6	22	6	9	8	5	3	1	6	9
17	17	1	1	11	18	8	7	17	9	22	3
11	9	7	7	3	6	12	3	7	11	8	7
26	3	14	5	11	7	17	0	5	10	4	16

	0	1	1	2	8	3	2	4	16	9	23	6	
16	7	5	5	5	6	8	6	6	0	22	22	4	
5	5	5	2	6	5	5	5	5	8	4	6	16	
1	10	10	16	10	10	10	0	8	7	6	6	2	
8	5	7	5	0	4	7	5	5	5	6	5	4	
28	6	5	8	16	7	7	6	2	7	7	0	5	
3	10	4	4	18	2	7	1	5	6	6	5	11	
8	8	5	16	5	9	6	6	5	6	6	5	2	
2	7	5	6	5	1	10	6	6	6	5	10	33	
5	2	6	7	8	4	3	7	8	3	8	8	8	
6	3	7	13	5	8	2	4	5	10	6	1	1	
9	4	8	22	6	0	5	8	5	5	5	5	6	
7	5	6	6	6	5	5	5	4	3	4	4	4	
4	4	8	8	13	3	8	9	5	5	5	5	4	
4	2	4	5	5	5	7	5	5	6	4	3	3	
1	5	6	5	5	4	2	5	8	7	6	5	5	
7	4	5	6	6	7	5	5	6	3	6	5	5	
5	7	7	3	3	6	2	5	6	6	9	1	1	
3	5	3	5	2	10	2	5	5	6	5	4	4	
5	5	9	8	5	5	5	8	2	4	5	4	4	
3	1	4	5	5	2	7	3	5	2	5	5	5	
5	4	0	4	4	1	5	3	2	0	5	4	4	
3	5	1	3	1	2	5	5	0	2	4	2	2	
5	1	0	2	4	5	3							
[5358	0	0	12	6	414	5	1	11	7	0	0	163	2
3	0	12	0	0	66	0	0	13	14	4	0	11	2
1	31	31	3	0	0	0	0	0	0	0	0	3	0
0	14	16	0	0	79	3	0	5	7	1	1	156	0
10	0	1	0	7	2	0	1	0	1	6	2	1	6
5	1	48	0	5	5	43	1	0	43	0	0	0	0
0	0	3	1	0	10	0	48	22	0	8	0	59	0
1	0	0	0	3	3	0	5	1	1	12	0	0	5
0	2	0	16	0	1	0	9	0	5	0	0	0	0
0	23	0	0	105	0	2	3	0	52	2	0	3	0
5	5	0	3	4	0	1	0	0	2	0	0	5	0
7	2	0	3	0	2	0	0	0	6	0	8	9	0
6	5	1	3	4	0	0	9	1	0	39	4	8	4
7	0	6	0	23	0	0	0	8	0	8	8	0	6
0	0	0	1	4	0	4	1	0	3	0	0	0	6
0	0	0	12	0	1	1	0	0	10	0	0	1	1
0	1	3	0	0	5	0	1	0	0	2	5	2	2
2	5	0	0	0	3	0	5	0	3	0	1	0	4
0	1	26	2	2	1	0	0	0	0	2	5	0	3
0	0	0	0	0	0	3	4	0	0	0	0	5	6
0	0	0	0	0	0	2	2	1	2	11	0	2	0
0	8	0	0	0	0	0	2	0	1	0	4	0	7
0	0	1	3	5	6	8	3	1	3	3	1	0	6
0	0	0	0	2	0	0	0	1	0	6	0	0	3
1	1	0	6	1	0	0	0	0	2	2	0	5	0
3	1	0	0	8	0	0	0	0	3	0	0	9	0
5	4	4	0	0	4	0	1	0	0	0	0	0	0
0	0	4	1	9	0	1	0	2	2	0	3	0	4
0	0	0	0	0	1	2	0	0	0	0	0	3	0
0	13	2	0	0	0	0	6	0	0	0	0	0	1
1	3	3	2	1	1	4	1	0	3	2	0	6	10
0	1	1	0	4	2	4	1	0	5	0	1	0	0
0	5	0	0	0	1	1	0	1	2	0	0	0	1
0	0	1	5	1	1	1	2	0	0	0	2	0	2
0	1	0	1	2	2	2	4	0	0	0	0	0	0

1	2	5	0	0	0	3	1	5	0	0	0	2	0
2	7	0	0	0	0	1	0	0	0	0	0	0	0
0	3	2	1	5	0	0	0	0	4	2	0	3	0
3	0	1	0	0	0	0	1	0	0	0	0	0	0
0	0	2	0	0	0	0	4	2	0	0	1	0	2
0	3	0	0	0	3	5	1	1	4	0	2	3	5
0	0	0	0	4	3	0	3	0	0	6	3	1	2
0	3	6	4	1	0	2]							

Silhouette Coefficient: -0.509

Overview over the Cluster

Cluster 0 has Size: 223

Silluete is: -0.9099651384164588

Cluster 1 has Size: 7

Silluete is: -0.4542331661800692

Cluster 2 has Size: 704

Silluete is: -0.5686356679343483

Cluster 3 has Size: 583

Silluete is: -0.8708745235776049

Cluster 4 has Size: 1464

Silluete is: -0.8765173927056469

Cluster 5 has Size: 17

Silluete is: -0.8688603709215722

Cluster 6 has Size: 15

Silluete is: -0.5434140676590435

Cluster 7 has Size: 306

Silluete is: -0.8047902791870272

Cluster 8 has Size: 8

Silluete is: -0.6610231568939334

Cluster 9 has Size: 34

Silluete is: -0.925518695598126

Cluster 10 has Size: 16

Silluete is: -0.7044206422167366

Cluster 11 has Size: 458

Silluete is: -0.8479870160392302

Cluster 12 has Size: 12

Silluete is: 0.7668145635755153

Cluster 13 has Size: 385

Silluete is: -0.7090129455288001

Cluster 14 has Size: 59

Silluete is: -0.8079412848771793

Cluster 15 has Size: 32

Silluete is: -0.883651099050557

Cluster 16 has Size: 14  
Silluete is: -0.8444601814506356

Cluster 17 has Size: 14  
Silluete is: -0.6199216996663627

Cluster 18 has Size: 181  
Silluete is: -0.7113981845005647

Cluster 19 has Size: 5  
Silluete is: -0.8317020383276225

Cluster 20 has Size: 47  
Silluete is: -0.7588511995276653

Cluster 21 has Size: 105  
Silluete is: -0.6460725347212048

Cluster 22 has Size: 161  
Silluete is: -0.8517385496541822

Cluster 23 has Size: 13  
Silluete is: -0.9350194414618332

Cluster 24 has Size: 39  
Silluete is: -0.8597553811806239

Cluster 25 has Size: 17  
Silluete is: -0.8679868152788289

Cluster 26 has Size: 12  
Silluete is: -0.6967417486147008

Cluster 27 has Size: 7  
Silluete is: -0.8290649236496961

Cluster 28 has Size: 43  
Silluete is: -0.7627910812048375

Cluster 29 has Size: 36  
Silluete is: -0.6876457376280307

Cluster 30 has Size: 21  
Silluete is: -0.7636172883933082

Cluster 31 has Size: 30  
Silluete is: -0.8689558973966761

Cluster 32 has Size: 23  
Silluete is: 0.5483444670647125

Cluster 33 has Size: 18  
Silluete is: 0.3939243828171768

Cluster 34 has Size: 17  
Silluete is: -0.8234140681292494

Cluster 35 has Size: 10  
Silluete is: -0.8837069908736864

Cluster 36 has Size: 128  
Silluete is: -0.5629309269926802

Cluster 37 has Size: 12  
Silluete is: -0.9131689924096165

Cluster 38 has Size: 16  
Silluete is: -0.860194275845499

Cluster 39 has Size: 17  
Silluete is: -0.8173225664035986

Cluster 40 has Size: 62  
Silluete is: -0.9286200632866639

Cluster 41 has Size: 10  
Silluete is: 0.6753246753246575

Cluster 42 has Size: 67  
Silluete is: -0.9158935396242225

Cluster 43 has Size: 27  
Silluete is: -0.7522360425510393

Cluster 44 has Size: 17  
Silluete is: 0.09332335671739145

Cluster 45 has Size: 19  
Silluete is: -0.828722119907001

Cluster 46 has Size: 101  
Silluete is: -0.9133612902072216

Cluster 47 has Size: 13  
Silluete is: -0.9111298677849174

Cluster 48 has Size: 5  
Silluete is: -0.8061656413795889

Cluster 49 has Size: 6  
Silluete is: -0.8591956577771345

Cluster 50 has Size: 31  
Silluete is: -0.8280589015295253

Cluster 51 has Size: 33  
Silluete is: 0.5972988047449396

Cluster 52 has Size: 9  
Silluete is: -0.7247984572101293

Cluster 53 has Size: 223  
Silluete is: -0.8402249412595735

Cluster 54 has Size: 10  
Silluete is: -0.8687010842492898

Cluster 55 has Size: 13

Silluete is: -0.6298761538837394

Cluster 56 has Size: 43  
Silluete is: 0.86824285898423

Cluster 57 has Size: 14  
Silluete is: -0.808479688847535

Cluster 58 has Size: 9  
Silluete is: -0.5740762092539733

Cluster 59 has Size: 154  
Silluete is: 0.3842688536716596

Cluster 60 has Size: 8  
Silluete is: -0.884990745548604

Cluster 61 has Size: 12  
Silluete is: -0.8960656378261528

Cluster 62 has Size: 18  
Silluete is: -0.6973229578502695

Cluster 63 has Size: 12  
Silluete is: -0.7868521450686506

Cluster 64 has Size: 14  
Silluete is: -0.8381783104117971

Cluster 65 has Size: 15  
Silluete is: -0.6744211083874132

Cluster 66 has Size: 52  
Silluete is: -0.6262477674563012

Cluster 67 has Size: 11  
Silluete is: -0.25600849450184515

Cluster 68 has Size: 6  
Silluete is: -0.89738566661074

Cluster 69 has Size: 67  
Silluete is: -0.05162780539116612

Cluster 70 has Size: 18  
Silluete is: -0.893600611301161

Cluster 71 has Size: 66  
Silluete is: -0.9020334190690789

Cluster 72 has Size: 15  
Silluete is: -0.7458530119927838

Cluster 73 has Size: 9  
Silluete is: -0.4623645927378718

Cluster 74 has Size: 14  
Silluete is: -0.5327950985904314

Cluster 75 has Size: 89  
Silluete is: -0.2370808005806802

Cluster 76 has Size: 65  
Silluete is: -0.06717180463686427

Cluster 77 has Size: 15  
Silluete is: -0.38987311484327564

Cluster 78 has Size: 117  
Silluete is: -0.955213194358859

Cluster 79 has Size: 7  
Silluete is: -0.8979648505414121

Cluster 80 has Size: 5  
Silluete is: 0.5156640305412908

Cluster 81 has Size: 24  
Silluete is: -0.8442031357646607

Cluster 82 has Size: 7  
Silluete is: -0.8827362004918089

Cluster 83 has Size: 5  
Silluete is: -0.906679437479186

Cluster 84 has Size: 50  
Silluete is: -0.6710456040418076

Cluster 85 has Size: 5  
Silluete is: 0.6487123375202337

Cluster 86 has Size: 7  
Silluete is: -0.5681348755217788

Cluster 87 has Size: 66  
Silluete is: 0.8483047231062564

Cluster 88 has Size: 19  
Silluete is: 0.68104427466784

Cluster 89 has Size: 19  
Silluete is: -0.7627616490693426

Cluster 90 has Size: 104  
Silluete is: -0.8658267710001022

Cluster 91 has Size: 58  
Silluete is: -0.612135725242718

Cluster 92 has Size: 5  
Silluete is: -0.7886450033577531

Cluster 93 has Size: 109  
Silluete is: -0.80734934651161

Cluster 94 has Size: 17  
Silluete is: 0.5240541297223896



Cluster 95 has Size: 105  
Silluete is: -0.798167624358604

Cluster 96 has Size: 24  
Silluete is: -0.665267188491076

Cluster 97 has Size: 12  
Silluete is: -0.8614135385735787

Cluster 98 has Size: 10  
Silluete is: -0.870625329809957

Cluster 99 has Size: 13  
Silluete is: -0.8430836748474299

Cluster 100 has Size: 8  
Silluete is: -0.6932074720912034

Cluster 101 has Size: 17  
Silluete is: -0.6453509451784527

Cluster 102 has Size: 126  
Silluete is: -0.818188870904933

Cluster 103 has Size: 17  
Silluete is: -0.7440055884628982

Cluster 104 has Size: 26  
Silluete is: -0.622796255819265

Cluster 105 has Size: 58  
Silluete is: -0.8175487663344505

Cluster 106 has Size: 7  
Silluete is: -0.44549188200934925

Cluster 107 has Size: 23  
Silluete is: -0.864407437394507

Cluster 108 has Size: 50  
Silluete is: -0.4791816727881902

Cluster 109 has Size: 16  
Silluete is: -0.484239133866574

Cluster 110 has Size: 6  
Silluete is: -0.5396242607381548

Cluster 111 has Size: 15  
Silluete is: -0.7979711284892056

Cluster 112 has Size: 6  
Silluete is: -0.8888965235536862

Cluster 113 has Size: 5  
Silluete is: -0.7791231409498989

Cluster 114 has Size: 65

Silluete is: -0.783426976025379

Cluster 115 has Size: 36  
Silluete is: -0.7854747942402225

Cluster 116 has Size: 9  
Silluete is: 0.5061408971557103

Cluster 117 has Size: 11  
Silluete is: -0.7466001702843669

Cluster 118 has Size: 45  
Silluete is: -0.8300939015168268

Cluster 119 has Size: 9  
Silluete is: -0.6221305003593631

Cluster 120 has Size: 9  
Silluete is: 0.653462538846902

Cluster 121 has Size: 17  
Silluete is: 0.5509808673638306

Cluster 122 has Size: 5  
Silluete is: 0.7963659518821911

Cluster 123 has Size: 11  
Silluete is: -0.9068414779370395

Cluster 124 has Size: 77  
Silluete is: -0.8597158158350431

Cluster 125 has Size: 18  
Silluete is: -0.8834561747577682

Cluster 126 has Size: 35  
Silluete is: -0.916773923582276

Cluster 127 has Size: 37  
Silluete is: 0.6869124447768737

Cluster 128 has Size: 33  
Silluete is: 0.8191234055688561

Cluster 129 has Size: 189  
Silluete is: -0.725405655728969

Cluster 130 has Size: 9  
Silluete is: -0.7551984863201913

Cluster 131 has Size: 5  
Silluete is: -0.872149558765597

Cluster 132 has Size: 79  
Silluete is: -0.7888493179767277

Cluster 133 has Size: 20  
Silluete is: 0.20921348129610756

Cluster 134 has Size: 64  
Silluete is: -0.8080158651398862

Cluster 135 has Size: 16  
Silluete is: 0.17162639164679155

Cluster 136 has Size: 11  
Silluete is: 0.5258165327067199

Cluster 137 has Size: 8  
Silluete is: -0.8979433419433056

Cluster 138 has Size: 17  
Silluete is: 0.19091063865876626

Cluster 139 has Size: 9  
Silluete is: -0.8112947089914702

Cluster 140 has Size: 29  
Silluete is: 0.6138291950826182

Cluster 141 has Size: 15  
Silluete is: -0.6390137127048887

Cluster 142 has Size: 8  
Silluete is: 0.17829130063773713

Cluster 143 has Size: 9  
Silluete is: -0.8985389964557936

Cluster 144 has Size: 14  
Silluete is: -0.8398103908668103

Cluster 145 has Size: 31  
Silluete is: -0.8300616931920664

Cluster 146 has Size: 18  
Silluete is: -0.8969856946976653

Cluster 147 has Size: 30  
Silluete is: -0.8306978267432769

Cluster 148 has Size: 17  
Silluete is: -0.8554883300750945

Cluster 149 has Size: 31  
Silluete is: -0.8612508471205056

Cluster 150 has Size: 7  
Silluete is: -0.867395161241805

Cluster 151 has Size: 23  
Silluete is: -0.6997030405589192

Cluster 152 has Size: 13  
Silluete is: -0.8653574225021081

Cluster 153 has Size: 11  
Silluete is: 0.6325527085523068

Cluster 154 has Size: 8  
Silluete is: -0.8978567137785599

Cluster 155 has Size: 8  
Silluete is: -0.6058177482742073

Cluster 156 has Size: 17  
Silluete is: 0.634243599972965

Cluster 157 has Size: 40  
Silluete is: -0.33584042534706743

Cluster 158 has Size: 43  
Silluete is: 0.6362837075182752

Cluster 159 has Size: 6  
Silluete is: -0.8091173713677381

Cluster 160 has Size: 8  
Silluete is: -0.6894312577161332

Cluster 161 has Size: 13  
Silluete is: -0.4438141026581775

Cluster 162 has Size: 28  
Silluete is: -0.812862261846835

Cluster 163 has Size: 5  
Silluete is: -0.8559940452248747

Cluster 164 has Size: 17  
Silluete is: -0.8346010087751555

Cluster 165 has Size: 12  
Silluete is: 0.48041955185752855

Cluster 166 has Size: 66  
Silluete is: -0.8098486038252551

Cluster 167 has Size: 20  
Silluete is: -0.8787817670674063

Cluster 168 has Size: 17  
Silluete is: 0.7001698075905928

Cluster 169 has Size: 7  
Silluete is: -0.5142303651838842

Cluster 170 has Size: 5  
Silluete is: -0.8461307845463407

Cluster 171 has Size: 8  
Silluete is: -0.6641260720927573

Cluster 172 has Size: 10  
Silluete is: -0.8681899359880922

Cluster 173 has Size: 18

Silluete is: 0.618055555555345

Cluster 174 has Size: 27  
Silluete is: -0.8833437065781883

Cluster 175 has Size: 9  
Silluete is: -0.9170085994401004

Cluster 176 has Size: 32  
Silluete is: -0.2907261984944486

Cluster 177 has Size: 59  
Silluete is: 0.7864392809249655

Cluster 178 has Size: 6  
Silluete is: -0.9230428939394392

Cluster 179 has Size: 28  
Silluete is: -0.8800441908349799

Cluster 180 has Size: 5  
Silluete is: -0.829944449984413

Cluster 181 has Size: 20  
Silluete is: -0.7393211166407109

Cluster 182 has Size: 13  
Silluete is: -0.8265440723328489

Cluster 183 has Size: 7  
Silluete is: 0.5

Cluster 184 has Size: 31  
Silluete is: -0.8659638551077699

Cluster 185 has Size: 55  
Silluete is: -0.3160496650450615

Cluster 186 has Size: 5  
Silluete is: -0.8326547588775387

Cluster 187 has Size: 5  
Silluete is: -0.8257259917978472

Cluster 188 has Size: 5  
Silluete is: -0.8969084050905335

Cluster 189 has Size: 11  
Silluete is: -0.4222603244515041

Cluster 190 has Size: 40  
Silluete is: -0.7342138335493825

Cluster 191 has Size: 10  
Silluete is: -0.8095030459560715

Cluster 192 has Size: 11  
Silluete is: -0.8643344161250361

Cluster 193 has Size: 9  
Silluete is: 0.4877813003215054

Cluster 194 has Size: 8  
Silluete is: -0.6206904135865129

Cluster 195 has Size: 7  
Silluete is: 0.4656920731082475

Cluster 196 has Size: 58  
Silluete is: -0.6876639265647856

Cluster 197 has Size: 15  
Silluete is: -0.6267126304615059

Cluster 198 has Size: 13  
Silluete is: -0.8172285179236066

Cluster 199 has Size: 5  
Silluete is: 0.7493154892373655

Cluster 200 has Size: 33  
Silluete is: -0.9471129094597793

Cluster 201 has Size: 41  
Silluete is: 0.35429018879540114

Cluster 202 has Size: 5  
Silluete is: -0.5326110532861955

Cluster 203 has Size: 5  
Silluete is: -0.8479906753596541

Cluster 204 has Size: 8  
Silluete is: -0.8268992628943972

Cluster 205 has Size: 5  
Silluete is: -0.9171054011617747

Cluster 206 has Size: 9  
Silluete is: -0.7540176512225185

Cluster 207 has Size: 8  
Silluete is: 0.5082591785213078

Cluster 208 has Size: 11  
Silluete is: -0.6492240413878702

Cluster 209 has Size: 7  
Silluete is: 0.8047440923230262

Cluster 210 has Size: 7  
Silluete is: -0.8400468861243022

Cluster 211 has Size: 18  
Silluete is: -0.6286021731400727

Cluster 212 has Size: 14  
Silluete is: -0.8351508421792903

Cluster 213 has Size: 14  
Silluete is: -0.8452242255900912

Cluster 214 has Size: 22  
Silluete is: -0.9136011084237099

Cluster 215 has Size: 5  
Silluete is: -0.8419644036660436

Cluster 216 has Size: 9  
Silluete is: -0.8646448338405193

Cluster 217 has Size: 20  
Silluete is: 0.27114997313937345

Cluster 218 has Size: 18  
Silluete is: 0.8336755279082723

Cluster 219 has Size: 5  
Silluete is: -0.9209553832252708

Cluster 220 has Size: 20  
Silluete is: 0.4457506286287099

Cluster 221 has Size: 38  
Silluete is: -0.862680471538885

Cluster 222 has Size: 5  
Silluete is: -0.5297276097347825

Cluster 223 has Size: 15  
Silluete is: -0.6348219483761302

Cluster 224 has Size: 15  
Silluete is: -0.7479634370493345

Cluster 225 has Size: 9  
Silluete is: -0.7659220755752143

Cluster 226 has Size: 11  
Silluete is: -0.46340186118121696

Cluster 227 has Size: 46  
Silluete is: -0.8091935640835889

Cluster 228 has Size: 6  
Silluete is: -0.5304787802081928

Cluster 229 has Size: 27  
Silluete is: -0.8638820905300316

Cluster 230 has Size: 18  
Silluete is: -0.8289616956420057

Cluster 231 has Size: 9  
Silluete is: -0.8408348255045778

Cluster 232 has Size: 13

Silluete is: 0.14815246577513444

Cluster 233 has Size: 13  
Silluete is: -0.8193331269588531

Cluster 234 has Size: 7  
Silluete is: -0.7724184814291968

Cluster 235 has Size: 7  
Silluete is: 0.6017130068688008

Cluster 236 has Size: 6  
Silluete is: -0.8570543569717421

Cluster 237 has Size: 6  
Silluete is: -0.5017459498677969

Cluster 238 has Size: 9  
Silluete is: -0.7752855871389762

Cluster 239 has Size: 59  
Silluete is: -0.6382439042062653

Cluster 240 has Size: 17  
Silluete is: -0.3537200580549233

Cluster 241 has Size: 8  
Silluete is: -0.6605100220915822

Cluster 242 has Size: 6  
Silluete is: -0.7468004428479046

Cluster 243 has Size: 5  
Silluete is: 0.7285285047090662

Cluster 244 has Size: 5  
Silluete is: -0.8647060234399312

Cluster 245 has Size: 25  
Silluete is: -0.8340961426441182

Cluster 246 has Size: 5  
Silluete is: 0.6891381102187042

Cluster 247 has Size: 26  
Silluete is: -0.32074128076255387

Cluster 248 has Size: 7  
Silluete is: -0.7586716882474851

Cluster 249 has Size: 6  
Silluete is: -0.6911835179594512

Cluster 250 has Size: 5  
Silluete is: -0.812372768149926

Cluster 251 has Size: 5  
Silluete is: -0.8220164886548537



Cluster 252 has Size: 6  
Silluete is: -0.7422772180123998

Cluster 253 has Size: 27  
Silluete is: 0.3108390339402811

Cluster 254 has Size: 6  
Silluete is: -0.5521250046416228

Cluster 255 has Size: 6  
Silluete is: -0.8115545617688908

Cluster 256 has Size: 18  
Silluete is: -0.5215294228414137

Cluster 257 has Size: 5  
Silluete is: 0.687500000000023

Cluster 258 has Size: 8  
Silluete is: -0.8139488506102905

Cluster 259 has Size: 7  
Silluete is: -0.9279711317248004

Cluster 260 has Size: 17  
Silluete is: -0.3973795803182371

Cluster 261 has Size: 5  
Silluete is: 0.7996024712114675

Cluster 262 has Size: 6  
Silluete is: 0.7106259828852867

Cluster 263 has Size: 7  
Silluete is: -0.8210042671134535

Cluster 264 has Size: 6  
Silluete is: 0.7280290226206722

Cluster 265 has Size: 6  
Silluete is: -0.7232985682018791

Cluster 266 has Size: 22  
Silluete is: 0.5178541875179697

Cluster 267 has Size: 6  
Silluete is: -0.8799387365183179

Cluster 268 has Size: 9  
Silluete is: 0.42956458134728365

Cluster 269 has Size: 8  
Silluete is: 0.6457992829176946

Cluster 270 has Size: 5  
Silluete is: -0.8672939703708855

Cluster 271 has Size: 6  
Silluete is: -0.8645361658738914

Cluster 272 has Size: 5  
Silluete is: -0.4470037636787296

Cluster 273 has Size: 6  
Silluete is: -0.8578520529647449

Cluster 274 has Size: 9  
Silluete is: -0.2596389818919638

Cluster 275 has Size: 17  
Silluete is: 0.714700643101309

Cluster 276 has Size: 17  
Silluete is: 0.7175673887133144

Cluster 277 has Size: 6  
Silluete is: -0.4669672090273292

Cluster 278 has Size: 7  
Silluete is: -0.5892697793054162

Cluster 279 has Size: 11  
Silluete is: -0.7415133296148073

Cluster 280 has Size: 18  
Silluete is: -0.9532898093784391

Cluster 281 has Size: 8  
Silluete is: -0.8259487943362418

Cluster 282 has Size: 7  
Silluete is: -0.7669528023303861

Cluster 283 has Size: 17  
Silluete is: -0.8682067839526894

Cluster 284 has Size: 9  
Silluete is: -0.3781818072493219

Cluster 285 has Size: 24  
Silluete is: -0.6627744649678068

Cluster 286 has Size: 5  
Silluete is: -0.8230544570912216

Cluster 287 has Size: 12  
Silluete is: -0.8283089112563989

Cluster 288 has Size: 11  
Silluete is: -0.4857329750514

Cluster 289 has Size: 18  
Silluete is: 0.6087340571553927

Cluster 290 has Size: 7  
Silluete is: -0.44286605172155913

Cluster 291 has Size: 5

Silluete is: -0.33484480636686625

Cluster 292 has Size: 6  
Silluete is: -0.8262309024680948

Cluster 293 has Size: 12  
Silluete is: 0.8369495984526584

Cluster 294 has Size: 11  
Silluete is: 0.7341167553036022

Cluster 295 has Size: 7  
Silluete is: -0.48449879196950213

Cluster 296 has Size: 11  
Silluete is: -0.8306139070611223

Cluster 297 has Size: 8  
Silluete is: -0.789887944870982

Cluster 298 has Size: 7  
Silluete is: -0.627832051219289

Cluster 299 has Size: 26  
Silluete is: -0.7814723036147897

Cluster 300 has Size: 5  
Silluete is: -0.8134914263413932

Cluster 301 has Size: 14  
Silluete is: -0.8705435724858268

Cluster 302 has Size: 6  
Silluete is: -0.6114661053174898

Cluster 303 has Size: 11  
Silluete is: 0.7753568920392486

Cluster 304 has Size: 11  
Silluete is: -0.8652358950968901

Cluster 305 has Size: 17  
Silluete is: -0.7201915682742652

Cluster 306 has Size: 7  
Silluete is: -0.8484989493660186

Cluster 307 has Size: 5  
Silluete is: -0.7647092171624664

Cluster 308 has Size: 10  
Silluete is: -0.43952281066758886

Cluster 309 has Size: 5  
Silluete is: -0.8502176654989091

Cluster 310 has Size: 19  
Silluete is: -0.46204246614831446

Cluster 311 has Size: 5  
Silluete is: -0.7466634123813641

Cluster 312 has Size: 7  
Silluete is: -0.8767713541473355

Cluster 313 has Size: 9  
Silluete is: -0.8199090978170017

Cluster 314 has Size: 5  
Silluete is: -0.3438799613283945

Cluster 315 has Size: 9  
Silluete is: -0.434624421792664

Cluster 316 has Size: 6  
Silluete is: -0.7314906139705472

Cluster 317 has Size: 5  
Silluete is: -0.9233193969048563

Cluster 318 has Size: 5  
Silluete is: -0.6251219960772723

Cluster 319 has Size: 16  
Silluete is: -0.6185559739796255

Cluster 320 has Size: 15  
Silluete is: 0.7741327851854204

Cluster 321 has Size: 23  
Silluete is: -0.53037822379955

Cluster 322 has Size: 6  
Silluete is: -0.5132799572337166

Cluster 323 has Size: 16  
Silluete is: -0.9084343431320785

Cluster 324 has Size: 7  
Silluete is: -0.4564414298016497

Cluster 325 has Size: 7  
Silluete is: -0.8451154415079075

Cluster 326 has Size: 5  
Silluete is: -0.7778656874519125

Cluster 327 has Size: 6  
Silluete is: -0.8728784633530159

Cluster 328 has Size: 8  
Silluete is: -0.8909448119107664

Cluster 329 has Size: 7  
Silluete is: -0.7929859864338851

Cluster 330 has Size: 6  
Silluete is: -0.8351936507714938

Cluster 331 has Size: 6  
Silluete is: -0.46204246614831446

Cluster 332 has Size: 22  
Silluete is: -0.7587718825135519

Cluster 333 has Size: 22  
Silluete is: -0.8543806426353832

Cluster 334 has Size: 7  
Silluete is: -0.8665254054897231

Cluster 335 has Size: 6  
Silluete is: -0.8458052952315203

Cluster 336 has Size: 6  
Silluete is: -0.5939040763773215

Cluster 337 has Size: 5  
Silluete is: -0.798088515242592

Cluster 338 has Size: 8  
Silluete is: -0.8795825102390483

Cluster 339 has Size: 7  
Silluete is: 0.5655198217248016

Cluster 340 has Size: 5  
Silluete is: 0.6464344662270608

Cluster 341 has Size: 5  
Silluete is: -0.7267005607345469

Cluster 342 has Size: 5  
Silluete is: -0.8503497026417542

Cluster 343 has Size: 8  
Silluete is: -0.8199049545319619

Cluster 344 has Size: 6  
Silluete is: -0.8632715942806963

Cluster 345 has Size: 8  
Silluete is: -0.859718224999263

Cluster 346 has Size: 16  
Silluete is: -0.41496471152172965

Cluster 347 has Size: 6  
Silluete is: -0.8500430127617036

Cluster 348 has Size: 10  
Silluete is: -0.5945189741544306

Cluster 349 has Size: 13  
Silluete is: -0.7720069121547397

Cluster 350 has Size: 17

Silluete is: -0.6121193416182213

Cluster 351 has Size: 10

Silluete is: -0.45006988677425375

Cluster 352 has Size: 10

Silluete is: -0.6121193416182213

Cluster 353 has Size: 8

Silluete is: -0.7123299680587782

Cluster 354 has Size: 8

Silluete is: 0.8260827478643177

Cluster 355 has Size: 7

Silluete is: -0.49420490153609004

Cluster 356 has Size: 6

Silluete is: -0.7982377613312176

Cluster 357 has Size: 6

Silluete is: -0.4846464989921374

Cluster 358 has Size: 5

Silluete is: -0.8689743005818696

Cluster 359 has Size: 8

Silluete is: -0.8801987620473122

Cluster 360 has Size: 5

Silluete is: -0.6233456344555509

Cluster 361 has Size: 16

Silluete is: -0.8765616136639918

Cluster 362 has Size: 5

Silluete is: -0.9261876550149907

Cluster 363 has Size: 5

Silluete is: -0.8929180817196725

Cluster 364 has Size: 8

Silluete is: 0.8587130177919302

Cluster 365 has Size: 11

Silluete is: -0.7965222256549315

Cluster 366 has Size: 5

Silluete is: -0.8456313104477882

Cluster 367 has Size: 5

Silluete is: -0.8120750289433398

Cluster 368 has Size: 10

Silluete is: -0.8397860177708973

Cluster 369 has Size: 5

Silluete is: -0.6882936472879004

Cluster 370 has Size: 5  
Silluete is: -0.853819495340141

Cluster 371 has Size: 28  
Silluete is: 0.8029131172101496

Cluster 372 has Size: 6  
Silluete is: -0.7969470877604639

Cluster 373 has Size: 5  
Silluete is: -0.6060788480850905

Cluster 374 has Size: 8  
Silluete is: 0.0019757984798283923

Cluster 375 has Size: 16  
Silluete is: 0.7311082209166323

Cluster 376 has Size: 7  
Silluete is: 0.6457992829176946

Cluster 377 has Size: 7  
Silluete is: -0.4623645927378718

Cluster 378 has Size: 6  
Silluete is: -0.722474232034476

Cluster 379 has Size: 6  
Silluete is: -0.7331939852008422

Cluster 380 has Size: 8  
Silluete is: -0.8523972024825485

Cluster 381 has Size: 9  
Silluete is: -0.8799715329313799

Cluster 382 has Size: 5  
Silluete is: -0.7946766195806345

Cluster 383 has Size: 4  
Silluete is: -0.7415617106005178

Cluster 384 has Size: 10  
Silluete is: -0.8321250308699257

Cluster 385 has Size: 6  
Silluete is: 0.7071132157069929

Cluster 386 has Size: 6  
Silluete is: -0.622796255819265

Cluster 387 has Size: 18  
Silluete is: -0.4711294825889623

Cluster 388 has Size: 5  
Silluete is: 0.2825619536452821

Cluster 389 has Size: 7  
Silluete is: -0.5308199543697611

Cluster 390 has Size: 5  
Silluete is: 0.8795646010890326

Cluster 391 has Size: 5  
Silluete is: 0.6180555555555345

Cluster 392 has Size: 6  
Silluete is: -0.8257065680903806

Cluster 393 has Size: 5  
Silluete is: -0.7889918414490235

Cluster 394 has Size: 11  
Silluete is: -0.31307498491536695

Cluster 395 has Size: 8  
Silluete is: -0.6701995713073519

Cluster 396 has Size: 9  
Silluete is: 0.6925427487194569

Cluster 397 has Size: 7  
Silluete is: -0.9000534364095155

Cluster 398 has Size: 16  
Silluete is: -0.8307617546982551

Cluster 399 has Size: 5  
Silluete is: -0.884114393594026

Cluster 400 has Size: 9  
Silluete is: -0.11064695076987134

Cluster 401 has Size: 6  
Silluete is: 0.607737074117838

Cluster 402 has Size: 6  
Silluete is: 0.6307692307693666

Cluster 403 has Size: 8  
Silluete is: -0.7415707857156244

Cluster 404 has Size: 6  
Silluete is: -0.907163951018375

Cluster 405 has Size: 5  
Silluete is: -0.855301178361202

Cluster 406 has Size: 15  
Silluete is: -0.7965222256549315

Cluster 407 has Size: 4  
Silluete is: -0.7617829174598479

Cluster 408 has Size: 7  
Silluete is: -0.8770468375861004

Cluster 409 has Size: 5



Silluete is: 0.5931182235579945

Cluster 410 has Size: 6  
Silluete is: -0.6630590695141265

Cluster 411 has Size: 5  
Silluete is: 0.25409497216202737

Cluster 412 has Size: 7  
Silluete is: -0.6631490652805022

Cluster 413 has Size: 10  
Silluete is: -0.6101510483428444

Cluster 414 has Size: 6  
Silluete is: -0.626778102033604

Cluster 415 has Size: 6  
Silluete is: -0.775624665729038

Cluster 416 has Size: 5  
Silluete is: -0.8433995327818006

Cluster 417 has Size: 10  
Silluete is: -0.6959951823114429

Cluster 418 has Size: 34  
Silluete is: -0.8142122138545589

Cluster 419 has Size: 6  
Silluete is: -0.4162702930629965

Cluster 420 has Size: 5  
Silluete is: -0.8985493356532457

Cluster 421 has Size: 9  
Silluete is: -0.9404208064977844

Cluster 422 has Size: 9  
Silluete is: -0.8934795735919402

Cluster 423 has Size: 9  
Silluete is: -0.622796255819265

Cluster 424 has Size: 5  
Silluete is: -0.8592801212406879

Cluster 425 has Size: 7  
Silluete is: -0.5078342770223603

Cluster 426 has Size: 8  
Silluete is: -0.8625057556866024

Cluster 427 has Size: 8  
Silluete is: -0.8774569299783984

Cluster 428 has Size: 6  
Silluete is: 0.8298917487968371

Cluster 429 has Size: 10  
Silluete is: -0.2618422569332066

Cluster 430 has Size: 8  
Silluete is: -0.560478576325294

Cluster 431 has Size: 12  
Silluete is: 0.8892533894672869

Cluster 432 has Size: 13  
Silluete is: -0.8231522668156769

Cluster 433 has Size: 7  
Silluete is: -0.6729522349788721

Cluster 434 has Size: 14  
Silluete is: -0.816347814471351

Cluster 435 has Size: 6  
Silluete is: 0.8483047231062564

Cluster 436 has Size: 8  
Silluete is: -0.7974795069302223

Cluster 437 has Size: 6  
Silluete is: -0.8904815773658586

Cluster 438 has Size: 6  
Silluete is: -0.8230566386826297

Cluster 439 has Size: 9  
Silluete is: -0.8537288683552113

Cluster 440 has Size: 11  
Silluete is: -0.10452315741099244

Cluster 441 has Size: 6  
Silluete is: -0.9155770109222179

Cluster 442 has Size: 6  
Silluete is: 0.3797913455487794

Cluster 443 has Size: 9  
Silluete is: -0.4803444660383727

Cluster 444 has Size: 5  
Silluete is: -0.9255995760046621

Cluster 445 has Size: 8  
Silluete is: 0.39068130622022407

Cluster 446 has Size: 22  
Silluete is: -0.8162669804324318

Cluster 447 has Size: 6  
Silluete is: -0.8859436428872651

Cluster 448 has Size: 5  
Silluete is: -0.7734876032636562

Cluster 449 has Size: 5  
Silluete is: -0.9145381647212272

Cluster 450 has Size: 8  
Silluete is: -0.22057746356622396

Cluster 451 has Size: 5  
Silluete is: -0.8571199714011767

Cluster 452 has Size: 6  
Silluete is: -0.8463514258395857

Cluster 453 has Size: 6  
Silluete is: -0.783114206429569

Cluster 454 has Size: 6  
Silluete is: -0.8339868407436737

Cluster 455 has Size: 8  
Silluete is: -0.8079412848771793

Cluster 456 has Size: 7  
Silluete is: -0.8707686295896376

Cluster 457 has Size: 6  
Silluete is: -0.815167101105446

Cluster 458 has Size: 6  
Silluete is: -0.6731623778915341

Cluster 459 has Size: 6  
Silluete is: -0.8851543678999455

Cluster 460 has Size: 6  
Silluete is: 0.5882240520478041

Cluster 461 has Size: 5  
Silluete is: -0.8352181057781937

Cluster 462 has Size: 5  
Silluete is: -0.47902788582783723

Cluster 463 has Size: 5  
Silluete is: -0.8547700160252326

Cluster 464 has Size: 8  
Silluete is: 0.6898991157173672

Cluster 465 has Size: 5  
Silluete is: -0.861197322387543

Cluster 466 has Size: 5  
Silluete is: -0.8210267403707575

Cluster 467 has Size: 5  
Silluete is: -0.6550691767940986

Cluster 468 has Size: 6

Silluete is: -0.8618489058870483

Cluster 469 has Size: 8  
Silluete is: -0.8191695196303496

Cluster 470 has Size: 8  
Silluete is: -0.8644023847256694

Cluster 471 has Size: 13  
Silluete is: 0.4955459644316835

Cluster 472 has Size: 5  
Silluete is: -0.5524557265261445

Cluster 473 has Size: 8  
Silluete is: 0.4929312091144001

Cluster 474 has Size: 11  
Silluete is: -0.5916538639635959

Cluster 475 has Size: 5  
Silluete is: 0.495405845398201

Cluster 476 has Size: 6  
Silluete is: -0.8635845349741126

Cluster 477 has Size: 5  
Silluete is: 0.03415669089692822

Cluster 478 has Size: 5  
Silluete is: 0.74356627443336

Cluster 479 has Size: 6  
Silluete is: -0.3077258446037129

Cluster 480 has Size: 4  
Silluete is: -0.8263593052367085

Cluster 481 has Size: 6  
Silluete is: -0.5959040365601384

Cluster 482 has Size: 9  
Silluete is: -0.8528521936245834

Cluster 483 has Size: 5  
Silluete is: -0.6528521799369364

Cluster 484 has Size: 5  
Silluete is: -0.757786830268989

Cluster 485 has Size: 7  
Silluete is: 0.4868911095273897

Cluster 486 has Size: 5  
Silluete is: -0.8580814541691858

Cluster 487 has Size: 5  
Silluete is: -0.8668823381037037

Cluster 488 has Size: 6  
Silluete is: 0.4540440049393229

Cluster 489 has Size: 5  
Silluete is: -0.6371702771721013

Cluster 490 has Size: 5  
Silluete is: -0.8175236338067977

Cluster 491 has Size: 6  
Silluete is: 0.4533030651772854

Cluster 492 has Size: 5  
Silluete is: -0.7463418138556571

Cluster 493 has Size: 6  
Silluete is: -0.8389461287484018

Cluster 494 has Size: 5  
Silluete is: -0.7511816607695643

Cluster 495 has Size: 8  
Silluete is: -0.8397879287755603

Cluster 496 has Size: 5  
Silluete is: -0.5901654593046027

Cluster 497 has Size: 7  
Silluete is: -0.9435388602984777

Cluster 498 has Size: 5  
Silluete is: -0.8119874158724789

Cluster 499 has Size: 8  
Silluete is: -0.5497261692006419

Cluster 500 has Size: 7  
Silluete is: 0.3891947607805698

Cluster 501 has Size: 8  
Silluete is: 0.6358212138904528

Cluster 502 has Size: 5  
Silluete is: -0.7182403802262908

Cluster 503 has Size: 9  
Silluete is: -0.861896709131026

Cluster 504 has Size: 11  
Silluete is: -0.22775082668161312

Cluster 505 has Size: 5  
Silluete is: -0.8557150663359446

Cluster 506 has Size: 6  
Silluete is: 0.30492973274165164

Cluster 507 has Size: 6  
Silluete is: -0.5438842172239032

Cluster 508 has Size: 7  
Silluete is: 0.5010327262952101

Cluster 509 has Size: 6  
Silluete is: -0.5562374247614323

Cluster 510 has Size: 5  
Silluete is: -0.7508567015532865

Cluster 511 has Size: 6  
Silluete is: -0.7659085864534404

Cluster 512 has Size: 3  
Silluete is: -0.7853281886812713

Cluster 513 has Size: 6  
Silluete is: 0.6172839506172493

Cluster 514 has Size: 5  
Silluete is: -0.46420310204179926

Cluster 515 has Size: 5  
Silluete is: -0.7603799768826477

Cluster 516 has Size: 7  
Silluete is: -0.8766044128700092

Cluster 517 has Size: 7  
Silluete is: 0.5358912345626545

Cluster 518 has Size: 6  
Silluete is: 0.695627230704427

Cluster 519 has Size: 5  
Silluete is: -0.4917071656492036

Cluster 520 has Size: 7  
Silluete is: -0.541580785000513

Cluster 521 has Size: 7  
Silluete is: -0.732349205309308

Cluster 522 has Size: 5  
Silluete is: -0.7820238441657831

Cluster 523 has Size: 6  
Silluete is: -0.7979825944799465

Cluster 524 has Size: 6  
Silluete is: -0.8739595385932849

Cluster 525 has Size: 9  
Silluete is: 0.6403065051008664

Cluster 526 has Size: 5  
Silluete is: -0.5128747607251802

Cluster 527 has Size: 5

Silluete is: -0.41739417866377504

Cluster 528 has Size: 5  
Silluete is: -0.012802137973359195

Cluster 529 has Size: 6  
Silluete is: -0.8441467549890757

Cluster 530 has Size: 5  
Silluete is: -0.7807071830978941

Cluster 531 has Size: 5  
Silluete is: -0.6279880639753819

Cluster 532 has Size: 10  
Silluete is: -0.877829560851212

Cluster 533 has Size: 3  
Silluete is: -0.7296475588337062

Cluster 534 has Size: 5  
Silluete is: -0.5683334245232518

Cluster 535 has Size: 5  
Silluete is: -0.7254115046358294

Cluster 536 has Size: 6  
Silluete is: -0.7873220093126375

Cluster 537 has Size: 5  
Silluete is: -0.10819317897524548

Cluster 538 has Size: 5  
Silluete is: -0.8508532351354974

Cluster 539 has Size: 5  
Silluete is: -0.8365383718819135

Cluster 540 has Size: 5  
Silluete is: -0.6701995713073519

Cluster 541 has Size: 9  
Silluete is: -0.7074757810803918

Cluster 542 has Size: 8  
Silluete is: 0.7545269758776081

Cluster 543 has Size: 5  
Silluete is: -0.6025783596962254

Cluster 544 has Size: 5  
Silluete is: -0.585486763028596

Cluster 545 has Size: 5  
Silluete is: -0.8135248428915485

Cluster 546 has Size: 8  
Silluete is: -0.43152610323773793

Cluster 547 has Size: 4  
Silluete is: -0.6326601577255568

Cluster 548 has Size: 4  
Silluete is: 0.4235539424658626

Cluster 549 has Size: 5  
Silluete is: -0.36275501049992737

Cluster 550 has Size: 4  
Silluete is: -0.8740745154069229

Cluster 551 has Size: 3  
Silluete is: -0.8138422756790566

Cluster 552 has Size: 5  
Silluete is: -0.6987299868540551

Cluster 553 has Size: 6  
Silluete is: -0.8303521224236986

Cluster 554 has Size: 5  
Silluete is: -0.7605508726218972

Cluster 555 has Size: 5  
Silluete is: -0.7421365619489617

Cluster 556 has Size: 3  
Silluete is: -0.6930549003919773

Cluster 557 has Size: 7  
Silluete is: -0.8692560226968646

Cluster 558 has Size: 5  
Silluete is: -0.852171360394041

Cluster 559 has Size: 5  
Silluete is: -0.8473993400754366

Cluster 560 has Size: 5  
Silluete is: -0.8691489378665326

Cluster 561 has Size: 5  
Silluete is: -0.820558350411874

Cluster 562 has Size: 5  
Silluete is: 0.8309661078309041

Cluster 563 has Size: 5  
Silluete is: 0.25409497216202737

Cluster 564 has Size: 7  
Silluete is: -0.905287766636378

Cluster 565 has Size: 5  
Silluete is: -0.8825837208382781

Cluster 566 has Size: 5  
Silluete is: -0.8270825210212371



Cluster 567 has Size: 5  
Silluete is: -0.8859730431091913

Cluster 568 has Size: 5  
Silluete is: -0.8491776103241117

Cluster 569 has Size: 5  
Silluete is: -0.9026398971280422

Cluster 570 has Size: 5  
Silluete is: -0.627832051219289

Cluster 571 has Size: 5  
Silluete is: 0.2909418068731431

Cluster 572 has Size: 5  
Silluete is: -0.9371200265376164

Cluster 573 has Size: 5  
Silluete is: -0.8221937541142293

Cluster 574 has Size: 4  
Silluete is: -0.6391161724923553

Cluster 575 has Size: 3  
Silluete is: -0.7054313635265819

Cluster 576 has Size: 5  
Silluete is: -0.8908830343950291

Cluster 577 has Size: 5  
Silluete is: -0.7769138907878146

Cluster 578 has Size: 6  
Silluete is: -0.9230285475398418

Cluster 579 has Size: 1  
Silluete is: -0.8010608442277681

Cluster 580 has Size: 5  
Silluete is: -0.7539015598965197

Cluster 581 has Size: 5  
Silluete is: -0.8164565437391424

Cluster 582 has Size: 5  
Silluete is: -0.8346318511749392

Cluster 583 has Size: 6  
Silluete is: -0.8088951236391294

Cluster 584 has Size: 5  
Silluete is: -0.8411563064927856

Cluster 585 has Size: 5  
Silluete is: 0.3089003991576065

Cluster 586 has Size: 4

Silluete is: -0.8522460907629409

Cluster 587 has Size: 5  
Silluete is: 0.7945246757187894

Cluster 588 has Size: 4  
Silluete is: -0.8377564714552814

Cluster 589 has Size: 6  
Silluete is: -0.5977996099392539

Cluster 590 has Size: 6  
Silluete is: -0.8413796518914854

Cluster 591 has Size: 5  
Silluete is: -0.8073888689108452

Cluster 592 has Size: 5  
Silluete is: -0.8432910833706859

Cluster 593 has Size: 5  
Silluete is: -0.8226660083379358

**Worse Silluete, better but still bad Homogeneity, Way way to many Clusters**

```
In [85]: # epsilon = 1 and min_samples = 10
clustering_05_25_e = DBSCAN(eps=0.5, min_samples=25, metric='euclidean').fit(
```

```
In [86]: # Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = clustering_05_25_e.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultDatalabels, labels))
# I have no idea why the Labels 0 and 1 are in places 1 and 2. Wierd, i think
print("Confusion Matrix")
print( metrics.confusion_matrix(adultDatalabels, labels)[1] )
print( metrics.confusion_matrix(adultDatalabels, labels)[2] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1
```

Estimated number of clusters: 37

Estimated number of noise points: 28208

Homogeneity: 0.041

Confusion Matrix

```
[21008  610  763  175  504  239  295  149  85  26  16  76
    16   28   53   27   78   30   34   30   18   34   34   35
    34   16   14   20   44   24   51   53   20   15   26   17
    16    7]
[7200    5  276    1    2  101    1    0    3    1  65    2  32    0
    0   39    1    0    0    0   17   14   13    1    2   19   12    9
    0    1    1    0   11    2    0    0    9    1]
```

Silhouette Coefficient: -0.563

Overview over the Cluster

Cluster 0 has Size: 615

Silluete is: -0.8130391736791582

Cluster 1 has Size: 1039

Silluete is: -0.36875083694352806

Cluster 2 has Size: 176

Silluete is: -0.7439560513925336

Cluster 3 has Size: 506

Silluete is: -0.5594592857368968

Cluster 4 has Size: 340

Silluete is: -0.8098981869823688

Cluster 5 has Size: 296

Silluete is: -0.7532507934543684

Cluster 6 has Size: 149

Silluete is: -0.26464297339836895

Cluster 7 has Size: 88

Silluete is: -0.7512309870395377

Cluster 8 has Size: 27

Silluete is: -0.5444066996276077

Cluster 9 has Size: 81

Silluete is: -0.8280884613969474

Cluster 10 has Size: 78

Silluete is: -0.3479321453589858

Cluster 11 has Size: 48

Silluete is: -0.7897875385428182

Cluster 12 has Size: 28

Silluete is: -0.7369470434394608

Cluster 13 has Size: 53

Silluete is: -0.5621776970918404

Cluster 14 has Size: 66

Silluete is: -0.8243099024798517

Cluster 15 has Size: 79  
Silluete is: -0.2799118797265719

Cluster 16 has Size: 30  
Silluete is: -0.7350407942055985

Cluster 17 has Size: 34  
Silluete is: -0.20982041791315487

Cluster 18 has Size: 30  
Silluete is: -0.5796517492592728

Cluster 19 has Size: 35  
Silluete is: -0.6990636415123418

Cluster 20 has Size: 48  
Silluete is: -0.4329879530110059

Cluster 21 has Size: 47  
Silluete is: -0.4671951659998537

Cluster 22 has Size: 36  
Silluete is: -0.33713844013407596

Cluster 23 has Size: 36  
Silluete is: -0.5040380618692636

Cluster 24 has Size: 35  
Silluete is: 0.8371327089736812

Cluster 25 has Size: 26  
Silluete is: -0.5654047237920684

Cluster 26 has Size: 29  
Silluete is: -0.39167873913042967

Cluster 27 has Size: 44  
Silluete is: -0.43542450004541344

Cluster 28 has Size: 25  
Silluete is: -0.3757453281732689

Cluster 29 has Size: 52  
Silluete is: -0.5950442889147818

Cluster 30 has Size: 53  
Silluete is: -0.5768294573250482

Cluster 31 has Size: 31  
Silluete is: -0.7919743741155639

Cluster 32 has Size: 17  
Silluete is: -0.8394153017701484

Cluster 33 has Size: 26  
Silluete is: -0.8313838183197959

Cluster 34 has Size: 17  
Silluete is: -0.7822879382840742

Cluster 35 has Size: 25  
Silluete is: -0.5536465202902989

Cluster 36 has Size: 8  
Silluete is: 0.2769913520787751

Cluster -1 has Size: 28208  
Silluete is: -0.7812756953296704

Way less clusters, but now everything is noise, also not good

```
In [74]: # epsilon = 1 and min_samples = 10
clustering_075_10_e = DBSCAN(eps=0.75, min_samples=10, metric='euclidean').fit
```

```
In [75]: # Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = clustering_075_10_e.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultDatalabels, labels))
# I have no idea why the Labels 0 and 1 are in places 1 and 2. Wierd
print("Confusion Matrix")
print( metrics.confusion_matrix(adultDatalabels, labels)[1] )
print( metrics.confusion_matrix(adultDatalabels, labels)[2] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1
```

Estimated number of clusters: 216  
Estimated number of noise points: 19702  
Homogeneity: 0.142  
Confusion Matrix

[14380	268	966	812	1137	125	19	497	439	2	28	102
443	60	31	20	26	133	57	113	150	56	7	240
21	6	19	53	6	47	33	23	12	36	22	96
18	58	64	17	11	19	114	28	127	13	173	18
58	10	58	82	104	22	17	10	18	94	9	228
63	12	41	41	21	61	15	10	18	147	17	8
39	58	12	19	22	37	20	59	12	6	11	88
59	67	40	16	15	12	40	28	30	13	67	13
6	10	22	10	22	21	12	7	22	12	34	21
14	3	35	14	6	46	2	13	56	52	13	12

2	22	10	11	12	27	10	20	14	7	47	26		
13	14	9	19	3	12	18	20	17	26	6	11		
24	10	14	11	13	13	3	27	16	21	19	17		
11	25	10	24	12	14	17	23	10	11	11	13		
4	8	1	9	10	10	16	8	0	18	13	11		
9	13	12	36	13	8	7	8	16	9	10	10		
2	12	6	7	10	6	7	13	2	2	6	10		
15	2	6	10	8	1	7	10	3	2	2	0		
10]													
[5322	0	20	8	469	16	2	13	272	8	0	159	4	0
19	0	0	73	0	21	15	0	14	0	49	45	2	0
5	1	0	0	0	0	4	0	0	58	20	0	2	0
354	8	0	0	9	0	3	0	43	4	6	1	0	6
0	50	0	0	0	16	0	23	0	69	1	0	4	3
0	3	10	2	15	0	1	0	0	20	0	11	0	0
0	0	3	75	2	0	0	5	0	5	2	0	8	0
0	0	6	0	5	6	0	0	64	11	0	12	24	0
8	0	9	0	0	7	1	0	13	1	1	3	0	0
0	0	1	2	0	1	0	2	2	0	38	0	1	0
0	0	15	0	6	0	1	1	5	0	13	0	0	4
0	0	6	0	2	0	0	0	1	0	0	0	0	12
6	2	14	6	0	2	0	0	10	0	9	0	7	0
0	2	4	2	7	3	1	4	0	0	7	1	5	7
0	2	3	0	8	8	4	0	0	6	4	2	2	9
1	0	7	3	8	8	0]							

Silhouette Coefficient: -0.551

Overview over the Cluster

Cluster 0 has Size: 268

Silluete is: -0.8529673874419746

Cluster 1 has Size: 986

Silluete is: -0.41560309767190184

Cluster 2 has Size: 820

Silluete is: -0.5159539023686794

Cluster 3 has Size: 1606

Silluete is: -0.7938241530470116

Cluster 4 has Size: 141

Silluete is: -0.8615976302138827

Cluster 5 has Size: 21

Silluete is: -0.8613445421057243

Cluster 6 has Size: 510

Silluete is: -0.5210640440982799

Cluster 7 has Size: 711

Silluete is: -0.7888530458386501

Cluster 8 has Size: 10

Silluete is: -0.6322289646484988

Cluster 9 has Size: 28

Silluete is: -0.7531982771403648

Cluster 10 has Size: 261

Silluete is: -0.566528841807837

Cluster 11 has Size: 447  
Silluete is: -0.8072741504718823

Cluster 12 has Size: 60  
Silluete is: -0.8187237784066282

Cluster 13 has Size: 50  
Silluete is: -0.7142735139508797

Cluster 14 has Size: 20  
Silluete is: -0.6838881961006975

Cluster 15 has Size: 26  
Silluete is: -0.8817784178389193

Cluster 16 has Size: 206  
Silluete is: -0.793656431103767

Cluster 17 has Size: 57  
Silluete is: -0.42654270249420345

Cluster 18 has Size: 134  
Silluete is: -0.6562780471561785

Cluster 19 has Size: 165  
Silluete is: -0.8086278716312948

Cluster 20 has Size: 56  
Silluete is: -0.6532514092872701

Cluster 21 has Size: 21  
Silluete is: -0.6486907245545904

Cluster 22 has Size: 240  
Silluete is: -0.7868955786548131

Cluster 23 has Size: 70  
Silluete is: 0.5079815317511085

Cluster 24 has Size: 51  
Silluete is: -0.8463886495992388

Cluster 25 has Size: 21  
Silluete is: -0.8373477587660954

Cluster 26 has Size: 53  
Silluete is: -0.7182475251458825

Cluster 27 has Size: 11  
Silluete is: -0.7320293138245859

Cluster 28 has Size: 48  
Silluete is: -0.5940040263936509

Cluster 29 has Size: 33  
Silluete is: -0.6679480660314226

Cluster 30 has Size: 23  
Silluete is: -0.7277176926907036

Cluster 31 has Size: 12  
Silluete is: -0.8019987735232512

Cluster 32 has Size: 36  
Silluete is: -0.5825273293647322

Cluster 33 has Size: 26  
Silluete is: 0.29195220484141493

Cluster 34 has Size: 96  
Silluete is: -0.8208169790324166

Cluster 35 has Size: 18  
Silluete is: -0.8732075538484676

Cluster 36 has Size: 116  
Silluete is: -0.5929033661856634

Cluster 37 has Size: 84  
Silluete is: -0.8796872227088212

Cluster 38 has Size: 17  
Silluete is: -0.6020660065181698

Cluster 39 has Size: 13  
Silluete is: -0.7623696403457536

Cluster 40 has Size: 19  
Silluete is: -0.8967835903697646

Cluster 41 has Size: 468  
Silluete is: 0.7123515315264952

Cluster 42 has Size: 36  
Silluete is: -0.8691652055236054

Cluster 43 has Size: 127  
Silluete is: -0.7914339558885055

Cluster 44 has Size: 13  
Silluete is: -0.3283822628320106

Cluster 45 has Size: 182  
Silluete is: -0.8094433166679745

Cluster 46 has Size: 18  
Silluete is: -0.7023652328463941

Cluster 47 has Size: 61  
Silluete is: 0.41902443259564426

Cluster 48 has Size: 10  
Silluete is: -0.6809313132410479

Cluster 49 has Size: 101  
Silluete is: -0.7786041208198067



Cluster 50 has Size: 86  
Silluete is: -0.8235334575681472

Cluster 51 has Size: 110  
Silluete is: 0.18172817859824408

Cluster 52 has Size: 23  
Silluete is: -0.7148197010470039

Cluster 53 has Size: 17  
Silluete is: -0.810174259061229

Cluster 54 has Size: 16  
Silluete is: -0.8146794703843915

Cluster 55 has Size: 18  
Silluete is: -0.5871295077073166

Cluster 56 has Size: 144  
Silluete is: -0.8505112637616207

Cluster 57 has Size: 9  
Silluete is: -0.7088081781917885

Cluster 58 has Size: 228  
Silluete is: -0.38770294044486275

Cluster 59 has Size: 63  
Silluete is: -0.49219174831694806

Cluster 60 has Size: 28  
Silluete is: -0.47682872106481394

Cluster 61 has Size: 41  
Silluete is: -0.8927145391134185

Cluster 62 has Size: 64  
Silluete is: -0.6134580248316931

Cluster 63 has Size: 21  
Silluete is: -0.6717998749993312

Cluster 64 has Size: 130  
Silluete is: -0.8263763363301615

Cluster 65 has Size: 16  
Silluete is: -0.4759850142263104

Cluster 66 has Size: 10  
Silluete is: -0.6139116689111327

Cluster 67 has Size: 22  
Silluete is: -0.6726246822026605

Cluster 68 has Size: 150  
Silluete is: -0.8769917660302752

Cluster 69 has Size: 17

Silluete is: 0.15515645132316566

Cluster 70 has Size: 11  
Silluete is: -0.8627507532756703

Cluster 71 has Size: 49  
Silluete is: 0.3204500269955088

Cluster 72 has Size: 60  
Silluete is: -0.746229023320032

Cluster 73 has Size: 27  
Silluete is: -0.5146824971735737

Cluster 74 has Size: 19  
Silluete is: -0.5115933629093619

Cluster 75 has Size: 23  
Silluete is: 0.0077047101896475035

Cluster 76 has Size: 37  
Silluete is: -0.21359451051255468

Cluster 77 has Size: 20  
Silluete is: -0.3890713373506783

Cluster 78 has Size: 79  
Silluete is: 0.5225984268165162

Cluster 79 has Size: 12  
Silluete is: -0.8830404671036339

Cluster 80 has Size: 17  
Silluete is: 0.00477290623381657

Cluster 81 has Size: 11  
Silluete is: -0.8399048613221712

Cluster 82 has Size: 88  
Silluete is: -0.7971425443749975

Cluster 83 has Size: 59  
Silluete is: -0.7466066760891547

Cluster 84 has Size: 67  
Silluete is: -0.540346114261257

Cluster 85 has Size: 43  
Silluete is: -0.6705141275918345

Cluster 86 has Size: 91  
Silluete is: -0.5864416713108669

Cluster 87 has Size: 17  
Silluete is: -0.36594026229604565

Cluster 88 has Size: 12  
Silluete is: -0.9048042339011052

Cluster 89 has Size: 40  
Silluete is: -0.663752615293407

Cluster 90 has Size: 33  
Silluete is: -0.7343316642843596

Cluster 91 has Size: 30  
Silluete is: -0.6346369337655416

Cluster 92 has Size: 18  
Silluete is: -0.7601198944275104

Cluster 93 has Size: 69  
Silluete is: -0.739238572618031

Cluster 94 has Size: 13  
Silluete is: -0.7892437605053085

Cluster 95 has Size: 14  
Silluete is: -0.7628071195377174

Cluster 96 has Size: 10  
Silluete is: -0.649754809531348

Cluster 97 has Size: 22  
Silluete is: -0.4826492398158582

Cluster 98 has Size: 10  
Silluete is: -0.8707482309958419

Cluster 99 has Size: 28  
Silluete is: -0.8138033144961996

Cluster 100 has Size: 21  
Silluete is: -0.5940324995250634

Cluster 101 has Size: 17  
Silluete is: -0.6422678007123336

Cluster 102 has Size: 13  
Silluete is: -0.7995598837907077

Cluster 103 has Size: 22  
Silluete is: -0.7372055170023022

Cluster 104 has Size: 12  
Silluete is: -0.4632734255845417

Cluster 105 has Size: 98  
Silluete is: -0.7380897547745481

Cluster 106 has Size: 32  
Silluete is: -0.42815321597198164

Cluster 107 has Size: 14  
Silluete is: -0.8237134872622701

Cluster 108 has Size: 15  
Silluete is: -0.3577035883790141

Cluster 109 has Size: 59  
Silluete is: -0.4587746074194894

Cluster 110 has Size: 14  
Silluete is: -0.45474652699294643

Cluster 111 has Size: 14  
Silluete is: -0.7856643316607974

Cluster 112 has Size: 46  
Silluete is: -0.8818271978259685

Cluster 113 has Size: 11  
Silluete is: -0.7101836562440158

Cluster 114 has Size: 13  
Silluete is: -0.4542694105488835

Cluster 115 has Size: 56  
Silluete is: -0.7302903445690997

Cluster 116 has Size: 59  
Silluete is: 0.7884542725423515

Cluster 117 has Size: 14  
Silluete is: -0.6721405305686613

Cluster 118 has Size: 12  
Silluete is: -0.8184953222592607

Cluster 119 has Size: 15  
Silluete is: -0.6269025466826531

Cluster 120 has Size: 23  
Silluete is: 0.47739104797468496

Cluster 121 has Size: 11  
Silluete is: 0.6389169708653387

Cluster 122 has Size: 14  
Silluete is: -0.835328820044222

Cluster 123 has Size: 12  
Silluete is: -0.8259383616161307

Cluster 124 has Size: 27  
Silluete is: -0.8280827549935854

Cluster 125 has Size: 10  
Silluete is: -0.8050681910323535

Cluster 126 has Size: 20  
Silluete is: -0.47518766646776106

Cluster 127 has Size: 15  
Silluete is: -0.11152174276514525

Cluster 128 has Size: 9

Silluete is: -0.8270680743526915

Cluster 129 has Size: 47  
Silluete is: -0.7494933058492264

Cluster 130 has Size: 27  
Silluete is: -0.7321449528906261

Cluster 131 has Size: 13  
Silluete is: -0.5461400954107166

Cluster 132 has Size: 16  
Silluete is: -0.7586509230694537

Cluster 133 has Size: 11  
Silluete is: 0.195083065714718

Cluster 134 has Size: 19  
Silluete is: -0.8128349501851211

Cluster 135 has Size: 41  
Silluete is: 0.110029118796487

Cluster 136 has Size: 12  
Silluete is: 0.5791404358386729

Cluster 137 has Size: 19  
Silluete is: -0.7328252632829141

Cluster 138 has Size: 20  
Silluete is: -0.041088166857417945

Cluster 139 has Size: 17  
Silluete is: -0.7722930102993416

Cluster 140 has Size: 26  
Silluete is: 0.5587720861973823

Cluster 141 has Size: 21  
Silluete is: -0.6004728561030199

Cluster 142 has Size: 11  
Silluete is: 0.13808012949687515

Cluster 143 has Size: 30  
Silluete is: -0.7187864927389694

Cluster 144 has Size: 10  
Silluete is: -0.8093512780672597

Cluster 145 has Size: 15  
Silluete is: -0.8239882886321076

Cluster 146 has Size: 12  
Silluete is: -0.8957735675695139

Cluster 147 has Size: 18  
Silluete is: -0.7827720330353187

Cluster 148 has Size: 13  
Silluete is: -0.8558038703677746

Cluster 149 has Size: 16  
Silluete is: -0.8524822172884021

Cluster 150 has Size: 27  
Silluete is: 0.3316872571142502

Cluster 151 has Size: 16  
Silluete is: -0.4512038470289284

Cluster 152 has Size: 25  
Silluete is: -0.8203223055631685

Cluster 153 has Size: 19  
Silluete is: -0.6131243045290131

Cluster 154 has Size: 17  
Silluete is: -0.8054666052317725

Cluster 155 has Size: 17  
Silluete is: -0.5931746814719336

Cluster 156 has Size: 25  
Silluete is: 0.5175996197644368

Cluster 157 has Size: 12  
Silluete is: -0.32820901539882097

Cluster 158 has Size: 24  
Silluete is: 0.7539356106946111

Cluster 159 has Size: 12  
Silluete is: -0.7322890631771418

Cluster 160 has Size: 14  
Silluete is: -0.6452370912217457

Cluster 161 has Size: 18  
Silluete is: -0.5362857289439465

Cluster 162 has Size: 23  
Silluete is: -0.8047015263701328

Cluster 163 has Size: 10  
Silluete is: -0.7872124908214417

Cluster 164 has Size: 11  
Silluete is: -0.809278215120613

Cluster 165 has Size: 11  
Silluete is: 0.5926770094385551

Cluster 166 has Size: 25  
Silluete is: -0.769258138060028

Cluster 167 has Size: 10  
Silluete is: -0.8209703568552926

Cluster 168 has Size: 10  
Silluete is: 0.23268788816186767

Cluster 169 has Size: 15  
Silluete is: -0.512824962875642

Cluster 170 has Size: 15  
Silluete is: -0.016628144169423674

Cluster 171 has Size: 10  
Silluete is: -0.6544526396612909

Cluster 172 has Size: 12  
Silluete is: -0.5825949217179032

Cluster 173 has Size: 16  
Silluete is: 0.44734206021650347

Cluster 174 has Size: 8  
Silluete is: -0.7413046332615669

Cluster 175 has Size: 10  
Silluete is: -0.8100974766537327

Cluster 176 has Size: 18  
Silluete is: -0.5330788730015685

Cluster 177 has Size: 22  
Silluete is: 0.6458490717097385

Cluster 178 has Size: 11  
Silluete is: -0.824973684165139

Cluster 179 has Size: 16  
Silluete is: -0.8796390879765816

Cluster 180 has Size: 13  
Silluete is: -0.7341287394384776

Cluster 181 has Size: 12  
Silluete is: -0.6932611027389297

Cluster 182 has Size: 38  
Silluete is: -0.8007670561965827

Cluster 183 has Size: 17  
Silluete is: -0.8915377057474301

Cluster 184 has Size: 10  
Silluete is: -0.19413594032734288

Cluster 185 has Size: 14  
Silluete is: -0.41202580215608947

Cluster 186 has Size: 11  
Silluete is: -0.821965893780716

Cluster 187 has Size: 17

Silluete is: -0.7776297017967156

Cluster 188 has Size: 13

Silluete is: -0.8251610190235037

Cluster 189 has Size: 10

Silluete is: -0.38334088001863464

Cluster 190 has Size: 10

Silluete is: -0.7312694976348155

Cluster 191 has Size: 9

Silluete is: -0.6926856206652443

Cluster 192 has Size: 13

Silluete is: -0.7826005827623552

Cluster 193 has Size: 11

Silluete is: -0.5275625577981652

Cluster 194 has Size: 14

Silluete is: -0.5651103831841762

Cluster 195 has Size: 10

Silluete is: 0.8294413484716268

Cluster 196 has Size: 8

Silluete is: -0.6649530543560551

Cluster 197 has Size: 10

Silluete is: -0.6117454123319679

Cluster 198 has Size: 13

Silluete is: -0.7611360710492495

Cluster 199 has Size: 10

Silluete is: -0.8471326005889828

Cluster 200 has Size: 10

Silluete is: 0.7817745671097147

Cluster 201 has Size: 10

Silluete is: 0.5263479822285716

Cluster 202 has Size: 10

Silluete is: -0.531090837936794

Cluster 203 has Size: 15

Silluete is: -0.8411451889219294

Cluster 204 has Size: 8

Silluete is: -0.7529378721062727

Cluster 205 has Size: 10

Silluete is: -0.8889871906160963

Cluster 206 has Size: 12

Silluete is: -0.6430063257304262



Cluster 207 has Size: 10  
Silluete is: 0.4972442210633498

Cluster 208 has Size: 10  
Silluete is: -0.5834193806012852

Cluster 209 has Size: 8  
Silluete is: -0.34693196735267257

Cluster 210 has Size: 10  
Silluete is: -0.8407878904423877

Cluster 211 has Size: 10  
Silluete is: -0.46160443798805484

Cluster 212 has Size: 5  
Silluete is: -0.8359653353562445

Cluster 213 has Size: 10  
Silluete is: -0.828198800121756

Cluster 214 has Size: 8  
Silluete is: -0.8767940000464154

Cluster 215 has Size: 10  
Silluete is: -0.8372632871719267

Cluster 216 has Size: 10  
Silluete is: -0.8372632871719267

Same Homogeneity, just barely less clusters.  
Normalizing the Data was a Misstake, i now  
changed that, but its still bad.

In [76]:

```
# epsilon = 1 and min_samples = 10  
clustering_2_10_e = DBSCAN(eps=2, min_samples=10, metric='euclidean').fit(adu
```

In [77]:

```

# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = clustering_2_10_e.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultData.labels, labels))
# I have no idea why the Labels 0 and 1 are in places 1 and 2. Wierd
print("Confusion Matrix")
print( metrics.confusion_matrix(adultData.labels, labels)[1] )
print( metrics.confusion_matrix(adultData.labels, labels)[2] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1

```

```

Estimated number of clusters: 33
Estimated number of noise points: 2412
Homogeneity: 0.112
Confusion Matrix
[1392 2993 8720  966  628  708  907  425  71  390 5805  94  238  735
   98  350   0   7  14  13  49  14  11   9  25   8  11   9
   0   0   9   9   3   9]
[1020 2030 1553  42  756  208  302  24  194  20 1266  245   8  46
   3   20  29   3   0   8  17   8   1   2   2   0   0   2
  11  14   1   1   4   1]
Silhouette Coefficient: -0.091
Overview over the Cluster

Cluster 0 has Size: 5023
Silluete is: 0.05421512154885544

Cluster 1 has Size: 10273
Silluete is: -0.7357614367507748

Cluster 2 has Size: 1008
Silluete is: 0.1607479371049945

Cluster 3 has Size: 1384
Silluete is: -0.16048500454344192

Cluster 4 has Size: 916
Silluete is: 0.010228705638036373

Cluster 5 has Size: 1209
Silluete is: 0.14477494306140185

```

Cluster 6 has Size: 449  
Silluete is: -0.6436407959863428

Cluster 7 has Size: 265  
Silluete is: 0.12997492127296406

Cluster 8 has Size: 410  
Silluete is: -0.7116850846646927

Cluster 9 has Size: 7071  
Silluete is: -0.06480866785840515

Cluster 10 has Size: 339  
Silluete is: -0.8145388284612641

Cluster 11 has Size: 246  
Silluete is: 0.02204339871308713

Cluster 12 has Size: 781  
Silluete is: -0.3177666971104614

Cluster 13 has Size: 101  
Silluete is: 0.07804626957951043

Cluster 14 has Size: 370  
Silluete is: 0.17524961239595782

Cluster 15 has Size: 29  
Silluete is: -0.034495632550664575

Cluster 16 has Size: 10  
Silluete is: -0.0994692198248882

Cluster 17 has Size: 14  
Silluete is: 0.12148920281381384

Cluster 18 has Size: 21  
Silluete is: -0.4382941272922792

Cluster 19 has Size: 66  
Silluete is: 0.14077706082984245

Cluster 20 has Size: 22  
Silluete is: 0.037488475661637766

Cluster 21 has Size: 12  
Silluete is: -0.31175483736560355

Cluster 22 has Size: 11  
Silluete is: 0.17391066757186122

Cluster 23 has Size: 27  
Silluete is: -0.050744722983282704

Cluster 24 has Size: 8  
Silluete is: -0.001829810398091363

Cluster 25 has Size: 11

Silluete is: -0.1978103490785356

Cluster 26 has Size: 11  
Silluete is: -0.12663900796585553

Cluster 27 has Size: 11  
Silluete is: -0.2546555722391723

Cluster 28 has Size: 14  
Silluete is: -0.7151359175262076

Cluster 29 has Size: 10  
Silluete is: 0.1770467666622712

Cluster 30 has Size: 10  
Silluete is: 0.06105893994713644

Cluster 31 has Size: 7  
Silluete is: 0.0018700860448950041

Cluster 32 has Size: 10  
Silluete is: -0.04781134359321427

Cluster -1 has Size: 2412

Less Clusters, silluete is better but still bad, homginity is worse, a  
few clusters dominate everything

In [82]: 

```
# epsilon = 2.5 and min_samples = 15
clustering_25_15_e = DBSCAN(eps=2.5, min_samples=15, metric='euclidean').fit(a
```

In [84]:

```

# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = clustering_25_15_e.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultDatalabels, labels))
# I have no idea why the Labels 0 and 1 are in places 1 and 2. Wierd
print("Confusion Matrix")
print( metrics.confusion_matrix(adultDatalabels, labels)[1] )
print( metrics.confusion_matrix(adultDatalabels, labels)[2] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1

```

```

Estimated number of clusters: 2
Estimated number of noise points: 1403
Homogeneity: 0.008
Confusion Matrix
[ 784 23895    41]
[ 619 7221     1]
Silhouette Coefficient: 0.287
Overview over the Cluster

```

```

Cluster 0 has Size: 31116
Silluete is: 0.48314608220866917

```

```

Cluster 1 has Size: 42
Silluete is: 0.1672592796590383

```

```

Cluster -1 has Size: 1403
Silluete is: 0.40893918550431496

```

Homoginty became way worse, silluete finaly in a reasonable area, but still low, on cluster completely dominates

Result: DBSCAN can not deal with the Adult dataset, no matter wether we have larger or smaller epsilons, wether we enforce more or less min-samples, etc. We even tried rescaling

the numerical variables with the binary ones to give it better chances, but the binary ones are apparently destroying everything

Lets try K-Means for Fucks sake, maybe that one works

```
In [89]: kmeans2 = KMeans(n_clusters=2, random_state=0).fit(adultData)
```

```
In [91]: # Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = kmeans2.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
#n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
# print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultData.labels, labels))
# Here we can use lines 0 and 1 of the confusion matrix, so i suspect its the
print("Confusion Matrix")
print( metrics.confusion_matrix(adultData.labels, labels)[0] )
print( metrics.confusion_matrix(adultData.labels, labels)[1] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1
```

```
Estimated number of clusters: 2
Estimated number of noise points: 0
Homogeneity: 0.090
Confusion Matrix
[19247  5473]
[3365 4476]
Silhouette Coefficient: 0.324
Overview over the Cluster
```

```
Cluster 0 has Size: 22612
Silluete is: 0.47747886670098477
```

```
Cluster 1 has Size: 9949
Silluete is: 0.22835314007001262
```

far Far better Silluete, but Homogenity is still bad.

```
In [93]: kmeans3 = KMeans(n_clusters=3, random_state=0).fit(adultData)
```

```
In [94]: # Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = kmeans3.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
#n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
# print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultData.labels, labels))
print("Confusion Matrix")
print( metrics.confusion_matrix(adultData.labels, labels)[0] )
print( metrics.confusion_matrix(adultData.labels, labels)[1] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1
```

```
Estimated number of clusters: 3
Homogeneity: 0.114
Confusion Matrix
[ 5211  6885 12624]
[4370  2156 1315]
Silhouette Coefficient: 0.311
Overview over the Cluster
```

```
Cluster 0 has Size: 9581
Silluete is: 0.4551881569089662
```

```
Cluster 1 has Size: 9041
Silluete is: 0.22157652797213948
```

```
Cluster 2 has Size: 13939
Silluete is: 0.23805365114963375
```

We lost a bit of silluete coefficient, but our homogeneity is growing, lets try 4 and 6

```
In [101]: kmeans4 = KMeans(n_clusters=4, random_state=0).fit(adultData)
```

In [102]...

```

# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = kmeans4.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
#n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
# print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultData.labels, labels))
print("Confusion Matrix")
print( metrics.confusion_matrix(adultData.labels, labels)[0] )
print( metrics.confusion_matrix(adultData.labels, labels)[1] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1

```

Estimated number of clusters: 4

Homogeneity: 0.131

Confusion Matrix

[11411 6122 2420 4767]

[1119 2379 171 4172]

Silhouette Coefficient: 0.316

Overview over the Cluster

Cluster 0 has Size: 12530

Silluete is: 0.43411018844041

Cluster 1 has Size: 8501

Silluete is: 0.11084094861645084

Cluster 2 has Size: 2591

Silluete is: 0.13240759100415167

Cluster 3 has Size: 8939

Silluete is: 0.1315496813930071

In [96]:

```
kmeans6 = KMeans(n_clusters=6, random_state=0).fit(adultData)
```



In [97]:

```

# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = kmeans6.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
#n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
# print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultDatalabels, labels))
print("Confusion Matrix")
print( metrics.confusion_matrix(adultDatalabels, labels)[0] )
print( metrics.confusion_matrix(adultDatalabels, labels)[1] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1

```

Estimated number of clusters: 6

Homogeneity: 0.141

Confusion Matrix

[9674 4545 4852 3276 2373    0]

[1667 3980 1810    60 165 159]

Silhouette Coefficient: 0.309

Overview over the Cluster

Cluster 0 has Size: 11341

Silluete is: 0.4405589264787726

Cluster 1 has Size: 8525

Silluete is: 0.09363437165420492

Cluster 2 has Size: 6662

Silluete is: 0.33233928852888267

Cluster 3 has Size: 3336

Silluete is: 0.14373386147437736

Cluster 4 has Size: 2538

Silluete is: 0.31521429942317547

Cluster 5 has Size: 159

Silluete is: 0.5152721004169173

Homoginity is growing to 14%, silluette did not get significantly worse, lets try 12 for good measure

In [103...

```
kmeans12 = KMeans(n_clusters=12, random_state=0).fit(adultData)
```

In [104...

```
# Get Metrics of the resulting Clustering
#core_samples_mask = numpy.zeros_like(clustering_1_10_e.labels_, dtype=bool)
#core_samples_mask[clustering_1_10_e.core_sample_indices_] = True
labels = kmeans12.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
#n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
# print("Estimated number of noise points: %d" % n_noise_)
print("Homogeneity: %0.3f" % metrics.homogeneity_score(adultData.labels, labels))
print("Confusion Matrix")
print( metrics.confusion_matrix(adultData.labels, labels)[0] )
print( metrics.confusion_matrix(adultData.labels, labels)[1] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, labels))
print("Overview over the Cluster")
silluets = metrics.silhouette_samples(adultData, labels, metric='euclidean' )
i = 0
for l in set(labels):
    print(" ")
    num = list(labels).count(l)
    print("Cluster " + str(l) + " has Size: " + str( num ) )
    print("Silluete is: " + str( silluets[i] ))
    i = i+1
```

Estimated number of clusters: 12

Homogeneity: 0.185

Confusion Matrix

```
[5735 4450 1265 2821 1681 841 0 1176 2465 2467 684 1135]
[ 431 1378 1737 1107 61 111 159 102 35 1066 1175 479]
```

Silhouette Coefficient: 0.277

Overview over the Cluster

Cluster 0 has Size: 6166

Silluete is: 0.25545787377857293

Cluster 1 has Size: 5828

Silluete is: 0.0918858999254379

Cluster 2 has Size: 3002

Silluete is: 0.3225301137284371

Cluster 3 has Size: 3928

Silluete is: 0.1676499688718426

Cluster 4 has Size: 1742

Silluete is: 0.5310883975981647

Cluster 5 has Size: 952

Silluete is: 0.28537889756613133

Cluster 6 has Size: 159

Silluete is: 0.16647983383910747

Cluster 7 has Size: 1278  
Silluete is: 0.23994918986092278

Cluster 8 has Size: 2500  
Silluete is: -0.0677099863258039

Cluster 9 has Size: 3533  
Silluete is: -0.03770742335020617

Cluster 10 has Size: 1859  
Silluete is: 0.4011112743094932

Cluster 11 has Size: 1614  
Silluete is: 0.5104820447508001

Homogenity still growing, Silluete still sinking

In [ ]: