We are again using the Adult Dataset

So we can use our results from Project 1 of how to preopare 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", "educa
                             "martial-stat", "occupation", "relationship", "ra
                             "cap-gains", "cap_loss", "hpw", "native_country"
adultDf = adultDf.dropna()
```

Now prepare Data:

In [105…

```
# >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(uncorelated) and education (doube 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, marrie
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Wi
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Hus
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Unm
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Oth
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Not
adultData['relationship'] = adultData['relationship'].replace(to_replace=' Own

# Remove martial-stat as this only doubes reationship status
adultData.drop("martial-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 dimensionaity 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, who
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
adultData["occField"] = numpy.where(adultData['occupation']== ' Prof-specialty
# 3 low paying fields, <10% over 50k
adultData["occField"] = numpy.where(adultData['occupation']== ' Handlers-clean
adultData["occField"] = numpy.where(adultData['occupation']== ' Other-service
adultData["occField"] = numpy.where(adultData['occupation']== ' Priv-house-ser
adultData.drop('occupation', axis=1, inplace=True)

# STANDARTIZE NOW: <--- Misstake, not snormalized Now, but standardized numeri
adultData['education_num'] = adultData['education_num'].apply(lambda x: x/13*1
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
column = adultData["hpw"]
max_hpw = column.max()
adultData['hpw'] = adultData['hpw'].apply(lambda x: x/max_hpw*30)

# Drop most binary stuff:
adultData.drop('isSelfemployed', axis=1, inplace=True)
adultData.drop('isPublic', axis=1, inplace=True)
adultData.drop('isPrivate', axis=1, inplace=True)
adultData.drop('isAsian', axis=1, inplace=True)
adultData.drop('isBlack', axis=1, inplace=True)
adultData.drop('isWhite', axis=1, inplace=True)
adultData.drop('sex', axis=1, inplace=True)

# Lets look at the Table Now
adultDatalabels = adultData["50k"]
adultData.drop('50k', axis=1, inplace=True)
print(adultData)
print(adultDatalabels)
```

```
             age  education_num  occField  relationship  cap-gains        hpw
0      13.000000      30.000000         0             0   0.652207  12.121212
1      16.666667      30.000000         1             1   0.000000   3.939394
2      12.666667      20.769231        -1             0   0.000000  12.121212
3      17.666667      16.153846        -1             1   0.000000  12.121212
4       9.333333      30.000000         1             1   0.000000  12.121212
...          ...            ...       ...           ...        ...        ...
32556   9.000000      27.692308         0             1   0.000000  11.515152
32557  13.333333      20.769231         0             1   0.000000  12.121212
32558  19.333333      20.769231         0             0   0.000000  12.121212
32559   7.333333      20.769231         0             0   0.000000   6.060606
32560  17.333333      20.769231         1             1   4.507245  12.121212

[32561 rows x 6 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

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

In [106…

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

In [107…

```
# 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, la
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: 55
Estimated number of noise points: 6791
Homogeneity: 0.097
Confusion Matrix
```

```
[4346 2678 7958  827  479  444  703   19  261 5290    3   95  299   32
   67  425   99   27  105    7  150  139   12    6   17    7   26   13
    4   23   26    8    4   15   11    4    2    3    8    7   11   10
    3    1    8    1    0    3    8    4    1    0    3    1   10    7]
[2445 1621 1363   19  553  115  190    1   11 1064   25    5   15   46
    2   32    4   22    0   38    0    8   37   15    9   11    1   58
    3    0    5    4   19    1    2   10    8    1    2    2    0    2
    2    7    2    9   10   11    3    4    9   10    0    0    0    5]
Silhouette Coefficient: -0.325
Overview over the Cluster

Cluster 0 has Size: 4299
Silluete is: -0.20860707457697614

Cluster 1 has Size: 9321
Silluete is: -0.6084224764517165

Cluster 2 has Size: 846
Silluete is: 0.009675074381455703

Cluster 3 has Size: 1032
Silluete is: -0.7025315343802458

Cluster 4 has Size: 559
Silluete is: -0.019345783963501537

Cluster 5 has Size: 893
Silluete is: -0.1232385591886929

Cluster 6 has Size: 20
Silluete is: -0.4930957882597632

Cluster 7 has Size: 272
Silluete is: -0.4981413777762802

Cluster 8 has Size: 6354
Silluete is: -0.735124652608269

Cluster 9 has Size: 28
Silluete is: -0.25177729926908626

Cluster 10 has Size: 100
Silluete is: -0.6738437371453738

Cluster 11 has Size: 314
Silluete is: -0.018528314886219893

Cluster 12 has Size: 78
Silluete is: -0.5848757502539855

Cluster 13 has Size: 69
Silluete is: -0.37134358495702846

Cluster 14 has Size: 457
Silluete is: 0.16863270194250407

Cluster 15 has Size: 103
Silluete is: -0.836968676433643
```

```
Cluster 16 has Size: 49
Silluete is: -0.31537800010673994

Cluster 17 has Size: 105
Silluete is: -0.19333387646726205

Cluster 18 has Size: 45
Silluete is: -0.6137393401343918

Cluster 19 has Size: 150
Silluete is: 0.004537964810785236

Cluster 20 has Size: 147
Silluete is: -0.8617480198929838

Cluster 21 has Size: 49
Silluete is: 0.6425819162177755

Cluster 22 has Size: 21
Silluete is: 0.14596747923918413

Cluster 23 has Size: 26
Silluete is: -0.4261461030421694

Cluster 24 has Size: 18
Silluete is: -0.6544906077193062

Cluster 25 has Size: 27
Silluete is: -0.4724436529699961

Cluster 26 has Size: 71
Silluete is: -0.3412864138530418

Cluster 27 has Size: 7
Silluete is: -0.5918393010664459

Cluster 28 has Size: 23
Silluete is: -0.7287536819767765

Cluster 29 has Size: 31
Silluete is: -0.4835358763023966

Cluster 30 has Size: 12
Silluete is: -0.6244093923485935

Cluster 31 has Size: 23
Silluete is: -0.03319830428075033

Cluster 32 has Size: 16
Silluete is: -0.29415158804660135

Cluster 33 has Size: 13
Silluete is: -0.07611970555159514

Cluster 34 has Size: 14
Silluete is: -0.43662422638450377

Cluster 35 has Size: 10
Silluete is: -0.6306880370011866
```

```
Cluster 36 has Size: 4
Silluete is: 0.0026519787019637048

Cluster 37 has Size: 10
Silluete is: -0.6062400942137605

Cluster 38 has Size: 9
Silluete is: -0.08568135882690088

Cluster 39 has Size: 11
Silluete is: -0.6141630538874976

Cluster 40 has Size: 12
Silluete is: 0.06302498036072321

Cluster 41 has Size: 5
Silluete is: -0.47856423470796494

Cluster 42 has Size: 8
Silluete is: -0.1200709703200619

Cluster 43 has Size: 10
Silluete is: -0.4825601330491392

Cluster 44 has Size: 10
Silluete is: -0.31263477794370553

Cluster 45 has Size: 10
Silluete is: -0.4470274327916476

Cluster 46 has Size: 14
Silluete is: -0.5819829570219789

Cluster 47 has Size: 11
Silluete is: 0.08731550532205064

Cluster 48 has Size: 8
Silluete is: 0.08724741604780938

Cluster 49 has Size: 10
Silluete is: 0.13334432634035406

Cluster 50 has Size: 10
Silluete is: -0.02114009209269771

Cluster 51 has Size: 3
Silluete is: -0.5055544687407639

Cluster 52 has Size: 1
Silluete is: 0.4395300665804318

Cluster 53 has Size: 10
Silluete is: 0.779962555637225

Cluster 54 has Size: 12
Silluete is: -0.568423347514986

Cluster -1 has Size: 6791
```

```
Silluete is: -0.15813623376702565
```

Thats Bad, lots of clusters, bad siluette, no homoginity

In [108…

```
# epsilon = 0.5 and min_samples = 5
clustering_05_5_e = DBSCAN(eps=0.5, min_samples=5, metric='euclidean').fit(adu
```

In [109…

```
# 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(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, la
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: 509
Estimated number of noise points: 11304
Homogeneity: 0.243
Confusion Matrix
[7657  646    6  156   49    3   16    7  121 2621   11    6   28  165
   61 1724  232  527 1466  113  150  105   98    2  582   11   11   13
    0  469   64  127  106  179   49   16   26  472   73    2   95  264
   23   46   42   29  116   70   36   19  220    7   32   55   69    2
   23  248  122   31   11    3   80   69    5   35    1   17   21  113
   60    0   13    8   31   73   16    6   18  137   49  153    0   12
  203   40   22   24   10    6    6   17   58   39    1   26    6    4
    4   30    8   19    5    5    9    3   48   15   13   28   31   28
    3   17    5   12   82    8    6   89   23   16    7   23   67  101
   18   18    3   27    5    6    5    6   18    8    6   25    4    0
   33   30    8    0   21    5    2   13   10    4   12   12    5   12
    4    7   11   16    9   25   10    7   46    6    0    1   12    2
    7   18    8   22    5    5    5    2    6   24   24   17   23    6
   62   29   13   13    2    8    4   11    0    8    2    2    0    6
   13   10   32   47    0   13   16   14    5   26   20    6    5    7
   10    7    6    8    9    2    5   42   57    3    2   10    5    8
    6   19    8   30    0   13    4    9    3    5    0   48    9   15
    1    5    3    8    6    8    2    9    9    5    5    0   33    7
```

```
     3    8    4   14   13    3   10    5   12   16    4    4   10    7
     5   13   15    1    5    3    0   17    9    2    5    6    6    2
     4    5   10    5    8    6    6    2    8    5    3    9   11    6
     2   14    7   18    5   10   16    0    5   11    9    4    0    5
     6   10   10    9    8    3   12   10    1    7    7    7    4    5
     0    5    3    5    0    6    2   10   18    5    0    9    7    9
     7    5    0    4    5   11    8    7    3    6   15    6    8    3
     8   14    6   10   18    3    7    9    4    5   14    9    0    4
     4    7    6    6    4    2    5    6   10    5    5    5    6    6
     4    6    6    1    4    9    0    7    9    6    8    6    2   14
     6    7    6    7    5    7    6    5    0    8    3    3    6    6
     3    5    7    3    0    6    1    4    4   10    9   14    0    8
     7    5    5    4    4    8    5    5    3   11    7    2    3    1
     2    0    6    1    6    8    5    5    4    3    0    5    0    7
     2    4    1    8    8    1    5    7    6    6    6    1    5    6
     5    0    5    5    5    6    0    5    5    5    1    5    5    1
     1    4    2    3    2    5    1    5    5    2    4    4    0    4
     5    5    5    4    4    2    5    1    5    5    2    0    5    5
     5    4    5    3    5    2]
[3647    2    1  320  166    5    0    4   79   46    8    0    3  197
   44  708   18   52   30    0   30  122    0    4  362    2    4    1
    6   27   42   10    0   31    3   72    0    0    0    7   70    3
   57  152    0   56   53    0    0    0   11    7    0    9    1    3
    0   15    0    5    0    8    1   14    1    6    4    0    1   83
    6    8    0    5  106    0   12    0   21    0   12    2    8    0
    6   10    1   59    0    9    0    0    0    0    4   29    0    4
    9    0    1    0    0    0    0    6    4    0    0    1   11    2
   17    0    1    0    0    1    2    0    0    0    0    0    0   35
    0    0    8    0    2    0    2    1    0    2    0    0    1    5
    0    0    1    6    0    3    3    2   16    2   15    0    0   38
    5   12   13    0    1    0    0    4    0    2   12   13    5   10
    6    0    2    0    1    0   10    9    0   13    1    0    0    0
    6    2    0    0    8    0   21    0    6    0    3    3    6    0
    0    2    0    1    6    0    2    0    0   12    0    0    3    6
   10    2    4    1    0    4    0    0    0   14    3    1    0    0
    0    0    3    5    5    0    3    1    2    2    5    0    0    1
    6    6    2    5    1    0    3    0    0    0    2   14    5    5
    3    1    5    0    0    4    0    1    0    0    1    2    0    0
    1    0    1   10    1    3    9    1    1    6    5    0    0    3
    3    0    3    0    0    0    0    3    4    1    2    0    0    1
    7    0    1    0    0    0    0    9    0    0    1    5    5    0
    0    0    0    1    0    4    0    6    5    1    5    0    2    1
   15    8    3    0    6    0    3    0    0    0    5    0    1    0
    1    0    7    1    0    1    1    0    3    1    0    0    0    2
    0    0    2    1    0    2    0    0    3    1    0    0    5    1
    3    0    0    1    1    5    0    0    0    0    0    0    0    0
    1    0    0    8    1    0    5    2    0    0    0    0    3    1
    0    0    1    2    0    0    1    0    5    0    5    2    0    1
    2    0    0    2    5    0    5    1    1    1    0    0    7    1
    0    0    0    1    3    0    2    1    4    1    0    3    0    2
    4    5    0    4    1    1    0    0    3    0    4    0    6    0
    4    2    5    0    0    4    0    2    0    0    1    4    0    0
    2    5    0    0    0    0    5    0    0    0    4    0    0    2
    4    1    3    0    3    0    4    0    0    4    3    1    6    1
    0    0    0    1    1    3    0    0    0    0    3    5    0    0
    0    1    1    1    0    3]
Silhouette Coefficient: -0.484
Overview over the Cluster
```

```
Cluster 0 has Size: 648
Silluete is: -0.6199541438938895

Cluster 1 has Size: 7
Silluete is: -0.568534613633992

Cluster 2 has Size: 476
Silluete is: -0.5490211474354176

Cluster 3 has Size: 215
Silluete is: 0.7387202689647895

Cluster 4 has Size: 8
Silluete is: -0.803120565088768

Cluster 5 has Size: 16
Silluete is: -0.553042045431671

Cluster 6 has Size: 11
Silluete is: -0.5696436929808584

Cluster 7 has Size: 200
Silluete is: 0.4019567997199122

Cluster 8 has Size: 2667
Silluete is: -0.7451909637422511

Cluster 9 has Size: 19
Silluete is: -0.9301222848773353

Cluster 10 has Size: 6
Silluete is: -0.7439649061563277

Cluster 11 has Size: 31
Silluete is: -0.8071248855489973

Cluster 12 has Size: 362
Silluete is: 0.37876212040143853

Cluster 13 has Size: 105
Silluete is: 0.721633732945388

Cluster 14 has Size: 2432
Silluete is: -0.3898186498580482

Cluster 15 has Size: 250
Silluete is: -0.8843041537660713

Cluster 16 has Size: 579
Silluete is: -0.7582972557267706

Cluster 17 has Size: 1496
Silluete is: -0.49293069403736933

Cluster 18 has Size: 113
Silluete is: -0.7322203565688052

Cluster 19 has Size: 180
Silluete is: 0.021102547144811846
```

```
Cluster 20 has Size: 227
Silluete is: -0.8961633295618644

Cluster 21 has Size: 98
Silluete is: 0.9065260561391866

Cluster 22 has Size: 6
Silluete is: -0.20202810571969035

Cluster 23 has Size: 944
Silluete is: -0.874605653674075

Cluster 24 has Size: 13
Silluete is: -0.8573731933285546

Cluster 25 has Size: 15
Silluete is: -0.6293321835176455

Cluster 26 has Size: 14
Silluete is: -0.5878413666673454

Cluster 27 has Size: 6
Silluete is: -0.8349322368120625

Cluster 28 has Size: 496
Silluete is: -0.7792920124144272

Cluster 29 has Size: 106
Silluete is: -0.5848828280324305

Cluster 30 has Size: 137
Silluete is: -0.8227512605415768

Cluster 31 has Size: 106
Silluete is: -0.6332724706969065

Cluster 32 has Size: 210
Silluete is: -0.6922175914708738

Cluster 33 has Size: 52
Silluete is: -0.4039757258466637

Cluster 34 has Size: 88
Silluete is: -0.9110567095682617

Cluster 35 has Size: 26
Silluete is: 0.4062088010954869

Cluster 36 has Size: 472
Silluete is: -0.48253797205229043

Cluster 37 has Size: 73
Silluete is: -0.8805038175478169

Cluster 38 has Size: 9
Silluete is: -0.847880720947595

Cluster 39 has Size: 165
```

```
Silluete is: 0.7170772103022749

Cluster 40 has Size: 267
Silluete is: -0.9139663051462008

Cluster 41 has Size: 80
Silluete is: -0.8165549626766186

Cluster 42 has Size: 198
Silluete is: -0.8633505263243207

Cluster 43 has Size: 42
Silluete is: -0.7703482353481858

Cluster 44 has Size: 85
Silluete is: -0.18741593637162407

Cluster 45 has Size: 169
Silluete is: -0.6529771193766751

Cluster 46 has Size: 70
Silluete is: -0.9173710644762941

Cluster 47 has Size: 36
Silluete is: -0.516387025508857

Cluster 48 has Size: 19
Silluete is: -0.45653720754232247

Cluster 49 has Size: 231
Silluete is: -0.8994103314914469

Cluster 50 has Size: 14
Silluete is: -0.75473109780599

Cluster 51 has Size: 32
Silluete is: -0.6914475300934971

Cluster 52 has Size: 64
Silluete is: -0.9413385529050211

Cluster 53 has Size: 70
Silluete is: 0.8208904882209581

Cluster 54 has Size: 5
Silluete is: -0.8334551260584541

Cluster 55 has Size: 23
Silluete is: -0.5774645419261352

Cluster 56 has Size: 263
Silluete is: 0.7166666666666445

Cluster 57 has Size: 122
Silluete is: -0.8765764969639191

Cluster 58 has Size: 36
Silluete is: -0.6524998501274112
```

```
Cluster 59 has Size: 11
Silluete is: 0.28977093686301797

Cluster 60 has Size: 11
Silluete is: -0.8873419643589454

Cluster 61 has Size: 81
Silluete is: 0.8021978021978066

Cluster 62 has Size: 83
Silluete is: -0.6039405824334294

Cluster 63 has Size: 6
Silluete is: -0.9000276359480859

Cluster 64 has Size: 41
Silluete is: -0.9141999504864203

Cluster 65 has Size: 5
Silluete is: -0.674109081985911

Cluster 66 has Size: 17
Silluete is: -0.4727503454861531

Cluster 67 has Size: 22
Silluete is: -0.7095259399263241

Cluster 68 has Size: 196
Silluete is: 0.7927668931418337

Cluster 69 has Size: 66
Silluete is: -0.5434368949381447

Cluster 70 has Size: 8
Silluete is: -0.9462878879362147

Cluster 71 has Size: 13
Silluete is: -0.37189580926523746

Cluster 72 has Size: 13
Silluete is: -0.8475319829791793

Cluster 73 has Size: 137
Silluete is: -0.49927278081737486

Cluster 74 has Size: 73
Silluete is: -0.7512203405774038

Cluster 75 has Size: 28
Silluete is: -0.46562657732992646

Cluster 76 has Size: 6
Silluete is: -0.48437407122022114

Cluster 77 has Size: 39
Silluete is: -0.5833267746055117

Cluster 78 has Size: 137
Silluete is: -0.9586682330991834
```

```
Cluster 79 has Size: 61
Silluete is: -0.8190964839759479

Cluster 80 has Size: 155
Silluete is: -0.39331265526171605

Cluster 81 has Size: 8
Silluete is: -0.8214589814187733

Cluster 82 has Size: 12
Silluete is: -0.5129485451533953

Cluster 83 has Size: 209
Silluete is: 0.6818181818182691

Cluster 84 has Size: 50
Silluete is: -0.722279859353705

Cluster 85 has Size: 23
Silluete is: -0.16140355533954162

Cluster 86 has Size: 83
Silluete is: -0.5848828280324305

Cluster 87 has Size: 10
Silluete is: -0.6504445791588628

Cluster 88 has Size: 15
Silluete is: 0.1680310577822056

Cluster 89 has Size: 6
Silluete is: -0.9115072649536536

Cluster 90 has Size: 17
Silluete is: -0.8385319698336303

Cluster 91 has Size: 58
Silluete is: -0.5114911691415444

Cluster 92 has Size: 39
Silluete is: -0.10077089720152961

Cluster 93 has Size: 5
Silluete is: 0.5013490234385646

Cluster 94 has Size: 55
Silluete is: -0.6008239260366661

Cluster 95 has Size: 6
Silluete is: -0.9415268101784691

Cluster 96 has Size: 8
Silluete is: -0.9360820282256848

Cluster 97 has Size: 13
Silluete is: -0.861202397350363

Cluster 98 has Size: 30
```

```
Silluete is: -0.9007696018946032

Cluster 99 has Size: 9
Silluete is: -0.36632441624258477

Cluster 100 has Size: 19
Silluete is: -0.7059564508999759

Cluster 101 has Size: 5
Silluete is: -0.7718801429487285

Cluster 102 has Size: 5
Silluete is: 0.6372994038942676

Cluster 103 has Size: 9
Silluete is: -0.5075976612715732

Cluster 104 has Size: 9
Silluete is: -0.5630584703848935

Cluster 105 has Size: 52
Silluete is: 0.7290044980546148

Cluster 106 has Size: 15
Silluete is: -0.47099772706620746

Cluster 107 has Size: 13
Silluete is: -0.09706048836512833

Cluster 108 has Size: 29
Silluete is: -0.49658012316233685

Cluster 109 has Size: 42
Silluete is: -0.4156409849828443

Cluster 110 has Size: 30
Silluete is: -0.6666674414110231

Cluster 111 has Size: 20
Silluete is: 0.7733614169160912

Cluster 112 has Size: 17
Silluete is: -0.8698684159766565

Cluster 113 has Size: 6
Silluete is: -0.7721114411374601

Cluster 114 has Size: 12
Silluete is: -0.6312525552380792

Cluster 115 has Size: 82
Silluete is: -0.8318994637655641

Cluster 116 has Size: 9
Silluete is: -0.7780827210771434

Cluster 117 has Size: 8
Silluete is: -0.6330179715263123
```

```
Cluster 118 has Size: 89
Silluete is: -0.8848750925236663

Cluster 119 has Size: 23
Silluete is: -0.6119410209620456

Cluster 120 has Size: 16
Silluete is: -0.478575069325517

Cluster 121 has Size: 7
Silluete is: 0.09844704862033783

Cluster 122 has Size: 23
Silluete is: -0.5186255988157654

Cluster 123 has Size: 67
Silluete is: -0.5646854688079233

Cluster 124 has Size: 136
Silluete is: 0.3678853265321031

Cluster 125 has Size: 18
Silluete is: -0.6978914170050841

Cluster 126 has Size: 18
Silluete is: -0.9023315110784336

Cluster 127 has Size: 11
Silluete is: -0.3874762580196019

Cluster 128 has Size: 27
Silluete is: -0.6821050926380126

Cluster 129 has Size: 7
Silluete is: -0.5596661567748388

Cluster 130 has Size: 6
Silluete is: -0.8120717183015856

Cluster 131 has Size: 7
Silluete is: -0.8490302062231033

Cluster 132 has Size: 7
Silluete is: -0.8369386529101237

Cluster 133 has Size: 18
Silluete is: -0.11300193592282631

Cluster 134 has Size: 10
Silluete is: -0.8478216891595818

Cluster 135 has Size: 6
Silluete is: -0.458660950896823

Cluster 136 has Size: 25
Silluete is: -0.4546730739629351

Cluster 137 has Size: 5
Silluete is: -0.9100039548581045
```

```
Cluster 138 has Size: 5
Silluete is: -0.41230419304203375

Cluster 139 has Size: 33
Silluete is: -0.6018696027111724

Cluster 140 has Size: 30
Silluete is: -0.8818267821123094

Cluster 141 has Size: 9
Silluete is: -0.5163859514822996

Cluster 142 has Size: 6
Silluete is: -0.10113572807855335

Cluster 143 has Size: 21
Silluete is: -0.19093049370252488

Cluster 144 has Size: 8
Silluete is: -0.945504691034468

Cluster 145 has Size: 5
Silluete is: -0.5131503054764521

Cluster 146 has Size: 15
Silluete is: -0.8990711874808923

Cluster 147 has Size: 26
Silluete is: 0.5262832658910364

Cluster 148 has Size: 6
Silluete is: -0.9481121719317833

Cluster 149 has Size: 27
Silluete is: -0.6441203157057357

Cluster 150 has Size: 12
Silluete is: -0.8761740836520384

Cluster 151 has Size: 5
Silluete is: -0.5787234695745913

Cluster 152 has Size: 50
Silluete is: -0.4265572984439192

Cluster 153 has Size: 9
Silluete is: -0.8233513919038691

Cluster 154 has Size: 19
Silluete is: -0.7698835891646616

Cluster 155 has Size: 24
Silluete is: -0.6322310309053402

Cluster 156 has Size: 16
Silluete is: 0.23326416991261037

Cluster 157 has Size: 10
```

```
Silluete is: -0.6919842621461754

Cluster 158 has Size: 25
Silluete is: 0.4934558235558666

Cluster 159 has Size: 10
Silluete is: -0.9062693358409426

Cluster 160 has Size: 11
Silluete is: -0.6886208797857913

Cluster 161 has Size: 46
Silluete is: -0.4490029678330182

Cluster 162 has Size: 8
Silluete is: -0.516387025508857

Cluster 163 has Size: 12
Silluete is: -0.5607188676135426

Cluster 164 has Size: 14
Silluete is: 0.6357248993845894

Cluster 165 has Size: 17
Silluete is: -0.42464491477119826

Cluster 166 has Size: 12
Silluete is: -0.9243678195853519

Cluster 167 has Size: 13
Silluete is: -0.8974939106199246

Cluster 168 has Size: 18
Silluete is: 0.7423284350892619

Cluster 169 has Size: 10
Silluete is: -0.8181073294525149

Cluster 170 has Size: 22
Silluete is: -0.8211640369273763

Cluster 171 has Size: 6
Silluete is: -0.7980579604307604

Cluster 172 has Size: 5
Silluete is: -0.5497751551431601

Cluster 173 has Size: 15
Silluete is: 0.3919729055268398

Cluster 174 has Size: 11
Silluete is: -0.803120565088768

Cluster 175 has Size: 6
Silluete is: -0.3172219107926928

Cluster 176 has Size: 37
Silluete is: -0.5780762589631997
```

```
Cluster 177 has Size: 25
Silluete is: -0.45429625738833435

Cluster 178 has Size: 17
Silluete is: -0.9378238999252122

Cluster 179 has Size: 23
Silluete is: 0.8721395289633752

Cluster 180 has Size: 6
Silluete is: -0.8138824701393353

Cluster 181 has Size: 68
Silluete is: -0.6441203157057357

Cluster 182 has Size: 31
Silluete is: -0.6659083686831812

Cluster 183 has Size: 13
Silluete is: -0.7442147949104033

Cluster 184 has Size: 13
Silluete is: -0.5752110790963658

Cluster 185 has Size: 10
Silluete is: -0.6658217104337046

Cluster 186 has Size: 8
Silluete is: -0.48452753839233675

Cluster 187 has Size: 25
Silluete is: -0.9146711465292969

Cluster 188 has Size: 11
Silluete is: -0.6547147155265056

Cluster 189 has Size: 6
Silluete is: -0.4349558330840379

Cluster 190 has Size: 8
Silluete is: -0.9287002020296649

Cluster 191 has Size: 5
Silluete is: -0.7950864694317825

Cluster 192 has Size: 5
Silluete is: -0.5805769885896717

Cluster 193 has Size: 6
Silluete is: -0.6216998450840526

Cluster 194 has Size: 6
Silluete is: -0.850616891384311

Cluster 195 has Size: 13
Silluete is: 0.3290655579744959

Cluster 196 has Size: 12
Silluete is: -0.4888115919991004
```

```
Cluster 197 has Size: 32
Silluete is: -0.7841229787978734

Cluster 198 has Size: 48
Silluete is: -0.8031222743544449

Cluster 199 has Size: 6
Silluete is: -0.5613772884018388

Cluster 200 has Size: 13
Silluete is: -0.9042179602618057

Cluster 201 has Size: 18
Silluete is: -0.5878413666673454

Cluster 202 has Size: 14
Silluete is: -0.6572376738981017

Cluster 203 has Size: 5
Silluete is: -0.516311310902087

Cluster 204 has Size: 38
Silluete is: -0.6446742234017827

Cluster 205 has Size: 20
Silluete is: -0.48616961412763093

Cluster 206 has Size: 6
Silluete is: -0.8586750640924935

Cluster 207 has Size: 8
Silluete is: 0.002012093593447026

Cluster 208 has Size: 13
Silluete is: -0.6975098329427187

Cluster 209 has Size: 20
Silluete is: 0.7989990092665831

Cluster 210 has Size: 9
Silluete is: -0.5447616369509303

Cluster 211 has Size: 10
Silluete is: -0.6401203019433335

Cluster 212 has Size: 9
Silluete is: -0.5046147275510258

Cluster 213 has Size: 9
Silluete is: -0.8970816607585363

Cluster 214 has Size: 6
Silluete is: -0.6311563677616048

Cluster 215 has Size: 5
Silluete is: -0.8805734597696542

Cluster 216 has Size: 42
```

```
Silluete is: -0.8384865749406177

Cluster 217 has Size: 57
Silluete is: -0.21839386987705248

Cluster 218 has Size: 17
Silluete is: 0.5247141064700079

Cluster 219 has Size: 5
Silluete is: -0.5080352341686114

Cluster 220 has Size: 11
Silluete is: 0.095564396853736

Cluster 221 has Size: 5
Silluete is: -0.8610710132426448

Cluster 222 has Size: 8
Silluete is: -0.5777779498737166

Cluster 223 has Size: 6
Silluete is: -0.8850081710597851

Cluster 224 has Size: 19
Silluete is: -0.7881343226290014

Cluster 225 has Size: 11
Silluete is: -0.5824475385326014

Cluster 226 has Size: 35
Silluete is: -0.5803554967550434

Cluster 227 has Size: 5
Silluete is: 0.60756105485827

Cluster 228 has Size: 13
Silluete is: -0.7757946216193742

Cluster 229 has Size: 7
Silluete is: -0.9041418981536243

Cluster 230 has Size: 10
Silluete is: -0.8566556938422596

Cluster 231 has Size: 5
Silluete is: -0.7572685291594525

Cluster 232 has Size: 7
Silluete is: -0.5690437141104037

Cluster 233 has Size: 5
Silluete is: -0.4854908133974673

Cluster 234 has Size: 48
Silluete is: 0.7834523554170819

Cluster 235 has Size: 9
Silluete is: 0.3358963196032987
```

```
Cluster 236 has Size: 16
Silluete is: -0.6111365980898361

Cluster 237 has Size: 7
Silluete is: -0.5261566460210068

Cluster 238 has Size: 11
Silluete is: -0.7979246485379788

Cluster 239 has Size: 5
Silluete is: -0.569119827173386

Cluster 240 has Size: 13
Silluete is: -0.5098105861518775

Cluster 241 has Size: 7
Silluete is: -0.8680570907159845

Cluster 242 has Size: 8
Silluete is: -0.8494694908483488

Cluster 243 has Size: 5
Silluete is: -0.41230419304203375

Cluster 244 has Size: 9
Silluete is: -0.8221778114047322

Cluster 245 has Size: 9
Silluete is: -0.91105394113584

Cluster 246 has Size: 5
Silluete is: -0.8415522609836315

Cluster 247 has Size: 7
Silluete is: -0.49683131730145597

Cluster 248 has Size: 14
Silluete is: -0.8001813660453106

Cluster 249 has Size: 38
Silluete is: -0.5825508468362384

Cluster 250 has Size: 12
Silluete is: 0.7829721920934555

Cluster 251 has Size: 6
Silluete is: -0.6213799526129478

Cluster 252 has Size: 9
Silluete is: -0.7848198138665992

Cluster 253 has Size: 9
Silluete is: -0.5225983695958978

Cluster 254 has Size: 14
Silluete is: -0.8241570219702272

Cluster 255 has Size: 13
Silluete is: -0.6177267520630952
```

```
Cluster 256 has Size: 7
Silluete is: -0.6130572917918546

Cluster 257 has Size: 10
Silluete is: -0.5563431095649539

Cluster 258 has Size: 6
Silluete is: -0.790130693186932

Cluster 259 has Size: 12
Silluete is: -0.9094222054176656

Cluster 260 has Size: 16
Silluete is: -0.45744889934159727

Cluster 261 has Size: 5
Silluete is: -0.4439146742880775

Cluster 262 has Size: 6
Silluete is: 0.8587803372087085

Cluster 263 has Size: 10
Silluete is: -0.8363906732081652

Cluster 264 has Size: 7
Silluete is: 0.7900569762775459

Cluster 265 has Size: 6
Silluete is: -0.794041533317998

Cluster 266 has Size: 13
Silluete is: -0.1319237734644076

Cluster 267 has Size: 16
Silluete is: -0.7677995488129427

Cluster 268 has Size: 11
Silluete is: -0.07786783188497079

Cluster 269 has Size: 6
Silluete is: -0.6396713575279741

Cluster 270 has Size: 6
Silluete is: -0.7873635147442607

Cluster 271 has Size: 9
Silluete is: -0.8712106858644201

Cluster 272 has Size: 18
Silluete is: -0.45679476705257277

Cluster 273 has Size: 10
Silluete is: -0.8666142453052434

Cluster 274 has Size: 8
Silluete is: -0.481047307494312

Cluster 275 has Size: 10
```

```
Silluete is: -0.4188865961869933

Cluster 276 has Size: 6
Silluete is: 0.6634971765147704

Cluster 277 has Size: 6
Silluete is: -0.34974643388489624

Cluster 278 has Size: 5
Silluete is: -0.5434368949381447

Cluster 279 has Size: 7
Silluete is: -0.7761242024514503

Cluster 280 has Size: 5
Silluete is: 0.4003646786715306

Cluster 281 has Size: 13
Silluete is: 0.21669270462339157

Cluster 282 has Size: 5
Silluete is: 0.5478712594170206

Cluster 283 has Size: 8
Silluete is: -0.8675310111841985

Cluster 284 has Size: 6
Silluete is: -0.6073036078946576

Cluster 285 has Size: 6
Silluete is: -0.7596756292361942

Cluster 286 has Size: 5
Silluete is: 0.7368421052630287

Cluster 287 has Size: 12
Silluete is: -0.8861921956195988

Cluster 288 has Size: 6
Silluete is: -0.5367396917110668

Cluster 289 has Size: 5
Silluete is: -0.4013552388456755

Cluster 290 has Size: 9
Silluete is: -0.7364015788987316

Cluster 291 has Size: 11
Silluete is: -0.45266997831555694

Cluster 292 has Size: 7
Silluete is: -0.8998837197661307

Cluster 293 has Size: 9
Silluete is: -0.6673912522395904

Cluster 294 has Size: 14
Silluete is: -0.4836886008325187
```

```
Cluster 295 has Size: 8
Silluete is: 0.45204362287590677

Cluster 296 has Size: 18
Silluete is: -0.8909744361825969

Cluster 297 has Size: 5
Silluete is: -0.8824969963171148

Cluster 298 has Size: 10
Silluete is: -0.5613772884018388

Cluster 299 has Size: 16
Silluete is: -0.7929737869091572

Cluster 300 has Size: 9
Silluete is: -0.8925244714562716

Cluster 301 has Size: 5
Silluete is: 0.5353783169314614

Cluster 302 has Size: 11
Silluete is: -0.5564804225442661

Cluster 303 has Size: 10
Silluete is: 0.469282042762279

Cluster 304 has Size: 9
Silluete is: -0.7950864694317825

Cluster 305 has Size: 5
Silluete is: -0.777674179562443

Cluster 306 has Size: 5
Silluete is: 0.8226695365202218

Cluster 307 has Size: 6
Silluete is: -0.80244771765981

Cluster 308 has Size: 10
Silluete is: -0.5293469284408279

Cluster 309 has Size: 10
Silluete is: -0.5089165739719598

Cluster 310 has Size: 10
Silluete is: -0.5165739563972105

Cluster 311 has Size: 8
Silluete is: -0.7861907508488184

Cluster 312 has Size: 7
Silluete is: -0.14396459517705976

Cluster 313 has Size: 12
Silluete is: -0.45744889934159727

Cluster 314 has Size: 16
Silluete is: 0.33160730895293933
```

```
Cluster 315 has Size: 6
Silluete is: -0.5279626030478072

Cluster 316 has Size: 8
Silluete is: -0.7836590593824128

Cluster 317 has Size: 12
Silluete is: 0.023621603368942982

Cluster 318 has Size: 7
Silluete is: -0.41382076383973176

Cluster 319 has Size: 6
Silluete is: -0.8961083800531492

Cluster 320 has Size: 6
Silluete is: -0.42803277578238336

Cluster 321 has Size: 15
Silluete is: -0.6382614748084598

Cluster 322 has Size: 13
Silluete is: -0.5978956686584389

Cluster 323 has Size: 6
Silluete is: -0.8606986601683324

Cluster 324 has Size: 5
Silluete is: -0.4699685696046727

Cluster 325 has Size: 6
Silluete is: -0.9204764870647957

Cluster 326 has Size: 6
Silluete is: -0.8391684559297571

Cluster 327 has Size: 5
Silluete is: -0.8771560089727324

Cluster 328 has Size: 10
Silluete is: 0.3469631490694923

Cluster 329 has Size: 18
Silluete is: -0.855063999625161

Cluster 330 has Size: 5
Silluete is: -0.687290805623564

Cluster 331 has Size: 5
Silluete is: -0.5165739563972105

Cluster 332 has Size: 9
Silluete is: -0.7963820735611878

Cluster 333 has Size: 8
Silluete is: -0.22176540003076176

Cluster 334 has Size: 9
```

```
Silluete is: -0.7095259399263241

Cluster 335 has Size: 8
Silluete is: -0.8645429740859338

Cluster 336 has Size: 5
Silluete is: -0.5157811426597819

Cluster 337 has Size: 7
Silluete is: -0.8522786447863554

Cluster 338 has Size: 5
Silluete is: -0.43386533996018417

Cluster 339 has Size: 5
Silluete is: -0.467442888765865

Cluster 340 has Size: 12
Silluete is: -0.6153648249546377

Cluster 341 has Size: 9
Silluete is: -0.7771448077441229

Cluster 342 has Size: 7
Silluete is: 0.7919986411067805

Cluster 343 has Size: 6
Silluete is: 0.7406200196567514

Cluster 344 has Size: 7
Silluete is: -0.46442404042956764

Cluster 345 has Size: 15
Silluete is: -0.9254805820030548

Cluster 346 has Size: 6
Silluete is: -0.6964768244459311

Cluster 347 has Size: 8
Silluete is: -0.4589378362672778

Cluster 348 has Size: 5
Silluete is: -0.624542040332358

Cluster 349 has Size: 8
Silluete is: -0.8494029462961608

Cluster 350 has Size: 14
Silluete is: -0.5917633348444943

Cluster 351 has Size: 8
Silluete is: -0.5854579486883835

Cluster 352 has Size: 11
Silluete is: -0.5917633348444943

Cluster 353 has Size: 18
Silluete is: -0.7473147331221983
```

```
Cluster 354 has Size: 5
Silluete is: -0.6847082290100535

Cluster 355 has Size: 7
Silluete is: -0.5851548230648576

Cluster 356 has Size: 9
Silluete is: -0.8966583022487802

Cluster 357 has Size: 7
Silluete is: -0.3898186498580482

Cluster 358 has Size: 6
Silluete is: -0.3689947576867405

Cluster 359 has Size: 14
Silluete is: 0.3837883181298198

Cluster 360 has Size: 9
Silluete is: -0.6629682560581603

Cluster 361 has Size: 5
Silluete is: -0.785561715508594

Cluster 362 has Size: 5
Silluete is: -0.5921370777678725

Cluster 363 has Size: 7
Silluete is: -0.029039595869856695

Cluster 364 has Size: 7
Silluete is: 0.8482704765823272

Cluster 365 has Size: 6
Silluete is: -0.7980579604307604

Cluster 366 has Size: 7
Silluete is: 0.6553279893971885

Cluster 367 has Size: 5
Silluete is: 0.8193223901890059

Cluster 368 has Size: 7
Silluete is: -0.8479305031952798

Cluster 369 has Size: 5
Silluete is: -0.7216878490142096

Cluster 370 has Size: 6
Silluete is: 0.19715191649402056

Cluster 371 has Size: 10
Silluete is: -0.23718826664524081

Cluster 372 has Size: 5
Silluete is: -0.8411016791287711

Cluster 373 has Size: 5
Silluete is: -0.5825508468362384
```

```
Cluster 374 has Size: 5
Silluete is: -0.7341098047857559

Cluster 375 has Size: 6
Silluete is: 0.6097350991636592

Cluster 376 has Size: 6
Silluete is: -0.6396713575279741

Cluster 377 has Size: 5
Silluete is: -0.49927278081737486

Cluster 378 has Size: 6
Silluete is: -0.8470877704283155

Cluster 379 has Size: 6
Silluete is: -0.665392857590311

Cluster 380 has Size: 9
Silluete is: -0.798795827369648

Cluster 381 has Size: 5
Silluete is: -0.8924518158666119

Cluster 382 has Size: 9
Silluete is: -0.8068611272534046

Cluster 383 has Size: 5
Silluete is: -0.7132722936495548

Cluster 384 has Size: 9
Silluete is: -0.5480703283700994

Cluster 385 has Size: 9
Silluete is: 0.41014136052959715

Cluster 386 has Size: 6
Silluete is: -0.5630584703848935

Cluster 387 has Size: 8
Silluete is: -0.6684658488271024

Cluster 388 has Size: 6
Silluete is: 0.20749406827604117

Cluster 389 has Size: 5
Silluete is: -0.6103477375681635

Cluster 390 has Size: 15
Silluete is: -0.3391876223202975

Cluster 391 has Size: 6
Silluete is: 0.3919729055268398

Cluster 392 has Size: 7
Silluete is: -0.7389824071929871

Cluster 393 has Size: 7
```

```
Silluete is: -0.8447854509237241

Cluster 394 has Size: 9
Silluete is: -0.2879170499402275

Cluster 395 has Size: 5
Silluete is: -0.5934969193216859

Cluster 396 has Size: 7
Silluete is: -0.7731137494650709

Cluster 397 has Size: 7
Silluete is: -0.8259280631338227

Cluster 398 has Size: 5
Silluete is: 0.7789887891448006

Cluster 399 has Size: 5
Silluete is: -0.7018792974325024

Cluster 400 has Size: 8
Silluete is: -0.5076152066679821

Cluster 401 has Size: 8
Silluete is: -0.433616950087064

Cluster 402 has Size: 5
Silluete is: -0.6473163464978193

Cluster 403 has Size: 6
Silluete is: -0.8605773830656029

Cluster 404 has Size: 7
Silluete is: -0.8825886561852239

Cluster 405 has Size: 5
Silluete is: -0.8687372340672649

Cluster 406 has Size: 5
Silluete is: -0.7980579604307604

Cluster 407 has Size: 7
Silluete is: -0.8431456813537522

Cluster 408 has Size: 5
Silluete is: -0.6305756431505473

Cluster 409 has Size: 5
Silluete is: -0.526093583084798

Cluster 410 has Size: 6
Silluete is: -0.8353097243310876

Cluster 411 has Size: 6
Silluete is: -0.5109331251459727

Cluster 412 has Size: 5
Silluete is: -0.39661290927989407
```

```
Cluster 413 has Size: 5
Silluete is: -0.7402347771862741

Cluster 414 has Size: 11
Silluete is: -0.6005942317609222

Cluster 415 has Size: 9
Silluete is: -0.7856016060520605

Cluster 416 has Size: 14
Silluete is: -0.8858627308502109

Cluster 417 has Size: 7
Silluete is: -0.7189864459362584

Cluster 418 has Size: 9
Silluete is: -0.49935193387246474

Cluster 419 has Size: 7
Silluete is: -0.5787234695745913

Cluster 420 has Size: 5
Silluete is: -0.9037441273576754

Cluster 421 has Size: 5
Silluete is: -0.9445019759998324

Cluster 422 has Size: 5
Silluete is: -0.6861469746000177

Cluster 423 has Size: 7
Silluete is: -0.5630584703848935

Cluster 424 has Size: 8
Silluete is: -0.8661691658601943

Cluster 425 has Size: 7
Silluete is: -0.6816215133465988

Cluster 426 has Size: 6
Silluete is: -0.40126362833529655

Cluster 427 has Size: 7
Silluete is: -0.886001808802081

Cluster 428 has Size: 12
Silluete is: -0.7647507965767253

Cluster 429 has Size: 7
Silluete is: -0.40495523156531876

Cluster 430 has Size: 5
Silluete is: -0.7539203635713919

Cluster 431 has Size: 3
Silluete is: -0.5252613670545415

Cluster 432 has Size: 3
Silluete is: -0.8378344094829476
```

```
Cluster 433 has Size: 6
Silluete is: -0.7646464735741132

Cluster 434 has Size: 5
Silluete is: -0.3944344691823939

Cluster 435 has Size: 6
Silluete is: -0.6504445791588628

Cluster 436 has Size: 5
Silluete is: -0.2585992506211323

Cluster 437 has Size: 7
Silluete is: -0.504549678811402

Cluster 438 has Size: 9
Silluete is: -0.7724043055937021

Cluster 439 has Size: 5
Silluete is: -0.6086143448514774

Cluster 440 has Size: 5
Silluete is: -0.5294875715057956

Cluster 441 has Size: 7
Silluete is: -0.9038222940945865

Cluster 442 has Size: 3
Silluete is: -0.5613772884018388

Cluster 443 has Size: 4
Silluete is: -0.42479153864423946

Cluster 444 has Size: 5
Silluete is: -0.21132897267382797

Cluster 445 has Size: 6
Silluete is: -0.7819222798734289

Cluster 446 has Size: 7
Silluete is: -0.5806537259806598

Cluster 447 has Size: 6
Silluete is: -0.663410886937318

Cluster 448 has Size: 6
Silluete is: -0.7217721040202102

Cluster 449 has Size: 6
Silluete is: 0.7687051711417687

Cluster 450 has Size: 8
Silluete is: -0.5878413666673454

Cluster 451 has Size: 8
Silluete is: 0.8082882564723025

Cluster 452 has Size: 5
```

```
Silluete is: -0.8970816607585363

Cluster 453 has Size: 5
Silluete is: 0.7019477245569219

Cluster 454 has Size: 9
Silluete is: -0.5447616369509303

Cluster 455 has Size: 6
Silluete is: -0.3898186498580482

Cluster 456 has Size: 6
Silluete is: -0.6457676457119815

Cluster 457 has Size: 7
Silluete is: -0.49678909421520984

Cluster 458 has Size: 5
Silluete is: -0.7791438884493177

Cluster 459 has Size: 5
Silluete is: -0.6280502481199148

Cluster 460 has Size: 6
Silluete is: -0.7116017134887628

Cluster 461 has Size: 7
Silluete is: -0.7111427560433754

Cluster 462 has Size: 5
Silluete is: -0.4652121863848102

Cluster 463 has Size: 5
Silluete is: -0.8858889257319036

Cluster 464 has Size: 5
Silluete is: 0.512219532836953

Cluster 465 has Size: 5
Silluete is: -0.6132567641575715

Cluster 466 has Size: 6
Silluete is: -0.8064079329390534

Cluster 467 has Size: 5
Silluete is: -0.5278154862524685

Cluster 468 has Size: 5
Silluete is: -0.8730490272778492

Cluster 469 has Size: 5
Silluete is: -0.9119477117059569

Cluster 470 has Size: 5
Silluete is: 0.6726190476191324

Cluster 471 has Size: 5
Silluete is: -0.46442404042956764
```

```
Cluster 472 has Size: 5
Silluete is: -0.6166137127420371

Cluster 473 has Size: 5
Silluete is: 0.36629366011433606

Cluster 474 has Size: 3
Silluete is: -0.7958013698759466

Cluster 475 has Size: 5
Silluete is: -0.7138198918997819

Cluster 476 has Size: 5
Silluete is: -0.8772337428276334

Cluster 477 has Size: 5
Silluete is: -0.002949974242627596

Cluster 478 has Size: 3
Silluete is: 0.47498972356342795

Cluster 479 has Size: 5
Silluete is: -0.19775390980764482

Cluster 480 has Size: 5
Silluete is: -0.8993222942806294

Cluster 481 has Size: 5
Silluete is: -0.8363049631788113

Cluster 482 has Size: 5
Silluete is: 0.16978800859376017

Cluster 483 has Size: 5
Silluete is: -0.8151879047237368

Cluster 484 has Size: 6
Silluete is: 0.7471753989868029

Cluster 485 has Size: 7
Silluete is: -0.24586395219881763

Cluster 486 has Size: 5
Silluete is: -0.8581412590327849

Cluster 487 has Size: 6
Silluete is: -0.8430124474335359

Cluster 488 has Size: 5
Silluete is: -0.7047558882831019

Cluster 489 has Size: 5
Silluete is: -0.7614253487187901

Cluster 490 has Size: 5
Silluete is: 0.959415584415579

Cluster 491 has Size: 5
Silluete is: -0.7726587353513585
```

```
Cluster 492 has Size: 5
Silluete is: -0.7809176666574279

Cluster 493 has Size: 5
Silluete is: -0.9103836216020675

Cluster 494 has Size: 5
Silluete is: -0.9105944139484612

Cluster 495 has Size: 5
Silluete is: -0.7475820044519411

Cluster 496 has Size: 1
Silluete is: -0.5075976612715732

Cluster 497 has Size: 5
Silluete is: 0.4328773798692447

Cluster 498 has Size: 5
Silluete is: -0.845603012254786

Cluster 499 has Size: 5
Silluete is: -0.8795559814479693

Cluster 500 has Size: 5
Silluete is: -0.569119827173386

Cluster 501 has Size: 5
Silluete is: -0.8545260569332116

Cluster 502 has Size: 5
Silluete is: -0.5888512047103527

Cluster 503 has Size: 5
Silluete is: -0.9356861017822521

Cluster 504 has Size: 5
Silluete is: -0.599051788353387

Cluster 505 has Size: 6
Silluete is: -0.5891924190070925

Cluster 506 has Size: 4
Silluete is: 0.4210384319643196

Cluster 507 has Size: 5
Silluete is: -0.4732809486494108

Cluster 508 has Size: 5
Silluete is: -0.7881875057871474
```

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

In [110…

```
# epsilon = 0.5 and min_samples = 25
clustering_05_25_e = DBSCAN(eps=0.5, min_samples=25, metric='euclidean').fit(a
```

In [111…

```
# 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, la
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: 67
Estimated number of noise points: 21787
Homogeneity: 0.111
Confusion Matrix
[15627   542   124  1957  1279  1259     7   485   416    52   107    27
   131   473   166    58    44    31   128   106    24    26    53    51
   112    36     6    11    88    45    13    35    37    34    59    73
    45    18    27    43    55    43    11    36    22   106    38    10
    65    36     8    42    38    43    25    14    44    25    14    28
    16    25    25    24    32     4    18    18]
[6160    1  268   22  502   26   14  299   20    0  135    6    0   35
    0    0    0    0    0    0   18    0    3    0   44    7   24   16
    0    0   59    0    2    4   18    0    0   11   19    0    0    0
   15    0    9    0    0   16    5    1   42    0    0    0    0   16
    0    3    3    2    0    1    0    2    5    0    8    0]
Silhouette Coefficient: -0.515
Overview over the Cluster

Cluster 0 has Size: 543
Silluete is: -0.3616277095866529

Cluster 1 has Size: 392
Silluete is: -0.41758033665275407

Cluster 2 has Size: 1979
Silluete is: -0.5395724521662205
```

```
Cluster 3 has Size: 1781
Silluete is: -0.5242163815809564

Cluster 4 has Size: 1285
Silluete is: -0.4184604583762632

Cluster 5 has Size: 21
Silluete is: -0.8784125047351619

Cluster 6 has Size: 784
Silluete is: -0.30777152455085477

Cluster 7 has Size: 436
Silluete is: -0.8081878782272945

Cluster 8 has Size: 52
Silluete is: -0.6279988217159784

Cluster 9 has Size: 242
Silluete is: -0.7310241005699176

Cluster 10 has Size: 33
Silluete is: -0.3461173971419233

Cluster 11 has Size: 131
Silluete is: -0.3387906147241542

Cluster 12 has Size: 508
Silluete is: -0.6853223609033137

Cluster 13 has Size: 166
Silluete is: -0.7198635561305384

Cluster 14 has Size: 58
Silluete is: 0.7258314083344124

Cluster 15 has Size: 44
Silluete is: -0.547711387268926

Cluster 16 has Size: 31
Silluete is: -0.8504223497623095

Cluster 17 has Size: 128
Silluete is: -0.5176335472689532

Cluster 18 has Size: 106
Silluete is: -0.5849507187932954

Cluster 19 has Size: 42
Silluete is: -0.8305995663359351

Cluster 20 has Size: 26
Silluete is: -0.5353046150131736

Cluster 21 has Size: 56
Silluete is: -0.40411352370136805

Cluster 22 has Size: 51
```

```
Silluete is: -0.5608645493223976

Cluster 23 has Size: 156
Silluete is: -0.572447394115716

Cluster 24 has Size: 43
Silluete is: -0.48350074998287795

Cluster 25 has Size: 30
Silluete is: -0.902403946816245

Cluster 26 has Size: 27
Silluete is: -0.6805828401435543

Cluster 27 has Size: 88
Silluete is: -0.6436339584853393

Cluster 28 has Size: 45
Silluete is: -0.3730625592693834

Cluster 29 has Size: 72
Silluete is: -0.2944332304753474

Cluster 30 has Size: 35
Silluete is: -0.7302820772797145

Cluster 31 has Size: 39
Silluete is: -0.827689587594939

Cluster 32 has Size: 38
Silluete is: -0.6605997029518254

Cluster 33 has Size: 77
Silluete is: -0.6206272631362629

Cluster 34 has Size: 73
Silluete is: -0.8980152180331903

Cluster 35 has Size: 45
Silluete is: -0.5368988287519139

Cluster 36 has Size: 29
Silluete is: -0.6546482098002979

Cluster 37 has Size: 46
Silluete is: -0.9138064601486907

Cluster 38 has Size: 43
Silluete is: -0.8157876892662217

Cluster 39 has Size: 55
Silluete is: -0.7376271272468082

Cluster 40 has Size: 43
Silluete is: -0.6442722486525415

Cluster 41 has Size: 26
Silluete is: -0.711101734558065
```

```
Cluster 42 has Size: 36
Silluete is: -0.8362406972094385

Cluster 43 has Size: 31
Silluete is: -0.48622441508204084

Cluster 44 has Size: 106
Silluete is: -0.8868215178964686

Cluster 45 has Size: 38
Silluete is: 0.9129286295030853

Cluster 46 has Size: 26
Silluete is: -0.6808365582358391

Cluster 47 has Size: 70
Silluete is: -0.8784292079249061

Cluster 48 has Size: 37
Silluete is: -0.8972852234635094

Cluster 49 has Size: 50
Silluete is: -0.8498871960644498

Cluster 50 has Size: 42
Silluete is: -0.8753220225827381

Cluster 51 has Size: 38
Silluete is: 0.6091485808362795

Cluster 52 has Size: 43
Silluete is: -0.5667279576488825

Cluster 53 has Size: 25
Silluete is: -0.6087391720386056

Cluster 54 has Size: 30
Silluete is: -0.6915757381914146

Cluster 55 has Size: 44
Silluete is: -0.6182504229447332

Cluster 56 has Size: 28
Silluete is: -0.35202013578535346

Cluster 57 has Size: 17
Silluete is: -0.8381941644942846

Cluster 58 has Size: 30
Silluete is: -0.11385445039365935

Cluster 59 has Size: 16
Silluete is: -0.8133784216468686

Cluster 60 has Size: 26
Silluete is: -0.7923162607177606

Cluster 61 has Size: 25
Silluete is: -0.5931448104209651
```

```
Cluster 62 has Size: 26
Silluete is: -0.327767335495985

Cluster 63 has Size: 37
Silluete is: -0.681654142610721

Cluster 64 has Size: 4
Silluete is: -0.6852646350273081

Cluster 65 has Size: 26
Silluete is: -0.6430041972867321

Cluster 66 has Size: 18
Silluete is: -0.6186844029321332

Cluster -1 has Size: 21787
```

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

In [112…

```
# epsilon = 0.75 and min_samples = 10
clustering_075_10_e = DBSCAN(eps=0.75, min_samples=10, metric='euclidean').fit
```

In [113…

```
# 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, la
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: 164
Estimated number of noise points: 10662
Homogeneity: 0.237
Confusion Matrix
[7275  664 1031  261   94   29    9  137 3211  201  257  847 2117 2526
  824   39  118   20   17   37   31   71   98  134  232   48   83    4
```

```
  645   24   41   76  231   11   16   24  288  128  128    4  138   87
   34  112  121   10  120  174  227   27    6   50   38   20   11   14
    2   75   16   19   36   24   41   18   24   46   16   14   24   16
    8    7    9   55    6    4   16   30   22   29    6   15   20   14
   22    9    7   10   14   59    3   12   13   24   33   13   14   17
   44    0    9   10    6   26    5   20    1    5   12   10   16   12
   22    5    0    9    5   18   17   10    7   21   15    8    9   10
   15   16    9    0   10   10    3    8    0   10   16   14   12   10
   14    0    8    8    6    4    7    8   10   12    0   14    7    9
    5    9    8    5    1    4    6   10    3    6    9]
[3387   48    2  685  368    0    4   93   56   51  389  577  915   54
  116    1  137    2    1    1   45   54    0   11   44    0    0   15
    5   64    0    5   14   31    6    0   17    0    0   12    1   21
    0    9  105   25   25    3    8    1   17    0    0    0    2    0
   32    0    0    2    0    0    0    0    0    0    0   21   33    0
    3    5    2    5    9   11    0   17    0    2   26    0    0    2
    0    0    6    3    0    0   21    1    0    0    0    0    3    1
    2   20    8    1    7    3    7    0   18    8    0    0    0    0
    0    1   10    1    5    0    3    1    8    0    1    2    1    3
    1    0    0   10    0    0    3    1    9    4    0    2    1    0
    1    7    2    2    0    2    3    5    2    0    8    0    1    1
    5    0    2    5    9    5    4    0    4    0    0]
Silhouette Coefficient: -0.502
Overview over the Cluster

Cluster 0 has Size: 712
Silluete is: -0.5173220598721631

Cluster 1 has Size: 1033
Silluete is: -0.5128680966525412

Cluster 2 has Size: 946
Silluete is: -0.22138327527798551

Cluster 3 has Size: 462
Silluete is: -0.8703738470934991

Cluster 4 has Size: 29
Silluete is: -0.5279700985171164

Cluster 5 has Size: 13
Silluete is: -0.5835188102263661

Cluster 6 has Size: 230
Silluete is: -0.5665244907769138

Cluster 7 has Size: 3267
Silluete is: 0.7236272345089299

Cluster 8 has Size: 252
Silluete is: -0.7025413425196062

Cluster 9 has Size: 646
Silluete is: 0.31796568488185134

Cluster 10 has Size: 1424
Silluete is: -0.749223367163023

Cluster 11 has Size: 3032
```

```
Silluete is: -0.5236086958995424

Cluster 12 has Size: 2580
Silluete is: 0.3868697694347258

Cluster 13 has Size: 940
Silluete is: 0.6875433611301465

Cluster 14 has Size: 40
Silluete is: -0.33389899063459666

Cluster 15 has Size: 255
Silluete is: -0.878354138324001

Cluster 16 has Size: 22
Silluete is: -0.34438153641796154

Cluster 17 has Size: 18
Silluete is: -0.30246795552195743

Cluster 18 has Size: 38
Silluete is: -0.6720784501385929

Cluster 19 has Size: 76
Silluete is: -0.45366968209848185

Cluster 20 has Size: 125
Silluete is: -0.8562113549190288

Cluster 21 has Size: 98
Silluete is: -0.703507475126047

Cluster 22 has Size: 145
Silluete is: 0.1294692444675187

Cluster 23 has Size: 276
Silluete is: -0.19565262384792592

Cluster 24 has Size: 48
Silluete is: -0.8419065004327314

Cluster 25 has Size: 83
Silluete is: -0.5418163657373134

Cluster 26 has Size: 19
Silluete is: -0.6246929457714683

Cluster 27 has Size: 650
Silluete is: -0.8014547257718954

Cluster 28 has Size: 88
Silluete is: -0.7858933221050335

Cluster 29 has Size: 41
Silluete is: -0.5060516915505644

Cluster 30 has Size: 81
Silluete is: -0.7744184473260355
```

```
Cluster 31 has Size: 245
Silluete is: -0.6541368562859318

Cluster 32 has Size: 42
Silluete is: -0.5204001187985342

Cluster 33 has Size: 22
Silluete is: -0.5642142104099691

Cluster 34 has Size: 24
Silluete is: -0.8192868950601521

Cluster 35 has Size: 305
Silluete is: -0.24220178521206231

Cluster 36 has Size: 128
Silluete is: -0.5197648992148161

Cluster 37 has Size: 128
Silluete is: -0.8221256368762633

Cluster 38 has Size: 16
Silluete is: -0.6251054883733813

Cluster 39 has Size: 139
Silluete is: -0.7721099078770371

Cluster 40 has Size: 108
Silluete is: 0.4718462791519387

Cluster 41 has Size: 34
Silluete is: -0.35344096306310296

Cluster 42 has Size: 121
Silluete is: -0.5308974079890619

Cluster 43 has Size: 226
Silluete is: -0.6078816765729922

Cluster 44 has Size: 35
Silluete is: -0.22975094291362902

Cluster 45 has Size: 145
Silluete is: -0.5729336168728151

Cluster 46 has Size: 177
Silluete is: -0.5004518745163475

Cluster 47 has Size: 235
Silluete is: -0.20659274773077824

Cluster 48 has Size: 28
Silluete is: -0.3837011823422918

Cluster 49 has Size: 23
Silluete is: -0.852419618564331

Cluster 50 has Size: 50
Silluete is: -0.6762165276774814
```

```
Cluster 51 has Size: 38
Silluete is: -0.4311185360618045

Cluster 52 has Size: 20
Silluete is: 0.5927465049985423

Cluster 53 has Size: 13
Silluete is: -0.8317027005756029

Cluster 54 has Size: 14
Silluete is: -0.7612548785274621

Cluster 55 has Size: 34
Silluete is: -0.47720295353569736

Cluster 56 has Size: 75
Silluete is: 0.7643165668737749

Cluster 57 has Size: 16
Silluete is: -0.805902092606364

Cluster 58 has Size: 21
Silluete is: -0.4029986870325597

Cluster 59 has Size: 36
Silluete is: -0.4094704753696985

Cluster 60 has Size: 24
Silluete is: -0.6884716733497398

Cluster 61 has Size: 41
Silluete is: 0.7816383386373715

Cluster 62 has Size: 18
Silluete is: -0.4571418416889901

Cluster 63 has Size: 24
Silluete is: -0.866155631987991

Cluster 64 has Size: 46
Silluete is: 0.41676388012069293

Cluster 65 has Size: 16
Silluete is: -0.3151201874018535

Cluster 66 has Size: 35
Silluete is: -0.5412035360504189

Cluster 67 has Size: 57
Silluete is: -0.5858161339758725

Cluster 68 has Size: 16
Silluete is: -0.47289480623360036

Cluster 69 has Size: 11
Silluete is: -0.537510183794097

Cluster 70 has Size: 12
```

```
Silluete is: 0.6577950492213798

Cluster 71 has Size: 11
Silluete is: -0.42027528312859014

Cluster 72 has Size: 60
Silluete is: -0.7760343341793379

Cluster 73 has Size: 15
Silluete is: -0.5457503707109074

Cluster 74 has Size: 15
Silluete is: -0.6873881121066796

Cluster 75 has Size: 16
Silluete is: -0.3329256631871694

Cluster 76 has Size: 47
Silluete is: -0.42226468183636473

Cluster 77 has Size: 22
Silluete is: -0.550109401226425

Cluster 78 has Size: 31
Silluete is: 0.4154513157422456

Cluster 79 has Size: 32
Silluete is: -0.8132820471859806

Cluster 80 has Size: 15
Silluete is: -0.5946182532527816

Cluster 81 has Size: 20
Silluete is: -0.7477339987407159

Cluster 82 has Size: 16
Silluete is: -0.13670704438491893

Cluster 83 has Size: 22
Silluete is: -0.8096378702903493

Cluster 84 has Size: 9
Silluete is: -0.6706296552396642

Cluster 85 has Size: 13
Silluete is: -0.5336190240891103

Cluster 86 has Size: 13
Silluete is: -0.5060516915505644

Cluster 87 has Size: 14
Silluete is: -0.7707043961010068

Cluster 88 has Size: 59
Silluete is: -0.24904809121105637

Cluster 89 has Size: 24
Silluete is: -0.8208623468460845
```

```
Cluster 90 has Size: 13
Silluete is: -0.8624868534005868

Cluster 91 has Size: 13
Silluete is: -0.5075268513425117

Cluster 92 has Size: 24
Silluete is: -0.5415264249434498

Cluster 93 has Size: 33
Silluete is: -0.4766577779060093

Cluster 94 has Size: 13
Silluete is: -0.6026095865189861

Cluster 95 has Size: 17
Silluete is: -0.852187757406142

Cluster 96 has Size: 18
Silluete is: -0.8607306013437268

Cluster 97 has Size: 46
Silluete is: -0.47714268903389734

Cluster 98 has Size: 20
Silluete is: -0.7824172154928144

Cluster 99 has Size: 17
Silluete is: -0.19889619336563646

Cluster 100 has Size: 11
Silluete is: -0.6960918080109545

Cluster 101 has Size: 13
Silluete is: -0.6561255181915656

Cluster 102 has Size: 29
Silluete is: 0.6697579543238872

Cluster 103 has Size: 12
Silluete is: -0.49991386742640204

Cluster 104 has Size: 20
Silluete is: -0.2494831966326535

Cluster 105 has Size: 19
Silluete is: -0.3439661839758353

Cluster 106 has Size: 13
Silluete is: -0.473724879334576

Cluster 107 has Size: 12
Silluete is: 0.03476443220333002

Cluster 108 has Size: 10
Silluete is: -0.2906848468700764

Cluster 109 has Size: 16
Silluete is: 0.003329983240252127
```

```
Cluster 110 has Size: 12
Silluete is: -0.5254927018761238

Cluster 111 has Size: 22
Silluete is: 0.6194313180426764

Cluster 112 has Size: 6
Silluete is: -0.791712422580809

Cluster 113 has Size: 10
Silluete is: -0.7742379046564314

Cluster 114 has Size: 10
Silluete is: -0.28004060221948257

Cluster 115 has Size: 10
Silluete is: -0.7885586882034252

Cluster 116 has Size: 18
Silluete is: -0.7513454441843195

Cluster 117 has Size: 20
Silluete is: -0.6213275472666235

Cluster 118 has Size: 11
Silluete is: -0.8241604856111854

Cluster 119 has Size: 15
Silluete is: -0.6512793988937184

Cluster 120 has Size: 21
Silluete is: -0.3453471997074231

Cluster 121 has Size: 16
Silluete is: 0.06029726624665034

Cluster 122 has Size: 10
Silluete is: -0.22153667118898968

Cluster 123 has Size: 10
Silluete is: -0.602277728106423

Cluster 124 has Size: 13
Silluete is: -0.36094453085377093

Cluster 125 has Size: 16
Silluete is: -0.3303666130315036

Cluster 126 has Size: 16
Silluete is: -0.7356132371799495

Cluster 127 has Size: 9
Silluete is: -0.3970374501909322

Cluster 128 has Size: 10
Silluete is: -0.3592528092736128

Cluster 129 has Size: 10
```

```
Silluete is: -0.5849585159981638

Cluster 130 has Size: 10
Silluete is: -0.8022874282893199

Cluster 131 has Size: 6
Silluete is: -0.5204089861243826

Cluster 132 has Size: 9
Silluete is: -0.7076864197243031

Cluster 133 has Size: 9
Silluete is: -0.5109864802376741

Cluster 134 has Size: 14
Silluete is: -0.4051949487241647

Cluster 135 has Size: 16
Silluete is: -0.4909523137775052

Cluster 136 has Size: 16
Silluete is: -0.6007529738170321

Cluster 137 has Size: 13
Silluete is: -0.7931736339497018

Cluster 138 has Size: 10
Silluete is: -0.8214703592165674

Cluster 139 has Size: 15
Silluete is: -0.545940966298877

Cluster 140 has Size: 7
Silluete is: -0.7828896379692754

Cluster 141 has Size: 10
Silluete is: -0.5126554164878249

Cluster 142 has Size: 10
Silluete is: -0.4522323959900166

Cluster 143 has Size: 6
Silluete is: -0.530149044033636

Cluster 144 has Size: 6
Silluete is: 0.8303963942002227

Cluster 145 has Size: 10
Silluete is: -0.371347352588387

Cluster 146 has Size: 13
Silluete is: -0.8318341635285217

Cluster 147 has Size: 12
Silluete is: -0.8424290846034599

Cluster 148 has Size: 12
Silluete is: 0.48687725121170505
```

```
Cluster 149 has Size: 8
Silluete is: -0.6342991499296188

Cluster 150 has Size: 14
Silluete is: -0.5994891250257242

Cluster 151 has Size: 8
Silluete is: -0.30469953744472433

Cluster 152 has Size: 10
Silluete is: -0.19668355988289912

Cluster 153 has Size: 10
Silluete is: -0.43660487303985707

Cluster 154 has Size: 9
Silluete is: -0.780489583873063

Cluster 155 has Size: 10
Silluete is: -0.640989347565757

Cluster 156 has Size: 10
Silluete is: 0.6061135258618094

Cluster 157 has Size: 10
Silluete is: -0.6557425120854568

Cluster 158 has Size: 9
Silluete is: -0.5184614266254268

Cluster 159 has Size: 10
Silluete is: -0.850374045576525

Cluster 160 has Size: 10
Silluete is: -0.688168274555593

Cluster 161 has Size: 7
Silluete is: -0.47498572068578193

Cluster 162 has Size: 6
Silluete is: -0.20659274773077824

Cluster 163 has Size: 9
Silluete is: -0.147405965305222

Cluster -1 has Size: 10662
```

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

In [114…

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

In [115…

```
# 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(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, la
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: 27
Estimated number of noise points: 1361
Homogeneity: 0.118
Confusion Matrix
[ 721 3100 8781 1044  704 5872  767  959  532   79  421  104  251  795
  104  375    0   14   26   13    7   20    0    7    0    9   10    5]
[ 640 2111 1606   49  876 1297  230  326   34  223   21  263   10   52
    3   26   36    1    0    0    3    2   10    3   14    1    0    4]
Silhouette Coefficient: -0.061
Overview over the Cluster

Cluster 0 has Size: 5211
Silluete is: 0.05371808383043013

Cluster 1 has Size: 10387
Silluete is: -0.6223957333079858

Cluster 2 has Size: 1093
Silluete is: 0.1677322398088815

Cluster 3 has Size: 1580
Silluete is: -0.17753048401224883

Cluster 4 has Size: 7169
Silluete is: 0.025675172412631744

Cluster 5 has Size: 997
Silluete is: 0.12594194439330283

Cluster 6 has Size: 1285
```

```
Silluete is: -0.7063306170939103

Cluster 7 has Size: 566
Silluete is: -0.20568572075862956

Cluster 8 has Size: 302
Silluete is: -0.33410061991243295

Cluster 9 has Size: 442
Silluete is: 0.017646788804516617

Cluster 10 has Size: 367
Silluete is: -0.5310128476673013

Cluster 11 has Size: 261
Silluete is: 0.03556730753336913

Cluster 12 has Size: 847
Silluete is: -0.056032835177888436

Cluster 13 has Size: 107
Silluete is: 0.0747661174415045

Cluster 14 has Size: 401
Silluete is: 0.19993974782542856

Cluster 15 has Size: 36
Silluete is: -0.18153039654360242

Cluster 16 has Size: 15
Silluete is: -0.09567182871761397

Cluster 17 has Size: 26
Silluete is: 0.12163581664779408

Cluster 18 has Size: 13
Silluete is: -0.03216456797126432

Cluster 19 has Size: 10
Silluete is: 0.11083942640689336

Cluster 20 has Size: 22
Silluete is: 0.014490845973672933

Cluster 21 has Size: 10
Silluete is: -0.38867683122599905

Cluster 22 has Size: 10
Silluete is: 0.2206895683167755

Cluster 23 has Size: 14
Silluete is: -0.042343636017809244

Cluster 24 has Size: 10
Silluete is: -0.20510741302496727

Cluster 25 has Size: 10
Silluete is: -0.11285515269719461
```

```
Cluster 26 has Size: 9
Silluete is: -0.12118657897049373

Cluster -1 has Size: 1361
```

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

In [116…

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

In [117…

```
# 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, la
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: 949
Homogeneity: 0.009
Confusion Matrix
[  494 24226     0]
[ 455 7371   15]
Silhouette Coefficient: 0.474
Overview over the Cluster

Cluster 0 has Size: 31597
Silluete is: 0.5176998366603035

Cluster 1 has Size: 15
Silluete is: 0.3792688833229348

Cluster -1 has Size: 949
Silluete is: 0.6036780851106535
```

Homoginty became way worse, silluete finaly in a reasonable area, but still low, on cluster completly 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 apperantly destroying everything

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

In [118…

```
kmeans2 = KMeans(n_clusters=2, random_state=0).fit(adultData)
```

In [119…

```
# 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(adultDatalabels, 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(adultDatalabels, labels)[0] )
print( metrics.confusion_matrix(adultDatalabels, labels)[1] )
print("Silhouette Coefficient: %0.3f" % metrics.silhouette_score(adultData, la
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
Homogeneity: 0.090
Confusion Matrix
[19244  5476]
```

```
[3366 4475]
Silhouette Coefficient: 0.330
Overview over the Cluster

Cluster 0 has Size: 22610
Silluete is: 0.49250051947956236

Cluster 1 has Size: 9951
Silluete is: 0.23122877079819149
```

far Far better Silluete, but Homogenity is still bad.

In [120…

```
kmeans3 = KMeans(n_clusters=3, random_state=0).fit(adultData)
```

In [121…

```
# 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(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, la
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: 0.115
Confusion Matrix
[ 6925 12588  5207]
[2172 1298 4371]
Silhouette Coefficient: 0.319
Overview over the Cluster

Cluster 0 has Size: 9097
Silluete is: 0.47080766932948265

Cluster 1 has Size: 13886
Silluete is: 0.2259135988712402

Cluster 2 has Size: 9578
Silluete is: 0.24210046971253524
```

We lost a bit of silluetee coeficcient, but our homogenity is growing, lets try 4 and 6

In [122…

```
kmeans4 = KMeans(n_clusters=4, random_state=0).fit(adultData)
```

In [123…

```
# 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(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, la
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
[11401  6127  4771  2421]
[1117 2380 4173  171]
Silhouette Coefficient: 0.326
Overview over the Cluster

Cluster 0 has Size: 12518
Silluete is: 0.4504095445488164

Cluster 1 has Size: 8507
Silluete is: 0.11329264596803394

Cluster 2 has Size: 8944
Silluete is: 0.1340192847498332

Cluster 3 has Size: 2592
Silluete is: 0.13803957697431551
```

In [124…

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

In [125…

```
# 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, la
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
[9671 4854 4544 2373    0 3278]
[1665 1810 3981  166  159   60]
Silhouette Coefficient: 0.320
Overview over the Cluster

Cluster 0 has Size: 11336
Silluete is: 0.4590226417888194

Cluster 1 has Size: 6664
Silluete is: 0.09539638340818919

Cluster 2 has Size: 8525
Silluete is: 0.34199242311146244

Cluster 3 has Size: 2539
Silluete is: 0.14979981006469928

Cluster 4 has Size: 159
Silluete is: 0.3410730381690188

Cluster 5 has Size: 3338
Silluete is: 0.5284341191588952
```

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

In [126…

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

In [127…

```
# 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(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, la
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
[1134 2459  684 4451 1681 1175    0 2812 2472 1274  840 5738]
[ 475   35 1179 1379   61  102  159 1107 1069 1733  111  431]
Silhouette Coefficient: 0.294
Overview over the Cluster

Cluster 0 has Size: 1609
Silluete is: 0.2821013555629567

Cluster 1 has Size: 2494
Silluete is: 0.0952647049778358

Cluster 2 has Size: 1863
Silluete is: 0.34860168919578927

Cluster 3 has Size: 5830
Silluete is: 0.17929718045952786

Cluster 4 has Size: 1742
Silluete is: 0.5788932760822273

Cluster 5 has Size: 1277
Silluete is: 0.29541244772336595

Cluster 6 has Size: 159
Silluete is: 0.17229258095376254
```

```
Cluster 7 has Size: 3919
Silluete is: 0.26612577287189776

Cluster 8 has Size: 3541
Silluete is: -0.06673998155704802

Cluster 9 has Size: 3007
Silluete is: -0.03142614327976197

Cluster 10 has Size: 951
Silluete is: 0.41447820200190927

Cluster 11 has Size: 6169
Silluete is: 0.5765624214193797
```

Homogenity still growing, Silluette still sinking

In [ ]: